13th International Symposium on Hearing

Dourdan, France, 24-29 August 2003

These papers will appear in a book with the provisional title:

Auditory signal processing: physiology, psychoacoustics, and models

Editors: Daniel Pressnitzer, Alain de Cheveigné, Stephen McAdams, and Lionel Collet

To be published by Springer Verlag, New York, Spring 2004

Supported by:

– U.S. Office of Naval Research - International Field Office
– Centre National de la Recherche Scientifique, Département SDV
– Délégation Générale pour l'Armement

Table of contents

Cochlear signal processing

Mary Ann Cheatham 1
Nonlinearities at the apex of the cochlea: Implications for auditory perception

Marcel van der Heijden and Philip X. Joris 7
Reconstructing the traveling wave from auditory nerve responses

Alberto Lopez-Najera, Ray Meddis, and Enrique A. Lopez-Poveda 14
A computational algorithm for computing cochlear frequency selectivity: Further studies

Roy D. Patterson, Masashi Unoki, and Toshio Irino 21
Comparison of the compressive-gammachirp and double-roex auditory filters

Mary Florentine, Søren Buus, and Mindy Rosenberg 28
Reaction-time data support the existence of Softness Imperception in cochlear hearing loss

Michael G. Heinz, Danilo Scepanovic, Murray B. Sachs, and Eric D. Young 35
Normal and impaired level encoding: Effects of noise-induced hearing loss on auditory-nerve responses

Stephen T. Neely, Kim S. Schairer, and Walt Jesteadt 42
Estimates of cochlear compression from measurements of loudness growth

Christopher J. Plack, Catherine G. O'Hanlon, and Vit Drga 49
Additivity of masking and auditory compression

Magdalena Wojtczak and Neal F. Viemeister 56
Psychophysical response growth under suppression

Brainstem signal processing

David W. Smith, E. Christopher Kirk, and Emily Buss 63
The function(s) of the medial olivocochlear efferent system in hearing

Katuhiro Maki and Masato Akagi 70
A computational model of cochlear nucleus neurons

Kazuhito Ito and Masato Akagi 77
Study on improving regularity of neural phase-locking in single neurons of AVCN via a computational model

Dries H. Louage, Marcel van der Heijden, and Philip X. Joris 84
Fibers in the trapezoid body show enhanced synchronization to broadband noise when compared to auditory nerve fibers

Pitch

Leonardo Cedolin and Bertrand Delgutte 91
Representations of the pitch of complex tones in the auditory nerve

Lutz Wiegrebe, Alexandra Stein, and Ray Meddis 98
Coding of pitch and amplitude modulation in the auditory brainstem: One common mechanism?

Andrew J. Oxenham, Joshua G. Bernstein, and Christophe Micheyl 105
Pitch perception of complex tones within and across ears and frequency regions

Laurent Demany, Gaspard Montandon, and Catherine Semal 112
Internal noise and memory for pitch

André Rupp, Stefan Uppenkamp, Jen Bailes, Alexander Gutschalk, and Roy D. Patterson 119
Time constants in temporal pitch extraction: A comparison of psychophysical and neuromagnetic data

Bernd Lütkenhöner, Christian Borgmann, Katrin Krumbholz, Stefan Seither, and Annemarie Seither-Preisler 126
Auditory processing at the lower limit of pitch studied by magnetoencephalography

Frequency modulation

Günter Ehret, Steffen R. Hage, Marina Egorova, and Birgit A. Müller 133
Auditory maps in the midbrain: the inferior colliculus

Craig Atencio, Fabrizio Strata, David Blake, Ben Bonham, Benoit Godey, Michael Merzenich, Christoph Schreiner, and Steven Cheung 140
Representation of frequency modulation in the primary auditory cortex of New World monkeys

Pierre L. Divenyi 147
Frequency change velocity detector: A bird or red herring?

Robert P. Carlyon, Christophe Micheyl, and John Deeks 154
Coding of FM and the continuity illusion

Streaming

Makio Kashino and Minae Okada 161
The role of spectral change detectors in sequential grouping of tones

Christophe Micheyl, Robert P. Carlyon, Rhodri Cusack, and Brian C.J. Moore 167
Performance measures of auditory organization

Nicolas Grimault, Sid P. Bacon, and Christophe Micheyl 174
Auditory streaming without spectral cues in hearing-impaired subjects

Amplitude modulation

Neal F. Viemeister, Mark A. Stellmack, and Andrew J. Byrne 181
The role of temporal structure in envelope processing

Christian Füllgrabe, Laurent Demany, and Christian Lorenzi 188
Detecting changes in amplitude-modulation frequency: A test of the concept of excitation pattern in the temporal-envelope domain

Erick Gallun, Ervin R. Hafter, and Anne-Marie Bonnel 195
Modeling the role of duration in intensity increment detection

Stanley Sheft and William A. Yost 202
Minimum integration times for processing of amplitude modulation

Responses to complex sounds

Ellen Covey and Paul A. Faure 209
Neural mechanisms for analyzing temporal patterns in echolocating bats

Diana B. Geissler and Günter Ehret 216
Time-critical frequency integration of complex communication sounds in the auditory cortex of the mouse

Israel Nelken, Nachum Ulanovsky, Liora Las, Omer Bar-Yosef, Michael Anderson, Gal Chechik, Naftali Tishby, and Eric Young 223
Transformation of stimulus representations in the ascending auditory system

Dennis L. Barbour and Xiaoqin Wang 230
AM and FM coherence sensitivity in the auditory cortex as a potential neural mechanism for sound segregation

Speech

Fan-Gang Zeng, Kaibao Nie, Ginger Stickney, and Ying-Yee Kong 237
Auditory perception with slowly-varying amplitude and frequency modulations

Steven J. Eliades and Xiaoqin Wang 244
The role of auditory-vocal interaction in hearing

Ingrid Johnsrude, Matt Davis, and Alexis Hervais-Adelman 251
From sound to meaning: Hierarchical processing in speech comprehension

John F. Culling and Julia S. Porter 258
Effects of differences in the accent and gender of interfering voices on speech segregation

Jont B. Allen 265
The Articulation Index is a Shannon channel capacity

Comodulation masking release

Ian M. Winter, Veronika Neuert, and Jesko L. Verhey 272
Comodulation masking release and the role of wideband inhibition in the cochlear nucleus

Georg M. Klump and Sonja B. Hofer 279
The relevance of rate and time cues for CMR in starling auditory forebrain neurons

Torsten Dau, Stephan D. Ewert, and Andrew J. Oxenham 285
Effects of concurrent and sequential streaming in comodulation masking release

Binaural hearing

Ray Meddis, Christian Sumner, and Susan Shore 292
Effects of contralateral sound stimulation on forward masking in the guinea pig

H. Steven Colburn, Yi Zhou, and Vasant Dasika 299
Inhibition in models of coincidence detection

Birger Kollmeier and Helmut Riedel 306
What can auditory evoked potentials tell us about binaural processing in humans?

Trevor M. Shackleton and Alan R. Palmer 313
Sensitivity to changes in interaural time difference and interaural correlation in the inferior colliculus

Leslie R. Bernstein and Constantine Trahiotis 320
Processing of interaural temporal disparities with both ”transposed” and conventional stimuli

D. Wesley Grantham, Daniel H. Ashmead, and Todd A. Ricketts 327
Sound localization in the frontal horizontal plane by post-lingually deafened adults fitted with bilateral cochlear implants

Steven van de Par, Armin Kohlrausch, Jeroen Breebaart, and Martin McKinney 334
Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise

Courtney C. Lane, Norbert Kopco, Bertrand Delgutte, Barbara G. Shinn-Cunningham, and H. Steven Colburn 341
A cat’s cocktail party: Psychophysical, neurophysiological, and computational studies of spatial release from masking

Brad Rakerd and William M. Hartmann 348
Localization of noise in a reverberant environment

Anthony J. Watkins 355
Listening in real-room reverberation: Effects of extrinsic context

Daniel J. Tollin, Micheal L. Dent, and Tom C.T. Yin 361
Psychophysical and physiological studies of the precedence effect and echo threshold in the behaving cat

Michael A. Akeroyd 368
Some similarities between the temporal resolution and the temporal integration of interaural time differences

Caroline Witton, Gary G.R. Green, and G. Bruce Henning 375
Binaural ”sluggishness” as a function of stimulus bandwidth

Temporal coding

Peter Heil and Heinrich Neubauer 382
Auditory thresholds re-visited

Marjorie Leek, Robert Dooling, Otto Gleich, and Micheal Dent 389
Discrimination of temporal fine structure by birds and mammals

Philip X. Joris, Marcel van der Heijden, Dries Louage, Bram Van de Sande, and Cindy Van Kerckhoven 396
Dependence of binaural and cochlear ”best delays” on characteristic frequency

Mounya Elhilali, David J. Klein, Jonathan B. Fritz, Jonathan Z. Simon, and Shihab A. Shamma 403
The enigma of cortical responses: Slow yet precise

Plasticity

Jean-Marc Edeline 410
Learning-induced sensory plasticity: Rate code, temporal code, or both?

Katrina M. MacLeod and Catherine E. Carr 416
Synaptic dynamics and intensity coding in the cochlear nucleus

Beverly A. Wright and Matthew B. Fitzgerald 423
Learning and generalization on five basic auditory discrimination tasks as assessed by threshold changes

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Nonlinearities at the apex of the cochlea: Implications for auditory perception

Mary Ann Cheatham

Department of Communication Sciences and Disorders, Northwestern University
m-cheatham@northwestern.edu

1 Introduction

Many aspects of auditory perception are thought to reflect basilar membrane responses, which are nonlinear by virtue of their feedback relationship with outer hair cells. This association between cochlear mechanics and perception (Oxenham and Plack 1997) is primarily based on comparisons with the well-characterized mechanical responses recorded from the base of the mammalian cochlea (Ruggero, Rich, Recio, Narayan and Robles 1997). Hence, the purpose of this report is to review what is known about apical inner hair cell (IHC) responses, which provide an indication of how signals are coded by the basilar membrane-outer hair cell-tectorial membrane complex. Because mammalian IHCs supply the inputs to the auditory nerve, their receptor potentials ultimately provide the substrate upon which central auditory processing is based. Therefore, results are evaluated for IHCs in turns 2 and 3 of the guinea pig cochlea, where best frequencies (BF) are ~1 and ~4 kHz, respectively. In other words, recordings are made from regions where the vast majority of psychophysical measurements are obtained.

Data indicate that intermodulation distortion is observed in the responses of IHCs with low and moderate BFs (Cheatham and Dallos 1997). When primary pairs, f1 and f2, are placed well above BF, responses are recorded at the cubic difference tone, $2f1\text{-}f2 \approx BF$. Because the primaries do not excite the IHC, the distortion is thought to originate at a location basal to the recording site. Energy is then redistributed to a more apical location, which is appropriate to the frequency of the distortion product (Goldstein 1967; Kim, Molnar and Matthews 1980; Robles and Ruggero 2001). Evidence for redistribution is also found for quadratic difference tones (f2-f1) in both IHC and mechanical (Cooper and Rhode 1997) responses obtained from the apical half of the cochlea. There is, however, no evidence that higher-order harmonics are redistributed with their own traveling waves, consistent with the physical properties of the cochlear partition (Lighthill 1981). This report provides additional information about harmonic and

intermodulation distortion at the IHC level and compares results with those on cochlear mechanics where appropriate. Implications for auditory perception are also discussed.

2 Methods

Inner hair cells were recorded from turns 2 and 3 of the guinea pig cochlea using the lateral approach (Dallos, Santos-Sacchi and Flock 1982; Cheatham and Dallos 1997). Peak values for various ac components were determined offline from FFTs of averaged response waveforms. By using a signal consisting of two partially overlapping tones, it is possible to measure harmonics, produced in response to one input, as well as combination tones produced in the region of overlap when both tones are presented simultaneously. Stimulus frequencies are the closest in our collection to those used in experiments on the perception of mistuned consonances. All procedures were approved by the NIH and by Northwestern University's Animal Care and Use Committee.

3 Results and discussion

The top panel of Fig. 1 shows the receptor potential recorded from a third-turn IHC in response to a two-tone input composed of f1 at 370 Hz and f2 at 910 Hz, with both presented at 70 dB. Energy at harmonic frequencies up to 8f1 at 2960 Hz and 3f2 at 2730 Hz is represented by the FFT at the bottom. Even though harmonics are not redistributed via their own traveling waves, there exist a large number of these components with sizable magnitudes at the generation site. In addition to the series of harmonics generated in response to each primary, combination tones are also observed at 170 Hz (f2-2f1) and 540 Hz (f2-f1). Because bandwidths are wide at moderate levels, this information from a single IHC is probably available in the temporal responses recorded from single units (Rose, Brugge, Anderson and Hind 1967; Kim et al. 1980).

Viemeister, Rickert and Stellmack (2001, Fig. 1) considered whether harmonic distortion could influence forward masking patterns for signals near the second harmonic of the masker frequency. For example, when a 500 Hz masker was presented at 80 dB, small fluctuations in threshold were observed for a 1 kHz signal. Viemeister and colleagues concluded that traveling harmonics did not influence these forward masking patterns. Assuming that threshold variations are due to beat detection, their conclusion is consistent with Plomp's (1967) results showing that high-pass masking did not influence beat perception. It is also unlikely that harmonic components at the generation site increase masker excitation and/or the neural adaptation that underlies forward masking. This is because the energy added by multiple components is related to the square root of the sums of the squares of their individual magnitudes (Beranek 1954).

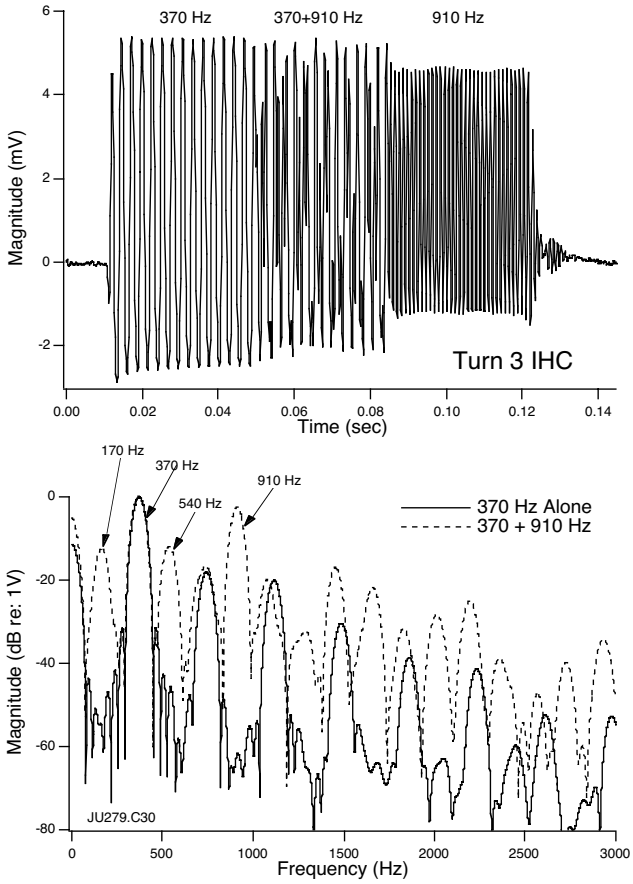


Fig. 1. The top panel shows the ac receptor potential generated in response to 370 and 910 Hz. The bottom panel shows the spectrum (0 dB corresponds to 5 mV peak) obtained when the two inputs were presented together at 70 dB. This third-turn IHC had a BF of 900 Hz.

Results in Fig. 2 were obtained from a second-turn IHC with a BF of 4100 Hz. In this case, f1 at 2040 Hz is ~1/2 octave below BF, while f2 at 4300 Hz is slightly above BF. The FFT in part A shows energy in the ac receptor potential at f2-2f1 (4300-4080=220 Hz), as well as at f1, 2f1 and f2. Also notice that the side bandssurrounding f1 represent amplitude modulation between f1 and f2-2f1. As the level of f2 increases relative to f1, the response at 220 Hz decreases, as shown in part B, and disappears into the noise, in part C. In addition, responses to f1 and 2f1 decrease as the level of f2 exceeds f1. Hence, mutual suppression influences the interaction between inputs, which was shown previously in single unit responses (Kim et al. 1980). If these distortion products influence perception, then the level dependence of the interactions may be important. For example, psychophysical data on the perception of a mistuned octave at 4 kHz (Viemeister et al. 2000, Fig. 3A)

indicate that performance decreases as the higher frequency primary increases in level. Although speculative, mutual suppression could contribute to these effects. The companion waveforms on the right show amplitude modulation when the two tones are presented together. These envelope variations might be useful in detecting a mistuned octave, as suggested by Viemeister and colleagues.

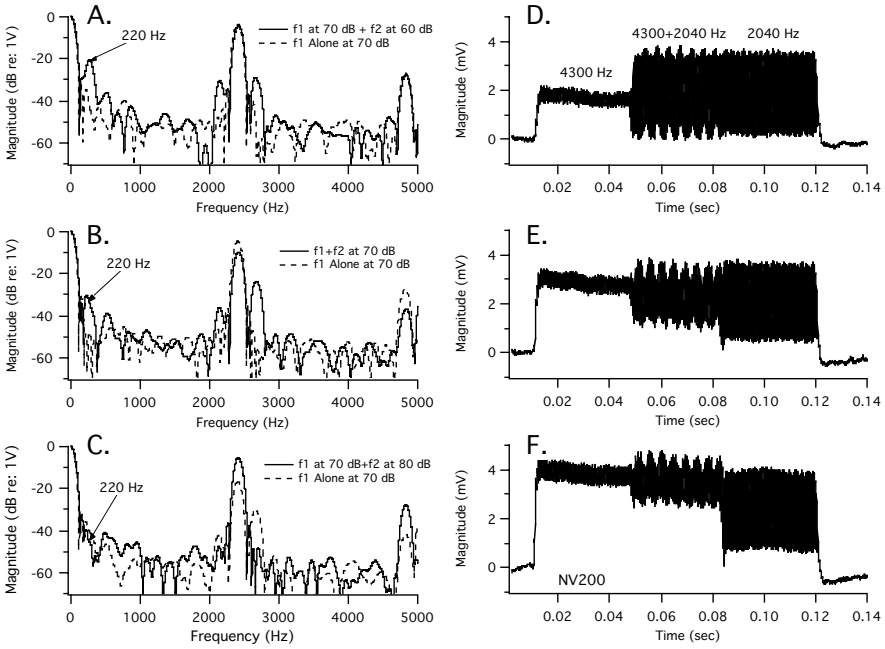


Fig. 2. Spectra are shown on the left for IHC responses to f1 at 2040 Hz and 70 dB measured in the presence of f2 at 4300 Hz. The latter input is presented at 60 dB in part A, 70 dB in part B and 80 dB in part C. Companion waveforms are shown in D, E and F.

In order to emphasize longitudinal variations in cochlear function, second harmonics recorded at the base of the cochlea (Cooper 1998) are shown in Fig. 3A. All responses were normalized to the largest value measured at the fundamental. By plotting results on a normalized frequency scale, it is possible to compare Cooper's data with those from third-turn IHCs, as in part B. The mechanical responses indicate that the second harmonic exhibits a bimodal frequency distribution such that one peak occurs when the stimulus is at BF/2 (8.5 kHz). In this case, the primary is below BF and, therefore, receives little amplification. The second harmonic, however, is amplified because it coincides with BF where the cochlear amplifier has its largest gain. Cooper refers to this as amplified distortion. When the input is at BF (17 kHz), responses are highly nonlinear but the second harmonic at 34 kHz receives no amplification because it is well above BF. Cooper refers to this second peak as distorted amplification.

In contrast to Cooper's results from the base of the cochlea, data from an IHC with BF at 1 kHz show only one peak. These different distribution patterns are consistent with the knowledge that apical mechanical and hair cell responses are

nonlinear throughout the response area (Rhode and Cooper 1996) and that filter shapes are relatively symmetrical with no well defined tip and tail segments at the 1 kHz location. Hence, it is not surprising that two peaks are not found in apical IHC responses because amplified distortion would not be expected in IHCs with low BFs. Although not shown here, data at 80 dB do not display individual peaks, i.e., the second harmonic is expressed throughout the response area at both locations.

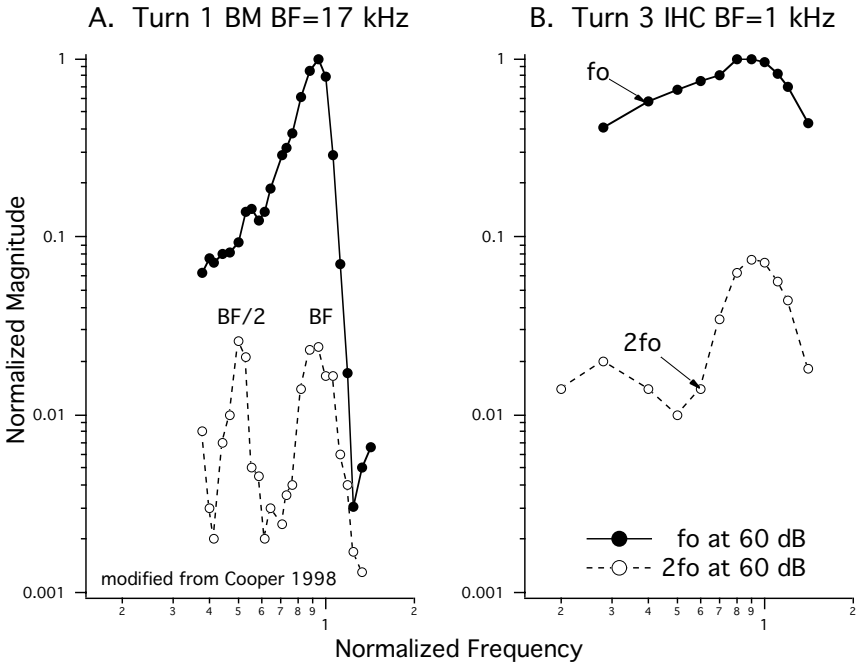


Fig. 3. In part A, basilar membrane data (Cooper 1998, Fig. 2A) are modified to show normalized responses for the fundamental (solid lines). Dashed lines indicate responses at the second harmonic also normalized to the largest fundamental value. Companion IHC data are plotted in panel B. All results were acquired at 60 dB.

4 Conclusions

Although hydromechanical filtering precludes aural harmonics with their own traveling waves, harmonic components are prevalent at the generation site (Dallos and Cheatham 1989). In the apex, where amplifier gain is weaker than in the base, the filters are broad and responses are nonlinear throughout the response area. In addition, cochlear nonlinearities result in envelope variations when two tones are presented at moderate signal levels. These observations are consistent with the Viemeister et al. (2000) explanation for the perception of mistuned octaves, which is based on the fact that compression, rectification and low-pass filtering are present at the IHC level. Because the IHC data presented here were obtained using short

tone bursts, envelope variations at really low frequencies are precluded. However, the 500 msec signals used in psychophysical experiments could very well develop slow modulations at 5 Hz. If human subjects do use envelope extraction, it would be interesting to know if the perception of mistuned octaves is affected when short-duration signals are used.

Acknowledgments

Work supported by NIH grant # DC00089 from the NIDCD. Comments on the manuscript by Peter Dallos, Jon Siegel and Neal Viemeister are appreciated.

References

Beranek, L.L. (1954) *Acoustics*. McGraw-Hill, New York.

Cheatham, M.A. and Dallos, P. (1997) Intermodulation components in inner hair cell and organ of Corti responses. J. Acoust. Soc. Am. 102, 1038-1048.

Cooper, N.P. (1998) Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea. J. Physiol. 509, 277-288.

Cooper N.P. and Rhode, W.S. (1997) Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea. J. Neurophysiol. 78, 261-270.

Dallos, P. and Cheatham, M.A. (1989) Nonlinearities in cochlear receptor potentials and their origins. J. Acoust. Soc. Am. 86, 1790-1796.

Dallos, P., Santos-Sacchi, J. and Flock A. (1982) Intracellular recordings from cochlear outer hair cells. Science 218, 582-584.

Goldstein, J.L. (1967) Auditory nonlinearity. J. Acoust. Soc. Am. 41, 676-689.

Kim, D.O., Molnar, C.E. and Matthews, J.W. (1980) Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure. J. Acoust. Soc. Am. 67, 1704-1721.

Lighthill, J. (1981) Energy flow in the cochlea. J. Fluid Mech. 106, 149-213.

Oxenham A.J. and Plack, C.J. (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. J. Acoust. Soc. Am. 101, 3666-3675.

Plomp, R. (1967). Beats of mistuned consonances. J. Acoust. Soc. Am. 42, 462-474.

Rhode, W.S. and Cooper, N.P. (1996) Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*. Aud. Neurosci. 3, 101-121.

Robles, L. and Ruggero, M.A. (2001) Mechanics of the mammalian cochlea. Physiol. Rev. 81, 1305-1352.

Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1967) Phase-locked responses to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. 30, 769-793.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S. and Robles, L. (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J. Acoust. Soc. Am. 101, 2151-2163.

Viemeister, N.F., Rickert, M. and Stellmack, M. (2001) Beats of mistuned consonances: Implications for auditory coding. In: D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs and R. Schoonhoven (Eds.), *Physiological and Psychophysical Bases of Auditory Function*. Shaker Publishing, Maastricht, pp. 113-120.

Reconstructing the traveling wave from auditory nerve responses

Marcel van der Heijden and Philip X. Joris

Laboratory of Auditory Neurophysiology, K.U.Leuven, Leuven, Belgium
Marcel.Vanderheyden@med.kuleuven.ac.be

1 Introduction

In a previous study, we presented gain and phase characteristics of a single, intact, cochlea of a cat (van der Heijden and Joris 2003). These curves were derived from responses of the auditory nerve (AN) to irregularly spaced tone complexes. Unlike traditional techniques like stimulation by single tones (Kim and Molnar 1979) and reverse correlation (de Boer 1967), this novel technique is not restricted to the range of phase locking of the nerve. Figure 1 shows a collection of gain and phase curves obtained from a collection of AN fibers innervating a single cochlea of the cat. Characteristic frequencies (CFs) of the fibers ranged from 700 Hz to 19.3 kHz; the curves cover almost all stimulus frequencies that excited these fibers.

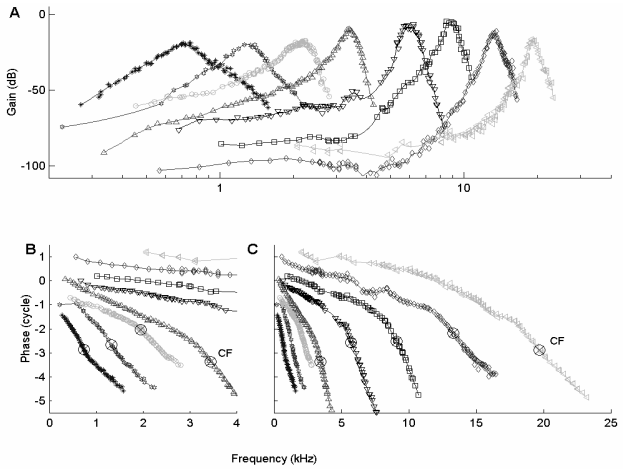


Fig 1. Collection of gain and phase curves obtained from the data of a single auditory nerve of a cat. (A) Gain curves. CFs range from 740 Hz to 19.7 kHz. (B) Blow-up of the low-frequency portion of the phase curves. The ⊗ indicate the fibers' CFs. (C) Entire phase curves, advanced by 1 ms. Relative vertical positions were adjusted to avoid overlap.

Figure 1 shows a subset of a collection of 17 fibers for which we were able to determine gain and phase curves. The present study offers an analysis of these data aimed at a reconstruction of the propagation of acoustic energy within the cochlea. In order to arrive at such a "panoramic view," we need to describe gain and phase as a function of CF – and thus of cochlear location – rather than stimulus frequency.

2 Determination of amplitude, phase and group delay

We now describe the steps that lead from the set of gain and phase curves (Fig. 1) to a more panoramic view of the pattern of cochlear vibration.

2.1 Limitations of the data; supplementary data

The gain and phase curves of Fig. 1 were derived from an analysis from the interaction between primaries of tone complexes (van der Heijden and Joris, 2003). Details need not concern us here but an essential limitation of this approach is the lack of *absolute* gain and phase data. Thus each of the individual curves is only defined up to an unknown offset, i.e., the relative positions of the individual curves are arbitrary. In order to fix the offsets we need supplementary data.

In the case of the gain curves, we used the thresholds to CF tones of all the fibers of the collection. The maxima of the gain curves were adjusted in such a way that their positions reflected the fibers' sensitivity to CF tones. This amounts to expressing them *re* a hypothetical threshold displacement needed to excite the fiber. The reliability of this approach for describing the amplitude of cochlear vibration *per se* is limited by two circumstances: a potential systematic variation of inner hair cell sensitivity with CF, and the variability of thresholds across same-CF fibers (Liberman 1978). We did not attempt to correct for these effects.

In the case of the phase curves, the supplementary information that fixes their relative positions is provided by the phase locking to low-frequency primary tones. This fixes the low-frequency portion of the phase curve and thereby the whole curve. For CFs above 4 kHz, primary phase (not shown here) varied smoothly with frequency but for CFs below 4 kHz unwrapping became problematic and we did not use the phases for these low CFs. In theory, unwrapping problems can always be solved by increasing the sampling density. Unfortunately, sampling along the CF axis is not under direct control of the experimenter.

The phase data in Fig. 1 are compensated for an estimated synaptic delay of 1 ms (Ruggero and Robles 1987). This compensation is intended to facilitate comparisons with mechanical measurements of the basilar membrane (Robles and Ruggero 2001). Note, however, that the range of phases is sensitive to the exact choice of synaptic delay; this is particularly true for the high-CF curves. For example, an extra delay of only 100 μs will increase the phase accumulation at 20 kHz by as much as 2 cycles. In any event, for our present purposes the estimate of synaptic delay is irrelevant since it does not affect the variation of phase with CF.

2.2 Interpolation and smoothing

The next step is an interpolation of the amplitude and phase data to arbitrary values of both stimulus frequency (SF) and CF within the range covered by the data. This was done by representing the data on a *log*(SF) x *log*(CF) matrix and applying a bilinear interpolation. A running average was then applied using boxcar windows of 0.3 octaves along both SF and CF axes. The resulting two-dimensional matrices are shown in Figs. 2 and 3 as contour plots.

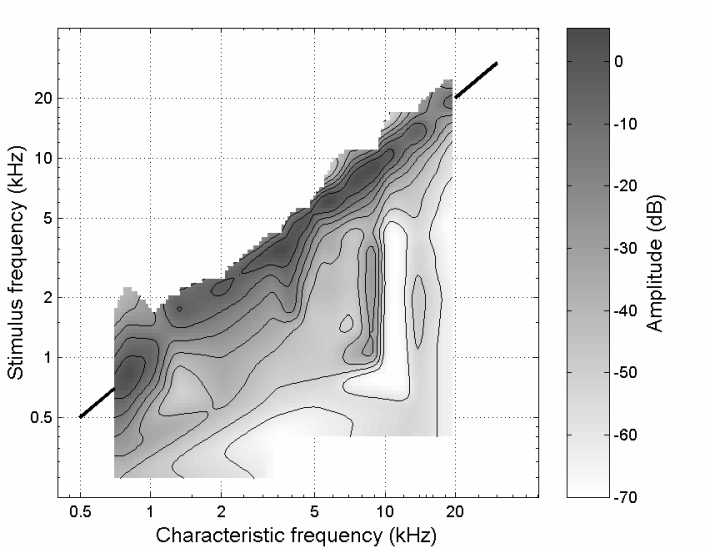


Fig. 2. Contour plot of the amplitude over a range of stimulus frequencies and CFs. The thick skewed lines indicate the diagonal (stimulus frequency=CF).

As expected, the amplitude generally peaks around the main diagonal, where stimulus frequency approaches CF. Phase is seen to vary quite systematically; iso-phase contours are roughly parallel to the main diagonal and accumulate near the diagonal. The phases cover a total range of about 7 cycles. Recall that the *total* variation of phase in the SF ("vertical") direction depends on the choice of synaptic delay whereas the variation of phase with CF is not affected by this choice. Variation of amplitude and phase along the latter dimension provides a representation of the vibration evoked by a single stimulus frequency.

2.3 Traveling waves

The interpolated data in Figs. 2 and 3 allow us to examine how amplitude and phase of *single* primaries vary along the length of cochlea. Figure 4 shows the gain and phase as a function of CF or, equivalently, distance from the apex for a set of stimulus frequencies ranging from 3 to 14 kHz. The empirical formula provided by Greenwood (1990) was used to convert CF to cochlear location.

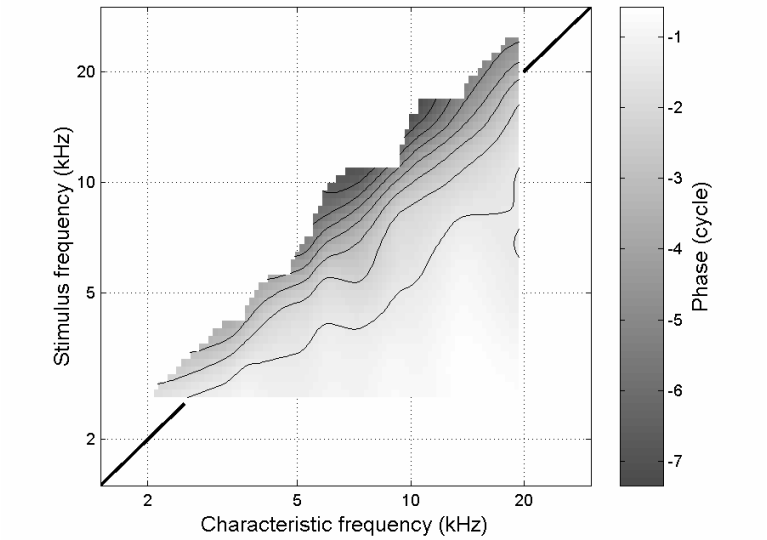


Fig. 3. Contour plot of the phase over a range of stimulus frequencies and CFs. Note that the range of stimulus frequencies and CFs is smaller than for the amplitude data of Fig. 2 .

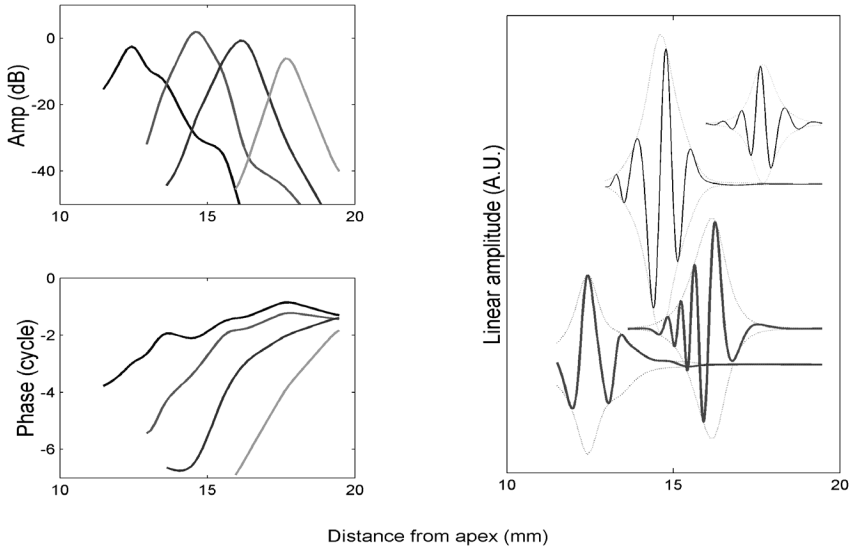


Fig. 4. Left panels: Amplitude and phase as a function of cochlear location for stimulus frequencies of 5, 7, 10 and 14 kHz. Right panel: Snapshot of the corresponding vibration patterns. Dotted lines indicate the envelopes. Vertical offsets were added to avoid overlap.

The amplitude peaks at the best site of the tone and the phase indicates that propagation of the tones is slowed down around the best site. The right panel of Fig 4 shows these same data in the form of "snapshots", i.e., the linear amplitude of cochlear vibration at an arbitrary instance of time. As time evolves, these waves propagate towards the apex while their spatial envelopes (dotted lines) remain in place. The spread of excitation can be obtained from an inspection of the envelopes; it is in the order of 2 mm. This narrow region of excitation contains only a few (2 to 3) cycles of the wave pattern

Figure 4 allows estimates of the wavelengths by an evaluation of the distance over which the phase advances by one cycle. The estimated wavelengths around best site were 1.2, 0.87, 0.80 and 0.69 mm for tones of 5, 7, 10 and 14 kHz. These wavelengths correspond to phase velocities of 6.0, 6.1, 8.0 and 9.7 m/s, respectively. The slowing down of the wave generally causes wavelength and phase velocity to become smaller towards the base.

2.4 Group delays

The phase pattern at low CFs could not be reliably reconstructed due to phase ambiguities (see 2.2). Phase seems to behave rather wildly at the apex, as can also be inferred by the low-CF phase curves in Fig. 1. The lack of knowledge about relative phase across CF at the apex prevents us from constructing the traveling wave in this region. It does not, however, prevent an analysis of group delay. Group delay is defined as the *derivative* of phase with respect to stimulus frequency. It is thus unaffected by arbitrary phase offsets across fibers. Physically, group delay represents the delay with which the acoustic energy of a given frequency is delivered to the site of the measurement. The underlying group *velocity* is the speed with which the energy is being transported (e.g., Elmore and Heald 1985).

In order to evaluate group delay, we extracted the slopes of the phase curves of Fig. 1. These data were interpolated to a SF x CF grid by the methods described above. The resulting contour plot is shown in Fig. 5. Figure 6 shows group delay as a function of cochlear position for a number of fixed stimulus frequencies.

Figures 5 and 6 suggest the distinction of two cochlear regions. The high-CF region (CF > 3 kHz) is traversed by low-frequency (< 3 kHz) energy with only a small delay. Energy above 3 kHz intially also travels fast but slows down when approaching its best site; it does not propagate far beyond the best site. The apical region is quite different. When crossing the region around CF = 2 kHz, energy at all frequencies is slowed down almost uniformly. On top of this uniform delay, frequencies accumulate an extra delay when they approach their best site. Moreover, in the apical region the energy at a given frequency *does* penetrate the region apical to its best site.

The 1.2-kHz curve in Fig. 6 seems to indicate that group delays fall off *on both sides* of the best site. Such nonmonotonic patterns of group delay would present a deviation from a "proper" traveling wave in which the acoustic energy travels in a single direction. Alternative patterns of cochlear energy transport could originate from reflections or from acoustic pressure waves in the cochlea.

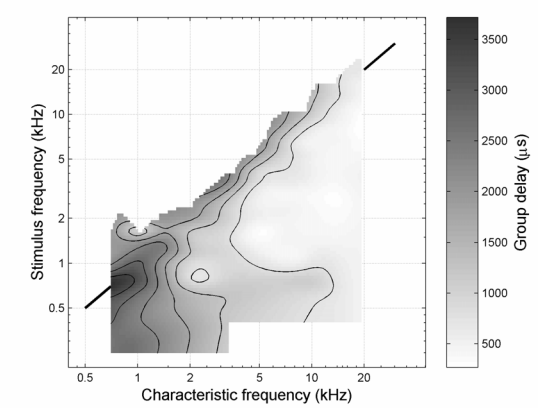


Fig. 5. Contour plot of the group delay over the entire range of stimulus frequencies and CFs employed in this study.

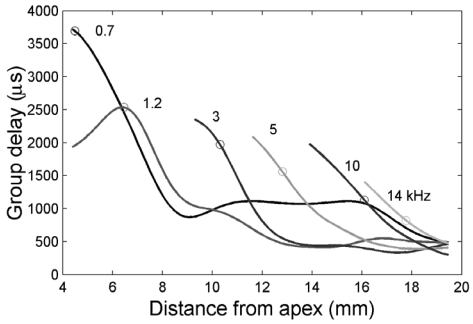


Fig. 6. Group delay as a function of cochlear location for tones of various stimulus frequencies. Circles indicate best sites.

Summary and discussion

The reconstruction of the vibration pattern in the cochlea resulted in the following observations on the basal region of the cochlea (CF>3 kHz):

- waves travel towards the apex
- waves are sharply peaked around their best site
- waves slow down when approaching their best site
- waves do not travel far beyond their best site
- wavelength at best site decreases with stimulus frequency
- phase velocity at best site increases with stimulus frequency

The patterning of group delays in the basal region is also consistent with the concept of traveling waves which slow down when approaching their best site and disappear beyond it. Due to insufficient data from low-CF fibers from the same cochlea, we could not extend the reconstruction of the vibration pattern to the apical region of the cochlea. The analysis of group delays, however, did not suffer from

this sampling problem; it indicated that the propagation of acoustic energy is qualitatively different at the apex than at the base.

Ren (2002) used a scanning laser interferometer to measure snapshots of cochlear vibration in the basal turn of the sensitive gerbil cochlea. He found a sharply peaked response to tones around their best site comprising about two cycles of a traveling wave over a limited (<1 mm) region of the cochlea. His estimates of wavelength and phase velocity at the 16 kHz region are 0.2 mm and 3.2 m/s, respectively.

Overall, our reconstructed snapshots of cochlear vibration patterns (Fig. 4) are similar to the patterns reported by Ren in that, at low levels, excitation is confined to a narrow cochlear region containing a few cycles of the wave. Also consistent with Ren's data is the decrease of wavelength towards the apex or, equivalently, the slowing down of the traveling wave. For the highest stimulus frequency considered in our snapshot analysis (14 kHz), our estimates of wavelength and phase velocity are 0.7 mm and 9.7 m/s, respectively. Both numbers are about 3 times larger than the values reported by Ren. Differences in species and cochlear location might underly these differences.

Overall, our high-CF reconstructions are consistent with current views on traveling waves in the cochlea (Robles and Ruggero 2001). Independent data are needed before one can judge the reproducibility and accuracy of our methods. As for the low-CF region, the group delays certainly suggest interesting deviations from the "ordinary" traveling wave at the apex. A lack of low-CF data prevented the reconstruction of apical vibration patterns but we can think of no fundamental reason why the method should not work for the low-CF region as well.

References

de Boer, E. (1967) Correlation studies applied to the frequency resolution of the cochlea. J. Aud. Res. 7, 209-217.

Elmore, W.C. and Heald, M.A. (1985) *Physics of Waves*, Dover, New York.

Greenwood, D.D. (1990) A cochlear frequency-position function for several species – 29 years later. J. Acoust. Soc. Am. 87, 2592-605.

van der Heijden and Joris (2003) in preparation.

Liberman, M.C. (1978) Auditory-nerve response from cats raised in a low-noise chamber. J. Acoust. Soc. Am. 63, 442-455.

Kim, D. O. and Molnar, C. E. (1979) A population study of cochlear nerve fibers: comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. J. Neurophysiol. 42, 16-30.

Ren, T. (2002) Longitudinal pattern of basilar membrane vibration in the sensitive cochlea. Proc. Natl. Acad. Sci. USA 99, 17101-17106.

Robles, L. and Ruggero, M. A. (2001) Mechanics of the mammalian cochlea. Physiol. Rev. 81, 1305-1352.

Ruggero, M. A. and Rich, N.C. (1987) Timing of spike initiation in cochlear afferents: dependence on site of innervation. J. Neurophys. 58, 379-403.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

A computational algorithm for computing cochlear frequency selectivity: Further studies

Alberto Lopez-Najera[1], Ray Meddis[2], and Enrique A. Lopez-Poveda[1§]

[1] Universidad de Castilla-La Mancha, {enrique.lopezpoveda, alberto.lopez}@uclm.es
[2] University of Essex, rmeddis@essex.ac.uk
[§] Corresponding author, presently at Universidad de Salamanca, ealopezpoveda@usal.es

1 Introduction

Phenomenological filter algorithms that compute nonlinear cochlear frequency selectivity are important for basic and applied hearing research. They constitute a crucial stage for modeling the physiology of processes higher in the auditory system (e.g., Sumner, Lopez-Poveda, O'Mard and Meddis 2002; Zhang, Heinz, Bruce and Carney 2001), or even to model behavioral data pertaining to auditory frequency selectivity (Lopez-Poveda and Meddis 2001; Irino and Patterson 2001). They also may serve as the basis for developing speech processors for auditory prostheses (Wilson, Brill, Cartee, Cox, Lawson, Schatzer and Wolford 2002).

Several algorithms of this sort have been proposed (Goldstein 1990; Irino and Patterson 2001; Zhang *et al*. 2001; Meddis, O'Mard and Lopez-Poveda 2001). Their success will finally depend on their ability to reproduce *accurately* as many features of the basilar-membrane (BM) response as possible, as well as on their computational *speed* and conceptual *simplicity*.

Meddis *et al*. (2001) presented one such algorithm termed dual-resonance nonlinear (DRNL) filter. Preceded by a middle-ear (ME) filter, the DRNL filter reproduced reasonably the measured response of point sites on the BM to single tones, clicks, and combination tones. However, it failed to model two prominent characteristics of the BM response. First, it did not reproduce the amplitude and phase plateaus observed at high levels for frequencies higher than the characteristic frequency (CF) of the measurement site (Ruggero, Rich, Recio, Narayan and Robles 1997). Instead, the gain of the DRNL filter decreased steeply above CF regardless of level. Second, the filter did not reproduce the long-lasting multiple-lobe aspect of the BM impulse response (IR) (Recio, Rich, Narayan and Ruggero 1998). Instead, the filter's IR showed two lobes only, consistent with its dual-resonance structure. Both these failures question the validity of the simple, dual-resonance architecture of the DRNL filter to model BM responses.

In search for an optimum algorithm, the present work investigates the reason for these failures and introduces improvements on the original DRNL filter to

overcome them. It is shown that the second of these two failures is a result of using an unrealistic middle-ear filter. Indeed, longer multiple-lobe IRs can be reproduced with the DRNL filter if empirical stapes IR waveforms are input directly to the filter. It is also shown, however, that a third, *all-pass* parallel filter path must be incorporated into the DRNL filter to model the high-frequency plateau.

The analysis below focuses on the ability of the improved filter to reproduce many aspects of the BM response to pure tones and clicks for a single chinchilla at a CF ~ 10 kHz (L113 of Ruggero *et al*. 1997 and Recio *et al*. 1998). The reason for this choice is that L113 is perhaps the case for which the most complete set of BM measurements has ever been reported.

2 The improved DRNL filter

The improved DRNL filter consists of the original dual-resonance nonlinear algorithm (described in full in Meddis *et al*., 2001) completed with a *third* parallel path that acts as a *linear*, *zero-phase*, *all-pass* filter. The output from the improved filter is the sum of the outputs from the original DRNL filter and the third path. The gain (scalar) of the third-path filter, *k*, is free to vary above zero, but is usually lower than the gain of the linear path, *g*, of the original DRNL filter. This makes the contribution of the third path to the total filter output become prominent only at high input levels. Furthermore, its contribution is most evident for frequencies higher than CF because in this region the output from both the linear and the nonlinear paths of the DRNL filter is highly attenuated (line "no-3rd" in Fig. 1B).

This third path allows modeling of the high-frequency amplitude and phase plateaus observed in BM tonal responses (Robles and Ruggero 2001). The idea of using a *zero-phase*, *all-pass* filter is based on a suggestion by Robles and Ruggero (p. 1313) that the plateaus "…reflect, more or less directly (…) stapes motion…"

3 Implementation and evaluation

The improved DRNL filter was implemented and evaluated *digitally* in the *time domain*. The part corresponding to the original DRNL filter was implemented as described by Lopez-Poveda and Meddis (2001, Appendix). The new zero-phase, all-pass filter was implemented digitally as suggested by Smith (2002).

A middle-ear stage was placed between the stimulus and the input to the filter. For any sound pressure (Pa) waveform, this stage produces stapes velocity (m/s), which is the assumed input to the DRNL filter. Special attention was paid to make sure that this ME stage preserves *all* aspects of the experimental stapes response. For this reason, it was implemented differently when evaluating the filter for tones and clicks. For pure tones, it was realized as an FIR filter whose coefficients match the empirical IR of the chinchilla stapes. For clicks, however, the measured IR of the stapes (in velocity units) was used directly as the input to the improved DRNL filter. Unfortunately, the stapes IR for chinchilla L113 was not available (Ruggero, pers. comm). Instead, we used the experimental waveform (Fig. 3) for a different

animal (CB063 of Rhode and Recio, 2000). This may explain some of the discrepancies between the experimental and the model responses discussed later.

The response of the improved DRNL filter was examined for stimuli identical to those used in the experiments. Its phase and amplitude responses for pure tones were obtained by fitting the output waveform to a sine waveform by means of a sine-fitting routine. The sampling rate was 6.25×10^4 Hz for tones and 2.5×10^5 Hz for clicks. All filters were implemented and evaluated in Matlab™ 6.5. The code is available from the authors on request.

4 Filter parameters

The parameters for the *improved* DRNL filter are given in Table I. The same set was used throughout this report. The optimization strategy differed from that used by Meddis *et al.* (2001) or Lopez-Poveda and Meddis (2001). They paid attention to reproduce the *amplitude* aspect of the BM response to pure tones only. However, their procedure is inadequate, as it does not set constraints on every parameter. For example, varying the order or the bandwidth of the gammatone filters in the nonlinear path may be equally valid to reproduce the tuning of the BM response at low levels, where the experimental data is usually scarce. However, both parameters have a different effect on the phase of the DRNL filter, as can be easily understood from the analytical description of the filter's response of Lopez-Poveda (submitted). As a result, the parameters in Table I were optimized considering both the *amplitude* and *phase* aspects of the experimental response simultaneously.

Table I. Parameters of the improved DRNL filter used throughout this report. The notation is identical as in Lopez-Poveda and Meddis (2001). (GT: Gammatone; LP: Lowpass)

Linear path		Nonlinear path		All-Pass path	
GT cascade	5	GT cascade	3	Gain k	1
LP cascade	7	LP cascade	4		
CF_{lin} (Hz)	9000	CF_{nl} (Hz)	10000		
BW_{lin} (Hz)	3500	BW_{nl} (Hz)	1800		
LP_{lin} (Hz)	8800	LP_{nl} (Hz)	10000		
Gain, g	89	Gain a	2900		
		Gain b[(m/s)$^{(1-c)}$]	0.04		
		Gain c	0.25		

5 Response to pure tones

Figure 1 compares the response of the improved DRNL filter against the experimental data for identical pure tone stimuli. Overall, the model (Fig. 1B) reproduces the sensitivity data (Fig. 1A) to a good approximation both qualitatively and quantitatively. The discrepancies are more evident at high levels (90 dB), and may be attributed to using a ME filter in the model that does not correspond to chinchilla L113.

Note that the improved DRNL filter now reproduces the plateau observed at high levels for frequencies higher than 13 kHz. It also reproduces the shift in best frequency (BF) from CF at low levels to around 0.6CF at high levels (the actual BF at high levels may be influenced by the presence of peaks in the stapes response). A closer look reveals, however, that in the model the shift occurs abruptly from the BF of its nonlinear path to the BF of its linear path, whereas the shift is more gradual in the data. Indeed, Recio *et al*. (1998, Fig. 12) suggested that the data may be better described if a third highly tuned, nonlinear filter were added between the two resonances already modeled by the DRNL filter (arrow in Fig. 1A).

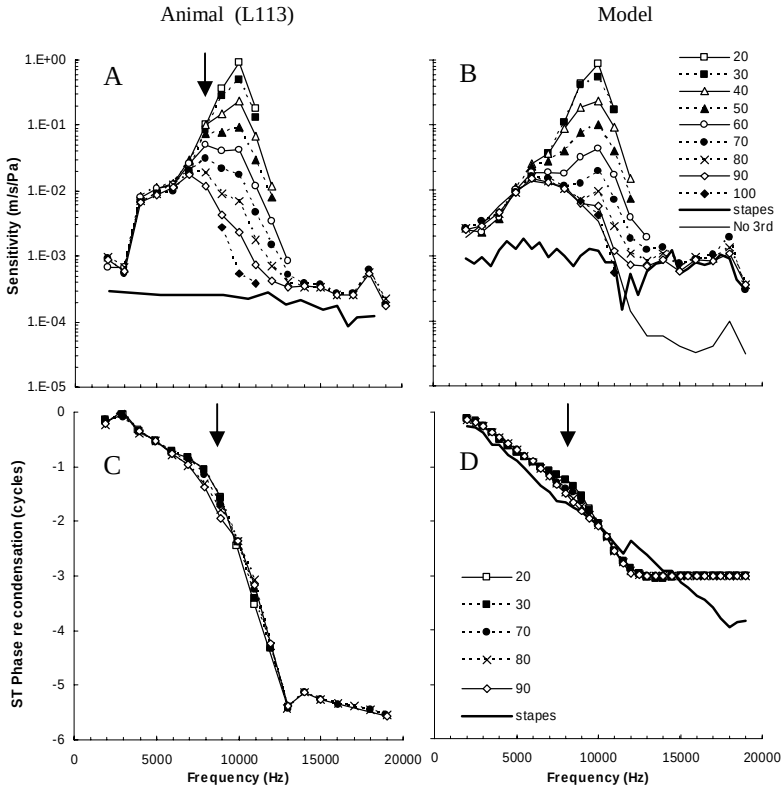


Fig. 1. Left panels: BM response for cochlea L113 (from Ruggero *et al*. 1997). (A) Sensitivity (m/s/Pa). (C) Phase (cycles) relative to input pressure (condensation). Right panels: model response. (B) Sensitivity (m/s/Pa). (D) Phase (cycles) relative to stapes phase. Different symbols (insets) illustrate different signal levels (dB SPL).

The bottom panels in Fig. 1 compare the *phase* responses. The agreement in shape is reasonable, including the presence of a high-frequency plateau. However, the phase lag in the plateau region is 2.5 cycles larger in the data. The reason may be that the data includes the lag introduced by the stapes, which is absent in the model results. Unfortunately, the phase of the stapes response for L113 was not

measured (Ruggero, pers. comm.), and the lag introduced by the ME stage in our model (shown in Fig. 2D) is unusually large compared with the values measured in Ruggero's laboratory (Temchin, Robles and Ruggero 2001).

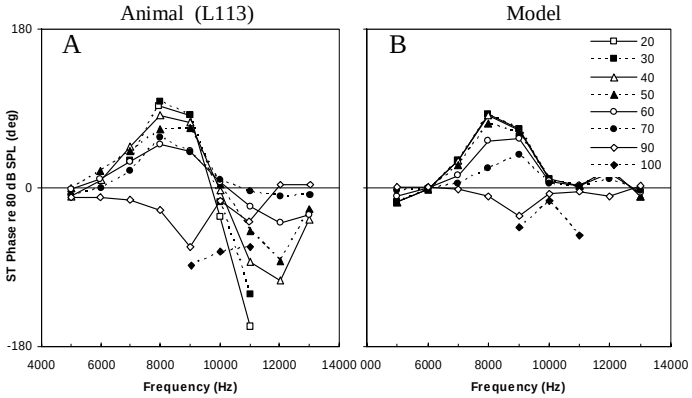


Fig. 2. Phase relative to 80 dB SPL. (A) Animal data (L113 from Ruggero *et al*. 1997). (B) Model response.

Figure 2 reproduces the phase data of Fig. 1 but relative to 80 dB SPL. Therefore, the contribution of the stapes to the BM phase, which is linear, is cancelled out in this representation. The model reproduces the systematic phase lead with level for frequencies lower than CF. It also reproduces the phase lag observed at high levels (≥ 90 dB) across frequencies. However, the model does *not* reproduce the phase lag at low levels for frequencies higher than CF. This failure could be corrected if different parameters were used (not shown), but then the fit to the sensitivity data in Fig. 1 would deteriorate slightly.

6 Response to clicks

Figure 3 compares the IR for cochlea L113 (Recio *et al*. 1998) and the response of the improved DRNL filter when the input is an empirical stapes IR (see Sec. 3). Note that the filter's response shows more than two lobes (as would be expected from a dual-resonance architecture) and its duration is comparable to the experimental IRs. This is a consequence of using a realistic stapes IR as input to the DRNL filter. It was not observed when the middle ear was modeled with a bandpass filter (compare Fig. 3B below with Fig. 7B of Meddis *et al*., 2001).

As shown by Recio *et al*. (1998), the BM responses to clicks in Fig. 3A are frequency modulated (Fig. 4A). Recio *et al*. wrote (their p. 1976) that the "…instantaneous frequency was influenced by level but its time trajectory retained its main features even at the highest level…" *and* post-mortem. The improved DRNL filter reproduces this behavior to a reasonable approximation (Fig. 4B), but only when a realistic stapes response is used as input. When the ME stage is removed ("no-ME" in Fig. 4), or when it is modeled by a simple Butterworth bandpass filter (not shown), the trajectory of the instantaneous frequency of the model differs considerably from the data. If the ME largely determined the trajectory, it would explain its invariance post-mortem. For the trajectory to settle

down at approximately the same frequency for all levels, it is also very important that the center frequencies of the gammatone filters in the linear and nonlinear paths of the DRNL filter are relatively close together (compare CF_{lin} and CF_{nl} in Table I). The shift in BF with level (illustrated in Fig. 1) may still be modeled by setting the cut-off frequency of the lowpass filter in the linear path (LP_{lin}) lower than CF_{lin}.

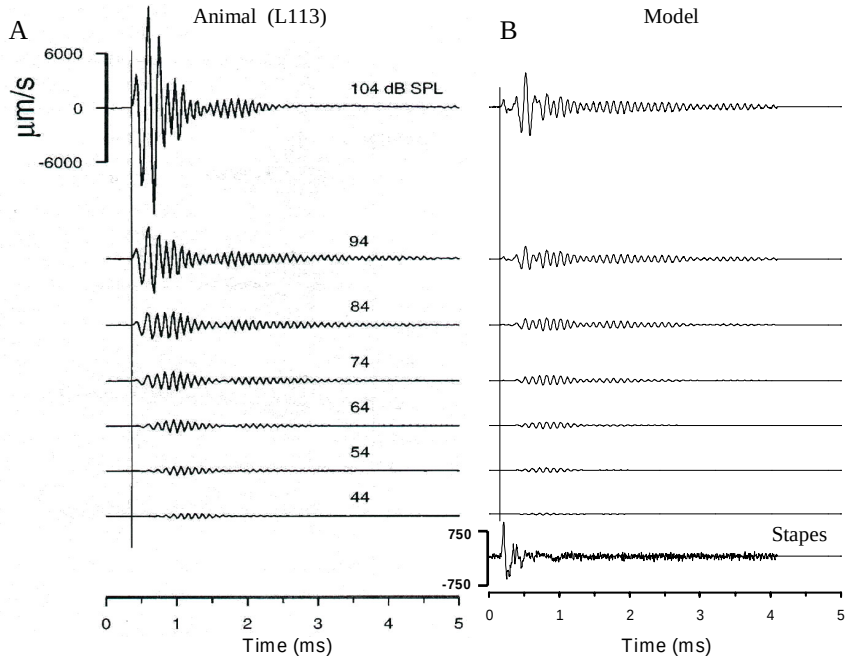


Fig. 3. (A) BM velocity click response for case L113 (from Recio *et al.*, 1998). (B) Response of the improved DRNL filter to an empirical stapes IR (bottom waveform).

7 Conclusions

The *improved* DRNL filter reproduces the characteristic high-frequency plateaus. The fits to the data improve when its parameters are optimized bearing in mind the amplitude and phase aspects of the response to tones simultaneously. The middle-ear stage contributes significantly to improving the fits, particularly at high levels, and is essential to explain and to model the waveforms and the instantaneous frequency of the BM response to clicks, both *in vivo* and *post mortem.*

Acknowledgments

Thanks to Mario Ruggero and Alberto Recio for their suggestions and for providing some of the experimental data. Work supported by FIS PI020343 and G03/203.

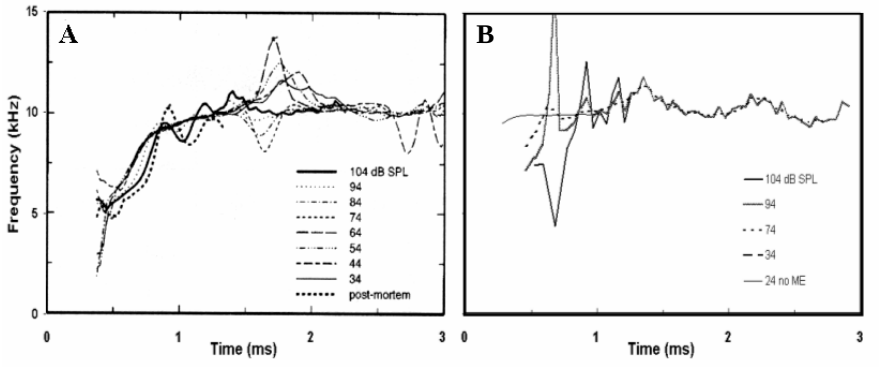


Fig. 4. Instantaneous frequency for different levels (insets). (A) Data for cochlea L113 (from Recio *et al*. 1998). (B) Results for the improved DRNL filter.

References

Goldstein, J. L. (1990) Modeling rapid wave form compression on the basilar membrane as multiple-band-pass-nonlinearity filtering. Hear. Res. 49, 39-60.

Irino, T., and Patterson, R.D. (2001) A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. 109, 2008-2022.

Lopez-Poveda, E.A. and Meddis R. (2001) A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. 110, 3107-3118.

Lopez-Poveda, E.A. (submitted). An approximate transfer function for the dual-resonance nonlinear filter model of auditory frequency selectivity. J. Acoust. Soc. Am.

Meddis, R. O'Mard, L. and Lopez-Poveda, E.A. (2001) A computational algorithm for computing nonlinear auditory frequency selectivity. J. Acoust. Soc. Am. 109, 2852-2861.

Recio, A. Rich, N.C. Narayan, and S. Ruggero, M.A. (1998) Basilar-membrane responses to clicks at the base of the chinchilla cochlea. J. Acoust. Soc. Am. 103, 1972-1989.

Rhode, W.S. and Recio, A. (2000) Study of mechanical motions in the basal region of the chinchilla cochlea. J. Acoust. Soc. Am. 107: 3317-3332.

Robles, L. Ruggero, M.A. (2001) Mechanics of the mammalian cochlea. Physiol. Rev. 81, 1305-1352.

Ruggero, M.A. Rich, N.C. Recio, A. Narayan, S. and Robles L. (1997) Basilar-membrane responses to tones at the base of chinchilla cochlea. J. Acoust. Soc. Am. 101: 2151-2163.

Smith, J. O. (2003). *Introduction to Digital Filters*, Stanford University. Web published at http://www-ccrma.stanford.edu/~jos/filters/.

Sumner, C.J., Lopez-Poveda, E.A., O'Mard, L.P., and Meddis, R. (2002) A revised model of the inner-hair cell and the auditory-nerve complex. J. Acoust. Soc. Am. 111, 2178-2188.

Temchin, A.N. Robles, L. and Ruggero, M.A. (2001) A re-examination of middle-ear transmission in chinchilla. Poster #586. Meeting of the Assoc. Res. Otolaryng.

Wilson B.S., Brill, S.M., Cartee, L.A., Cox, J.H., Lawson, D.T., Schatzer, R. and Wolford, R.D. (2002) Speech processors for auditory prostheses. Final Report. NIH project N01-DC-8-2105.

Zhang, X., Heinz, M.G., Bruce, I.C., and Carney, L.H. (2001) A phenomenological model for the responses of auditory nerve fibers. I. Non-linear tuning with compression and suppression. J. Acoust. Soc. Am. 109, 648-670.

Comparison of the compressive-gammachirp and double-roex auditory filters

Roy D. Patterson[1], Masashi Unoki[2], and Toshio Irino[3]

[1] Centre for the Neural Basis of Hearing, Physiology Dept. Cambridge University, Cambridge, U.K. rdp1@cam.ac.uk
[2] School of Information Science, Japan Advanced Institute of Science and Technology, Tatsunokuchi, Nomi, Ishikawa, 923-1292 Japan. unoki@jaist.ac.jp
[3] Faculty of Systems Engineering, Wakayama University, Wakayama, Japan. irino@sys.wakayama-u.ac.jp

1 Introduction

The comparison of psychophysical and physiological measures of frequency selectivity is often described either in terms of the rounded exponential, or roex, auditory filter, which was introduced to explain tone-in-noise masking in humans, or the gammatone auditory filter, which was introduced to characterize the reverse-correlation, or 'revcor' data in cats. Recently, a descendent of the gammatone filter, referred to as the gammachirp (GC) auditory filter, has been developed to describe both the level-dependent physiological data from small mammals and the level-dependent notched-noise masking data from humans. The architecture of the compressive GC filter is intended to reflect recent developments in cochlear physiology, and in particular, how the tip and tail components of the cochlear filter interact as a function of stimulus level. In this paper, we describe a double roex filter system with comparable functionality to the GC filter. It has separate tip and tail filters that interact to produce the appropriate level-dependent gain in the passband of the filter. Both the double roex filter and the compressive GC were fitted to two large sets of simultaneous masking data to provide a basis for comparing the filter systems in detail.

2 The compressive gammachirp auditory filter

Irino and Patterson (1997, 2001) developed the gammachirp filter as an asymmetric level-dependent version of the gammatone filter with the advantages of both the gammatone and roex filters. They showed that an 'analytic' form of the GC filter could explain level-dependent tone-in-noise masking data (e.g., Rosen and Baker 1994), and that a 'compressive' form of the GC filter could explain the frequency

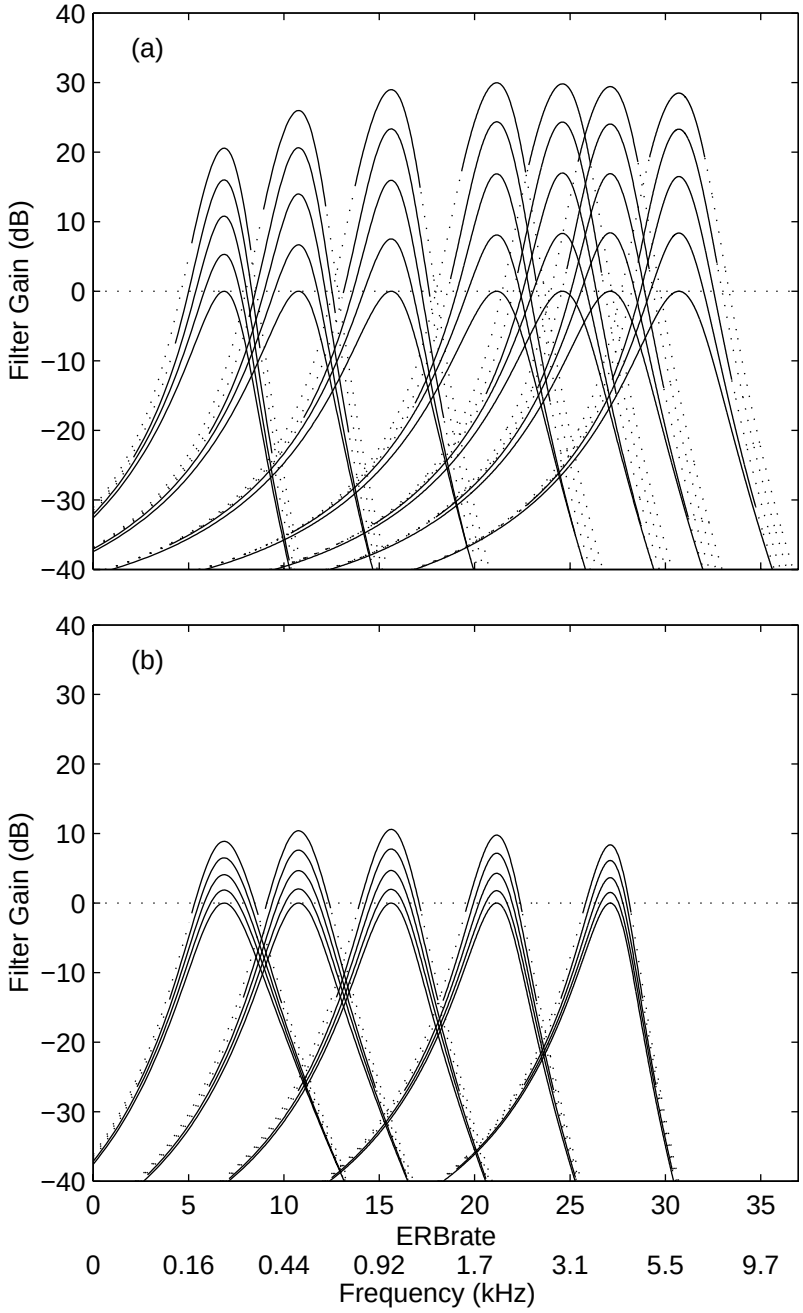


Fig. 1. Compressive gammachirp filters for the threshold data of (a) Baker *et al.* (1998) with probe frequencies 0.25, 0.5, 1.0, 2.0 3.0, 4.0 and 6.0 kHz, and (b) Glasberg *et al.* (2000) with probe frequencies 0.25, 0.5, 1.0, 2.0 and 4.0 kHz.

glide, or chirp, observed at the onset of the impulse response of the cat's cochlear filter (de Boer and Nuttall 2000). Recently, Patterson, Unoki and Irino (2003) fitted the compressive GC filter to the data of Baker, Rosen and Darling (1998) and Glasberg and Moore (2000), both of whom measured threshold for a probe tone in the presence of an asymmetric notched noise over, what is effectively, the entire range of frequencies and levels encountered in everyday hearing. A version of the PolyFit procedure (Baker *et al,* 1998) was used to fit the data for all probe frequencies simultaneously. The result for the data of Baker *et al.* (1998) was the seven families of compressive GC filters shown in Fig. 1a. The five functions in each family show filters at probe levels from 30 to 70 dB in 10-dB steps. The families of filters all have the same general form; the bandwidth is roughly uniform across frequency on this ERBrate scale (Glasberg and Moore 1990), and the gain of the filter *decreases* monotonically with increasing stimulus level in the passband of the filter, as would be expected. The range of the gain, or compression, is a little under 30 dB at the lower probe frequencies (0.25, 0.5 and 1.0 kHz) and a little over 30 dB at the higher probe frequencies (2.0, 3.0, 4.0 and 6.0 kHz). When the compressive GC was fitted to all five probe frequencies in the data of Glasberg *et al.* (2000), the result was the five families of compressive GC filters shown in Fig. 1b. The families of filters have the same general form; the bandwidth is roughly uniform across frequency and the gain of the filter *decreases* monotonically with increasing stimulus level. However, the two data sets produce quite different estimates of the asymmetry and maximum gain in the system. The filters in Fig. 1b are more symmetric and the maximum gain is limited to about 10 dB. Moreover, the maximum gain is slightly less at the higher probe frequencies.

3 The double-roex auditory filter

The term 'roex auditory filter' actually describes a family of ***ro***unded-***ex***ponential filters developed to explain the masking of a sinusoidal probe by a notched noise (Patterson 1976). When the noise bands are close to the signal, threshold is dominated by masker components in the passband of the auditory filter and the data can be explained with a simple, one-parameter roex; that is, a filter composed of a pair of back-to-back exponentials, $\exp(-pg)$, multiplied by a term, $(1+pg)$ that rounds the peak and makes the filter continuous at its center frequency (Patterson, Nimmo-Smith, Weber and Milroy 1982); g is a normalized frequency variable that describes the distance in frequency from the probe frequency f_0 to the edge of the noise in terms of the center frequency, so $g = |f - f_0| / f_0$. The parameter, p, describes the rate of decay of the exponentials that form the ***p***assband of this roex(***p***) filter. When the notch is very wide, the roex(*p*) filter underestimates threshold badly, indicating that the tails of the auditory filter at frequencies remote from the center are much shallower. So, the model was extended to include a second roex to represent the ***t***ails of the auditory filter. The two filters operate in parallel and their relative ***w***eighting is ***w***.

In experiments like those performed by Baker *et al*. (1998) and Glasberg and Moore (2000), the noise bands are positioned both symmetrically and asymmetrically about the signal frequency, and the level of the masker is varied.

Such experiments reveal that the auditory filter becomes progressively more asymmetric and the tip filter becomes less prominent as stimulus level increases. These effects can be accommodated by allowing p and t to have different values in the ***l***ower and ***u***pper halves of the filter, and allowing w to vary with level in which case there are five filter parameters, t_l, t_u, w, p_l, p_u. The centre frequency of the tip filter is also allowed to shift relative to the centre frequency of the tail filter. If the normalized frequency variables of the tail and tip filters are g_1 and g_2, and f_{rat} is the ratio of the two centre frequencies, then $g_1 = |f - f_0| / f_0$, $g_2 = |f - f_{\text{rat}} \cdot f_0| / f_0$, the tail and tip filters are

$$\begin{aligned} W_{\text{tail}}(g_1) &= \begin{cases} (1 + t_l g_1)\exp(-t_l g_1) & f < f_0 \\ (1 + t_u g_1)\exp(-t_u g_1) & f \geq f_0 \end{cases} \\ W_{\text{tip}}(g_2) &= \begin{cases} (1 + p_l g_2)\exp(-p_l g_2) & f < f_{\text{rat}} \cdot f_0 \\ (1 + p_u g_2)\exp(-p_u g_2) & f \geq f_{\text{rat}} \cdot f_0 \end{cases} \end{aligned} \tag{1}$$

and, the frequency response of the double roex auditory filter is

$$W(f) = W_{\text{tail}}(f) + w \cdot W_{\text{tip}}(f). \tag{2}$$

This double roex auditory filter was fitted to the data of Baker *et al.* (1998) and Glasberg *et al.* (2000) using the PolyFit procedure (Baker *et al.* 1998) which enabled us to fit all probe frequencies simultaneously. The result for the Baker *et al.* (1998) study is the seven families of double roex filters shown in Fig. 2a. The five functions in each family show filters at probe levels from 30 to 70 dB in 10-dB steps. The families of filters all have the same general form; the bandwidth is roughly uniform across frequency and the gain of the filter *decreases* monotonically with increasing stimulus level in the passband of the filter. The maximum gain is around 20 dB. Both the bandwidth and the maximum gain are less than estimated with the compressive GC; the asymmetry is similar for the two different filters.

Figure 2b shows the tail filters and sets of tip filters that together produced the families of double roex filters in Fig. 2a. Each of the double roex filters in the family associated with one probe frequency is produced by combining the fixed tail filter for that probe frequency with the appropriate tip filter, weighted by the factor w. The weight, or gain, is largest at low stimulus levels and decreases as stimulus level increases. The low-frequency side of the tail filter is much shallower than the high-frequency tail, as on the passive basilar membrane observed at high stimulus levels. However, low-frequency side becomes shallower, and the filter more asymmetric, as probe frequency increases. The five tip filters shown are for levels from 30 to 70 dB in 10 dB steps. The low-frequency slope of the tip filter becomes slightly steeper as probe frequency increases, and as a result, the passband of the double roex filter becomes more symmetric as probe frequency increases.

The simultaneous fit for the Glasberg *et al.* (2000) data is shown by the five families of double roex filters in Fig. 3a. The families of filters all have the same general form and the gain of the filter *decreases* monotonically with increasing stimulus level in the passband of the filter. However, the gain provided by the tip filter is greatest at 1.0 kHz and decreases at both lower and higher probe frequencies. The tail filters in Fig. 3b have a uniform shape across frequency and

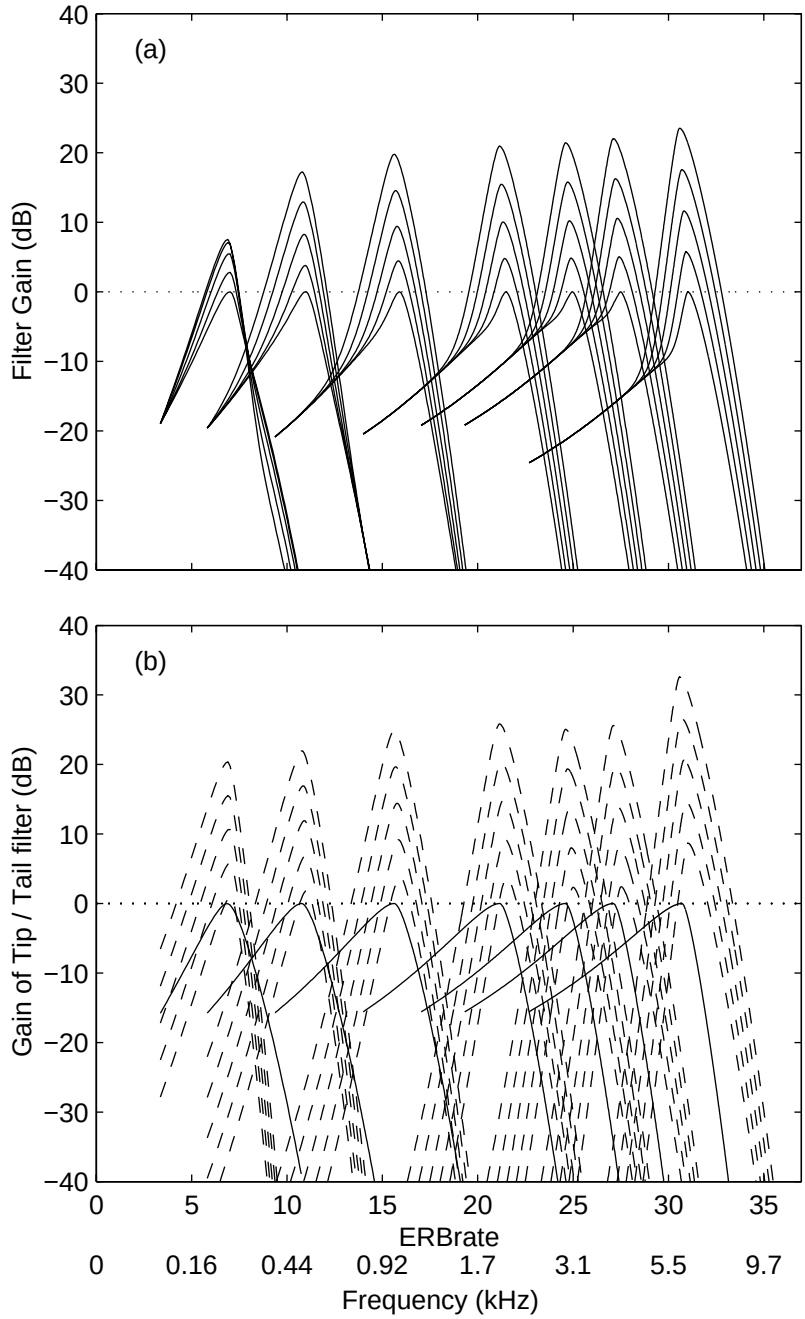


Fig. 2. Families of (a) double roex filters for the data of Baker *et al.* (1998), with (b) the corresponding tail filters (solid line) and sets of tip filters (dashed lines).

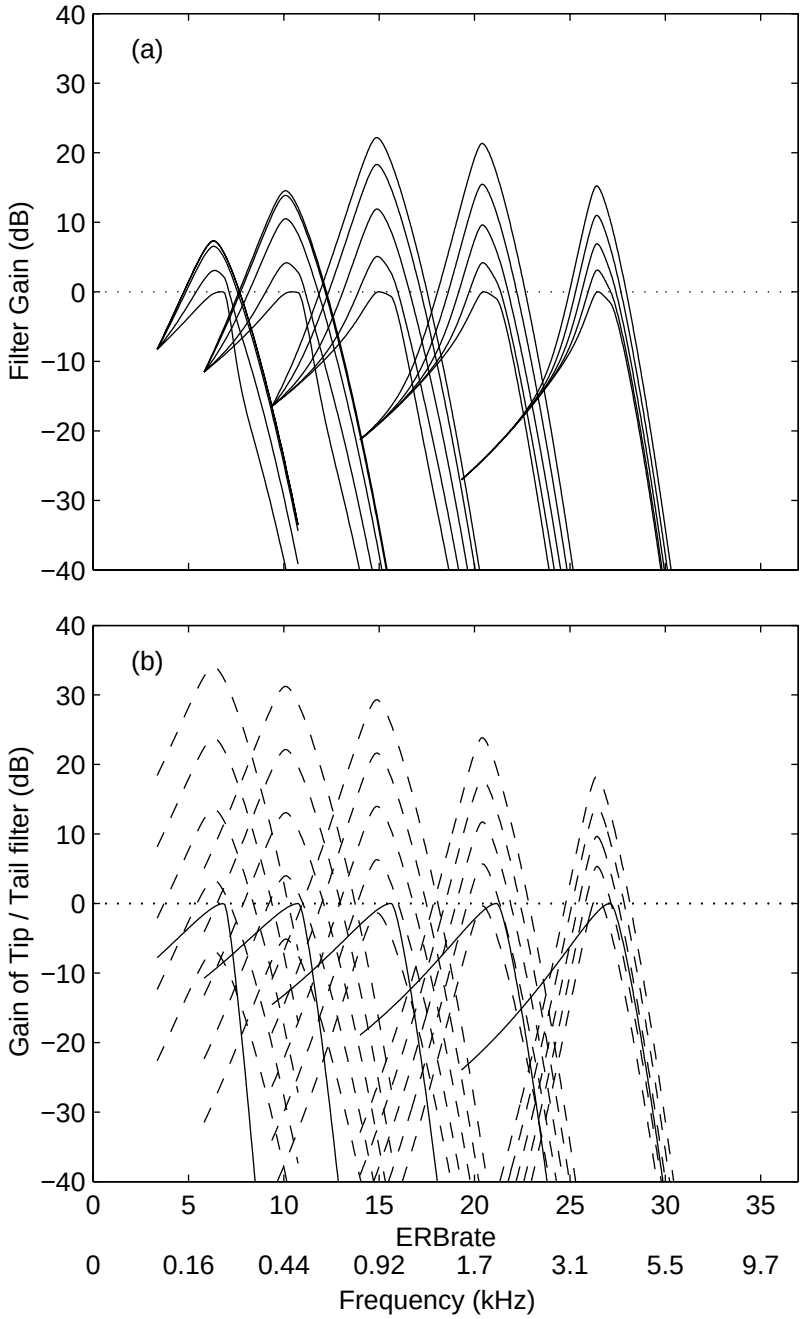


Fig. 3. Families of (a) double roex filters for the data of Glasberg *et al.* (2000), with (b) the corresponding tail filters (solid line) and tip filters (dashed lines).

are reasonably asymmetric. The tip filters are much broader at the lower probe frequencies than those in Fig 2b, and the maximum gain of the tip filter is much greater at the lower probe frequencies than for the Baker *et al*. data (Fig. 2b).

In summary, even when the PolyFit procedure is used to fit all probe frequencies simultaneously, the estimates of filter bandwidth, asymmetry and maximum gain vary considerably from study to study and from the double roex filter to the compressive GC auditory filter. The fitting process includes two additional parameters, for the efficiency constant, K, and the threshold asymptote observed with wide notches at low levels. These parameters reduce the gain of the filters at the lowest and highest probe frequencies in Figs 2a and 3a, relative to what might have been expected from the corresponding tip and tail filters in Figs 2b and 3b, respectively. However, this does not explain the variability of bandwidth, asymmetry and maximum gain across studies and filter types.

Acknowledgments

Brian Glasberg and Stuart Rosen kindly provided convenient files with the threshold data from Glasberg and Moore (2000) and Baker *et al.* (1998). The research was supported by the UK Medical Research Council (G9900369, G9901257) and a Special Coordination Fund for Promoting Science and Technology of young researchers with fixed-term appointments from the Japanese Ministry of Education.

References

Baker, R. J., Rosen, S. & Darling, A. M. (1998) An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable. In: *Psychophysical and physiological advances in hearing: Proceedings of the 11th International Symposium on Hearing.* A. Palmer, A. Rees, Q. Summerfield and R. Meddis (Eds). Whurr, London, 81-88.

de Boer, E. and Nuttall, A. L. (2000). The mechanical waveform of the basilar membrane. III. Intensity effects. J. Acoust. Soc. Am., **10**7, 1497-1507.

Glasberg, B. R. and Moore, B. C. J. (1990) Derivation of auditory filter shapes from notched-noise data. Hear. Res. **47**, 103-138.

Glasberg, B. R. and Moore, B. C. J. (2000) Frequency selectivity as a function of level and frequency measured with uniformly exciting noise. J. Acoust. Soc. Am., **108**, 2318-2328.

Irino, T. and Patterson, R.D. (1997) A time-domain, level-dependent auditory filter: The gammachirp. J. Acoust. Soc. Am. **101**, 412-419.

Irino, T., and Patterson, R. D. (2001) A compressive gammachirp auditory filter for both physiological and psychophysical data. J. Acoust. Soc. Am. **109**, 2008-2022.

Patterson, R.D. (1976) Auditory filter shapes derived with noise stimuli. J. Acoust. Soc. Am. **59**, 640-654.

Patterson, R.D., Nimmo-Smith, I., Weber, D. L., and Milroy, R., (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788-1803.

Patterson, R.D., Unoki, M. and Irino, T. (2003) Extending the domain of center frequencies for the compressive gammachirp auditory filter. J. Acoust. Soc. Am. (submitted).

Rosen S. and Baker, R.J. (1994) Characterizing auditory filter nonlinearity. Hear. Res. **73**, 231-243.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Reaction-time data support the existence of Softness Imperception in cochlear hearing loss

Mary Florentine[1,2], Søren Buus[1,3], and Mindy Rosenberg[1,2]

[1] Institute for Hearing, Speech, and Language
[2] SLPA Dept. (106A FR), Northeastern University, Boston, MA, U.S.A., florentin@neu.edu
[3] ECE Dept. (440 DA), Northeastern University, Boston, MA, U.S.A., buus@neu.edu

1 Introduction

The purpose of this paper is to test the hypothesis that listeners with hearing losses of primarily cochlear origin have Softness Imperception, SI. The concept of SI arose from recent modeling of loudness summation in listeners with cochlear hearing losses (Buus 1999; Buus and Florentine 2001). These studies found that loudness grows at similar rates between threshold and approximately 15 dB SL, whether the threshold is normal or is elevated by cochlear hearing loss. The modeling also indicated that loudness at threshold is larger than normal when threshold is elevated. This result led us to hypothesize that listeners with cochlear hearing losses have SI (Florentine and Buus 2002). In other words, they are unable to hear some low loudnesses that are audible to normal listeners.

The surprising findings that listeners with cochlear hearing losses have greater-than-normal loudness at threshold and rates of growth of loudness that are within the normal range overturn 60 years of thinking in the field. Although the new concept of softness imperception contradicts the classical notion of recruitment, it provides a coherent view of loudness perception in cochlear hearing losses and agrees with loudness functions measured by magnitude estimation and cross-modality matching in over 100 listeners with cochlear hearing losses (Hellman and Meiselman 1990). It also agrees with modern findings in auditory physiology (Heinz, Sachs, and Young 2003).

Although some investigators have suggested that reaction time does not assess loudness (e.g., Kohfeld, Santee, and Wallace 1981) and some have reported ambiguous results (e.g., Humes and Ahlstrom 1984), the majority of studies indicate that simple reaction time, RT, may be used as an indirect measure of loudness for tones (e.g., Chocholle 1940; Pfingst, Hienz, Kimm, and Miller 1975; Marshall and Brandt 1980; Buus, Greenbaum, and Scharf 1982). Because RT varies inversely with loudness, our SI hypothesis implies that RTs to tones at threshold should be faster for listeners with cochlear hearing losses than for normal listeners. In fact, clinical group data indicate that this is true. Hustert, Kumpf, and Stoll

(1996) found shorter RTs in groups of patients with hearing losses than in a group of normal listeners. Unfortunately, their data vary greatly within groups and thresholds were determined by a clinical method, which is sensitive to the listeners' response criterion. Therefore, Hustert et al.'s results need to be confirmed in a careful laboratory study. The present experiment was designed to address this need by assessing the effect of hearing loss on RTs to tones presented at and near carefully measured forced-choice thresholds.

2 Method

2.1 Experimental design

Because RTs can vary considerably among listeners and may be affected by age (Burke, Creston, and Shutts 1965), the experiment was designed so that each listener could serve as his or her own control by testing two frequencies with different amounts of hearing loss within each listener. One test frequency was chosen to have normal or near-normal threshold and the other to have a substantially elevated threshold. To determine whether the resulting within-listener, across-frequency comparison was unduly influenced by a simple effect of frequency not related to hearing loss, a control group of normal listeners was tested at several frequencies.

2.2 Listeners

Seven listeners with sloping hearing losses of primarily cochlear origin were tested. They ranged in age from 49 to 76 years. Data from one of these listeners was excluded because of inconsistent responses even to stimuli well above threshold. In addition, two normal listeners served as controls. They were 42 and 73 years old. Our diagnostic criteria were the same as those previously published (e.g., Buus, Müsch, and Florentine 1998; Buus and Florentine 2001).

2.3 Stimuli

Listeners were presented tones with an equivalent rectangular duration of 200 ms and rise and fall times of 6.67 ms. Levels ranged from 0 dB SL to 100 dB SPL. Two test frequencies were chosen individually for each impaired listener as explained above. In addition, normal listeners were tested at 0.5, 1, and 4 kHz.

2.4 Procedure

The procedure consisted of two parts. In the first part, absolute thresholds were carefully measured for each listener using a two-interval, two-alternative forced choice (2I, 2AFC) paradigm and a three-down one-up adaptive method. (For further information regarding the procedure, see Florentine, Buus, and Poulsen 1996).

In the second part, RTs were obtained by asking the listeners to press a telegraph key as soon as they heard a tone. The RT was defined as the time elapsed

between the onset of the stimulus and the listener's response. Each trial started by flashing an LED, which marked the beginning of a random-duration foreperiod that was inserted between the response and the presentation of the next tone. The duration of the foreperiod was the sum of a fixed one-second interval plus an exponentially distributed random variable with a mean of one second. RTs were retained for analysis only if the response occurred between 125 ms and 4 s after the onset of the stimulus. Responses occurring during the first 125 ms of the stimulus were counted as false positives. If no response occurred within 4 s, the trial was counted as a miss and the LED was flashed once to indicate to the listener that the stimulus was missed and that a new trial was starting.

Within each block of trials, 5- and 200-ms tones at a single frequency were presented with levels in random order (without replacement) for a total of five presentations at each level and duration. (Only data for the 200-ms tones will be presented here.) At least four blocks of trials were obtained at each frequency. Missed trials were repeated twice at most. Each repeat presentation was preceded by a dummy trial at the next higher level in an attempt to cue the listener to the signal.

2.5 Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli through a 16-bit D/A converter (TDT DD1) with a 41.67-kHz sample rate, recorded the listeners' responses, and executed the experimental procedure. The output of the D/A was attenuated (TDT PA4), lowpass filtered (TDT FT5, f_c = 20 kHz, 135 dB/octave), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of the Sony MDR-V6 headset. The listeners were seated in a sound-attenuating booth with the response box, telegraph key, and the headphones; all other equipment was positioned outside the booth.

3 Results

Median RTs from the two age-matched normal listeners are shown in separate panels in Fig. 1. For each test frequency, the RTs decrease as the sensation level of the tone increases from threshold to 65 dB SL. The range of RTs is consistent with other data in the literature for normal listeners. Although the RTs for one frequency may be slightly faster or slower than the RTs for other frequencies within a single listener, the effect of frequency is not consistent across listeners. Even within listeners any effects of frequency that do exist are small. Additional measurements in young, normal listeners yielded similar results. Altogether, these data indicate no consistent effect of frequency on RTs in normal listeners.

Median RTs from each of the six hearing-impaired (HI) listeners are shown in Fig. 2. All six impaired listeners show faster RTs at their elevated thresholds than at their normal or near-normal thresholds. Above threshold, the RTs decrease more or less monotonically and remain faster for the frequency with elevated threshold than for the frequency with normal or near-normal threshold for most listeners. The exception is HI-6 whose RTs for the frequency with elevated threshold decrease

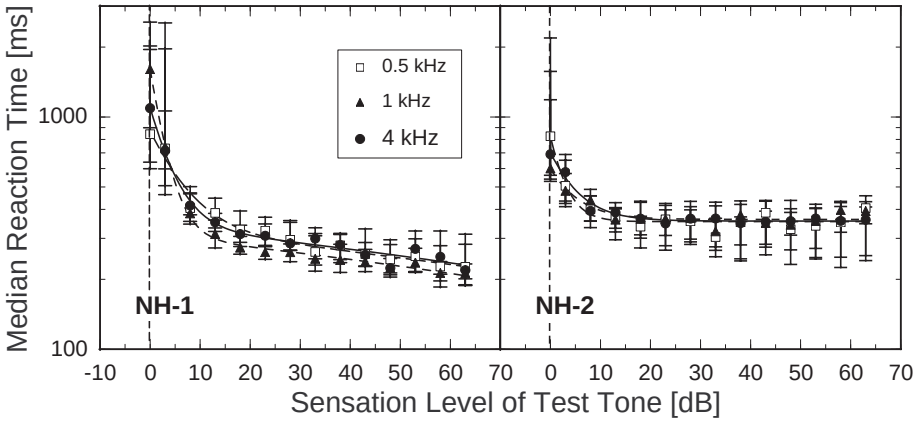


Fig. 1. Each panel shows data for one normal-hearing (NH) listener. Median RTs to tones at three frequencies are plotted as a function of level. The vertical bars show the inter-quartile ranges.

with SL only up to about 10 or 15 dB SL, above which they increase such that the RT is approximately equal to that for the frequency with relatively mild hearing loss at 18 dB SL. A similar increase in RTs is also apparent for HI-6 at 4 kHz for levels above 50 dB SL. In other words, the RTs for HI-6 appear to increase when the tone exceeds approximately 90 dB SPL. Whatever the source of this non-monotonicity may be, it is not important to the answer of the question posed in the present paper. The data for all the impaired listeners show that the RTs to tones at and near threshold are faster at frequencies with substantial hearing losses than at frequencies with normal or near-normal hearing.

4 Discussion

Because the normal listeners in the present study did not show a consistent difference in RTs across frequencies, any effects of frequency obtained in the listeners with cochlear hearing losses can be attributed to their hearing losses. All six impaired listeners showed faster RTs at their elevated thresholds than at their normal or near-normal thresholds. These data are consistent with the idea that loudness at threshold is not constant but increases when threshold is elevated by cochlear hearing losses. This finding supports recent work (Buus and Florentine 2001) and indicates that there is a range of low loudnesses that can be heard by normal listeners but not by listeners with hearing losses. In other words, listeners with cochlear hearing losses have SI.

Although all six of the impaired listeners in the present study appear to have SI, a note of caution should be made here. Listeners with cochlear hearing losses can differ greatly in their abilities to process auditory stimuli at supra-threshold levels. For example, two listeners with matched audiograms can differ in their abilities to detect a pause in a noise (e.g., Florentine and Buus 1984). Likewise, impaired

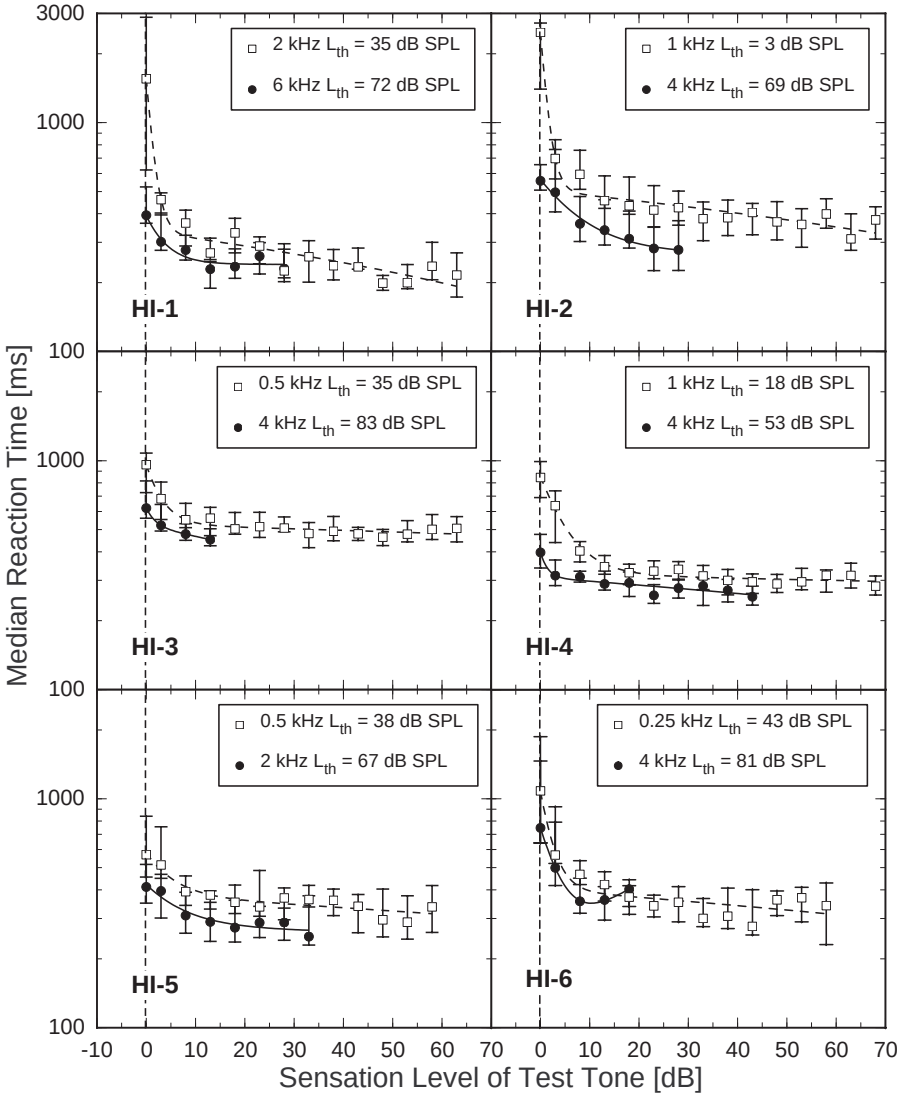


Fig. 2. Each panel shows data for one hearing-impaired (HI) listener. Median RTs to tones are plotted as a function of level. The test frequencies and thresholds of the tones for each listener are shown in the legends. The vertical bars show the inter-quartile ranges.

listeners vary in their ability to hear one sound in the presence of others (e.g. Florentine 1992; Moore 1995). Given such variation, it is likely that not every person with cochlear hearing loss may experience SI. The present work will continue to uncover diagnostic classifications in which impaired listeners experience SI.

Variation among individuals aside, the present data indicate that loudness at threshold generally is greater than normal when threshold is elevated by cochlear hearing losses. There are at least two possible explanations for this finding. One is that hearing loss may increase internal noise in the loudness domain. If so, a larger-than-normal loudness will be required to attain the d' corresponding to threshold (e.g., d'=1). Another possibility follows from the idea that excitation in a single frequency-selective auditory channel is coded by a combination of fibers working together to encompass the entire dynamic range (e.g., Delgutte 1987; Viemeister 1988). According to this view, loudness may be thought to depend on a central response derived from weighted sum of spikes in the auditory nerve, where the weights are high for the low-SR fibers that are effective in coding high stimulus levels and low for the high-SR fibers that are effective in coding low stimulus levels. Whereas high-SR fibers probably mediate detection of tones in quiet for normal listeners, it is possible that low- or medium-SR fibers mediate the threshold response in cochlear losses. This might occur if hearing loss causes high-SR fibers that normally would be most sensitive to become unresponsive or if hearing loss lowers the spontaneous rate of some high-SR fibers. It may seem paradoxical that loudness at threshold should increase in the latter case. The straightforward interpretation of a lowered spontaneous rate is that the internal noise is reduced, but such an interpretation presumes that the weights are static. If the weight for each fiber is set dynamically according to its activity when the stimulus is absent, the weight would increase if the spontaneous rate is decreased by hearing loss, which could cause the loudness of a just-detectable response to increase.

5 Conclusions

Whereas the physiological explanation for softness imperception remains unclear, some conclusions are quite clear. The RT measurements indicate that loudness at threshold should not be assumed to be constant. It can vary across listeners and frequency. Certainly, the present data are consistent with the idea that loudness at threshold increases when threshold is elevated by cochlear hearing losses, at least in most cases. This finding indicates that there is a range of low loudnesses that can be heard by normal listeners but not by listeners with hearing losses. In other words, listeners with cochlear hearing losses often have softness imperception and reduced dynamic ranges both in terms of SPL and in terms of loudness.

Acknowledgments

The authors would like to thank Candice Costa and Sandra Cleveland for providing audiological testing and Joe McCormack for providing technical support. This work was supported by NIH/NIDCD Grant No. R01DC02241.

References

Burke, K.S., Creston, J.E. and Shutts, R.E. (1965) Hearing loss and reaction time. Arch. Otolaryngol. 81, 49-56.

Buus, S. (1999) Loudness functions derived from measurements of temporal and spectral integration of loudness. In: A. N. Rasmussen, P. A. Osterhammel, T. Andersen and T. Poulsen (Ed.) *Auditory Models and Non-linear Hearing Instruments*, GN ReSound, Taastrup, Denmark, pp. 135-188.

Buus, S. and Florentine, M. (2001) Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. J. Assoc. Res. Otolaryngol. 3, 120-139.

Buus, S., Greenbaum, H. and Scharf, B. (1982) Measurements of equal loudness and reaction times. J Acoust Soc Am 72 Suppl. 1, S94.

Buus, S., Müsch, H. and Florentine, M. (1998) On loudness at threshold. J. Acoust. Soc. Am. 104, 399-410.

Chocholle, R. (1940) Variation des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. L'Année Psychologique 41, 65-124.

Delgutte, B. (1987) Peripheral auditory processing of speech information: Implications from a physiological study of intensity discrimination. In: M. E. H. Schouten (Ed.) *The Psychophysics of Speech Perception*, Nijhoff, Dordrecht, The Netherlands, pp. 333-353.

Florentine, M. (1992) Effects of cochlear impairment and Equivalent-Threshold Masking on psychoacoustic tuning curves. Audiology 31, 241-253.

Florentine, M. and Buus, S. (1984) Temporal gap detection in sensorineural and simulated hearing impairment. J. Speech Hear. Res. 27, 449-455.

Florentine, M. and Buus, S. (2002) Evidence for normal loudness growth near threshold in cochlear hearing loss. In: L. Tranebjærg, J. Christensen-Dalsgaard, T. Andersen and T. Poulsen (Ed.) *Genetics and the Function of the Auditory System*, GN ReSound, Tåstrup, Denmark, pp. 411-426.

Florentine, M., Buus, S. and Poulsen, T. (1996) Temporal integration of loudness as a function of level. J. Acoust. Soc. Am. 99, 1633-1644.

Heinz, M.G., Sachs, M.B. and Young, E.D. (2003) Activity growth rates in auditory-nerve fibers following noise-induced hearing loss. Abs. 26th Ann. Midwinter Res. Mtng. Assoc. Res. Otolaryngol. 46-47.

Hellman, R.P. and Meiselman, C.H. (1990) Loudness relations for individuals and groups in normal and impaired hearing. J. Acoust. Soc. Am. 88, 2596-2606.

Humes, L.E. and Ahlstrom, J.B. (1984) Relation between reaction time and loudness. J Speech Hear Res 27, 306-10.

Hustert, B., Kumpf, W. and Stoll, W. (1996) Reaktionszeiten and der Hörschwelle: Eine Gegenüberstellung von normakusischen und hörgeschädigten Probanden. Laryngorhinootologie 75, 135-40.

Kohfeld, D.L., Santee, J.L. and Wallace, N.D. (1981) Loudness and reaction time: I. Percept Psychophys 29, 535-49.

Marshall, L. and Brandt, J.F. (1980) The relationship between loudness and reaction time in normal hearing listeners. Acta Otolaryngol 90, 244-9.

Moore, B.C.J. (1995) *Perceptual Consequences of Cochlear Damage*, Oxford University Press, Oxford.

Pfingst, B.E., Hienz, R., Kimm, J. and Miller, J. (1975) Reaction-time procedure for measurement of hearing. I. Suprathreshold functions. J Acoust Soc Am 57, 421-30.

Viemeister, N.F. (1988) Psychophysical aspects of auditory intensity coding. In: G. M. Edelman, W. E. Gall and W. M. Cowan (Ed.) *Auditory Function: Neurobiological Bases of Hearing*, Wiley, New York, pp. 213-241.

Normal and impaired level encoding: Effects of noise-induced hearing loss on auditory-nerve responses

Michael G. Heinz, Danilo Scepanovic, Murray B. Sachs, and Eric D. Young

Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
mgheinz@bme.jhu.edu

1 Introduction

A major constraint for listeners with sensorineural hearing loss (SNHL) is the reduced acoustic dynamic range associated with loudness recruitment; however, the physiological correlates of loudness and recruitment are still not well understood. Recent psychophysical results challenge some of the classic concepts of recruitment (Buus and Florentine 2001), and suggest that the dynamic range of loudness is also reduced. These results have potentially significant implications for hearing aids that have renewed an interest in how SNHL affects level encoding in the auditory nerve (AN). The present study compares several AN response properties related to level encoding in normal-hearing cats and in cats with a noise-induced hearing loss.

2 General methods

Single-unit AN responses were measured as a function of level in 1-dB steps in anesthetized (pentobarbital) cats. Stimuli ranged from simple to complex, including best-frequency (BF) tones, 1- and 2-kHz tones, broadband noise, and a speech token (/bɛsh/).

A noise-induced SNHL was produced by presenting a 50-Hz-wide noise band centered at 2 kHz for 4 hours at 103-108 dB SPL (e.g., Miller, Schilling, Franck and Young 1997). Hearing-impaired cats were pooled into two populations with differing degrees of hearing loss, as shown in Fig. 1. The mild-loss population had a ~30-dB loss near 2 kHz, with some fibers showing significantly broadened tuning (lower Q10) between 2-4 kHz. The moderate/severe-loss population had ~50-60 dB loss near 2 kHz, with all fibers between 2-4 kHz showing broadened tuning. The tuning-curve characteristics in Fig. 1 suggest that the SNHL consisted of mixed hair-cell damage (i.e., broadened tuning suggests OHC damage, while elevated thresholds with normal tuning suggests IHC damage; Liberman and Dodds 1984).

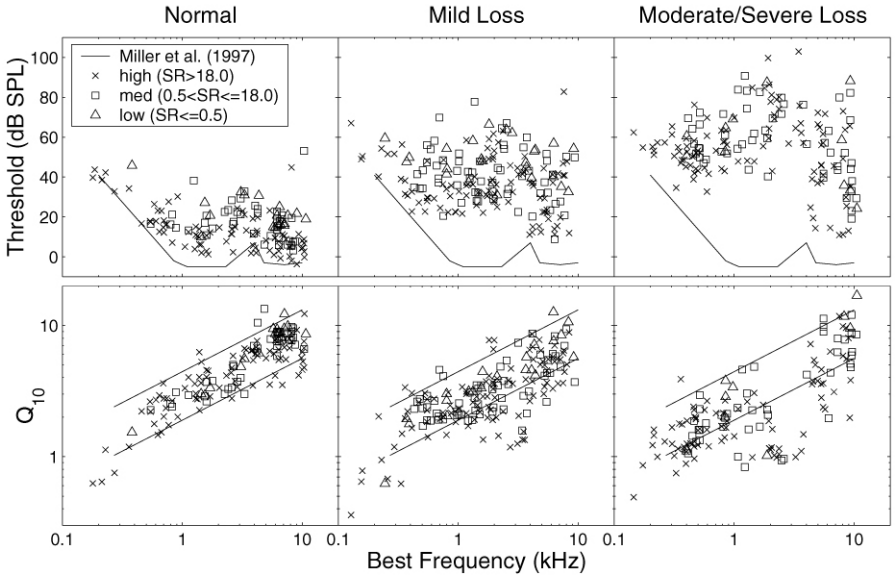


Fig. 1. Distributions of AN tuning-curve parameters illustrate the hearing loss for the two impaired populations. Solid lines are from a larger normal population (Miller et al. 1997): best thresholds (top); 5^{th} & 95^{th} percentiles of Q10 (=BF/BW). SR: spontaneous rate.

3 Impaired rate-level functions are not consistently steeper

It has been hypothesized that AN-fiber rate-level functions (RLFs) are steeper in impaired ears because impaired basilar-membrane (BM) responses are steeper (e.g., Pickles 1988; Moore 1991; Schroder, Viemeister and Nelson 1994). However, previous studies comparing BF-tone RLFs have reported mixed results (e.g., steeper: Harrison 1981; not steeper: Kiang, Moxon and Levine 1970; Salvi, Hamernik and Henderson 1983). This hypothesis was directly evaluated for a variety of stimuli by calculating RLF slopes from two-line fits (see inset of Fig. 2).

The slopes of individual RLFs for broadband noise are shown in Fig. 2. There is a large amount of variability within each population, and thus smoothed averages are also shown. Impaired RLFs were not steeper than normal near threshold (low-level slopes, left panel) for broadband noise or for tones (not shown). Sloping saturation occurred less often in the moderate/severe population (indicated by the absence of high-level slopes, right panel). For BFs between 0.5-4 kHz, sloping saturation was observed in 50%, 49%, and 13% of the noise RLFs in the normal, mild, and moderate/severe populations, respectively. The limited effect of SNHL on RLFs is consistent with the standard model for how BM compression affects RLF shapes (Sachs and Abbas 1974; Yates 1990). Sharply saturating RLFs (often low-threshold, high-SR fibers) should not be affected by impairment because the BM response is linear at low levels. Sloping-saturation RLFs (often higher-threshold, low-SR fibers) should become steeper, but not near threshold.

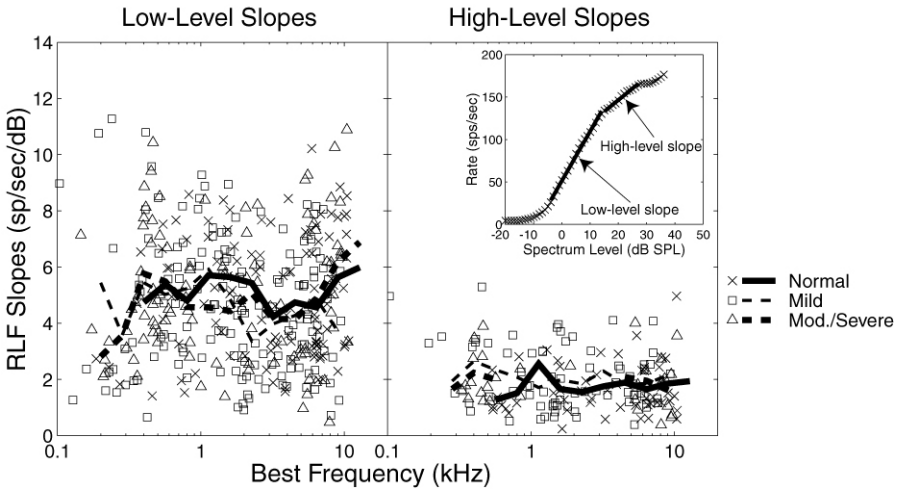


Fig. 2. Rate-level slopes for broadband noise for the three AN populations. Curves are smoothed averages. Inset: two-line fits (*low-* and *high-level* slopes) to sloping-saturation RLFs; high-level slopes were not fitted for sharply saturating RLFs.

The normal variation in RLF slopes across simple and complex stimuli was reduced by SNHL. In particular, RLF slopes for the vowel /ɛ/ were more similar to BF-tone slopes in impaired fibers (not shown). This effect would be expected from a reduction in peripheral suppression; however, much of this effect results from a number of impaired fibers having BF-tone RLF slopes that were shallower than normal. Impaired RLFs for noise were also sometimes shallower than normal (e.g., 1-3 kHz BFs, left panel, Fig. 2). AN fibers with shallow slopes above threshold typically retained sharp tuning, consistent with primarily IHC damage (Liberman and Dodds 1984). Shallower slopes could arise if IHC damage elevated threshold so that the entire AN-fiber dynamic range overlapped the BM compression region.

RLFs were sometimes steeper than normal for severe impairments (> ~80 dB SPL), in association with high-level irregularities (e.g., Liberman and Kiang 1984).

Overall, impaired AN RLF slopes were not consistently steeper than normal slopes. However, impaired RLFs were steeper in limited conditions, some of which can account for previous studies that reported steeper RLFs (e.g., Harrison 1981).

4 Response variability is unaffected by noise-induced hearing loss

Various theories predict that intensity JNDs should be different in normal-hearing and hearing-impaired listeners due to loudness recruitment; however, intensity JNDs are often normal for listeners with SNHL (e.g., Schroder et al. 1994; Neely and Allen 1997). A change in internal noise due to SNHL is often hypothesized as a possible explanation for this discrepancy. Thus, the variability in AN discharge counts was estimated for each RLF in the normal and impaired populations.

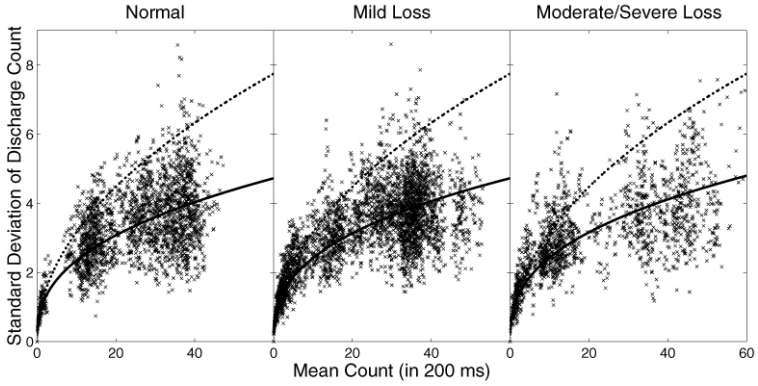


Fig. 3. Scatter plots of estimated discharge-count statistics. Dotted line: Poisson model. Solid lines: Power-function fits to each population. Estimates were made at each level of every tone RLF by pooling across 5 dB when the RLF slope at that level was < 2 sp/sec/dB. Only BFs in the impaired region (0.5-4 kHz) were included. At least two RLF reps were required.

Figure 3 shows estimated standard deviations as a function of mean discharge count for each AN population. Estimates for normal fibers were similar to previous studies, i.e., AN variability was less than Poisson (dotted line) at high rates (e.g., Young and Barta 1986). There was no apparent effect of noise-induced hearing loss on the variability in AN discharges. Thus, the present results do not support the hypothesis that normal intensity JNDs often measured for impaired listeners result from increased variance counteracting steeper RLFs associated with recruitment.

5 Growth of total impaired AN activity is normal near threshold

While individual RLFs were not consistently steeper following noise-induced hearing loss, broadened tuning could potentially produce steeper growth in the total AN activity due to abnormal spread of excitation. Estimates of total AN activity are shown in Fig. 4 as a function of level for 1- and 2-kHz tones. Impaired fibers with BFs near the 1-kHz tone had normal tuning, while impaired fibers near the 2-kHz tone often had broadened tuning (Fig. 1). Total AN activity grew at similar rates near threshold in the normal and impaired populations for both tones. Thus, broadened tuning did not have a significant effect on total AN activity growth.

AN activity growth rates are consistent with some aspects of a new view of loudness growth in impaired listeners (Buus and Florentine 2001). The similarity between normal and impaired AN growth rates near threshold is consistent with normal loudness growth near threshold in impaired listeners. Total AN activity in the impaired populations “caught up” to the normal population at high levels (at least partly due to reduced sloping saturation), consistent with normal loudness at high levels in impaired listeners.

Despite the similarities in the effect of SNHL on AN activity growth and on loudness growth, these comparisons must be made with caution. Pickles (1983) and Relkin and Doucet (1997) have provided evidence that contradicts the hypothesis

that loudness is related to total AN activity. Consistent with Relkin and Doucet (1997), our estimates of total AN activity grew slower than loudness (dotted lines in Fig. 4), at least at suprathreshold levels. Alternative hypotheses have been proposed for the neural correlates of loudness recruitment, including theories based on the temporal aspects of AN responses (e.g., Carney 1994).

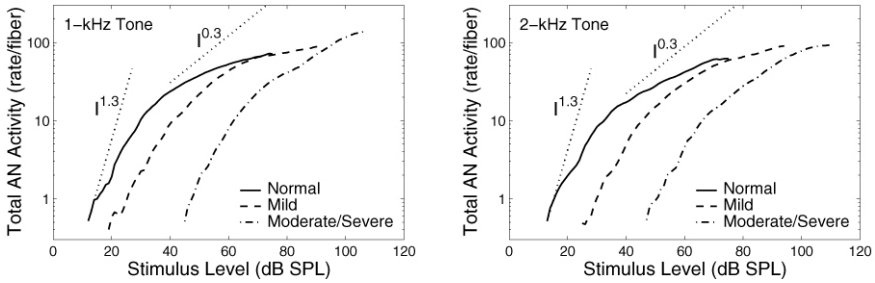


Fig. 4. Estimated growth of total AN activity with level for each AN population. Dotted lines: typical normal loudness growth near threshold and at supra-threshold levels (Buus and Florentine 2001). Total AN activity was estimated by summing the driven rate (rate minus SR) across all fibers within each of our three populations (normalized to rate per fiber).

6 The level dependence of phase is reduced by impairment

The systematic level dependence of the phase response in normal AN fibers (Anderson, Rose, Hind and Brugge 1971) has been proposed as an alternative neural code for level that could produce recruitment-like effects (Carney 1994). Heinz, Colburn and Carney (2001) demonstrated that nonlinear phase cues provide significant information near BF at medium-high levels and that an across-frequency, monaural coincidence detector could decode these cues based on the level dependence of relative phase across BF. These nonlinear phase responses are thought to be related to nonlinear cochlear tuning associated with OHC function (e.g., Ruggero, Rich, Recio, Narayan and Robles 1997).

The effect of SNHL on nonlinear phase responses is illustrated in Fig. 5, which shows the rate of change in phase for a 1-kHz, 60 dB SPL tone as a function of BF for our normal and mild-loss populations. A clear systematic "bowtie" effect is seen in the normal population, with increasing phase lags for BFs > 1 kHz and increasing phase leads below 1 kHz. This systematic effect is consistent with individual AN-fiber responses (Anderson et al. 1971) and BM responses (Ruggero et al. 1997). In the mild-loss population, the normal bowtie effect is absent, with slopes near zero on average above 1 kHz. The few large negative slopes above 1 kHz are primarily from a single impaired cat (triangles). The negative slopes below 1 kHz were not consistent with the normal responses.

This significant effect of SNHL on nonlinear phase responses occurred for a population of AN fibers with a mild hearing loss (~25-30 dB). Many of the AN fibers with reduced nonlinear phase had Q10 values that fell within the normal range (Fig. 1). Thus, the systematic nonlinear phase response across the AN population appears to be very sensitive to cochlear damage, consistent with BM

phase responses (Ruggero et al. 1997). The level dependence of phase near BF may provide an additional physiological window on OHC function, which could be more sensitive than the common use of Q10.

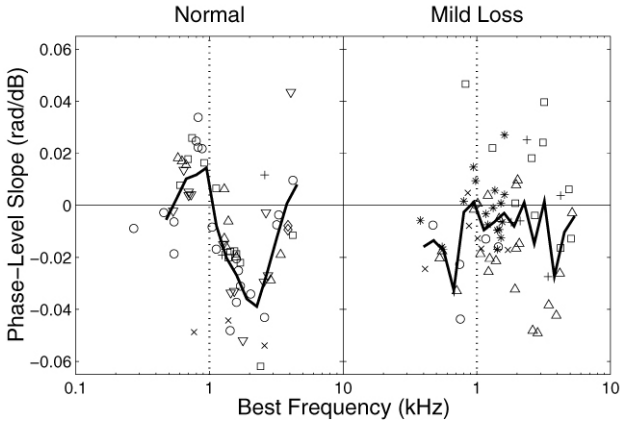


Fig. 5. The rate of change in phase with level as a function of BF for the normal and mild-loss populations for a 1-kHz tone at 60 dB SPL. Symbols: different experiments. Solid curves: smoothed averages. Phase was estimated at each level from period histograms formed by pooling spikes from all reps across a 5-dB range (100 spikes minimum). The phase-level slope was estimated from a linear regression over an 11-dB range.

7 Discussion

The present results apply to a mixed cochlear hair cell lesion, like that produced by acoustic trauma. While recruitment is often assumed to result from steeper BM responses associated with OHC damage, AN fibers do not provide a simple representation of the BM I/O function. The present results suggest that both OHC and IHC damage can affect AN response growth. Damage to OHCs only produces steeper AN RLFs in limited conditions. IHC damage can affect AN response growth in several ways, e.g., by producing shallower responses above threshold and/or steeper responses at very high levels.

It is intriguing that estimates of BM compression can be made from perceptual responses (e.g., Oxenham and Plack 1997), given that the complicated representation in the AN forms a bottleneck for information to the brain. The question of how level is encoded in the AN is critical for understanding limitations of current hearing aids and for designing improved algorithms. For example, the normal patterns of level-dependent phase responses, which could be perceptually relevant for level encoding and/or sound localization, would not be reproduced in the AN by hearing-aid algorithms that only correct for BM magnitude compression.

Acknowledgments

Supported by NIH/NIDCD grants: T32DC00023, R01DC00109, and F32DC05521.

References

Anderson, D.J., Rose, J.E., Hind, J.E. and Brugge, J.F. (1971) Temporal position of discharges in single auditory nerve fibers within the cycle of a sinewave stimulus: Frequency and intensity effects. J. Acoust. Soc. Am. 49, 1131-1139.

Buus, S. and Florentine, M. (2001) Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. J. Assoc. Res. Otolaryngol., 3, 120-139.

Carney, L.H. (1994) Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. Hear. Res. 76, 31-44.

Harrison, R.V. (1981) Rate-versus-intensity function and related AP responses in normal and pathological guinea pig and human cochleas. J. Acoust. Soc. Am. 70, 1036-1044.

Heinz, M.G., Colburn, H.S. and Carney, L.H. (2001) Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection. J. Acoust. Soc. Am. 110, 2065-2084.

Kiang, N.Y.S., Moxon, E.C. and Levine, R.A. (1970) Auditory nerve activity in cats with normal and abnormal cochleas. In: G.E.W. Wolstenholme and T. Knight (Eds.), *Sensorineural Hearing Los*s. Churchill, London, pp. 241-273.

Liberman, M.C. and Dodds, L.W. (1984) Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves. Hear. Res. 16, 55-74.

Liberman, M.C. and Kiang, N.Y.S. (1984) Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions. Hear. Res. 16, 75-90.

Miller, R.L., Schilling, J.R., Franck, K.R. and Young, E.D. (1997) Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers. J. Acoust. Soc. Am. 101, 3602-3616.

Moore, B.C.J. (1991) Characterization and simulation of impaired hearing: Implications for hearing aid design. Ear Hear. 12, 154S-161S.

Neely, S.T. and Allen J.B. (1997) Relation between the rate of growth of loudness and the intensity DL. In: W. Jesteadt (Ed.), *Modeling Sensorineural Hearing Loss.* Erlbaum, Mahwah NJ, pp. 213-222.

Oxenham, A.J. and Plack, C.J. (1997) Behavioral measure of basilar-membrane non-linearity in listeners with normal and impaired hearing. J. Acoust. Soc. Am. 101, 3666-3675.

Pickles, J. (1983) Auditory-nerve correlates of loudness summation with stimulus bandwidth in normal and pathological cochleae. Hear. Res. 12, 239-250.

Pickles, J. (1988). *An Introduction to the Physiology of Hearing.* Academic Press, New York.

Relkin, E.M. and Doucet, J.R. (1997) Is loudness simply proportional to the auditory nerve spike count? J. Acoust. Soc. Am. 101, 2735-2740.

Ruggero, M.A., Rich, N.C., Recio, A. Narayan, S.S. and Robles, L. (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J. Acoust. Soc. Am. 101, 2151-2163.

Sachs, M.B. and Abbas, P.J. (1974) Rate versus level functions for auditory nerve fibers in cats: Tone burst stimuli. J. Acoust. Soc. Am. 56, 1835-1847.

Salvi, R.J., Hamernik, R.P. and Henderson, D. (1983) Response patterns of auditory nerve fibers during temporary threshold shift. Hear. Res. 10, 37-67.

Schroder, A.C., Viemeister, N.F. and Nelson, D.A. (1994) Intensity discrimination in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 96, 2683-2693.

Yates, G.K. (1990) Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions. Hear. Res. 50, 145–162.

Young, E.D. and Barta, P.E. (1986) Rate responses of auditory-nerve fibers to tones in noise near masked threshold. J. Acoust. Soc. Am. 79, 426-442.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Estimates of cochlear compression from measurements of loudness growth

Stephen T. Neely, Kim S. Schairer, and Walt Jesteadt

Boys Town National Research Hospital, neely@boystown.org

1 Introduction

The growth of loudness can be measured by matching the loudness of a pure tone to that of a multi-tone complex. This method was originally used by Fletcher and Munson (1933) and has been used more recently by Buus et al. (1998). If the tones in the complex are equally loud and separated in frequency far enough that they do not mask each other, then the ratio of the loudness of the complex to the loudness of each component will be equal to the number of components. This empirical fact can be used to construct a relative loudness function without any additional assumptions about the form of this function. We have obtained loudness growth measurements using four-tone complexes with one-octave separation between equally-loud components. We assume that these four-fold complexes are four times as loud as any of their components. We consider loudness to be an intensity-like variable, so that a four-fold increase in loudness is equivalent to 6 dB. Therefore, we estimate cochlear compression by determining the number of dB (of physical intensity) required to achieve a four-fold increase in loudness and dividing that number by the corresponding 6-dB increase (in perceptual intensity).

2 Methods

Six adults with normal hearing served as subjects. Stimuli were 500-ms duration tones with cosine-squared envelopes (20-ms rise/fall) generated digitally (TDT AP2) at 0.5, 1, 2, and 4 kHz. Data were collected in three stages. (1) Quiet thresholds were obtained for the four individual tones. (2) The loudness of tones at 0.5, 2, and 4 kHz was matched to the loudness of a 1 kHz tone at 10-dB SL. (3) These four equally-loud tones were combined as a “chime” and the loudness of each component tone was matched to the loudness of this chime.

Threshold in quiet for each tone was estimated using a one-track paradigm, a two-interval forced-choice (2IFC) procedure and a 2-down, 1-up decision rule with 4-dB step size to estimate the 71% correct point on the psychometric function (PF). The mean threshold across four repetitions was calculated for each frequency.

In the next stage, the 0.5, 2, and 4 kHz tones were individually matched in loudness to a 1 kHz tone at 10 dB SL using a 2IFC, six-track paradigm (two tracks for each of the three matching tones). The subject's task was to identify the interval that contained the louder sound, regardless of pitch. There were 300 trials in each repetition (50 trials per track), and two repetitions were completed for each subject. One of the two tracks per matching tone estimated the 71% point on the PF; the other track estimated the 29% point track. The level at which each matching tone was judged to be equally loud to the 1-kHz tone was estimated from the mean of these two points (Jesteadt et al. 1980). The matching tone level started at 30 dB SPL for the 71% track and at threshold for the 29% track. Step sizes were 8 and 2 dB SPL. The matched level was calculated as the mean of reversal points after the fourth reversal. The average matched level was calculated across the two tracks for each frequency (i.e., across ascending and descending tracks) and across the two repetitions. The matched levels for the 0.5, 2, and 4-kHz tones, along with the 10-dB SL 1-kHz tone defined the first set of equally-loud tones for the final stage of data collection.

The four equally-loud tones were combined to create a fixed-level standard or "chime" that was four times as loud as its components. The chime was presented randomly in one interval of each trial, and one of the individual tones was presented in the other interval. In general, the starting stimulus level was about 16 dB above the chime level for the descending tracks and 16 dB SPL below for the ascending tracks. Step sizes were 4 and 2 dB. The average matched level was calculated across two tracks for each frequency (i.e., across ascending and descending tracks) and across two repetitions. This resulted in a new set of four equally-loud tones that were combined to create a new, fixed-level chime. This process was repeated until the subject provided responses that would have required a matching tone to exceed 90 dB SPL. Thus, subjects were not presented with the same number of loudness matching conditions. The time required to complete data collection on each subject ranged from 12 to 15.5 hours.

3 Results

We will describe individual results for the subject with the least intra-subject variability, known as our "best" subject, and describe average results for the entire group. In addition to our primary results for estimates of loudness, we consider Weber fraction estimates derived from the same data.

3.1 Loudness

Our most basic loudness result is the determination of equally-loud tones at the four stimulus frequencies. Iso-loudness contours for our best subject are shown in Fig. 1A and for the group average in Fig. 1B. The solid-black symbols at the lowest levels represent quiet thresholds. The gray symbol at 1 kHz is at 10 dB SL, which is our reference stimulus for all loudness values. The gray symbols at other frequencies have the same loudness as the reference tone. The open symbols above the gray symbols all have the same loudness as the combined loudness of the first

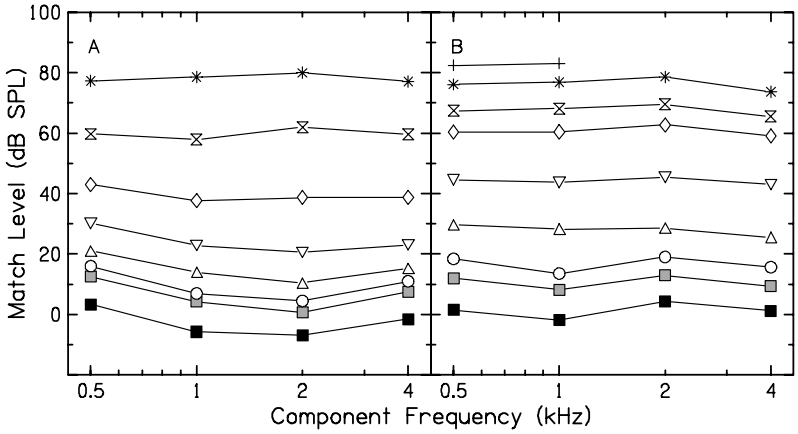


Fig. 1. Equal loudness contours. Black squares represent threshold for each component in quiet. Shaded squares represent the condition in which tones were matched to a 10-dB SL 1-kHz tone. A: Best subject. B: Average across subjects.

four equally-loud tones. Assuming that the loudness of these four tones is additive (Fletcher and Munson 1933), the loudness of each subsequent iso-loudness contour is four times the loudness of the iso-loudness contour below it.

Using the sum of four, equally-loud tones as the standard for loudness matching allows us to determine the number of dB required to quadruple the loudness of each of the component tones. This result is shown in Fig. 2. As in Fig. 1, panel A represents our best subject and panel B represents the group. In Fig. 2B, the

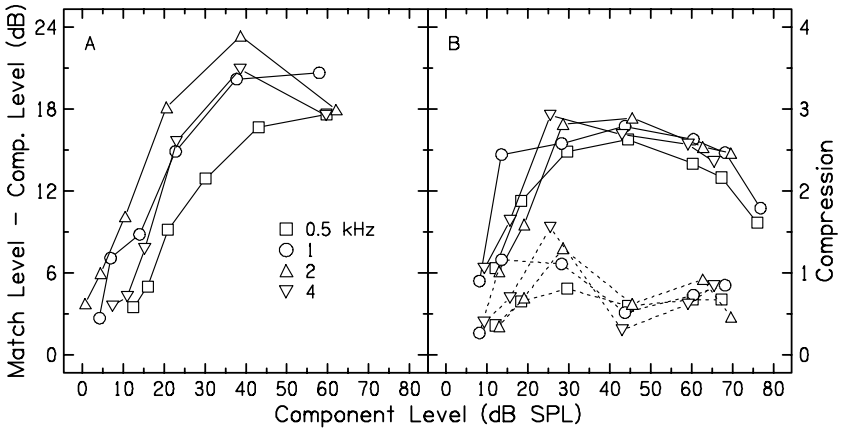


Fig. 2. Number of dB required to quadruple loudness of a tone as a function of its level. Compression is estimated by dividing the number of dB to quadruple loudness by 6 dB. A: Best subject. B: Average across subjects. Dashed lines represent the corresponding standard deviations.

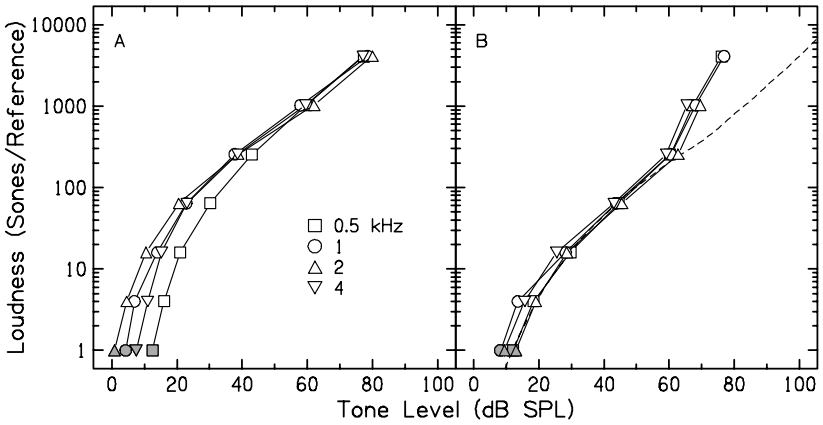


Fig. 3. Loudness functions. These functions are derived from the same data shown in Fig. 1 assuming that the equal-loudness contours are separated by a factor of four. The shaded symbols represent tones that have the same loudness as 1 kHz tone at 10 dB SL. A: Best subject. B: Average across subjects. The dashed line is from Fletcher and Munson (1933), but divided by 20 to account for the difference in reference value

symbols connected by solid lines indicate the group average and the symbols connected by dashed lines the standard deviation across subjects. The number of dB to quadruple loudness is equal to the vertical separation between symbols in Fig. 1. In general, about 6-dB increase in tone level is required to quadruple the loudness of the lowest level tones. The number of dB increases rapidly as level increases reaching 15 to 18 dB at moderate levels. At the highest levels, the number of dB often decreases.

Another way to plot the data shown in Fig. 1 is to take vertical slices through iso-loudness contours and plot the relative loudness of each contour as a function of tone level separately for each frequency. This representation of the data is shown in Fig. 3A for our best subject and Fig. 3B for the group average. The loudness values plotted in Fig. 3 are relative to the loudness of a 1 kHz tone at 10 dB SL; hence the units sones/reference on the ordinate. No assumption beyond loudness additivity was required to construct these loudness functions. For comparison, the loudness function (for a 1 kHz tone) of Fletcher and Munson (1933) is superimposed in Fig. 3B.

We can define *compression* as the number of dB increase in physical intensity for each dB increase in perceptual intensity. If we treat loudness as proportional to the perceptual intensity, then a four-fold increase in loudness is equivalent to 6 dB increase in perceptual intensity. We can estimate compression conveniently from our loudness measurements by taking the number of dB required to quadruple loudness (as shown in Fig. 2) and dividing that value by 6 dB. The right-hand axis in Fig. 2 shows the compression values that correspond to the dB values on the left-hand axis. Our compression estimates start around 1 for the lowest level tones and

increase to between 2.5 and 3 at moderate levels. In some cases, compression decreases abruptly at the highest levels to between 1.5 and 2.

The compression values in Fig. 2A are equal to the reciprocal of the slope of the loudness functions in Fig. 3A. The correspondence between compression and slope is not maintained exactly between Figs. 2B and 3B because the group averages were computed differently. Fig. 2B is based on the average of the compression values whereas Fig. 3B is based on the average of the iso-loudness contours.

3.2 Weber fraction

If we assume that a just noticeable difference (JND) in loudness is determined by the variability imposed by Poisson internal noise, then the loudness JND will be proportional to the square root of the average loudness (McGill and Goldberg 1968, Hellman and Hellman 1990). Using the variable N to represent loudness, this can be written as follows (e.g., Allen and Neely 1997).

$$\Delta N = h\sqrt{N} \tag{1}$$

The factor h is a constant that depends on the reference selected for the loudness scale. If we let α denote compression and assume that JNDs are small, our definition of compression allows us to relate the perceptual JND ΔN to the corresponding physical JND ΔI (e.g., Allen and Neely 1997).

$$\frac{\Delta I}{I} = \alpha \frac{\Delta N}{N} \tag{2}$$

Combining these two equations gives us a way to estimate the Weber fraction from loudness matching data.

$$\frac{\Delta I}{I} = \frac{h\alpha}{\sqrt{N}} \tag{3}$$

This estimate is plotted in Fig. 4A for our best subject using the compression values in Fig. 2 and the loudness values in Fig. 3. The unknown scale factor h has arbitrarily been set to 1 for this figure. Any other value of h would shift the plotted values vertically.

Because our procedure for obtaining loudness matches involved two separate tracks for each condition that converged to the 29% and 71% points on the PF for comparison of the loudness of a tone relative to the loudness of a fixed-level chime, we have sufficient data to provide an estimate of the slope of this PF. Relative to a tone at the midpoint of the PF, we can consider the 29% and 71% points on the PF as JNDs in loudness. If the intensity of the lower-level point is I_1 and the higher-level point is I_2, then we can estimate the Weber fraction by the following formula.

$$\frac{\Delta I}{I} = \frac{I_2 - I_1}{I_2 + I_1} \tag{4}$$

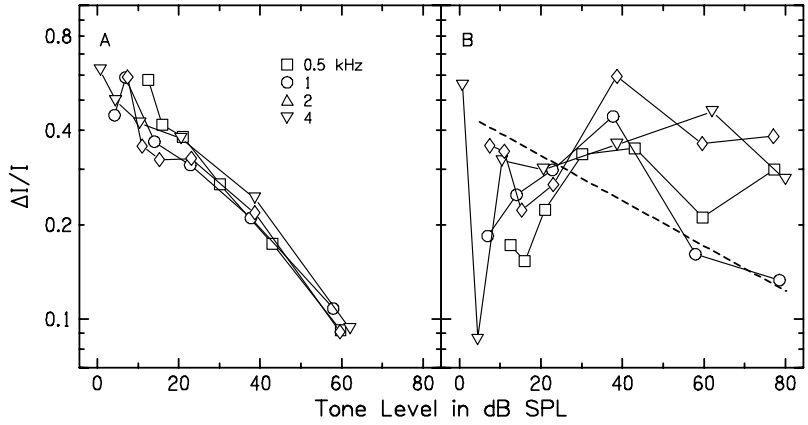


Fig. 4. Weber fraction estimates for best subject. A: Estimated directly from the loudness data in Fig. 3A. B: Estimated from the 29% and 71% points on psychometric function for loudness matching task. The dashed line is the estimate from Jesteadt et al. (1977) as a function of sensation level.

This Weber fraction estimate for our best subject is shown in Fig. 4B. For comparison, the fit by Jesteadt et al. (1977) to their intensity discrimination data, which was independent of frequency, is superimposed as a dashed line in Fig. 4B.

4 Discussion

The method of "loudness matching to the sum of equally-loud tones" was first described as a way to quantify loudness by Fletcher and Munson (1933). One advantage of this method over cross-modality scaling methods (Hellman, 1999) is its ability to more accurately characterize the rate of loudness growth. Buus et al. (1998) used a similar loudness-matching method to measure loudness; however, they combined equal-SL tones for their loudness matches instead of combining equally-loud tones. One advantage of using equally-loud tones is being able to determine the loudness growth rate directly without any additional modeling assumptions.

The group average loudness functions in Fig. 3B show an increase in slope at high levels that is not observed in the loudness functions for an individual subject in Fig. 3A. In fact, none of the individual subjects showed such an abrupt increase in slope. We consider this feature of the average loudness function to be an artifact of the averaging process. Individual loudness functions typically showed a slight increase in slope at the highest levels.

The group average compression functions in Fig. 2B appear to be representative of the individual compression functions. On average, compression appears to have a relatively constant value of a little less than 3. This is consistent with Steven's Law. At the lowest levels, the compression falls to about 1, consistent with energy detection, where perception growth is proportional to sound intensity. At high

levels, compression falls to about 2, consistent with a displacement detector, where perception growth is proportional to sound pressure.

We had hoped to see better agreement between the two estimates of the Weber fraction shown in Fig. 4, since both estimates were based on the same loudness measurements. Whether the disagreement indicates a failure of the Poisson internal noise model or reflects measurement errors remains to be determined.

In future work, we hope to correlate individual differences in the compression estimates based on loudness measurements with compression estimates based on otoacoustic emission measurements (OAE). We expect to see similar amounts of compression in OAE measurements as we observe in loudness measurements.

Acknowledgements

This work was funded by grants R01 DC02251, R01 DC00136, T32 DC00013, and P30 DC04662 from the National Institutes of Health.

References

Allen and Neely (1997) The relation between the intensity JND and loudness for pure tones and wide-band noise. J. Acoust. Soc. Am. 102, 3628-3646.

Buus, S., Musch, H., Florentine, M. (1998) On loudness at threshold. J Acoust Soc Am. 104, 399-410

Fletcher and Munson (1993) Loudness, its definition, measurement, and calculation. J. Acoust. Soc. Am. 5, 82-108.

Hellman, R. P. (1999) Cross-modality matching: a tool for measuring loudness in sensorineural impairment. Ear Hear. 20, 193-213.

Hellman, W. and Hellman, R. (1990) Intensity discrimination as the driving force of loudness: application to pure tones in quiet. J. Acoust. Soc. Am. 87, 1255-1271.

Jesteadt, W. (1980) An adaptive procedure for subjective judgments. Percept. Psychophys. 28, 85-88.

Jesteadt, W., Wier, C., and Green, D. (1977) Intensity discrimination as a function of frequency and sensation level. J. Acoust. Soc. Am. 61, 169-177.

McGill, W. and Goldberg J. (1968) Pure-tone intensity discrimination as energy detection. J. Acoust. Soc. Am. 44, 576-581.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Additivity of masking and auditory compression

Christopher J. Plack, Catherine G. O'Hanlon, and Vit Drga

Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.
cplack@essex.ac.uk.

1 Introduction

Over the last few years there have been several attempts to estimate the non-linear response function of the human basilar membrane (BM) using psychophysical procedures. The experimental designs have been based on *forward masking*, in which a signal is presented after the offset of a masker. Forward masking is used so that the masker and the signal do not interact on the BM, thus avoiding suppressive effects that complicate the interpretation of the results. Two popular techniques are the growth of masking (GOM) technique, in which the masker level required to mask a signal is found as a function of signal level (Moore, Vickers, Plack, and Oxenham 1999; Oxenham and Plack 1997; Plack and Oxenham 2000); and the temporal masking curve (TMC) technique, in which the masker level required to mask a signal is found as a function of the masker-signal interval (Lopez-Poveda, Plack, and Meddis 2003; Nelson, Schroder, and Wojtczak 2001; Plack and Drga 2003). Both these techniques estimate the BM response function to a tone at characteristic frequency (CF) by assuming that the BM response to a masker well below CF is *linear*. By comparing the GOM or TMC for a masker below the signal frequency (assumed to be processed linearly) with the masking function for a masker at the signal frequency, it is possible to derive the compressive CF response.

While these procedures have been useful for estimating the response function at high frequencies, it seems unlikely that the assumption of below-CF linearity is valid at low CFs. Measurements of BM vibration in the apex suggest that what little compression there is may affect a broad range of frequencies relative to CF (Rhode and Cooper 1996). Although the TMC technique can be adapted by using the off-frequency TMC at high frequencies as a linear yardstick at low frequencies (Lopez-Poveda et al. 2003; Plack and Drga 2003), this requires the further assumption of a constant internal decay of forward masking. An alternative is to use a technique that does not depend on the assumption of below-CF linearity. One such technique is the *additivity of forward masking* technique.

Two equally effective non-overlapping non-simultaneous maskers, when combined, should produce a 3-dB increase in signal threshold compared to the single-masker condition, if the auditory system is linear with respect to stimulus

intensity. Indeed, roughly linear additivity is observed in normally hearing listeners at low signal levels and in hearing-impaired listeners at all levels (Oxenham and Moore 1995). However, normally hearing listeners at moderate signal levels show a combined-masker threshold that is much more than 3 dB above the single-masker threshold (Cokely and Humes 1993; Oxenham and Moore 1995; Penner and Shiffrin 1980). This "excess masking" can be attributed to auditory compression. If two equally effective maskers are combined linearly, then the increase in internal masking will be 3 dB. However, if the signal is compressed by the auditory system, then the *physical* signal level will have to increase by more than 3 dB to compensate for the compression. For example, 5:1 compression (compression exponent of 0.2) would require a 15-dB increase in physical signal level to produce a 3-dB increase in the internal level. It follows that the amount of excess masking can be used as a measure of the compression applied to the signal.

Plack and O'Hanlon (in press) have used the additivity of two non-overlapping forward maskers to estimate auditory compression at 250, 500, and 4000 Hz in normally hearing listeners. At each frequency, they measured excess masking at sensation levels of 10 dB and at 30 dB. Although little excess masking was observed at the lower sensation level, at the higher sensation level the estimated compression exponents were 0.29, 0.34, and 0.17 at the three frequencies respectively. The results suggest that the auditory system is much more compressive at low CFs than would be expected from physiological measurements of BM vibration (Rhode and Cooper 1996; Zinn, Maier, Zenner, and Gummer 2000).

The present study extends this research to measure additivity over a range of signal levels at 250 and 4000 Hz. It will be shown how the masked thresholds can be used to estimate auditory compression as a function of level, and hence to derive the response function of the system. The aim of the study was to provide a further test of the hypothesis that the response to a tone at CF does not vary substantially with CF in humans.

2 Method

The experiment was conducted in two stages. In the first stage, the levels of two non-overlapping forward maskers were found that were roughly equally effective at masking the signal. The process was repeated for each signal frequency and level. In the second stage, thresholds for the signal were measured in the presence of each equally effective masker presented individually, and in the presence of the two maskers presented together. For each signal frequency and (nominal) level, the compression exponent was then calculated on the basis of these three signal thresholds.

2.1 Stimuli

The sinusoidal signal had a frequency of 250 or 4000 Hz. The noise maskers were low-pass filtered at 1000 Hz (3-dB cutoffs, 90 dB/octave) for the 250-Hz signal and band-pass filtered between 2800 and 5600 Hz (3-dB cutoffs, 90 dB/octave) for the

4000-Hz signal. At 250 Hz, the signal had a total duration of 10 ms, which consisted of 5-ms raised-cosine onset and offset ramps (no steady state). Masker 1 (M1) had a total duration of 200 ms, including 5-ms onset and offset ramps and 190-ms steady state. Masker 2 (M2) had a total duration of 10 ms, which consisted of 5-ms onset and offset ramps (no steady state). At 4000 Hz, the signal had a total duration of 4 ms, which consisted of 2-ms raised-cosine onset and offset ramps (no steady state). M1 had a total duration of 200 ms, including 2-ms onset and offset ramps and 196-ms steady state. M2 had a total duration of 6 ms, including 2-ms onset and offset ramps and 2-ms steady state. At both frequencies, the offset of M1 coincided with the onset of M2, and the silent interval between the end of M2 and the start of the signal (0 V points) was 4 ms.

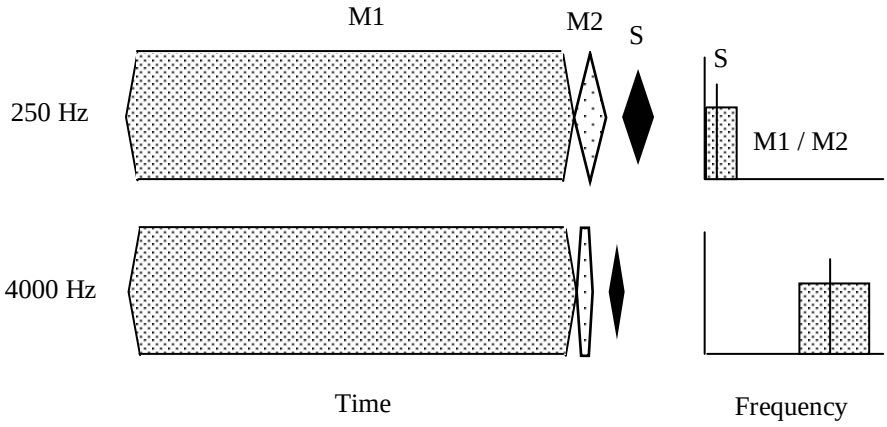


Fig. 1. A schematic illustration of the temporal and spectral characteristics of the stimuli.

The parameters of the stimuli were chosen to provide a reasonable amount of masking while minimizing the overlap between the stimuli on the BM resulting from the temporal response of the auditory filters. Figure 1 shows a schematic illustration of the temporal and spectral characteristics of the stimuli for the two frequencies.

2.2 Procedure

A three-interval three-alternative forced-choice paradigm was used, with an inter-stimulus interval of 300 ms. The individual or combined masker was presented in all three intervals. The signal was presented following the masker in only one of the three intervals, chosen at random with probability 1/3. Threshold was determined using a two-up one-down (masker thresholds) or a two-down one-up (signal thresholds) adaptive procedure that tracked the 70.7 percent correct point on the psychometric function (Levitt 1971). The step-size was 4 dB for the first four turnpoints, which reduced to 2 dB for twelve subsequent turnpoints. The mean of the last twelve turnpoints was taken as the threshold estimate for each block of trials. At least four estimates were made for each condition and the results averaged.

In the first stage of the experiment, the signal was presented at levels between 10 and 70 dB above absolute threshold, determined independently for each listener (limited by the need to avoid clipping when the masker level approached the maximum output of the apparatus). For each signal level, the spectrum levels of M1 and M2 were varied adaptively to find the levels needed to mask the signal. These masker levels were then used in the second stage of the experiment. For each pair of equally effective maskers, the signal threshold was measured in the presence of M1 alone, M2 alone, and M1 and M2 combined. At each frequency the conditions were presented in a random order.

Stimuli were presented to the right ear. Each listener sat in a soundproof booth and decisions were recorded via a computer keyboard. Listeners viewed a computer monitor through a window in the sound booth. Lights on the monitor display flashed on and off concurrently with each stimulus presentation and provided feedback at the end of each trial. Three normally hearing listeners were employed at 250 Hz, and an additional listener was added to make four listeners at 4000 Hz.

3 Results and analysis

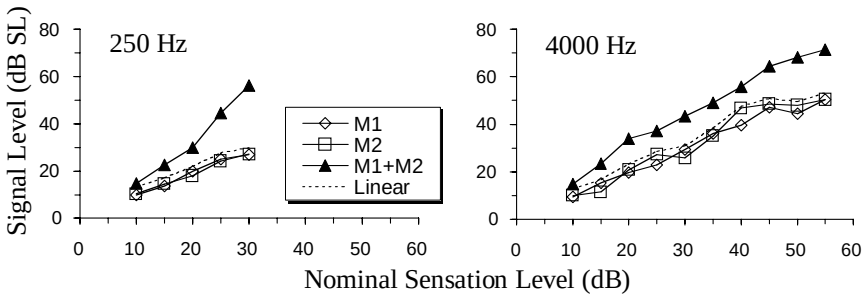


Fig. 2. The results of the second stage of the experiment for listener CO.

The results of the second stage of the experiment for listener CO are presented in Fig. 2. The figure shows the signal sensation level at threshold as a function of the "nominal" sensation level (the signal sensation level used to derive the masker levels in the first stage of the experiment) for M1 alone, M2 alone, and the combined condition (M1+M2). As expected, the signal thresholds in the individual masker conditions were similar to the nominal sensation levels, indicating little change in performance from the first stage of the experiment. The dashed line shows the combined-masker thresholds that would be obtained if the system were linear. It can be seen that the combined-masker thresholds measured experimentally deviate substantially from the linear prediction at moderate to high levels (excess masking), suggestive of a compressive response.

Because signal thresholds were measured in the presence of the individual maskers, as well as in the combined-masker condition, it was not necessary to

assume that there was no change in performance from the first to the second stage of the experiment, or that the maskers were equally effective. An estimate of compression could be obtained from an analysis of the three thresholds from the second stage (Plack and O'Hanlon in press). It was assumed that the ratio of the internal (i.e., post-cochlear) signal intensity to the internal (or effective) masker intensity is a constant at signal threshold. It was also assumed that the internal signal intensity is a power-law transformation of physical signal intensity. Hence:

$$I_M = kS_M^c \tag{1}$$

where I_M represents the internal effect of the masker, S_M is the physical signal intensity at masked threshold, c is the compression exponent, and k is a constant. It was assumed further that the effect of combining two maskers is a linear summation of their individual effects. Hence:

$$I_{M1+M2} = I_{M1} + I_{M2} \tag{2}$$

Substituting Eq 1 in Eq 2, and factoring out the constant k, leaves:

$$S_{M1+M2}^c = S_{M1}^c + S_{M2}^c \tag{3}$$

If S_{M1+M2} (the signal intensity at threshold for the combined maskers), S_{M1} (the signal threshold for M1 alone), and S_{M2} (the signal threshold for M2 alone) are all known, it is possible to determine the compression exponent c. For conditions in which the individual thresholds for M1 and M2 differed by more than 10 dB, the data were excluded. In these situations the combined threshold is often similar to the larger of the two individual thresholds, and the compression estimate is inaccurate.

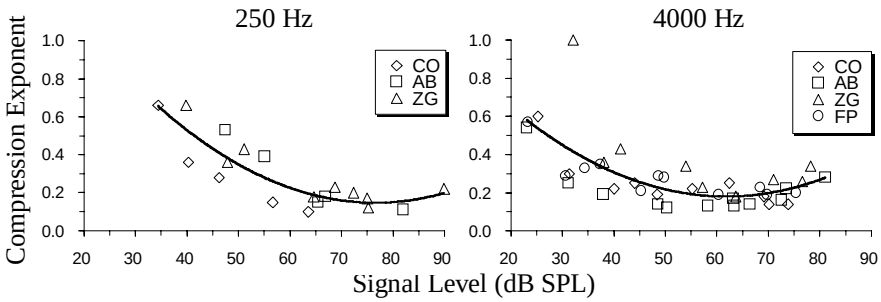


Fig. 3. The compression exponents derived from the masking data.

The compression exponents derived from the data are shown in Fig. 3. The horizontal axis for these plots is the mean of the single masker thresholds, *plus* the combined masker threshold, divided by two (all values expressed in dB SPL and calculated independently for each listener). This is just an estimate of the effective signal level to which the compression exponent applies. Figure 3 shows that the

compression exponents are broadly similar for the two frequencies. At both frequencies, the values are highest at low levels, decreasing to a minimum value of around 0.2 (5:1 compression) at higher levels.

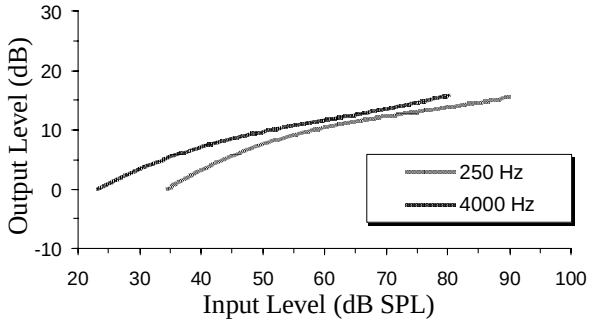


Fig. 4. The response functions derived from the compression exponents.

At each frequency, the compression exponent data from all the listeners were fit (least squares) with a 2nd-order polynomial. The fitted curves are shown by the solid lines in Fig. 3. These functions were then *integrated* to derive an estimate of the response (input/output) function at each frequency. The response functions derived in this way are shown in Fig. 4. These functions have been normalized so that the lowest output level was 0 dB in each case.

4 Conclusions

The results confirm recent psychophysical studies of forward masking (Lopez-Poveda et al. 2003; Plack and Drga 2003; Plack and O'Hanlon in press), and phase effects in simultaneous masking (Oxenham and Dau 2001), that suggest that the human auditory system is highly compressive at low frequencies. The study extends that of Plack and O'Hanlon to show that compression at 250 Hz occurs over a wide range of input levels. Although the derived response function at 250 Hz is shifted to higher input levels, the overall shape of the function is very similar to that at 4000 Hz. The present results suggest a compression exponent of around 0.2 at both frequencies, consistent with physiological findings at high CFs (Ruggero, Rich, Recio, Narayan, and Robles 1997; Russell and Nilsen 1997), but *inconsistent* with physiological findings at low CFs, that suggest a much more linear response than reported here (Rhode and Cooper 1996; Zinn et al. 2000). Assuming that the physiological and psychophysical results accurately reflect auditory processing, two possible reasons for the discrepancy may be suggested. First, the results could represent a between-species difference in the processing of sound, between the human cochlea on the one hand, and the chinchilla and guinea pig cochleae on the other. Second, the human results at *low* frequencies may reflect the effects of compression at a location or locations more central than the cochlea (see Plack and Drga 2003; Plack and O'Hanlon in press).

Acknowledgments

The research was supported by EPSRC Grant GR/N07219.

References

Cokely, C.G. and Humes, L.E. (1993) Two experiments on the temporal boundaries for the nonlinear additivity of masking. J. Acoust. Soc. Am. 94, 2553-2559.

Levitt, H. (1971) Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467-477.

Lopez-Poveda, E.A., Plack, C.J. and Meddis, R. (2003) Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing. J. Acoust. Soc. Am. 113, 951-960.

Moore, B.C.J., Vickers, D.A., Plack, C.J. and Oxenham, A.J. (1999) Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism. J. Acoust. Soc. Am. 106, 2761-2778.

Nelson, D.A., Schroder, A.C. and Wojtczak, M. (2001) A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 110, 2045-2064.

Oxenham, A.J. and Dau, T. (2001) Towards a measure of auditory filter phase response. J. Acoust. Soc. Am. 110, 3169-3178.

Oxenham, A.J. and Moore, B.C.J. (1995) Additivity of masking in normally hearing and hearing-impaired subjects. J. Acoust. Soc. Am. 98, 1921-1934.

Oxenham, A.J. and Plack, C.J. (1997) A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. J. Acoust. Soc. Am. 101, 3666-3675.

Penner, M.J. and Shiffrin, R.M. (1980) Nonlinearities in the coding of intensity within the context of a temporal summation model. J. Acoust. Soc. Am. 67, 617-627.

Plack, C.J. and Drga, V. (2003) Psychophysical evidence for auditory compression at low characteristic frequencies. J .Acoust. Soc. Am. 113, 1574-1586.

Plack, C.J. and O'Hanlon, C. (in press) Forward masking additivity and auditory compression at low and high frequencies. J. Assoc. Res. Otolaryngol.

Plack, C.J. and Oxenham, A.J. (2000) Basilar-membrane nonlinearity estimated by pulsation threshold. J. Acoust. Soc. Am. 107, 501-507.

Rhode, W.S. and Cooper, N.P. (1996) Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo. Auditory Neurosci. 3, 101-121.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S. and Robles, L. (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J. Acoust. Soc. Am. 101, 2151-2163.

Russell, I.J. and Nilsen, K.E. (1997) The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane. Proc. Nat. Acad. Sci. 94, 2660-2664.

Zinn, C., Maier, H., Zenner, H.-P. and Gummer, A.W. (2000) Evidence for active, nonlinear, negative feedback in the vibration response of the apical region of the in-vivo guinea-pig cochlea. Hear. Res. 142, 159-183.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Psychophysical response growth under suppression

Magdalena Wojtczak and Neal F. Viemeister

Department of Psychology, University of Minnesota, {wojtc001, nfv}@umn.edu

1 Introduction

A reduction of a response to a tone in the presence of another tone, called two-tone suppression, has been demonstrated at different levels of the auditory processing and linked to the operation of the active mechanism in the cochlea. Since two-tone suppression is a product of nonlinear processing, it is of interest to study the effects of level on the size of this phenomenon. Duifhuis (1980) measured level effects in psychophysical suppression and observed that suppression is not exclusively dependent on the level of the suppressor nor does it stay constant for a constant ratio of the suppressor and suppressee amplitudes. Data of Ruggero, Robles, and Rich (1992) provided more insight into the effect of an off-frequency suppressor on basilar-membrane (BM) responses to characteristic frequency (CF) tones. They demonstrated that the BM response to a CF tone grows in a compressive manner when the tone is presented alone. Adding an off-frequency suppressing tone changes the shape of the response function measured at CF by shifting the threshold for detecting the tone toward higher levels and increasing the slope of the function. Consequently, for a fixed-level suppressor, the magnitude of suppression decreases with increasing level of the suppressed (CF) tone. This finding is inconsistent with Duifhuis's observation that two-tone suppression "increases as the overall level increases". It may also be inconsistent with the auditory-nerve data of Javel, Geisler, and Ravindran (1978), which showed a parallel shift of rate-level functions in the presence of a suppressor. Their data, however, were obtained from high-spontaneous-rate fibers with low thresholds and small dynamic ranges, and thus they only covered a range of levels, for which BM processing is linear.

A reduction of the response at CF in the presence of a suppressing tone presumably reflects a reduction in gain. An important question is whether the effective reduction in gain is simply attenuative or whether it is proportional to the gain at CF in the absence of the suppressor. These two gain reduction schemes lead to different predictions about the magnitude of suppression across levels of the suppressed tone. In the former case, a constant magnitude of suppression across levels of the suppressee would be observed whereas in the latter case, the magnitude of suppression would decrease with increasing suppressee level.

Previous psychophysical studies of suppression did not provide data that could be used to make inferences about the magnitude of suppression across suppressee

levels since they varied the suppressor level while keeping the suppressee level constant or they covaried both levels keeping the difference between them constant (Shannon 1976; Duifhuis1980). In contrast, the present experiment measured suppression for different levels of the tone to be suppressed presented with a fixed-level suppressor. The purpose of this experiment was to determine whether psychophysical suppression measured for a fixed-level suppressor decreases with increasing level of the suppressee, consistent with the mechanical data of Ruggero *et al*. (1992), or stays constant across suppressee levels, consistent with the auditory-nerve data of Javel *et al*. (1978).

2 Methods

Using a three-interval forced-choice procedure, detection of a 10-ms 4-kHz probe was measured in forward masking for a 100-ms 4-kHz masker in one condition and a two-tone complex consisting of a 4-kHz masker and a simultaneous 4.8-kHz suppressor in another condition. The probe was temporally separated from the masker (and the suppressor when present) by a 2-ms gap. Within an experimental run, the level of the probe was fixed and the level of the masker was varied adaptively using a 3-up, 1-down stepping rule. When the suppressor was present, its level was fixed within a run. Several levels of the probe, ranging from 45 to 80 dB SPL (for one subject a 40-dB SPL probe was also included) were used to obtain growth-of-masking functions. The growth-of-masking functions were obtained for five suppressor levels between 40 and 80 dB SPL in steps of 10 dB. A band of noise extending from 3 to 5 kHz was presented simultaneously with the probe. This was done to reduce the advantage of using spread of excitation toward higher frequencies for probe detection in the masker-alone condition (the spread of excitation was at least partially masked by the excitation produced by the suppressor, in the masker-plus-suppressor condition). For each probe level, the noise level was set 15 dB below the level that would mask the probe. Additionally, a 50-dB-SPL 1-ERB band of noise centered at 4 kHz was presented to the non-test ear for the duration of the masker, to help listeners temporally resolve the probe from the masker. All stimuli were gated with 5-ms raised-cosine ramps. Visual feedback indicating the correct response was provided after each trial. Final threshold estimates were obtained by averaging threshold masker levels obtained in six separate runs.

Three normal-hearing listeners participated in the study.

2 Results

Figure 1 shows data for the individual listeners. Masker levels necessary to mask the probe are plotted as a function of the level of the probe for the masker-alone (open circles) and the masker-plus-suppressor (filled symbols) conditions. These plots will be referred to as growth-of-maskability (GMB) functions. Different symbols in each panel represent results for a different suppressor level. For lower probe levels, all three subjects required higher masker levels to mask the probe

when the suppressor was present, consistent with the notion that the suppressor had the effect of reducing the gain. At higher probe levels, similar masker levels had to be used in the presence and absence of the suppressor. In some cases, lower masker levels were sufficient to mask the probe with the suppressor present (subjects S2 and S3 at two highest probe levels). Possible explanations for this effect will be discussed below.

Generally, the GMB functions measured with a suppressor present have a shallower slope over a wide range of probe levels than the GMB functions measured without a suppressor. For a given probe level, the masker levels necessary to just mask the probe in the presence and absence of a suppressor presumably produce the same internal response. Thus, any increase in probe level would require the same increase in internal response to the masker in the presence and absence of a suppressor. A shallower slope of the GMB function implies that to produce the same increase in internal response, a smaller increase in masker level was required when the suppressor was present. This, in turn, implies a steeper growth of response to the masker in the presence of a suppressor, consistent with the notion that the BM response to a CF tone is more linear in the presence of an off-frequency suppressor.

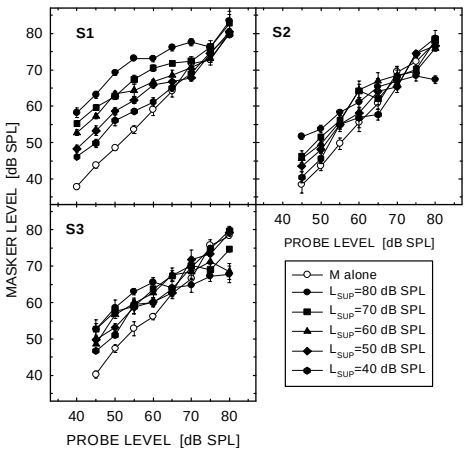


Fig. 1. Masker levels necessary to mask the probe in the presence (filled symbols) and absence (open symbols) of a suppressor.

Figure 2 shows the amount of suppression computed by taking the difference between masker levels required to mask a fixed-level probe, in the presence and absence of the suppressor. Different symbols represent the estimated suppression for different suppressor levels. For a fixed suppressor level, the magnitude of suppression was the largest for low masker/suppressee levels. As the level of the masker increased, the estimated magnitude of suppression decreased. At the highest levels, negative magnitudes of suppression were obtained,

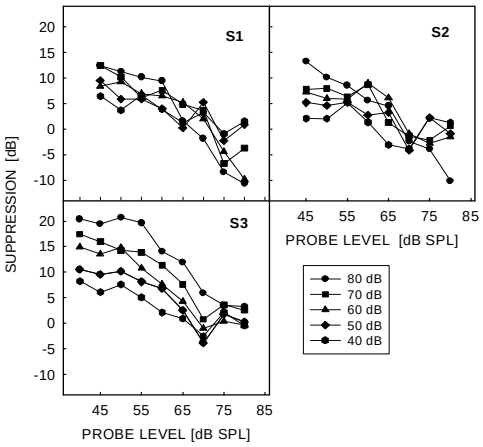


Fig. 2. The estimated magnitude of suppression.

indicating that at these levels the suppressor aided forward masking.

For subject S1, the magnitude of suppression increased with increasing level of the suppressor, except for the highest probe levels. The other two subjects show the same result for the lowest probe (and masker) levels, but this trend is not clearly represented by their data for medium and high levels, mainly because they generally exhibited much less suppression than S1.

3 Model

Since all the masked thresholds were measured for the same short probe-masker delay, a simple power spectrum model representing peripheral processing can be used to account for the data under the assumption that the temporal integrator that is operating at a higher processing stage is linear. The most commonly used power spectrum model is based on a sum of ROEX filters. The recent version of this model proposed by Glasberg and Moore (2000) included nonlinear processing. Although their model can be used to successfully predict data from simultaneous masking, it should not be used to account for forward masking by complex maskers because it does not allow for separation of excitatory and suppressive masking. An alternative approach proposed here assumes the presence of two cascaded filters in the peripheral processing. The first, "passive filter", is linear and is assumed to control the gain and bandwidth of the second, level-dependent "active filter". The idea of modeling the peripheral processing at a given place on the BM by two cascaded filters with a nonlinearity sandwiched in-between has been used in the past (e.g., Pfeiffer 1970; Duifhuis 1980; Plack, Oxenham, and Drga 2002). The difference between the present approach and the past studies is that in our model, the second filter has a bandwidth that co-varies with the gain applied by this filter to the input stimulus. A filter with variable gain and bandwidth was used in the auditory-nerve model of Zhang, Heinz, Bruce, and Carney (2001). The bandwidth of the active filter is controlled by making the damping time of this filter a function of gain, which in turn is a function of the total level at the output of the passive filter. The function is such that the damping time decreases as the output of the passive filter increases. The decreasing time constant results in a broadening of the active filter with increasing level.

Predictions by the proposed model are shown in Fig. 3 along with the data for listeners S1 and S2 (for clarity, predictions are shown only for selected levels). Data from S3 were generally similar to those from S2, and therefore predictions for S3 are not shown here. The predictions were produced with two 3rd-order gammatone filters. The frequency response of the passive filter was

$$G_{pf}(2\pi f) = \frac{(n-1)!}{2[1 + j2\pi \cdot \tau_{pf} \cdot (f - cf)]^n}, \quad (1)$$

where n is the order of the filter (n=3), τ_{pf} is the time constant describing the damping of the passive filter. The frequency response of the active filter was

$$G_{af}(2\pi f) = \frac{g \cdot (n-1)!}{2[1 + j2\pi \cdot \tau_{af}(g) \cdot (f - cf)]^n}, \quad (2)$$

where g is a function of the level at the output of the passive filter and is expressed in linear units. The dependence of time constant τ_{af} on the output of the passive filter was defined through the gain function using the following formula

$$\tau_{af}(g) = (g^{1/12}) \cdot \tau_{pf} \,. \tag{3}$$

This formula was chosen arbitrarily to ensure that when there is no gain applied by the system (g=1), the active filter is identical to the passive filter and that for large gain values, the bandwidth of the active filter is not unrealistically narrow.

In pilot simulations, the gain function proposed by Glasberg and Moore (2000) was used. Although their function led to reasonable qualitative agreement with our data, the quantitative agreement was poor for lower probe levels. More accurate predictions were obtained using a three-segment function to describe the response growth on the BM, with each segment represented by a straight line. The first and third segment had a slope of 1, and the middle segment had a slope less than 1 and represented compressive nonlinearity. The following parameters defining the nonlinear input-output (I/O) function were changed iteratively to obtain predictions shown in Fig. 3: the maximum gain, G_{max}, which is defined as the gain applied to the 0-dB SPL input stimulus; the breaking points (x-values) between the linear and compressive sections of the I/O function, B_1 and B_2; and the slope of the compressive segment, α. The gain function in dB was derived for the iterated parameters of the I/O function. It was then converted to gain g in linear units. Table I shows values of all the parameters used in the model and the ERB values for the active filter computed for $g=G_{max}$.

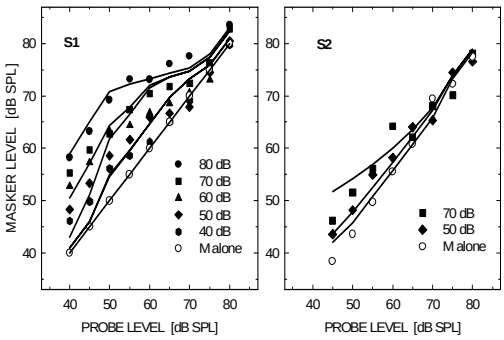


Fig. 3. Model predictions and data for S1 and S2.

Subj.	τ_{pf}[ms]	G_{max}[dB]	B_1 [dB SPL]	B_2 [dB SPL]	α [dB/dB]	ERB_{af}[Hz]
S1	0.4	60	45	70	0.18	264
S2	0.4	35	40	70	0.4	335

Table I. Parameter values used to produce model predictions shown in Fig. 3.

The predictions are reasonable although their agreement with the data could be improved by optimizing the choice of the function relating the damping time of the active filter to the gain function.

4 Discussion

The experimental data presented in Fig. 1 revealed that the response to a 4-kHz tone can be reduced in the presence of a 4.8-kHz tone when the tones are presented at appropriate levels. Generally, for a fixed-level suppressor, the difference between the response to the suppressee presented without and with the suppressor is greater when the suppressee level is low. As the suppressee level increases, that difference becomes smaller indicating less suppression of the response. This finding disagrees with the observation by Duifhuis (1980) that suppression increases as the overall stimulation level increases. Duifhuis varied the probe level to find its forward-masked thresholds for a fixed-level masker presented with and without the suppressor, and his estimated suppression was, therefore, affected by the nonlinear processing of the probe. The decreasing suppression with increasing suppressee level is consistent with the mechanical data of Ruggero *et al*. (1992).

At the highest probe levels, negative suppression was observed for subjects S2 and S3, when the suppressor level was high. One explanation for this result is in terms of spread of excitation. Since the probe has shorter duration than the masker, it produces a broader excitation pattern due to energy splatter across frequency. If the probe energy falling into higher-frequency channels contributes to probe detection, the presence of the high-level higher-frequency suppressor may reduce the detectability of the probe. In this case, lower masker levels would be required to mask the probe. To test the plausibility of this explanation, thresholds for detecting the probe in forward masking were measured in the presence of the suppressor presented alone. The highest level of the suppressor used in this study (80 dB SPL) elevated threshold for detecting the probe by 7-8 dB. The usability of spread of excitation was also limited by the bandpass noise presented simultaneously with the probe. Therefore, it seems unlikely that the suppressor would have such a big effect on the detectability of the probe when the probe is presented at levels that are well above its threshold. It is possible that some other, non-sensory, factors played a role when the probe and the suppressor levels were high (e.g. a greater perceptual similarity between the probe and the suppressor).

Generally, the data are consistent with the notion that the effect of a suppressor is to reduce the gain applied at the place corresponding to the frequency of the suppressee. A power spectrum model assuming two cascaded filters, one (linear) representing passive processes, and the other (level dependent) representing the active process in the cochlea, produces reasonable predictions for the level effects. The variable bandwidth of the active filter allows for suppressive only and combined excitatory and suppressive masking. The parameter values used to produce predictions shown in Fig. 3 are reasonable and are in agreement with the values that were used in other studies of the effects of nonlinear processing on masking. The ratio of the ERB and center frequency for the active filter was 0.07 for S1 and 0.08 for S2. These values are in agreement with the ratio of 0.08 used in the dual-resonance nonlinear window model of Plack *et al.* (2002).

5 Conclusions

The following conclusions may be drawn from this psychophysical experiment:

(1) The response to a tone becomes less compressive in the presence of a suppressing tone, consistent with the mechanical data of Ruggero *et al*. (1992).

(2) The magnitude of suppression for a fixed-level suppressor decreases with increasing level of the suppressed tone.

(3) The observed level effects are consistent with a reduction of gain that is proportional to the amount of gain applied to the test tone in the absence of a suppressor.

(4) A power spectrum model implementing two filters, a linear filter followed by a level-dependent filter with varying gain and bandwidth, reasonably predicts the observed level effects in psychophysical suppression.

Acknowledgments

This work was supported by Grant No. DC00683 from NIDCD.

References

Duifhuis, H. (1976) Cochlear nonlinearity and second filter: Possible mechanism and implications. J. Acoust. Soc. Am. 59, 408-423.

Duifhuis, H. (1980) Level effects in psychophysical two-tone suppression. J. Acoust. Soc. Am. 67, 914-927.

Glasberg, B. R. and Moore, B. C. J. (2000) Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. J. Acoust. Soc. Am. 108, 2318-2328.

Houtgast, T. (1972) Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Am. 29, 168-179.

Javel, E., Geisler, C. D., and Ravindran, A. (1978) Two-tone suppression in auditory nerve of the cat: rate-intensity and temporal analysis. J. Acoust. Soc. Am. 63, 1093-1104.

Pfeiffer, R. R. (1970) A model for two-tone inhibition of single cochlear nerve fibers. J. Acoust. Soc. Am. 48, 1373-1378.

Plack, C. J., Oxenham, A. J., and Drga, V. (2002) Linear and nonlinear processes in temporal masking. Acta Acustica/Acustica 88, 348-358.

Ruggero, M. A., Robles, L., and Rich, N. C. (1992) Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression. J. Neurophysiology 68, 1087-1099.

Shannon, R. V. (1976) Two-tone unmasking and suppression in a forward-masking situation. J. Acoust. Soc. Am. 59, 1460-1470.

Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (2001) A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. J. Acoust. Soc. Am. 109, 648-670.

The function(s) of the medial olivocochlear efferent system in hearing

David W. Smith[1], E. Christopher Kirk[2], and Emily Buss[3]

[1] Division of OHNS, Duke University Medical Center, Durham, david.w.smith@duke.edu
[2] Department of Biological Anthropology and Anatomy, Duke University Medical Center, Durham, eck@duke.edu
[3] Department of OHNS, University of North Carolina at Chapel Hill, ebuss@med.unc.edu

1 Introduction

The physiological consequences of medial olivocochlear (MOC) activity are typically described in terms of suppression of activity at the auditory periphery. Depending on the stimulus conditions employed, however, the effects of the MOC suppression may be shown to increase, or *enhance*, certain physiological responses (cf. Kawase and Liberman 1993; Winslow and Sachs 1988). We now understand that the observed efferent effects result from attenuations in the driven mechanical response of OHCs (i.e., reductions in the "gain" of the cochlear amplifier). When stimulated, the MOC activity reduces the resistance of the OHC basolateral membrane, effectively shunting the OHC receptor potential and reducing the "gain" of the cochlear amplifier (cf. Geisler 1998).

The ultimate goal, of course, is to understand how MOC activity influences what animals, especially humans, *hear*. In general cutting the MOC has had few perceptual consequences in non-human animals or human subjects after they undergo a vestibular neurotomy. Stimulus detection and the discrimination of frequency or intensity in quiet are unaffected by surgical de-efferentation (Igarashi, Alford, Nakai and Alford 1972; Scharf, Magnam and Chays 1997; Trahiotis and Elliot 1970; Zeng, Martino, Linthicum and Soli 2000).

The most consistent reports of auditory deficits that follow efferent lesions are associated with the detection and discrimination of signals in noise. In noise, intensity discrimination thresholds, as well as the vowel discrimination, are worsened following MOC lesions (Dewson 1968; Heinz, Stiles and May 1998; May and McQuone 1995; Zeng et al. 2000). The psychophysical "overshoot" effect is also reduced after de-efferentation (Zeng et al. 2000) and "central masking" for continuous contralateral noise maskers is significantly reduced after the MOC tracts are cut (Smith, Turner and Henson 2000). By contrast, Scharf et al. (1997) showed no noise-related threshold deficits after vestibular neurotomy in Ménière's patients.

The paucity and inconsistency of psychophysical data has led to few coherent theories concerning the evolved, biological function of the MOC system. Two roles are most frequently attributed to the MOC. First, a number of reports show that the MOC system suppresses the response of the cochlea to concurrent noise and that, in so doing, helps to "unmask" transient acoustic stimuli (cf. Kawase and Liberman 1993; Winslow and Sachs 1988). These data support the hypothesis that the MOC system evolved as a mechanism to improve the signal-to-noise ratio of the encoded signal and the dynamic range of the peripheral auditory system (Winslow and Sachs 1988). This function is consistent also with the literature showing psychophysical deficits in processing signals in noise following de-efferentation (May and McQuone 1995; Smith et al. 2000; Zeng et al. 2000). Secondly, a number of investigations (cf. Rajan 1995) have shown that noise-induced threshold shifts resulting from high-frequency tone exposures are *increased* after deactivation of the MOC system, and are *decreased* when the MOC system is stimulated. Implicit in these studies is the belief that the MOC system serves to protect the ear against damage from intense noise exposures.

This chapter proposes a coherent conceptual description of the function of the MOC system in *hearing*. It employs an evolutionary perspective in reviewing the extant efferent literature in the context of acoustic environmental conditions that likely gave rise to the development of the MOC system.

2 The MOC as a protective mechanism

The hearing organs of all vertebrates receive an efferent innervation (Roberts and Meredith 1992), but a differentiated MOC efferent system appears to be a derived feature of mammals (Kirk and Smith, *in press*). While only six orders of mammals have been definitively shown to possess an MOC system (primates, bats, carnivorans, rodents, and two marsupial orders), the common ancestor of these groups lived approximately 170 million years ago (Kirk and Smith, *in press*; Kumar and Hedges 1998). This early date of divergence suggests that the MOC system is an ancient feature of mammals that may have evolved concomitantly with the other specialized features of the mammalian cochlea (elongation of the auditory epithelium, differentiated hair cell populations, etc.). Furthermore, of the numerous mammals that have been studied, the only species that appear to have lost an MOC innervation of the cochlea (two species of bats and the naked mole rat) demonstrate auditory systems that are highly specialized for either high- or low-frequency hearing.

To test the proposed biological role of the MOC system (i.e., the function(s) for which the MOC system evolved), it is necessary to examine the acoustic stimuli that mammals encounter under natural conditions. Because the MOC system is widespread among mammals that occupy a diverse range of habitats, one would expect the selective factors that initially favored the evolution of the MOC system to be nearly universal. Accordingly, the unmasking hypothesis predicts that natural acoustic environments (i.e., those that lack significant anthropogenic noise sources)

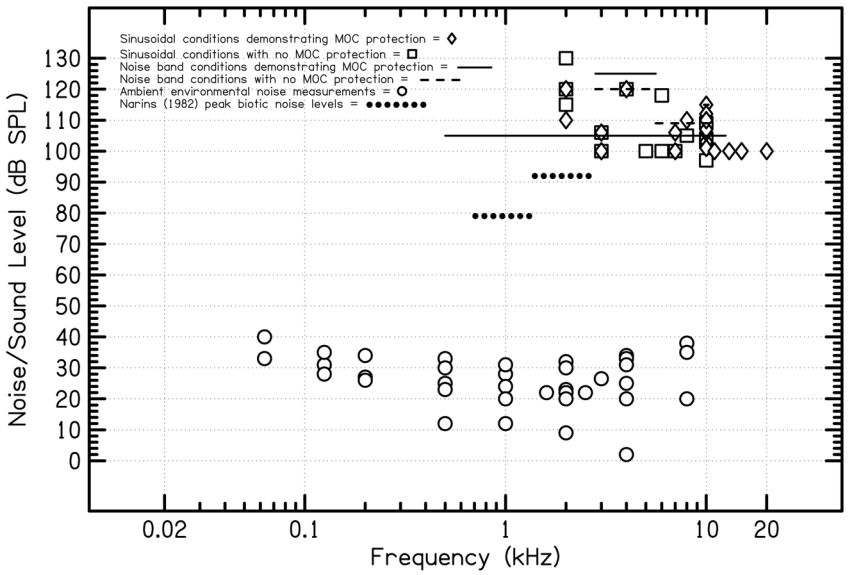


Fig. 1. Comparison of the stimulus conditions employed in experimental studies of MOC-mediated protection from acoustic trauma and naturally occurring ambient noise conditions (from Kirk and Smith, in press).

should exhibit numerous sources of acoustic masking stimuli. The "protection hypothesis," by contrast, predicts that natural acoustic environments contain noise sources that are sufficiently intense to *potentially damage* the cochlea. In other words, if traumatic acoustic stimuli are very rare or nonexistent under natural conditions, it is unlikely that the MOC system could have evolved to fulfill a protective function.

A survey of the available literature documenting noise levels in 23 localities that are largely free of anthropogenic noise clearly demonstrates one irrefutable point: the natural world is a very noisy place (Kirk and Smith *in press*). Whether the noise sources are primarily *abiotic* (e.g., wind, rain, and running water) or *biotic* (e.g., bird song, insect stridulations, and frog choruses), low to moderate noise levels ranging up to about 60 dB SPL appear to be a universal feature of natural acoustic environments (Fig. 1.). In rare instances (most notably in the presence of chorusing frogs and insects), ambient noise levels may be very intense, ranging up to about 90 dB SPL. Nonetheless, based on the extensive acoustic trauma literature (Borg, Canlon and Engström 1995), even the most intense sources of sustained natural noise are probably *insufficient* to activate the putative protective effects of the MOC system. Most experiments that have demonstrated a protective role for the MOC system used sinusoidal traumatizing stimuli (an acoustic waveform that does not exist in nature) in excess of 100 dB SPL (Kirk and Smith, *in press*). By contrast, nearly all noise present in natural acoustic environments is broadband and substantially less intense (Fig. 1.). These observations suggest that the MOC system is unlikely to play a protective role under natural conditions, but combined with the

unmasking literature, they provide strong indirect support for the hypothesis that the MOC system evolved to help detect transient acoustic stimuli in the presence of sustained masking noise.

3 The role of the MOC in optimizing the cochlea for the perception of transient signals

There are many physiological and psychophysical data showing that the MOC system can reduce the effects of concurrent noise on the reception and processing of target stimuli (cf., May and McQuone 1995; Winslow and Sachs 1988). The time course and manner in which this is accomplished, however, has only recently become apparent (Liberman, Puria and Guinan 1996). The Liberman study showed in cats that the level of the $2f_1$-f_2 distortion product otoacoustic emission (DPOAE) adapts rapidly after the onset of the primary tones. This rapid adaptation, illustrated in Fig. 2, with a time constant of 50-100 ms, is lost when the MOC tracts are sectioned. This response shows that the MOC selectively passes brief, transient stimuli while acting to suppress the OHCs response to relatively long duration (>50-100 ms) stimuli. More recent studies have shown that the rapid adaptation response is qualitatively similar in humans (Bassim, Miller, Buss and Smith, *in press*; Kim, Dorn, Neely and Gorga 2001).

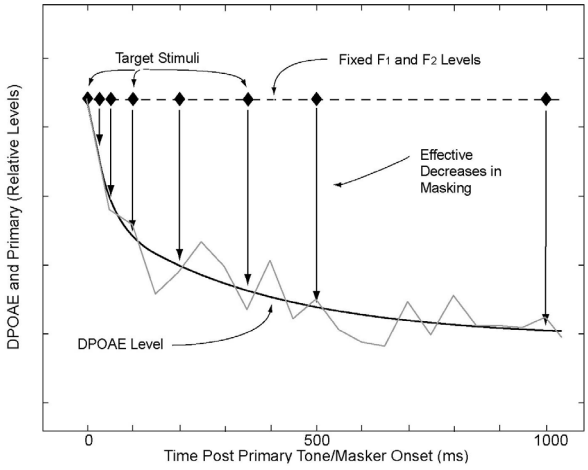


Fig. 2. Relationship between rapid adaptation and "overshoot." Averaged DPOAE waveform (light, gray line) and a two-exponential has been fitted to the curve (solid, dark line, labeled DPOAE level). This example, recorded in a human listener, shows approximately 3 dB of rapid adaptation. (See text for full description.)

Since this chapter seeks to describe the function of the MOC system in hearing, and since the animal's response to sound is behavioral, whether this physiological adaptation process as a psychophysical analog is of significant interest. In Fig. 2 the difference between the fixed level of the acoustic primaries and the adapted DPOAE level represents the predicted decrease in the masking of a transient signal

produced by MOC-mediated rapid adaptation of the long-duration masker. This relationship predicts that relative thresholds would *decrease* as the relative onset of the target signal is delayed relative to the onset of the masker (which, owing to its duration, adapts).

As should be immediately apparent, this stimulus paradigm is the same as that used to measure the psychophysical "overshoot" phenomenon, and the predicted results are observed (Zwicker 1965). Indeed, overshoot has been explained physiologically as being related to adaptive (though not MOC-based) mechanisms (Smith 1979). Interestingly, overshoot effects can be as great as 10-20 dB (Zwicker 1965), far greater effects than any other MOC-based phenomenon.

Confirmation that overshoot in human listeners is dependent on normal MOC activity was provided recently by Zeng et al. (2000), who showed that overshoot was significantly reduced in most subjects after vestibular neurotomy surgery for vertigo. The investigators, however, did not offer a physiological explanation for their findings.

4 Conclusions and implications for the study of MOC function

The biological role of the MOC system in hearing can be surmised based on the environmental acoustic conditions that likely supported the evolution of the system (Kirk and Smith *in press*). The maintained presence of the MOC system in nearly all mammals suggests that the selective factors that favored its evolution must be widespread. The literature describing natural acoustic environments shows, first, that extreme acoustic conditions sufficient to support evolution of an MOC-based protective mechanism are rare and are discontinuously distributed. These findings argue that protection from acoustic trauma is *not* a primary function of the MOC system. Second, our review also shows that the presence of low- to moderate-intensity broadband noise is a universal feature of natural acoustic environments. This fact suggests that mammals have experienced significant selective pressure to segregate biologically relevant acoustic signals from irrelevant background noise. In other words, selective pressures would place a very great value on a mechanism that could assist in *unmasking* transient stimuli in the presence of concurrent background noise. In the cochlea, the MOC system uses *adaptation* – a general feature of all sensory systems - to accentuate the properties of changing, versus constant, acoustic stimulation. Other sensory systems accomplish adaptation by either neural (e.g., amacrine cell interactions at the bipolar cell level of the retina) or non-neural (e.g., rapidly adapting Pacinian corpuscles in touch) mechanisms. The MOC adapts by altering the active behavior of OHCs.

For animals in natural environments, perhaps the most important transient acoustic signals are those used to locate predators and prey in three-dimensional space. Indeed, it has been argued that this task has been responsible for the development of the auditory system (Masterton, Heffner and Ravizza 1969). *It is our contention that the biological role of the MOC system is to play a critical role in sound localization.* Under naturally occurring noise conditions, the MOC system suppresses the OHC response to the constant low-level ambient noise, thereby

unmasking, or lowering, the detection threshold for the detection of transient target stimuli. Precisely this effect was first shown by Winslow and Sachs (1988).

It also should be noted that the source of the noise that might otherwise act to mask the transient stimulus need not be external in origin. Indeed, because the ear is sufficiently sensitive, the receptor is capable of responding to sounds generated by blood flow, respiration or myogenic activity (Cazals and Huang 1996). The MOC system will act to suppress a response to internal as well as external noise in order to enhance responses to transient stimuli (Kawase and Liberman 1993). Additionally, unattended stimuli, within or across sensory modalities also function like noise to hinder the reception of biologically relevant signals. While there is not space in this presentation to address this issue in detail, a number of studies have provided evidence that the MOC system might play a role in this process (Bassim and Smith *unpublished observations*; May, Prosen, Weiss and Vetter 2002; Scharf et al. 1997).

That the MOC system exerts little direct influence over transient stimuli also has significant implications for parametric studies, both physiological and psychophysical, of MOC function. Experimental studies employing brief target stimuli (<~50 ms, unless delivered at a sufficiently high pulse rate) are unlikely to show significant MOC effects, especially when measured in quiet backgrounds. For example, cutting the MOC and measuring "tonic" effects of de-efferentation on brief signals in quiet, or measuring contralateral suppression of compound action potentials elicited by tone pips in quiet, are unlikely to provide much direct evidence for, or at best, underestimate their actual influence on cochlear function.

The overshoot phenomenon, however, takes unique advantage of the critical ("natural") stimulus features underlying MOC evolution (the detection of transient signals in noise). For this reason, the paradigm might represent the ideal stimulus paradigm for investigating normal MOC function as it highlights the value of employing more natural stimulus conditions in characterizing *function.*

Acknowledgements

Support for this work was provided by the NIH-NIDCD grant DC01692 to DWS and James B. Duke and NSF graduate fellowships to ECK

References

Borg, E., Canlon, B. and Engström, B. (1995) Noise induced hearing loss: literature review and experiments in rabbits. Scand. Audiol. 24(suppl.) 40, 1-147.

Cazals, Y. and Huang, Z.W. (1996) Average spectrum of cochlear activity: A possible synchronized firing, its olivo-cochlear feedback and alterations under anesthesia. Hear. Res. 101, 81-92.

Dewson, J.H., III. (1968) Efferent olivocochlear bundle: Some relationships to stimulus discrimination in noise. J. Neurophysiol. 31, 122-130.

Geisler, C.D. (1998) *From Sound to Synapse: The Physiology of the Mammalian Ear*. Oxford University Press, New York.

Guinan, J.J. Jr. (1996) Physiology of olivocochlear efferents. In: P. Dallos, A.N. Popper, R.R. Fay (Eds.), *The Cochlea*. Springer-Verlag, New York, pp. 435-502.

Heinz, R.D., Stiles, P. and May, B.J. (1998) Effects of bilateral olivocochlear lesions on vowel formant discrimination in cats. Hear. Res. 116, 10-20.

Igarashi, M., Alford, B.R., Nakai, Y. and Gordon, W.P. (1972) Behavioral auditory function after transection of crossed olivocochlear bundle in cat . I. Pre-tone threshold and perceptual signal-to-noise ratio. Acta Otolaryngol. 73, 455-466.

Kawase, T. and Liberman, M.C. (1993) Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones. J. Neurophysiol. 70, 2519-2532.

Kim, D.O., Dorn, P.A., Neely, S.T. and Gorga, M.P. (2001) Adaptation of distortion product otoacoustic emission in humans. J. Assoc. Res. Otolaryngol. 2, 31-40.

Kirk, E.C. and Smith, D.W. (*in press*) Protection from acoustic trauma is not a primary function of the medial olivocochlear system. J. Assoc. Res. Otolaryngol.

Kumar, S. and Hedges, S.B. (1998) A molecular timescale for vertebrate evolution. Nature 392, 917-920.

Liberman, M.C., Puria, S. and Guinan, J.J. Jr. (1996) The ipsilaterally evoked olivocochlear reflex causes rapid adaptation of the 2*f*1-*f*2 distortion product otoacoustic emission. J. Acoust. Soc. Am. 99, 2572-3584.

Lima da Costa, D., Chibois, A., Erre, J.-P., Blanchet, C., Charlet de Sauvage, R. and Aran, J.-M. (1997) Fast, slow, and steady-state effects of contralateral acoustic activation of the medial olivocochlear efferent system in awake guinea pigs: action of gentamicin. J. Neurophysiol. 78, 1826-1836.

May, B.J. and McQuone, S.J. (1995) Effects of bilateral olivocochlear lesions on pure-tone intensity discrimination in cats. Aud. Neurosci. 1, 385-400.

May, B.J., Prosen, C.A., Weiss, D. and Vetter, D. (2002) Behavioral investigation of some possible effects of the central olivocochlear pathways in transgenic mice. Hear. Res. 171, 142-157.

Masterton, B., Heffner, H. and Ravizza, R. (1969) The evolution of human hearing. J. Acoust. Soc. Am. 45, 966-985.

Rajan, R. (1995) Frequency and loss dependence of the protective effects of the olivocochlear pathway in cats. J. Neurophysiol. 74, 598-615.

Roberts, B.L. and Meredith, G.E. (1992) The efferent innervation of the ear: variations on an enigma. In: D.B. Webster, R.R. Fay and A.N. Popper (Eds.), *The Evolutionary Biology of Hearing*. Springer-Verlag, New York, pp. 185-210.

Scharf, B., Magnan, J. and Chays, A. (1997) On the role of the olivocochlear bundle in hearing: 16 case studies. Hear. Res. 103, 101-122.

Smith, D.W., Turner, D.A. and Henson, M.M. (2000) Psychophysical correlates of contralateral efferent suppression. I. The role of the medial olivocochlear system in "central masking" in non-human primates. J. Acoust. Soc. Am. 107, 933-941.

Smith, R.L. (1979) Adaptation, saturation, and physiological masking in single auditory-nerve fibers. J. Acoust. Soc. Am. 65, 166-178.

Trahiotis, C. and Elliott, D.N. (1970) Behavioral investigation of some possible effects of sectioning the crossed olivocochlear bundle. J. Acoust. Soc. Am. 47, 592-596.

Winslow, R.L. and Sachs, M.B. (1988) Single tone intensity discrimination based on auditory-nerve responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle. Hear. Res. 35, 165-190.

Zeng, F.-G., Martino, K.M., Linthicum, F.H. and Soli, S.D. (2000) Auditory perception in vestibular neurectomy subjects. Hear. Res. 142, 102-112.

Zwicker, E. (1965) Temporal effects in simultaneous masking by white-noise bursts. J. Acoust. Soc. Am. 37, 653-663.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

A computational model of cochlear nucleus neurons

Katuhiro Maki[1] and Masato Akagi[2]

[1] Human and Information Science Laboratory, NTT Communication Science Laboratories, NTT Corporation, maki@avg.brl.ntt.co.jp
[2] Department of Information Processing, Graduate School of Information Science, Japan Advanced Institute of Science and Technology, akagi@jaist.ac.jp

1 Introduction

The cochlear nucleus (CN), the first processing center in the auditory central nervous system, consists of diverse neuronal types that differ in their patterns of temporal response to short tone bursts (Blackburn and Sachs 1989). Several functional models have been proposed and have demonstrated some degree of success in simulating the temporal response patterns of CN neurons (Arle and Kim 1991; Banks and Sachs 1991; Hewitt and Meddis 1993; Cai, Walsh and McGee 1997; Levy and Kipke 1997; Eriksson and Robert 1999). Those models, however, involve solving systems of differential equations based on the Hodgkin-Huxley equation (Hodgkin and Huxley 1952) and therefore tend to have complex structures and many parameters (typically > 10). A simple model with relatively few parameters would be useful not only for future implementation in a large-scale computer simulation, but also for estimating the underlying mechanisms generating the response of CN neurons to a variety of complex sounds, such as vocalizations. In this paper, we propose a model that represents the synaptic transduction mechanisms and membrane properties with stochastic processes. The model is expressed with only eight parameters. By fitting appropriate parameter values, the model can successfully simulate various temporal response types found in the anteroventral CN (AVCN), such as the primary-like (with or without notch), the chopper (regular and irregular), and the onset types, as well as the degree of phase-locking and response latencies.

2 Model

The input to the model is a train of discrete pulses that simulates inputs from multiple auditory nerve fibers (ANFs). The arrival time of the jth pulse in the train of pulses from the ith ANF input is denoted by t_{ij}. The post-synaptic potential (PSP) of a CN neuron at time t is modeled by:

$$V(t) = \sum_{i=1}^{N} \sum_{\{j|t_{ij}+t_c<t\}} a_i(t - t_c - t_{ij})\, e^{-(t-t_c-t_{ij})/\tau_i} \tag{1}$$

where N($\in \mathbf{Z}^+$) is the total number of input pulses, and t_c is random jitter from the normal distribution with a mean μ_c and a variance ${\sigma_c}^2$ ($t_c \sim N\left(\mu_c, {\sigma_c}^2\right)$). The values of μ_c and σ_c, respectively, determine the response latency and the degree of phase-locking of the CN neurons. The $\tau_i(\in \mathbf{R}^+)$ is the time constant of a PSP elicited by a single input pulse. The a_i is a coefficient of synaptic strength that determines the magnitude of PSP elicited by a single pulse input. A positive a_i indicates that the inputs is excitatory, and vice versa.

The output S at time t is represented by,

$$S(t) = \begin{cases} 1 & V(t) \geqq U(\alpha, \beta) \text{ and } S(t') = 0 \text{ for } t' \in [t - t_r, t], \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The output of the modeled neuron is a train of all-or-none (1 or 0) action potentials with unit amplitudes. Action potentials of the model are generated when the membrane potential V crosses a threshold U. The threshold U is a random variable from a uniform distribution in the range from α to β. The modeled neuron does not generate a spike during the refractory period after the previous spike. Refractory period t_r is modeled by a random variable from a normal distribution with a mean μ_r and a variance ${\sigma_r}^2$ ($t_r \sim N\left(\mu_r, {\sigma_r}^2\right)$).

3 Methods of evaluation

Figure 1 shows classification results for actual AVCN units (Blackburn and Sachs 1989). The model was evaluated with respect to its ability to reproduce the temporal response patterns in Fig. 1 and phase-locking properties of actual AVCN units. As in physiological experiments (Blackburn and Sachs 1989), two stimulus durations were used in this study: short tone bursts (STBs) of 25 ms durations with 1.6 ms rise and fall times; and long tone bursts (LTBs) of 400 ms duration with 10 ms rise and fall times. Post-stimulus time histograms (PSTHs) and first-spike latency histograms were computed based on 200 presentations of the STBs at the best frequency (BF), and were represented with 0.2 ms time bins. In evaluating PSTHs, the BF of the modeled CN neuron was fixed at 4.8 kHz. All simulations were run at a sampling rate of 48 kHz.

3.1 Inputs to the modeled CN neurons

Inputs of the modeled CN neuron, i.e., discrete pulse trains, were simulated responses of ANFs at the same BF of the modeled CN neuron, produced by an auditory peripheral model (Maki, Akagi and Hirota 1998). Temporal response patterns and phase-locking properties of simulated AN responses used as inputs to the modeled CN neurons are shown in Figs. 2(a) and 2(b), respectively.

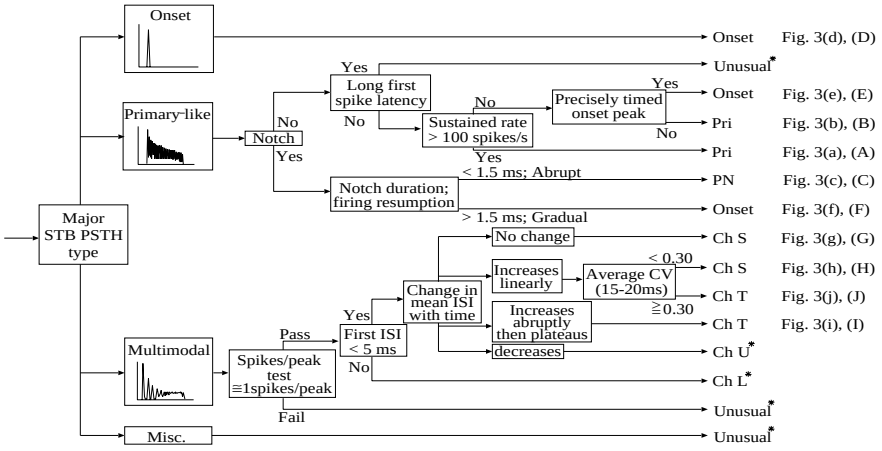


Fig. 1. Classification scheme proposed by Blackburn and Sachs for categorizing AVCN units (Blackburn and Sachs 1989). Graphs (*lefthand column*) depict major PSTH categories from which defined populations are refined. "Primary-like" PSTHs resemble those of auditory nerve fibers (ANFs). "Onset" PSTHs have a sharp peak at stimulus onset, followed by little or no sustained activity. Units are placed in the "onset" limb only if their sustained rates are $\leq$ 25 spikes/s. "Chopper" PSTHs have regularly spaced peaks of discharge whose period is not related to the stimulus frequency. Response types discarded in the study of Blackburn and Sachs (*) were also excluded in present study.

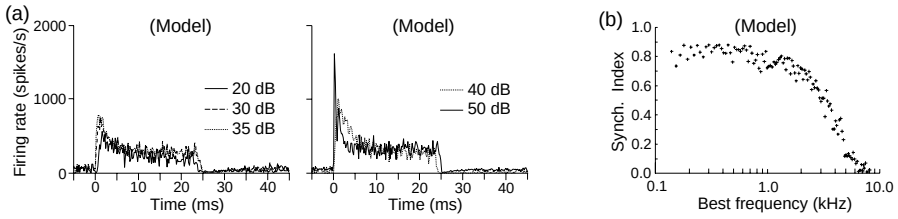


Fig. 2. Temporal response patterns and phase-locking properties of inputs to the CN model. (a) PSTHs of simulated ANF for five stimulus levels (20, 30, 35, 40 and 50 dB). PSTHs are represented with 0.2-ms time bins. (b) Synchronization index (Goldberg and Brown 1969) of simulated AN firing as a function of BF.

In Fig. 2(a), PSTHs calculated from the simulated ANF firing show similar temporal properties to actual ANFs (Johnson 1980). That is, there is an initial rapid increase of firing rate at the response onset, which strongly depends on the stimulus level, followed by a sustained discharge at a lower spike rate during stimulation. The simulated ANF responses show that the degree of phase-locking to BF tones decreases gradually with BF in the range from 0.1 to 1.5 - 2.0 kHz, and then drops sharply to reach the noise level above 5.0 - 6.0 kHz. This tendancy is also observed in actual ANFs (Johnson 1980).

4 Results

We adjusted the parameters of the CN model and selected the CN model inputs (i.e., simulated AN responses in Fig. 2) so that the CN model simulates the ten kinds of response patterns in Fig. 1. The parameter values used in this study are summarized in Table 1.

Table1. The parameter of the CN model for simulating each response type. "N" shows the total number of simulated AN inputs. "dB" indicates the stimulus sound level. Important parameter boundaries for differentiating the response types are indicated by vertical lines.

Type (Fig.)	Pri (3A)	Pri (3B)	PN (3C)	Onset (3D)	Onset (3E)	Onset (3F)	Ch S (3G)	Ch S (3H)	Ch T (3I)	Ch T (3J)
a_i ([value]$\times 10^6$)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
τ_i ([value]$\times 10^{-6}$)	6.7	6.7	4.8	6.7	4.8	4.8	11	25	66	50
μ_c (s)	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
σ_c ([value]$\times 10^{-6}$ s)	5	5	4	5	5	4	17	14	16	15
μ_r ([value]$\times 10^{-4}$s)	8	8	12	15	16	30	18	21	13	16
σ_r ([value]$\times \mu_r$ s)	0.050	0.025	0.025	0.025	0.003	0.230	0.025	0.025	0.025	0.040
β	5.38	7.69	2.81	3.09	2.13	0.90	12.9	18.0	243	153
α	0.84	1.00	0.17	2.75	0.49	0.57	2.29	18.0	96.0	33.3
N	30	30	30	50	50	65	50	30	50	30
dB	20	30	50	50	50	50	50	35	40	40

4.1 Response types

The simulation results for the CN model are shown in Figs. 3(A)-(J). Comparable physiological data are shown in Figs. 3(a)-(j). All the responses of the model meet the classification requirements of the decision tree in Fig. 1. The modeled and actual Pri units are similar in a broad distribution of first-spike latency, in contrast to other response types, which show restricted distribution [(A) and (a), (B) and (b)]. PSTHs of both modeled and actual PN units show a firing notch after the initial response peak [(C) and (c)]. The three modeled Onset units shown in (D), (E) and (F) differ in pattern after the initial response peak as found in actual Onset units [(d), (e), and (f)]. That is, the unit in (D) shows essentially no firing ($<$ 25 spikes/s) after the onset response; the unit in (E) shows slight sustained firing ($\geq$25 spikes/s and $<$100 spikes/s); the unit shown in (F) has a firing notch at around 6 ms. The data in (G)-(J) and (g)-(j) indicate that PSTHs of the two Ch S and two Ch T units modeled agree with the physiological data in the number of regular spaced peaks for the first 10 ms after the response onset. The modeled and the actual Ch S unit in (G) and (g) have a mean inter-spike interval (ISI) that is nearly constant throughout the entire response to the STB [(G') and (g')]. Similar to the actual Ch T unit in (i), the modeled Ch T unit in (I) has a mean ISI that stabilizes after the initial transient response [(I')]. Both the actual Ch T and Ch S units in (h) and (j) have a mean ISI that increases more or less linearly with time for the first 15 ms of the response to STB. The difference between these two units is in the average CV (Young, Robert and Shofner 1988, see caption of Fig. 3) between 15-20 ms after the response onset, i.e., the Ch T unit has larger average CV than the Ch S unit. These characteristics are also found in the modeled Ch T and Ch S units in (H) and (J).

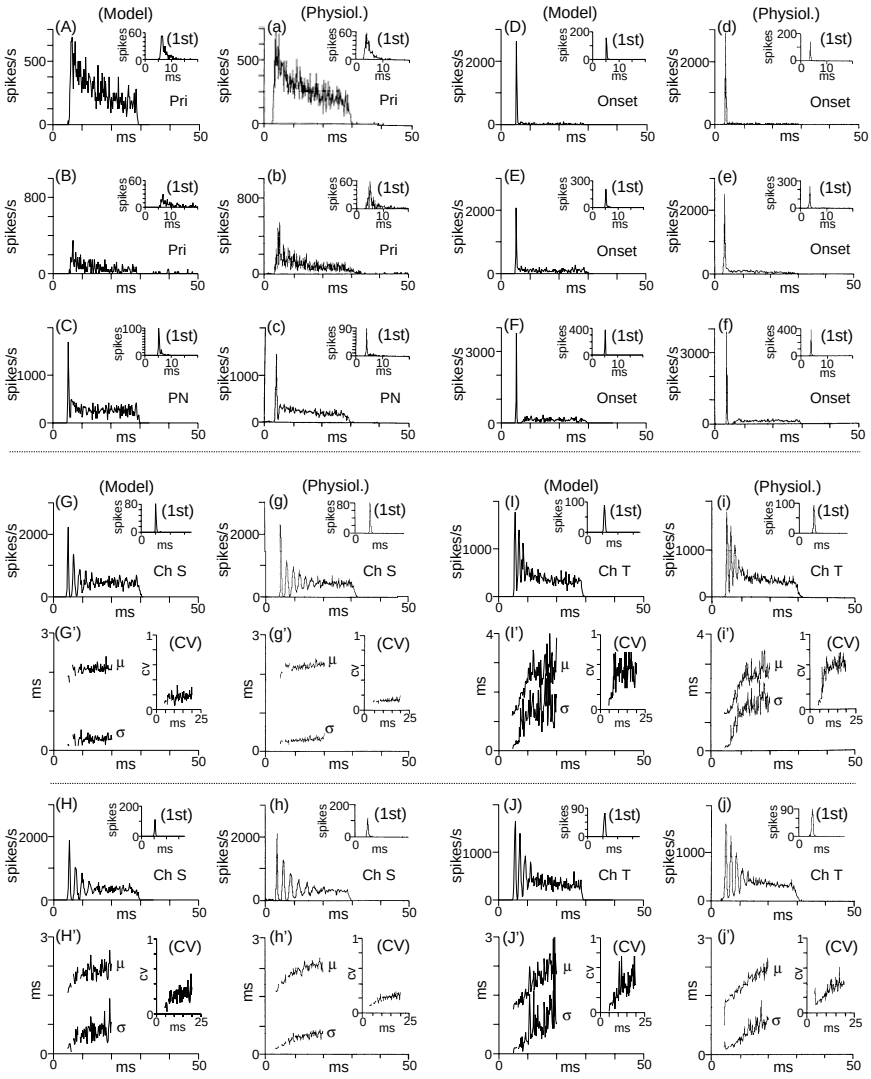


Fig. 3. Temporal properties of modeled and actual CN units' responses to STBs. (Model) Data from the model. (Physiol.) Physiological data redrawn from the original paper (Blackburn and Sachs 1989). Response types are indicated at right bottom in each PSTH figure. First-spike latency histograms (1st) are shown in the insets. The time scale for the first-spike latency is the same as that for PSTH. (G')-(J'), (g')-(j') The μ and σ respectively indicate the mean and standard deviation of inter-spike interval of spike data shown above (PSTH) (regularity analysis; Young, *et al* 1988). CV is calculated by dividing σ by μ. (Young, *et al* 1988).

4.2 Phase-locking properties

The synchronization index (Goldberg and Brown 1969) calculated for individual response types of the model in response to a BF tone are plotted as a function of BF in

Fig. 4(a) and (b), where the lines indicate least-squares fits to actual Pri and PN units data, and actual Ch S and Ch T units data, respectively (Blackburn and Sachs 1989).

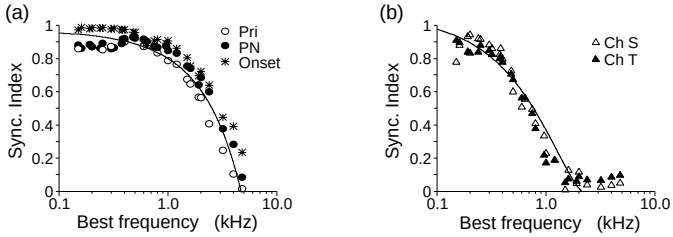


Fig. 4. Synchronization index of modeled responses as a function of BF. Dots represent data from the model. Lines in (a) and (b) indicate second-order polynomial functions fit to actual primary-like (Pri and PN) and chopper (Ch S and Ch T) units data, respectively, by use of least-squares (Blackburn and Sachs 1989). As in the case of the physiological data, calculations for the model were based on the entire response (10-400 ms) to a 400-ms long tone burst.

The pattern of the BF dependency of the phase-locking for the actual Pri and PN units resemble those for the ANFs [see Fig. 2(b)] in the fall-off of synchronization with increasing frequency between 1 and 5 kHz. The degree of phase-locking of actual Ch T and Ch S units falls more rapidly as a function of BF than does that of the Pri and PN units, approaching the noise level BF $>$ 1.5-2.5 kHz. The phase-locking degree of the two modeled primary-like units (Pri and PN) and two chopper units (Ch S and Ch T) closely follows the polynomial curve fit to the degree of phase-locking in actual CN units. Similarly to the actual Onset units, the modeled Onset units show a phase-locking ability resembling that of the Pri and PN units.

5 Discussion

We examined the eight parameters of the CN model and two parameters of the model inputs shown in Table 1, and found that only five of them are relevant for differentiating the response types. The two chopper (Ch S and Ch T) units modeled need a larger time constant of the PSPs (τ_i) than Pri, PN and Onset units. The modeled Ch T units have higher firing thresholds (α and β) than the modeled Ch S units. Consequently, the modeled Ch T units show more irregular discharge than the modeled Ch S units. Parameter values for the two primary-like units (Pri and PN) and Onset units are similar, except that Onset units receive a relatively larger number of inputs from the simulated ANFs (parameter N) and have a slightly higher firing threshold (α) than the primary-like units. The values of parameter σ_c determine the decay curves of phase-locking in Fig. 4. For a larger value of σ_c, the ability of the modeled CN neurons to phase-lock to BF tones falls more rapidly as a function of BF.

6 Conclusion

We proposed a computational model that functionally models the firing mechanisms of CN neurons. The model consists of eight parameters. We compared the responses of the model with physiological data from the AVCN (Blackburn and Sachs 1989), and found that the model could successfully simulate all temporal response types in term of the PSTHs, the degree of phase-locking, and spike latencies. Since the assumptions adopted are simple and are not specific to CN neurons, the proposed model can be easily extended to model neurons in higher-level auditory nuclei.

Acknowledgments

The authors thank Dr. Shigeto Furukawa for very helpful comments on an earlier version of this manuscript.

References

Arle, J.E. and Kim, D.O. (1991) Neural modeling of intrinsic and spike-discharge properties of cochlear nucleus neurons. Biol. Cybern. 64, 273-283.

Banks, M.I. and Sachs, M.B. (1991) Regularity analysis in a compartmental model of chopper units in the anteroventral cochlear nucleus. J. Neurophysiol. 65, 606-629.

Blackburn, C.C. and Sachs, M.B. (1989) Classification of unit types in the anteroventral cochlear nucleus: Histograms and regularity analysis. J. Neurophysiol. 62, 1303-1329.

Cai, Y., Walsh. E.J. and McGee, J. (1997) Mechanisms of onset responses in octopus cells of the cochlear nucleus: Implications of a model. J. Neurophysiol. 78, 872-883.

Eriksson, J.L., Robert, A. (1999) The representation of pure tones and noise in a model of cochlear nucleus neurons. J. Acoust. Soc. Am. 106, 1865-1879.

Goldberg, J.M. and Brown, P.B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization. J. Neurophysiol. 32, 613-636.

Hewitt, M.J. and Meddis, R. (1993) Regularity of cochlear nucleus stellate cells: A computational modeling study. J. Acoust. Soc. Am. 93, 3390-3399.

Hodgkin, A.L. and Huxley, A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. 117, 500-544.

Johnson, D.H. (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J. Acoust. Soc. Am. 68, 1115-1122.

Levy, K.L. and Kipke, D.R. (1997) A computational model of the cochlear nucleus octopus cell. J. Acoust. Soc. Am. 102, 391-402.

Maki, K., Akagi, M. and Hitota, K. (1998) A functional model of the auditory peripheral system: Responses to simple and complex stimuli. Proc. of NATO/ASI Computational Hearing Conference, 13-18.

Westerman, L.A., and Smith, R.L. (1984) Rapid and Short Term Adaptation in Auditory-Nerve Responses. Hear. Res. 15, 249-260.

Young, E.D., Robert, J.M. and Shofner, W.P. (1988) Regularity and latency of units in ventral cochlear nucleus: Implications for unit classification and generation of response properties. J. Neurophysiol. 60, 1-29.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Study on improving regularity of neural phase locking in single neurons of AVCN via a computational model

Kazuhito Ito and Masato Akagi

Japan Advanced Institute of Science and Technology

1 Introduction

One way of expressing temporal information in the auditory system is through the regularity of the interspike interval (ISI). This results from the firing of auditory nerve fibers (ANFs) being phase locked to the stimulation waveform (Johnson 1980). Consequently, ISIs represent periods corresponding to the reciprocals of their own characteristic frequencies (CFs). The temporal information of ISIs is thought to be used in various auditory perceptions such as sound localization based on the interaural time difference (ITD) and pitch perception. There may be a mechanism for transmitting accurate temporal information in the auditory system and this has been indicated by several psychophysical studies where humans can detect small changes in the ITD within microseconds (Mills 1958). However, ANFs do not always fire in synchronization with a certain phase of the stimuli, and impulses on the fibers fluctuate slightly (similar to jitter) and occasionally pause. The degree of jitter in ANFs ranges from hundreds of microseconds to a few milliseconds (Johnson 1980). Consequently, there must be a mechanism that improves temporal information in the auditory pathway. Recent physiological studies have found that bushy cells in the anteroventral cochlear nucleus (AVCN) have improved phase locking to low frequency tones (< 1 kHz) when compared to ANFs (Joris, Carney, Smith, and Yin 1994). In this paper, we demonstrate the mechanism through which a single cell improves the regularity of the ISI from the viewpoint of entrainment as well as synchronization using a computational model. We then discuss the relationship between the number of input terminals and the number of input events to maintain the ISI more regularly against the primary-like behavior of inputs from ANFs.

2 Synchronization and entrainment

The degree of phase locking in ANFs and neurons can be estimated with two scales, i.e., synchronization and entrainment. Synchronization is associated with how accurately spikes occur at a certain phase angle, and entrainment is associated with

how regularly spikes tune to each stimulus cycle. The synchronization index is calculated with the period histograms of spikes to tones at CF on individual cycles of the stimulus and defined through the following equations by Johnson (1980).

$$\hat{S}_f = (\hat{S}_{s,f}^2 + \hat{S}_{c,f}^2)^{1/2}, \quad \hat{S}_{s,f} = \frac{1}{N}\sum_{m=0}^{M-1} h_m \sin\frac{2\pi m}{M}, \quad \hat{S}_{c,f} = \frac{1}{N}\sum_{m=0}^{M-1} h_m \cos\frac{2\pi m}{M}, \tag{1}$$

where $\hat{S}_f$ indicates the estimated index, $\hat{S}_{s,f}$ and $\hat{S}_{c,f}$ denote the sine and cosine components of the period histogram, h_m (m = 0, 1, …, M-1) denotes the content of the mth bin of a period histogram with M bins, and N denotes the number of spikes contained in the period histograms. Perfect alignment of all spikes in one bin yields an index value of 1, whereas a uniform distribution of spikes throughout the stimulus cycle yields a value of 0. The entrainment index, defined by Joris et al. (1994), is derived from the interspike interval histograms of spikes to measure the ability of fibers and neurons to respond with one spike for every stimulus cycle.

$$E = h_F / N_s, \tag{2}$$

where E is the entrainment index, N_s denotes the total number of intervals of spikes during stimulation and h_F denotes the number of ISIs falling within a window equaling one stimulus period (1/CF). Perfect entrainment is achieved when a single spike occurs with each stimulus cycle and yields an index value of 1. If spikes are skipped on some stimulus cycles, the index value decreases. It is thus important to enhance temporal information so that it satisfies both synchronization and entrainment.

3 Physiological and anatomical properties of bushy cells

Mills (1958) found that some listeners could detect interaural time differences within microseconds in his psychophysical experiments. This suggests there is a capability to transmit temporal information accurately in the auditory system. Johnson (1980) reported that the synchronization indexes of ANFs with CFs below 1 kHz were between 0.7 and 0.9. These indexes sound high but could cause jitter on the order of hundreds of microseconds at lower CFs. These temporal values are larger than those for human perception found by Mills.

In early physiological studies, the responses of bushy cells in the AVCN, which receive impulses from ANFs, were thought to be similar to ANFs and behave like simple repeaters (e.g. Blackburn and Sachs 1989). However, Joris et al. (1994) revealed the higher tendency of cells to phase lock to low frequency tones such as enhanced synchronization and nearly perfect entrainment. Fig. 1 (Left) shows the synchronization for cells in the AVCN. The data are plotted on an expansive logarithmic scale that has an equal variance axis. The solid lines represent the range of indexes observed in ANFs by Johnson (1980). ANFs have indexes of less than 0.9 while 75% of presumed bushy cells with CFs below 700 Hz have indexes that are greater than 0.9. Higher indexes over 0.9 can decrease jitter in impulses. Fig. 1 (Right) shows the entrainment for the cells discussed above. ANFs are represented as plots of crosses (+) and have entrainment indexes that are about 0.8 at lower CFs decreasing systematically with frequency, while many of the presumed bushy cells have nearly perfect entrainment indexes up to about 700 Hz. This means the bushy

cells tend to fire with each stimulus cycle at lower CFs.

Bushy cells are connected by numerous synaptic terminals, which are of several types (Cant 1996). Especially striking terminals are a few large endbulbs of Held on spherical bushy cells (SBCs). Ryugo and Sento (1996) reported that each fiber gave rise to a single endbulb of Held. A single SBC has at most two endbulbs of Held on it, and the convergent endbulbs arise from fibers of the same SR (spontaneous discharge rate) group. As the physiological properties of a single SBC were thought to be similar to those of an ANF, its properties have also been considered to be dominated by a few endbulbs of Held despite the presence of many smaller terminals from ANFs (Cant 1996; Young 1998). However, globular bushy cells (GBCs) located more posteriorly in the AVCN receive a greater number of smaller somatic terminals, i.e. modified endbulbs, from ANFs (Cant 1996, Yin 2002).

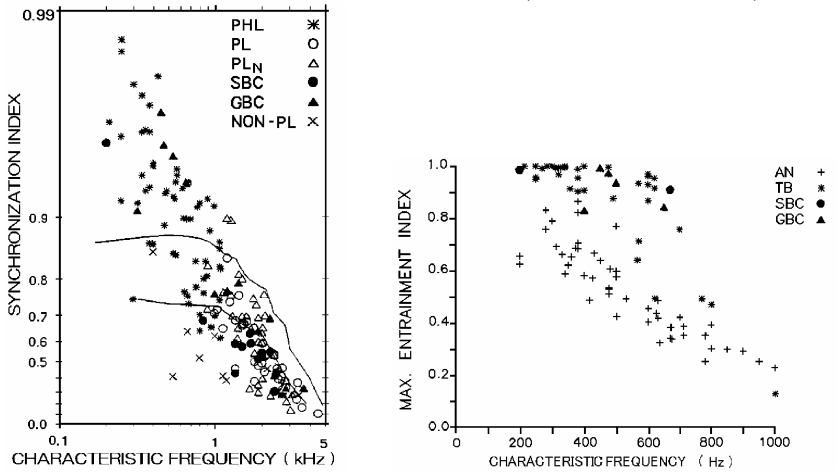


Fig. 1. *Left*: Synchronization index in ANFs and AVCN cells plotted against CF. The ordinate designates (1-syn.) on a logarithmic scale. The area between solid lines indicates the range of indexes of ANFs from Johnson (1980). *Right*: Entrainment index in ANFs (crosses,+) and AVCN cells (the others). These data are from Joris et al. (1994).

4 Simulations using computational model

If it is clear that bushy cell response differs from that of ANFs, a large endbulb does not function as a one-to-one relay even at low CFs where the endbulbs are largest. An important question is the mechanism by which bushy cells achieve enhanced phase locking. One way to obtain the enhanced phase locking is to have the cell behave as a coincidence detector with a number of presynaptic auditory nerve inputs with similar CFs (Yin 2002). Also, a postsynaptic bushy cell requires coincident input spikes before it generates an output spike (Joris et al. 1994). They showed that a simple “shot-noise” model (Colburn 1996) could mimic the enhanced phase locking of bushy cells to pure tones at CF. Rothman, Young, and Manis (1993) and Rothman and Young (1996) studied various combinations of inputs and strength to simulate the physiological responses of bushy cells using a computational model similar to a point-neuron model (Colburn 1996). These

models suggested that utilizing subthreshold inputs and required coincidence to evoke a spike is useful in enhancing synchronization. Although they demonstrated various cases, these were not sufficient for perfect entrainment.

This study demonstrates the mechanism by which a single neuron attains perfect entrainment as well as synchronization through using a computational model. In particular, we derive the quantitative relation between the number of input terminals (n) and the number of required input events (k) to evoke a spike by analyzing the relation between input and output entrainment indexes. The model cell used in this study was a type of shot-noise model. The model cell was connected by n input-terminals with the same CF. It generated an output spike with at least k coincident input events. An input event means a firing at a presynaptic fiber. The model cell potential in response to an input is described by the following equation.

$$V(t) = a \cdot e^{-t/\tau}, \tag{3}$$

where $V(t)$ denotes the postsynaptic potential (PSP) at time t. The PSP decays exponentially to the resting potential. τ denotes the time constant of PSP and a denotes the subthreshold amplitude of PSP. Output firing occurs when the total PSP potential exceeds a threshold level. This model cell operates as a kind of coincidence detector because input events must occur close together to cause the cell to cross the threshold. The membrane potential is reset to zero following output firing and the model cell does not respond to any input during the refractory period. The model cell receives input spike trains that simulate the primary-like pattern of ANFs including their synchronization and entrainment properties.

5 Results

We conducted some simulations to investigate the mechanism that improves phase locking in a single cell with various combinations of parameters n and k. We will discuss simulation results where $k = 2$ which means that the model cell needed at least 2 simultaneous input events to evoke an output spike. Other values of k yielded similar results to this case. The model cell had τ set to 0.5 (ms) and the refractory period to 1.5 (ms), which were the same values as those used by Joris et al. Fig. 2 has the results for synchronization (Left) and entrainment (Right) of the model cell ($k = 2$) with various parameters of n. The indexes of input spike trains that simulate the primary-like properties of ANFs at CFs are plotted as filled circles "●". The number of input terminals (n) is $n = 2$, 3, 5, and 10. The indexes of the model cell in response to each n are represented by the symbols "▽", "○", "□", "◊", respectively. Every plot, including input, represents the mean and standard deviation of each index obtained over a hundred trials of simulation.

In Fig. 2 (Left), the abscissa indicates CFs and the ordinate designates (1 – sync.) on a logarithmic scale. The simulations revealed that the synchronization indexes of all outputs were higher than those of inputs at CFs below 700 Hz, and the indexes become higher as parameter n became larger. The synchronization simulation was influenced by the fixed refractory period (1.5 ms) because the cell fired according to the refractory period at CFs above 700 Hz. Therefore, the results for CFs above 700 Hz were not significant. In Fig. 2 (Right), the abscissa

designates CFs and the ordinate designates entrainment indexes. The simulation revealed that the entrainment indexes for output with n = 2, "▽", were lower than those for input at all CFs, and the indexes for output with n = 3, "○", were similar to those for input at low CFs below 600 Hz. When n = 10, "◊", simulation revealed nearly perfect entrainment indexes at CFs below 600 Hz. Entrainment was also influenced by the fixed refractory period (1.5 ms) due to the same reasons as synchronization. The indexes declined rapidly at CFs above 600 Hz.

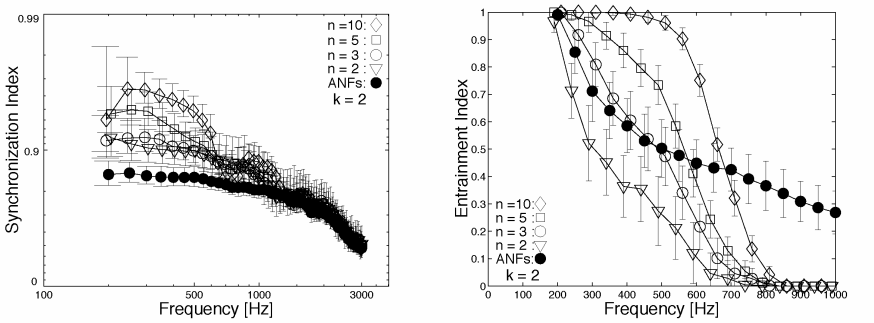


Fig. 2. Means and standard deviations of the synchronization and entrainment indexes in the model cell with n = 2, 3, 5, and 10 and k = 2. Filled circles denote the indexes for ANFs. *Left*: Synchronization indexes at CFs below 700 Hz are higher as n becomes larger. *Right*: Entrainment indexes at CFs below 600 Hz are nearly perfect when n = 10.

6 Discussion

We investigated transitions in synchronization and entrainment of the model cell in these simulations by changing parameter n while fixing parameter k to other values (k = 1, 2, 3,…). The synchronization results indicated that at low CFs below 700 Hz, the indexes increase monotonically from the index level of ANFs as parameter n increases, even when n = 2. These plots were good fits for the physiological data in this CF range. For other k parameters, simulations appeared to have a similar tendency. The mechanism for coincidence detection with larger n to improve synchronization is useful as Joris et al. (1994) predicted in their study. However, in enhancing temporal information it is important to satisfy both synchronization and entrainment. Entrainment results revealed that indexes do not increase monotonically as they do for synchronization. When parameter k = 2, only the entrainment indexes of n = 10,"◊", were nearly perfect at CFs below 600 Hz and this was a good fit with the physiological data. It seems that a mechanism for coincidence detection with a much larger n is needed compared with k to obtain perfect entrainment. These results suggest the possibility of a firing mechanism with multiple inputs on a single neuron in the AVCN that enhances temporal information. This does not seem to be consistent with the general agreement obtained from anatomical studies. If results obtained through simulations are plausible, we have to consider the possibility that bunches of smaller synaptic terminals may also be involved in the mechanism that enhances temporal information on a single neuron in the AVCN.

Then we considered a way of mathematically estimating the values of n and k to enhance temporal information. To achieve this, we modified the equation for the entrainment index so we could add the information for the stimulus period. The modified entrainment index G is defined by the following equation:

$$G = h_F / N_f, \tag{4}$$

where N_f denotes the number of times the stimulus period (1/CF) occurred during the entire stimulation, instead of the total number of spike intervals (N_s). h_F is the same as in Eq. 2. Hence, G denotes the ratio for the number of times ISI occurred over the number of times the stimulus period (1/CF) occurred. Fig. 3 (Left) shows modified entrainment converted from the normal one in Fig. 2 (Right). Then, we can define the modified entrainment index G of the ith input terminal as the probability $0 \le P_{G_i} \le 1,\ i \in \mathrm{N}$. Firing events at input terminals are regarded as independent. An output event at a cell is considered to be generated by the joint firing event at the input terminals. In the case of two input terminals (n = 2) and two input events (k = 2) simultaneously required to output a spike, output probability $P_{G_{out}}$ is given by the joint probability $P_{G_1} \cdot P_{G_2}$ of input probability P_{G_1} and P_{G_2} because coincidence of ISIs from both input terminals allows a cell to output an ISI. As the simulations show, the entrainment of output is not satisfied when n = 2 and k = 2. We need to add other input terminals to increase $P_{G_{out}}$. Since a parameter n greater than k increases the number of occasions to generate ISIs of output with no ISI from input terminals, output probability $P_{G_{out}}$ is approximated to the worst case. The input probabilities are regarded as having the same probability P_{G_i} because terminals on a single cell probably belong to the same SR group (Ryugo and Sento 1996). Therefore, the quantitative relation between n and k is given by the following estimate.

$$P_{G_{out}} = \sum_{m=k}^{n} \binom{n}{m} \cdot {P_{G_i}}^{m} \cdot \left(1 - P_{G_i}\right)^{n-m}. \tag{5}$$

This provides us with an approximation for the least number of input terminals (n) on a single cell when the number of required input events (k) to evoke a spike is given. Fig. 3 (Right) denotes output probability and shows good approximation to the modified entrainment G in Fig. 3 (Left) at CFs below 600 Hz. It suggests that a single cell with a CF of 400 Hz should have at least 10 input terminals to maintain the modified entrainment index so that is greater than 0.9. The output probability does not fit to the modified entrainment index at CFs above 600 Hz because this estimate does not take the refractory period into account.

7 Conclusion

This study demonstrated the mechanism through which a single neuron enhanced temporal information from the viewpoint of entrainment as well as synchronization using a computational model. As a result of simulation, a model with multiple-input configurations yielded enhanced synchronization and nearly perfect entrainment at low CFs. The relation between the number of input terminals and the number of

input events to output a spike was derived quantitatively to maintain enhanced temporal information against the primary-like behavior of inputs from ANFs. The results suggested the possibility of a firing mechanism that had multiple inputs on a single neuron in the AVCN.

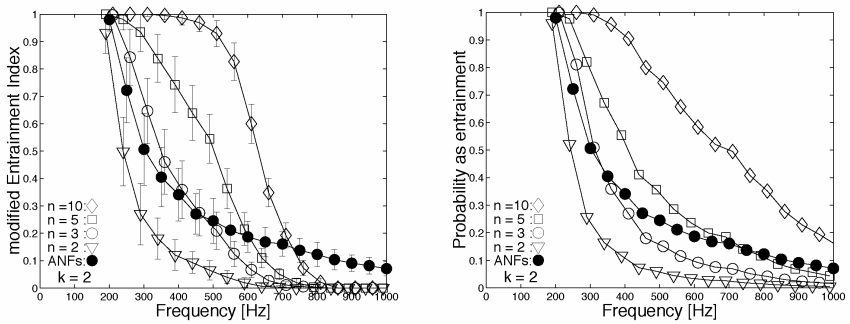


Fig. 3. *Left:* Modified entrainment indexes are converted from the normal entrainment indexes. *Right:* Output probability is calculated by Eq. 5 with n = 2, 3, 5, and 10 and k = 2. Input probability refers to the modified entrainment indexes of the ANFs.

References

Blackburn, C.C. and Sachs, M.B. (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. J. Neurophysiol., 62, 1303-1329.

Cant, N.B. (1996) The cochlear nucleus: neuronal types and their synaptic organization. In: D.B. Webster, A.N. Popper and R.R. Fay (Eds.), *The Mammalian Auditory Pathway: Neuroanatomy*, Springer-Verlag, New York. pp. 66-116.

Colburn, H.S. (1996) Binaural Models. In: H.L. Hawkins, T.A. McMullen, A.N. Popper and R.R. Fay (Eds.), *Auditory Computation*. Springer-Verlag, New York. pp.332-400.

Johnson, D.H. (1980) The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones. J. Acoust. Soc. Am. 68, 1115-1122.

Joris, P.X., Carney, L.H., Smith, P.H. and Yin, T.C.T. (1994) Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. J. Neurophysiol. 71, 1022-1036.

Mills, A.W. (1958) On the minimum audible angle. J. Acoust. Soc. Am. 30, 237-246.

Rothman, J.S., Young, E.D. and Manis, P.B. (1993) Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: implications of a computational model. J. Neurophysiol., 70, 2562-2583.

Rothman, J.S. and Young,E.D.(1996) Enhancement of neural synchronization in computational models of ventral cochlear nucleus bushy cells. Aud. Neurosci. 2, 47-62.

Ryugo, D.K. and Sento, S. (1996) Auditory nerve terminals and cochlear nucleus neurons: endbulbs of Held and spherical bushy cells. In: W.A. Ainsworth, E.F. Evans and C.M. Hackney (Eds.) *Advances in Speech, Hearing, and Language Processing, vol .3*, JAI Press, Connecticut

Yin, T.C.T. (2002) Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In: D. Oertel, R.R. Fay and A.N. Popper (Eds.), *Integrative Functions in the Mammalian Auditory Pathway*, Springer-verlag, New York, Chap. 4, pp. 99-159.

Young, E.D. (1998) Cochlear nucleus. In: G.M. Shepard (Ed.) *Synaptic Organization of the Brain (4th Ed.).* Oxford Press, London, pp. 131-157.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Fibers in the trapezoid body show enhanced synchronization to broadband noise when compared to auditory nerve fibers

Dries H. Louage, Marcel van der Heijden, and Philip X. Joris

Laboratory of Auditory Neurophysiology, K.U.Leuven, dries.louage@med.kuleuven.ac.be

1 Introduction

In the cat, many fibers in the trapezoid body (TB) show better synchronization to low-frequency pure tones than do auditory nerve (AN) fibers (Joris, Carney, Smith, and Yin 1994, Joris, Smith and Yin 1994). Because "hi-sync" fibers are axons of bushy cells and project to binaural nuclei in the superior olivary complex (SOC), enhanced synchronization may constitute monaural preprocessing critical for the sensitivity to interaural time differences (ITDs) in the SOC. As most natural stimuli are broadband and behavioral ITD-sensitivity is best for broadband stimuli, we examined whether enhanced synchronization also occurs to broadband noise. The traditional synchronization measure (vector strength, Goldberg and Brown 1969) is not applicable to non-periodic stimuli, and, to our knowledge, a simple metric to quantify synchronization to broadband signals has not been described. We derive two metrics based on the shuffled autocorrelograms first described by Joris (2001), and apply these metrics to responses of TB and AN fibers to broadband noise. We observe that enhancement of synchronization of TB fibers is not restricted to pure tones, but extends to broadband stimuli.

2 Methods

Single-unit recordings were obtained with glass micro-pipettes from the TB and the AN in barbiturate-anesthetized cats. Stimuli were generated digitally (Tucker Davis Technology) and delivered through a closed acoustic system. A standard pseudorandom broadband noise (100 to 30000 Hz, duration 1000 ms, repeated every 1200 ms) was presented (typically at least 50 repetitions) at 70 dB SPL overall level. Responses to repeated presentations of the same noise token were used to construct shuffled autocorrelograms (Joris, 2001).

Fig. 1 illustrates the construction of a shuffled all-order interval histogram. Intervals between a spike and all following spikes from a *different* spike train to the same stimulus are measured. Spikes from *different* spiketrains are taken rather than

spikes *within* spiketrains (e.g. Ruggero 1973), to avoid refractory effects. The histograms had bins of 50 μs. The resulting all-order interval histogram is referred to as a shuffled autocorrelogram. "Shuffled" refers to the exclusion of pairs of identical spiketrains, "auto" refers to the use of responses obtained from a single fiber, and "correlation" refers to the formal equivalence between this procedure and

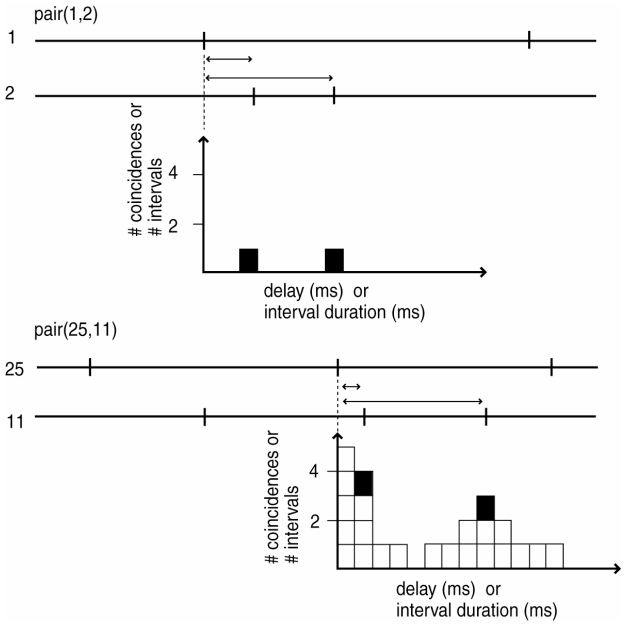


Fig. 1. Construction of a shuffled autocorrelogram. First, all permutations of *pairs* of non-identical spike trains are listed, of which 2 pairs are shown here. For example, 50 spike trains yielded 2450 pairs. Next, we measured intervals between all spikes of the first spike train and all spikes of the second spike train of each pair, and tallied these intervals in a histogram. The resulting histogram graphs the total number of intervals for all interval durations. Because counting intervals is identical to counting coincident spikes between two spiketrains shifted over different delays, the histogram can be graphed as number of intervals versus interval duration, or as number of coincidences versus delay.

the cross-correlation of two spike trains. Note that the procedure of tallying intervals across spiketrains gives identical results to the counting of coincident spikes across spiketrains for different delays between the two spiketrains (Joris 2001).

We obtained dimensionless normalized shuffled autocorrelograms (NSACs) by dividing by $N.(N-1).duration.(firing\ rate)^2.(binwidth)$. NSACs obtained from spike trains with uncorrelated temporal structure are flat and equal unity. Any deviation of NSAC values from unity indicates temporal correlation between spiketrains.

3 Results

Figure 2 shows an example of a typical NSAC obtained from a low-CF AN fiber. It has the shape of a damped oscillation with an oscillation frequency near the CF of the fiber. From the peak at delay 0 the NSAC damps out to unity on either side, indicating that spiketrains are uncorrelated at long time intervals. To quantify the central peak, we measured its height and width at half height.

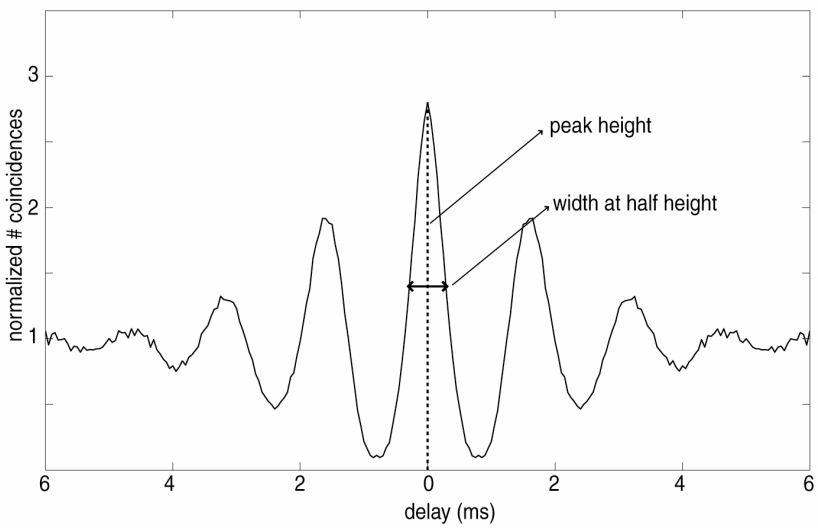


Fig. 2. Normalized shuffled autocorrelogram obtained from an auditory nerve (AN) fiber with a characteristic frequency (CF) of 600 Hz and a spontaneous rate (SR) of 60 spikes/sec. Two measures are extracted. The vertical dashed line indicates the peak height. The horizontal double arrow shows the width at halfheight.

The bottom panels in Figure 3 illustrate data for 6 AN fibers of different CFs. All NSACs are graphed on the same scale, and clearly differ for different CF regions. In low-CF fibers (left column) the shape is that of a damped oscillation, as in Fig. 2, but at high CFs (right column) the NSACs show a single peak at 0 delay which reflects envelope synchronization (Joris, 2001). At intermediate CFs (middle column) there is an oscillatory component superimposed on a broader peak.

The top panels in Fig. 3 show TB data on the same scale as the AN data. The most striking difference with AN fibers is in the central peak at low CFs, which tended to be larger and narrower in TB fibers (Fig. 3A,B) when compared to AN fibers (Fig. 3C,D).

Figure 4 shows NSAC peak heights for a population of TB and AN fibers. Every datapoint represents the response of one fiber to broadband noise at 70 dB SPL. Clearly, TB fibers tend to have larger peaks, particularly at CFs below a few kHz, even though there is overlap with the AN population. This indicates that, in response to repeated stimulation with a token of pseudorandom broadband noise, TB fibers have a strong tendency to discharge a spike in the same temporal position

with every repetition, and this tendency is stronger than in AN fibers. Peak height decreased with CF in both the TB and AN population. Moreover, in low-CF AN fibers there was a clear segregration between different spontaneous rate (SR) classes with low-medium fibers showing larger peaks than high-SR fibers. The

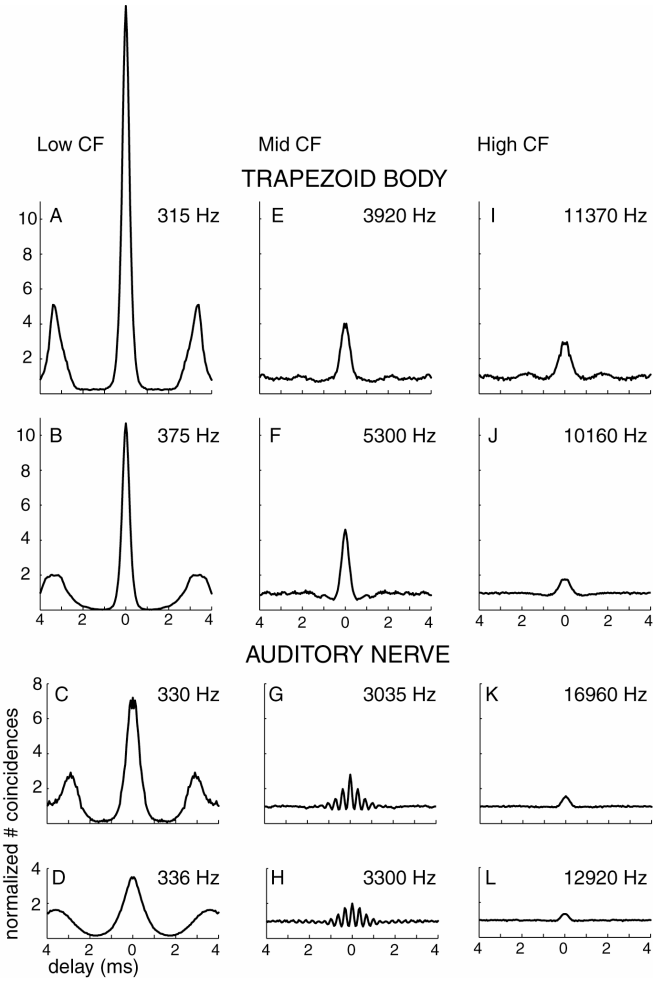


Fig. 3. Examples of NSACs obtained from TB and AN fibers of different CF ranges. Columns separate low-, mid-, and high-CF fibers. CF of fiber is shown in each panel.

spread of peak height was larger in the TB than in the AN population.

Figure 5 shows the width at half height (cf. Fig. 2) of the central peak of NSACs obtained from the same population of fibers as Fig. 4. Here, we focus on the low-CF population and are not further concerned with fibers of CF near 2 kHz and above, for which the NSAC can be dominated by envelope timing. At low CFs, halfwidth decreased with increasing CF. Moreover, for CFs < 1 kHz halfwidth was

generally smaller for TB fibers than for AN fibers. Thus, the locking of a spike to the stimulus was more precise in TB fibers than in AN fibers.

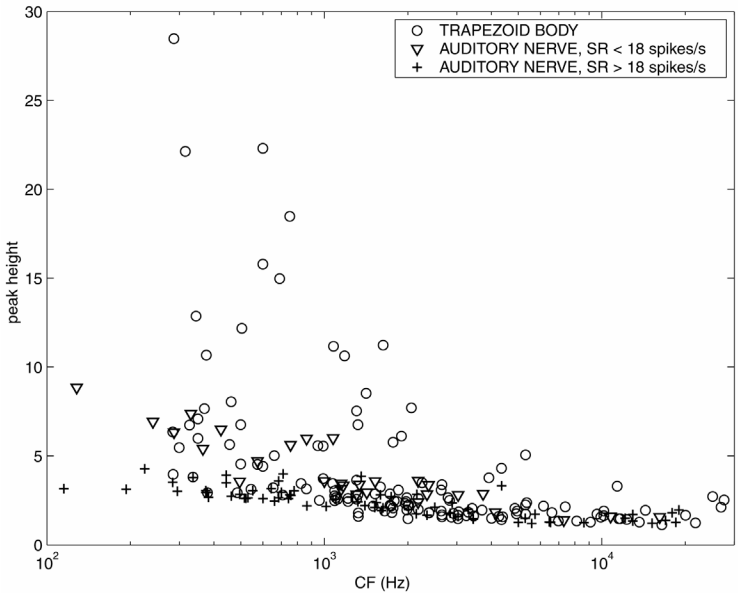


Fig. 4. Height of the central peak of NSACs for a population of TB fibers (○) and AN fibers. The AN fibers are grouped in 2 spontaneous rate (SR) classes (∇ low-medium SR fibers; + high-SR fibers). Each datapoint represents one fiber.

4 Discussion

We used autocorrelograms to compare temporal behavior to an aperiodic stimulus in AN and TB fibers. The NSACs obtained all had a central peak which we quantified by measuring height and width at halfheight. In comparison to AN fibers, NSACs of TB fibers tended to have larger and narrower central peaks. The discharge pattern of these fibers to broadband noise is therefore more stereotyped from repetition to repetition than that of AN fibers, and temporally more precise.

These results are consistent with those obtained with pure tones (Joris et al. 1994), which showed that TB responses differed in two ways from AN responses. TB fibers showed higher temporal precision, as measured with vector strength. Also, they entrained at very low frequencies, i.e. they tend to fire a spike at every stimulus cycle, as measured with an entrainment index. An exact correspondence between the tonal and noise responses can not be drawn, but nevertheless the NSACs seem to capture two similar aspects of the TB responses: entrainment roughly corresponds to larger central NSAC peaks, enhanced synchronization to narrower central peaks.

One of the striking findings in the study by Joris et al (1994) was the nearly complete segregation of AN and TB distributions at very low-CFs, in contrast to the

nearly complete overlap that had been reported in previous studies (e.g. Bourk 1976; Palmer, Winter and Darwin 1986; Blackburn and Sachs 1989; Winter and Palmer 1990). In the data reported here, there is some overlap between AN and TB responses, even at very low CFs. One contributing factor may be the long duration (1 s) of our noise bursts. The tonal data of Joris et al. (1994) were obtained to short

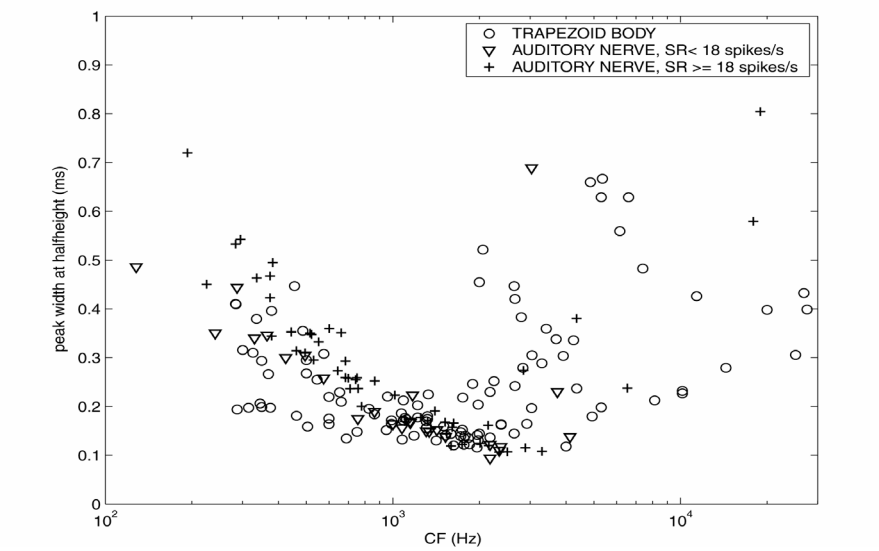


Fig. 5. Population scatter plot of the width at half height of the central peak of NSACs as a function of the CF of each fiber. Each datapoint represents one fiber.

(25 ms) CF tone pips, and these authors showed that even within that short time window TB responses showed temporal adaptation. The present data demonstrate that for stimuli that are more representative of natural environments, both in duration and bandwidth, TB fibers show an enhancement of synchronization.

Acknowledgements

Supported by the Fund for Scientific Research - Flanders (G.0083.02) and Research Fund K.U.Leuven (OT/10/42).

References

Blackburn, C. C. and Sachs, M.B. (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. J. Neurophysiol. 62, 1303-1329.

Bourk, T. R. (1976) Electrical responses of neural units in the anteroventral cochlear nucleus of the cat. Ph.D. thesis, MIT, Cambridge, Mass.

Goldberg, J. M. and Brown, P.B. (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J. Neurophysiol. 22, 613-636.

Joris, P. X. (2001) Sensitivity of inferior colliculus neurons to interaural time differences of broadband signals: comparison with auditory nerve firing patterns. In: D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven (Eds.) *Physiological and Psychophysical Bases of Auditory Function*, Shaker Publishing BV, Maastricht, pp. 177-183.

Joris, P. X., Carney, L.H.C., Smith, P.H. and Yin, T.C.T. (1994) Enhancement of synchronization in the anteroventral cochlear nucleus. I. Responses to tonebursts at characteristic frequency. J. Neurophysiol. 71, 1022-1036.

Joris, P. X., Smith, P.H. and Yin, T.C.T. (1994) Enhancement of synchronization in the anteroventral cochlear nucleus. II. Responses to tonebursts in the tuning-curve tail. J. Neurophysiol. 71, 1037-1051.

Palmer, A. R., Winter, I.M. and Darwin, C.J. (1986) The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. J. Acoust. Soc. Am. 79, 100-113.

Ruggero, M. A. (1973) Response to noise of auditory nerve fibers in the squirrel monkey. J. Neurophysiol. 36, 569-587.

Winter, I. M. and Palmer, A.R. (1990) Responses of single units in the anteroventral cochlear nucleus of the guinea pig. Hear. Res. 44, 161-178.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Representations of the pitch of complex tones in the auditory nerve

Leonardo Cedolin[1,2] and Bertrand Delgutte[1,2,3]

[1] Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary
[2] Harvard-MIT Division of Health Science and Technology, Speech and Hearing Bioscience and Technology Program, cedro@mit.edu
[3] Research Laboratory of Electronics, MIT, bard@epl.meei.harvard.edu

1 Introduction

Previous studies of the coding of the pitch of complex tones in the auditory nerve and cochlear nucleus have documented a robust temporal representation based on interspike interval distributions (Cariani and Delgutte, 1996; Rhode, 1995; Palmer and Winter, 1993). However, these studies have largely neglected possible rate-place cues to pitch available when individual harmonics are resolved by the peripheral auditory system. Stimuli used in these studies had fundamental frequencies in the range of human voice (100-300 Hz), which may produce few, if any, resolved harmonics in typical experimental animals, which have a poorer cochlear frequency selectivity compared to humans (Shera, Guinan and Oxenham, 2002). Human psychophysical studies suggest that the low pitch produced by stimuli with resolved harmonics is stronger and less dependent on phase relationships among the partials than the pitch based on unresolved harmonics (Shackleton and Carlyon, 1994).

Here, we investigate the resolvability of harmonics of complex tones in the cat auditory nerve, and compare the effectiveness of rate-place and interval-based representations of pitch over a much wider range of fundamental frequencies (110-3520 Hz) than in previous studies.

2 Method

2.1 Stimuli and recording techniques

We recorded from auditory-nerve (AN) fibers in dial-anesthetized cats using glass micropipettes filled with 2-M KCl. Upon contact with a fiber, we measured the pure-tone tuning curve to determine the characteristic frequency (CF).

Stimuli were complex tones whose fundamental frequency (F0) stepped up and down over a two-octave range. Each of the 25 F0 steps lasted 200 ms. The harmonics of each complex spanned a two-octave range around the fiber's CF, and were all of equal amplitude, in cosine phase. The fundamental frequency was al-

ways missing. The sound pressure level of each harmonic was usually 15-20 dB above a fiber's threshold, and ranged from 10 to 70 dB SPL.

2.2 Data analysis

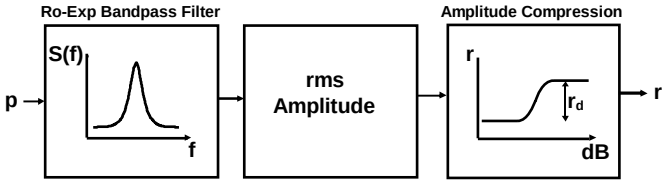


Fig.1. Block diagram of single-fiber rate model.

Simple phenomenological models were used to analyze average-rate responses to the complex-tone stimuli. Specifically, a single-fiber model was used to fit responses of a given fiber as a function of stimulus F0, while a population model was used to fit profiles of average rate against CF for a given F0. The model parameters provide quantitative measures of the ability of AN fibers to resolve individual harmonics.

The single-fiber model (Fig. 1) is a cascade of 3 stages. The band-pass filtering stage, representing cochlear frequency selectivity, is implemented by a symmetric rounded exponential function (Patterson, 1976). The Sachs and Abbas (1974) model is used to express the mean discharge rate as a function of the r.m.s. amplitude at the ouput of the band-pass filter. The model has 6 free parameters, considerably fewer than the 25 F0 values for which responses are obtained in each fiber.

The population model is an array of single-fiber models indexed on CF so as to predict the entire auditory-nerve rate response as a function of cochlear place. The bandwidths of the band-pass filters are constrained to be a power function of the CF (Shera et al., 2002). The population model has no free parameters; rather, parameters of the stimulus (F0 and SPL) are selected to fit the measured "rate-place profiles" expressing the normalized driven discharge rate as a function of CF (Sachs & Young, 1979). The resulting best-fitting F0 gives a rate-based estimate of pitch that does not require *a priori* knowledge of the actual F0.

3 Results: Rate-place representation

3.1 Single-fiber rate responses

Figure 2 shows the average discharge rate plotted against complex-tone F0 for two auditory-nerve fibers. The horizontal axis represents the dimensionless “harmonic number” CF/F0, so that higher F0s are towards the left. For both fibers, the mean rate shows a peak when the CF is a small integer multiple of F0, and a valley when the CF falls halfway between two harmonics. However, the oscillations are more pronounced for the higher-CF fiber on the right.

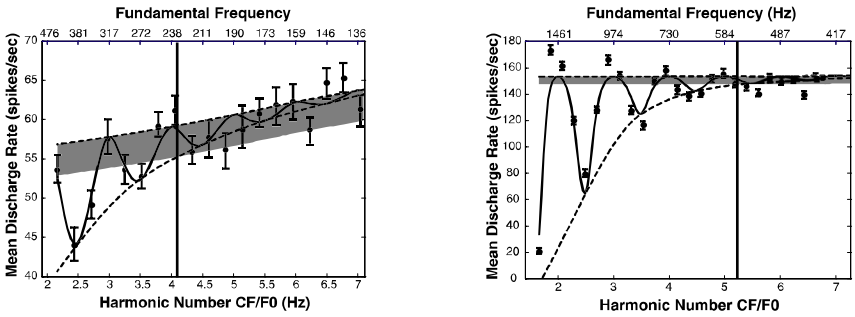


Fig. 2. Average discharge rate as a function of F0 for two fibers with CFs of 952 Hz (left) and 2922 Hz (right). Error bars show ±1 standard deviation of the discharge rate obtained by bootstrap resampling over stimulus trials. The solid lines are least-squares fits to the data using the model of Fig. 1.

For both fibers, the response of the best-fitting single-fiber model captures the main trend in the data. The harmonics of F0 are considered to be resolved so long as the oscillations in the fitted curve exceed two typical standard deviations (gray shading). We call N_{max} the maximum resolved harmonic number. Here, N_{max} for the low-CF fiber is 4.1, smaller than N_{max} for the high-CF fiber (5.3). The ratio N_{max}/CF gives $F0_{min}$, the lowest fundamental frequency whose harmonics can be resolved in the rate response of a given fiber.

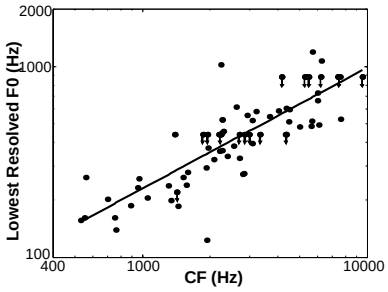


Fig. 3. Lowest F0 whose harmonics can be resolved as a function of CF. For some fibers (arrows), $F0_{min}$ was bounded by the lowest F0 presented and was therefore somewhat overestimated.

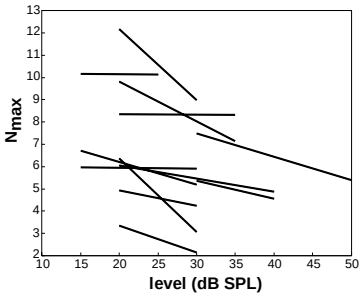


Fig. 4. Maximum resolved harmonic number for 11 fibers as a function of level expressed in dB SPL per component.

Figure 3 shows that $F0_{min}$ increases systematically with characteristic frequency. It also suggests that harmonics of F0s in the range of human voice (100-200 Hz) are rarely, if ever, resolved by AN fibers in the cat. The increase in $F0_{min}$ is well fit by a power function of the CF with an exponent of 0.63 (solid line). This result is consistent with the progressive sharpening of peripheral tuning with increasing CF when expressed as a Q factor. The exponent for Q would be 0.37, which closely matches the 0.37 exponent found by Shera et al. (2002), based on pure-tone tuning curves from AN fibers in the cat.

Rate responses of AN fibers to complex stimuli are known to depend strongly on stimulus level (Sachs and Young, 1979). To address this issue, responses to complex tones were recorded at two different stimulus levels in a few fibers. In these cases, the maximum resolved harmonic number N_{max} tended to decrease with increasing stimulus level (Fig. 4). This decrease could reflect either broadened cochlear tuning with increasing level, or rate saturation. Preliminary analysis suggests that the latter may be the dominant factor, since the bandwidths of the model auditory filters stay essentially constant with level.

3.2 Pitch estimation from rate-place profiles

Figure 5 shows the normalized driven discharge rate in response to a complex tone with an F0 of 541.5 Hz as a function of CF. Despite some scatter in the data, the normalized rate tends to show a local maximum when the CF is an integer multiple of F0 and a minimum when the CF falls halfway between two harmonics.

The oscillatory pattern in the rate-place profile of Fig. 5 can be used to estimate the fundamental frequency of the stimulus. To quantitatively derive such estimates, we determined the F0 of a complex tone with equal-amplitude harmonics for which the response of the population model (see Method) best fits the observed rate-place profile. Here the estimated pitch was 547.4 Hz, only 1.1% above the actual F0.

Figure 6 shows measures of the accuracy and precision of rate-based pitch estimates as a function of F0. With few exceptions, median pitch estimates only deviate by a few percent from the true F0 (top panel). For F0s above 400-500 Hz, the interquartile range of the pitch estimates over 100 bootstrap resamplings of the data are all below 5% (middle panel). However, the model produces few reliable estimates for F0s below 400 Hz. The bottom panel of Fig. 6 shows a measure of pitch salience derived from the amplitude of the oscillations in the model rate-place profiles. The salience is very low below 800 Hz, then increases rapidly to saturate above 2 kHz. Overall, pitch estimation from rate-place

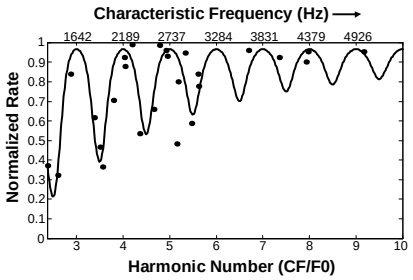


Fig. 5. •: normalized discharge rate as a function of CF; **—:** model rate-place profile. F0 = 541.5 Hz.

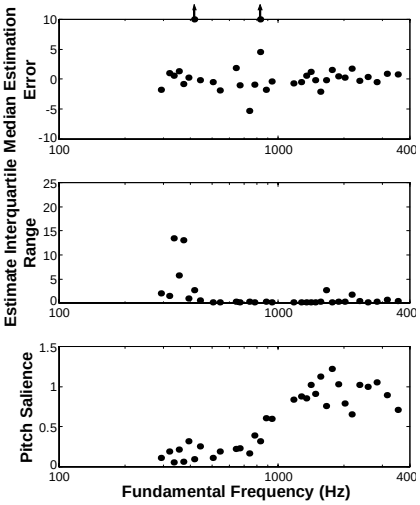


Fig. 6. Median pitch estimation error, interquartile range of the estimates, and pitch salience as a function of fundamental frequency.

profiles works best for F0s above 400 Hz.

4 Pitch estimation from pooled interspike interval distributions

As in previous studies of the neural coding of pitch (Cariani and Delgutte, 1996; Rhode, 1995), we derived pitch estimates from pooled interspike interval distributions. The pooled interval distribution is the sum of the all-order interspike intervals for all sampled auditory-nerve fibers, and is closely related to the summary autocorrelation in the Meddis and Hewitt (1991) model.

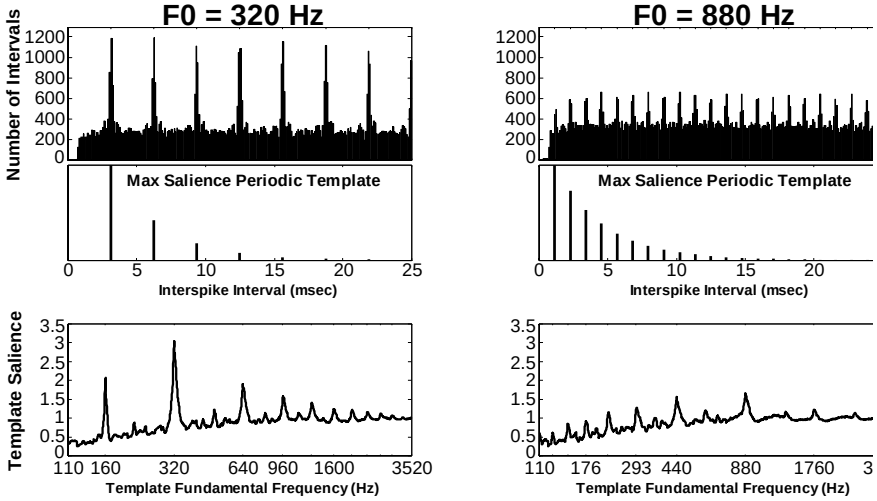


Fig. 7. Top row: pooled interval distributions for complex tones at two F0s. Middle row: periodic template with maximum salience. Bottom row: salience of periodic template as a function of its fundamental frequency. Results are shown for two F0s: 320 Hz (left column) and 880 Hz (right column).

Figure 7 (top) shows pooled interval distributions for two complex-tone stimuli with F0s of 320 and 880 Hz. For both stimuli, the pooled distributions show modes at the period of F0 and its integer multiples (dotted lines). However, these modes are less prominent at the higher F0.

To derive pitch estimates from pooled interval distributions, we used "periodic templates" that select intervals at a given period and its multiples (middle panels of Fig. 7). Specifically, the salience of a periodic template is defined as the ratio of the weighted mean number of intervals within the template to the mean number of intervals per bin. A pitch estimate is obtained by finding the template whose fundamental period maximizes the salience. Templates with exponentially-decaying weights were found to give fewer octave and suboctave errors than flat templates. The bottom panels of Fig. 7 show the template salience as a function of template F0 for the same two stimuli as on top. For both stimuli, the salience reaches an abso-

lute maximum when the template F0 is very close to the actual stimulus F0. However, the maximum salience is larger for the lower F0.

Figure 8 shows measures of the accuracy, precision and strength of the interval-based pitch estimates as a function of F0. The estimates are very accurate below 1300 Hz, where their medians are within 1-2% of the true F0. The bootstrap interquartile ranges of the estimates are also below 1-2% for F0s up to 1600 Hz. However, the interval-based estimates of pitch abruptly break down above 1300 Hz due to the degradation of phase locking at harmonic frequencies near the CF. Since the stimuli have missing fundamentals, the lowest harmonic actually present is always above 2600 Hz, in a range where the degradation of phase locking is already substantial (Johnson, 1980).

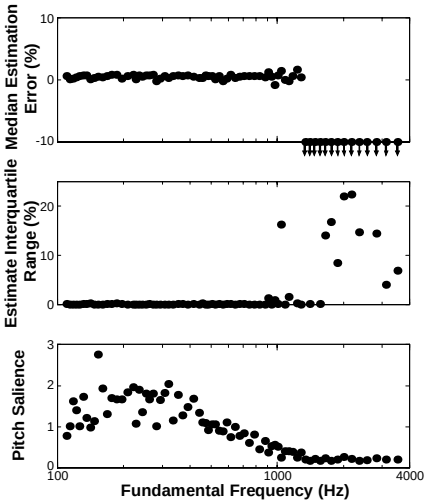


Fig. 8. Median pitch estimation error, interquartile range of the estimates and pitch salience as a function of F0.

The salience of the estimated pitch is highest below 400 Hz, then decreases gradually with increasing F0, to reach essentially zero at 1300 Hz. Thus, the salience of interval-based estimates of pitch is highest in the F0 range where rate-based pitch estimates are the least reliable due to the lack of strongly resolved harmonics. Conversely, the salience of rate-based estimates of pitch is highest above 2000 Hz, where the interval-based estimates break down.

5 Conclusions

We examined two possible representations of the pitch of complex tones with a missing fundamental in the cat auditory nerve for low and moderate sound levels.

1. A rate-place representation based on resolved harmonics was found to be viable for fundamental frequencies above 400 Hz.
2. Fundamental frequencies up to 1300 Hz were precisely represented in pooled interspike-interval distributions of the entire auditory nerve.
3. In the 400-1300 Hz range where both representations overlap, estimated pitch salience varies in opposite directions with F0 for the two representations.

The range of F0 over which the rate-place representation is viable in the cat does not include the 100-300 Hz region which is the most important for human voice. This failure may reflect the poorer frequency selectivity of the cat cochlea compared to the human (Shera et al., 2002). On the other hand, the frequency range of rate-based pitch estimates does include the F0 range of most cat vocalizations, which is centered around 600 Hz. An interesting question is whether this relationship between frequency ranges of vocalizations and resolved harmonics would hold in other species.

The range of F0s over which interval-based estimates of pitch are reliable in the cat roughly covers the entire perceptual range of the missing fundamental in humans (Moore, 1997). However, the salience of these estimates is strongest below 400 Hz, where individual harmonics are not strongly resolved in the cat. Thus, interval-based models of pitch may have trouble predicting the greater salience of pitch based on resolved harmonics compared to that based on unresolved harmonics (Shackleton & Carlyon, 1994).

In conclusion, neither representation of pitch is entirely satisfactory: the rate-place representation degrades at high sound levels and low frequencies, while the interval representation may have trouble accounting for the salience of pitch from resolved harmonics. This conclusion suggests a search for alternative neural codes for pitch such as those based on spatio-temporal patterns of discharge that would not rely on long interspike intervals (e.g. Shamma and Klein, 2000).

References

Cariani, P. A. and Delgutte, B. (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J. Neurophysiol. 76, 1698-1716.

Johnson, D.H. (1980) The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J. Acoust. Soc. Am. 68, 1115-1122.

Meddis, R. and Hewitt, M.J. (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery. J. Acoust. Soc. Am. 89, 2866-2882.

Moore, B. C. J. (1997) *An introduction to the psychology of hearing.* Academic Press.

Palmer, A.R. and Winter, I.M. (1993) Coding of the fundamental frequency of voiced speech sounds and harmonic complexes in the cochlear nerve and ventral cochlear nucleus. In Merchán, M.A., Juiz, J.M., Godfrey, D.A. and Mugnaini, E. (Eds.) *The mammalian cochlear nuclei: Organization and Function.* Plenum Press, New York, pp. 373-384.

Patterson, R. D. (1976) Auditory filter shapes derived with noise stimuli. J. Acoust. Soc. Am. 59, 640-654.

Rhode, W.S. (1995) Interspike intervals as correlates of periodicity pitch in cat cochlear nucleus. J. Acoust. Soc. Am. 97, 2414-2429.

Sachs, M. B. and Abbas, P. J. (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. J. Acoust. Soc. Am. 56, 1835-1847.

Sachs, M. B. and Young, E. D. (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. J. Acoust. Soc. Am. 66, 470-479.

Shackleton, T.M. and Carlyon, R.P. (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J. Acoust. Soc. Am. 95, 3529-3540.

Shamma, S. and Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. J. Acoust. Soc. Am. 107, 2631-2644.

Shera, C. A., Guinan, J. J., Jr. and Oxenham, A.J. (2002) "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements". Proc. Natl. Acad. Sci. USA 99, 3318-23.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Coding of pitch and amplitude modulation in the auditory brainstem: One common mechanism?

Lutz Wiegrebe,[1] Alexandra Stein,[1] and Ray Meddis[2]

[1] Dept. Biologie II, Universität München, Germany;
[2] Dept. of Psychology, University of Essex, UK
lutzw@lmu.de; stein@zi.biologie.uni-muenchen.de; rmeddis@essex.ac.uk;

1 Introduction

The neural mechanisms underlying the perception of pitch and amplitude modulation (AM) are still unclear. Current computer models designed to understand the functional mechanisms underlying the extraction of periodicity share a relatively detailed implementation of cochlear processing, but they diverge in their implementation of the neural processing strategy. Specifically, the modulation filterbank model of Dau, Kollmeier, and Kohlrausch (1997) recruits modulation filters not unlike the cochlear band-pass filters. A current model of pitch perception (Meddis and O'Mard 1997) recruits autocorrelation to extract periodicity from the information provided by the auditory nerve. Both these models lack a direct physiological correlate. Cariani and Delgutte (1996a,b) have shown that autocorrelation of auditory-nerve activation results in a good estimate of the perceived pitch for a comprehensive set of stimuli. Here, we investigate the hypothesis that sustained-chopper (Chop-S) units in the ventral cochlear nucleus (VCN) may play an important role in pitch and AM processing. Specifically, populations of Chop-S units with different best frequencies (BFs) and chopping periods (CPs) may serve to create a temporal place code of periodicity which, possibly at the level of the inferior colliculus (IC), is transformed into a rate-place code. This hypothesis is evaluated using a refined computer model of Chop-S units receiving their input from a state-of-the-art computer model of the auditory periphery.

2 Model structure

Stimuli are first subjected to a complex filter simulating the transformations of the outer- and middle ear (Glasberg and Moore 2002). Cochlear processing is implemented with a filterbank consisting of dual-resonance, non-linear band-pass filters (Lopez-Poveda and Meddis 2001). Inner hair cells (IHCs) are simulated by a module converting basilar-membrane motion into IHC depolarisation followed by a

refined model of the synaptic transmission into the auditory nerve (AN, (Sumner, Lopez-Poveda, O'Mard, and Meddis 2002). Each IHC is connected to 15 AN fibres with Poisson-like firing characteristics and a 0.75-ms refractory period. These 15 AN fibres converge on the dendrites of a Chop-S unit simulated as a McGregor cell (Hewitt, Meddis, and Shackleton 1992). The CP of the Chop-S unit is adjusted by manipulating the potassium recovery time constant of the McGregor cell.

3 Response to pure tones

The BF tone PSTH recorded 50 dB above threshold and a regularity analysis are shown in Fig. 1A from a Chop-S unit with a BF of 3 kHz recorded in the guinea-pig VCN. Figure 1B shows the simulation.

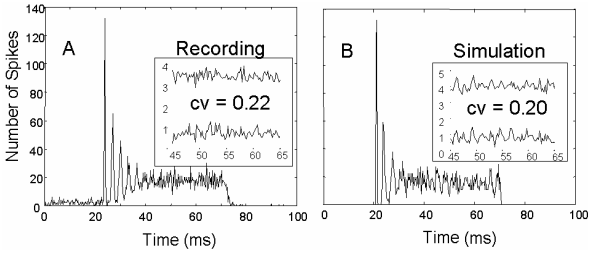


Fig. 1. Recording and simulation of a VCN Chop-S unit with a BF of 3 kHz in response to a 50-ms, 3-kHz pure tone starting with a delay of 20 ms. The level was 50 dB above firing threshold.

4 Responses to complex tones

Winter, Wiegrebe, and Patterson (2001) showed recordings of VCN Chop-S units in response to harmonic complexes in cosine phase (CPH) and random phase (RPH) and to iterated rippled noise (IRN) with a large range of fundamental periods (for IRN equivalent to the delay). IRN stimuli are generated using the 'add-same' algorithm (Yost 1996a). They are designated 'IRNS(d,g,n)' where 'd' is the delay in ms, 'g' is the linear gain in the delay loop, and 'n' is the number of iterations. Recordings and simulations of the above unit in response to complex tones and IRNS(d,1,16) are shown in Fig. 2. The simulated unit has a CP of 4 ms and simulates the general shape of the inter-spike interval histograms (ISIHs) in most cases with some exceptions for CPH complexes with higher f0s (shorter fundamental periods). Note that with IRNS or RPH stimuli and a fundamental period of 4 ms, the ISIHs show a pronounced mode, i.e., a concentration of intervals at the CP when the fundamental period is the same as the CP. This redistribution is quantified as 'Interval Enhancement' (see below). In contrast, the rate response is independent of the stimulus period for RPH and IRN stimuli (not shown).

The following sections investigate how arrays of simulated Chop-S units with different BFs and CPs encode the pitch of AM stimuli, IRN and harmonic complexes with different fundamentals and phase relations. To this end, the change these stimuli evoke in the ISIH is quantified. Specifically, we measure Interval Enhancement (Wiegrebe and Winter 2001) which represents the percentage of intervals corresponding to the stimulus period reduced by the percentage of intervals of the same length in response to Gaussian noise, i.e., in the absence of

periodicity. Wiegrebe and Winter (2001) have suggested that for IRN stimuli, this measure of temporal synchrony, applied to Chop-S unit recordings, is linearly related to the pitch strength perceived by humans.

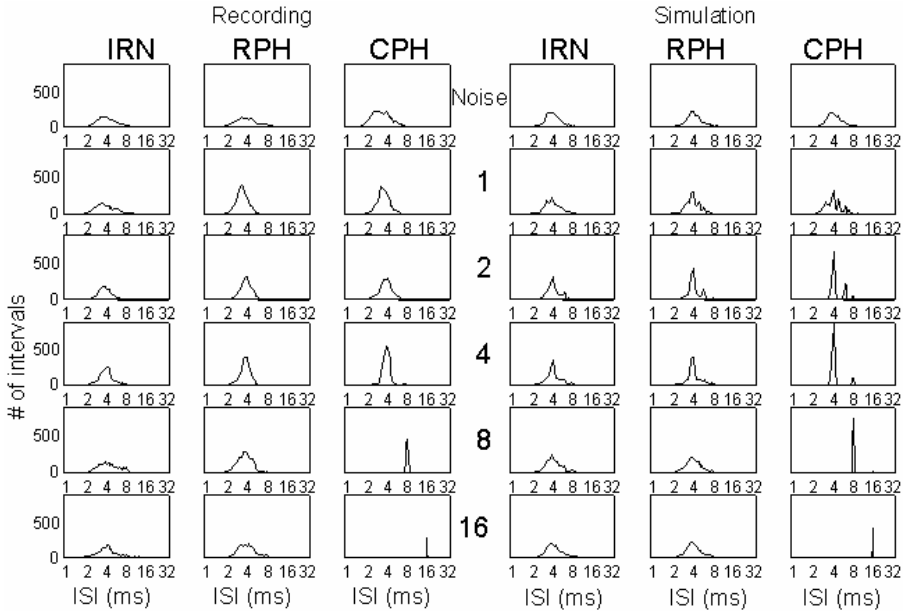


Fig. 2. Inter-spike interval histograms of a VCN Chop-S unit (BF = 3 kHz; CP = 4 ms, left) and its simulation (right). Stimuli were noise and IRN, RPH, and CPH complexes with a fundamental period in ms given by the number in the central column.

5 Simulations of pitch phenomena with the chopper model

5.1 Simulation paradigm

All subsequent simulations use populations of 20 simulated Chop-S units with BFs equally spaced on a log frequency axis between 200 and 6000 Hz. A population is characterised by its CP. Using the potassium recovery time constant, the CP was variable between 2 and 10 ms. Unless stated otherwise, all the following simulations are based on 25 repetitions of 409.6-ms stimuli with 20-ms raised-cosine ramps. Sound pressure level was fixed at 75 dB SPL.

5.2 The pitch strength of iterated rippled noise

At the psychophysical level, the pitch strength of IRNS increases with the number of iterations, while, at the physiological level, Interval Enhancement measured in VCN Chop-S units increases linearly with increasing pitch strength. The first simulation shows that Interval Enhancement can be used as an indicator of pitch strength in the subsequent demonstrations. The Chop-S unit population had a CP equal to the IRNS delay (4.2 ms). Interval Enhancement averaged across the 20

Chop-S units is shown by the fine line and error bars in Fig. 3. The thick line in the right panel shows how pitch strength increases (based on Yost, 1996b). Both Interval Enhancement and perceived pitch strength grow similarly with each doubling of the number of iterations. This confirms the linear relationship between pitch strength and Interval Enhancement.

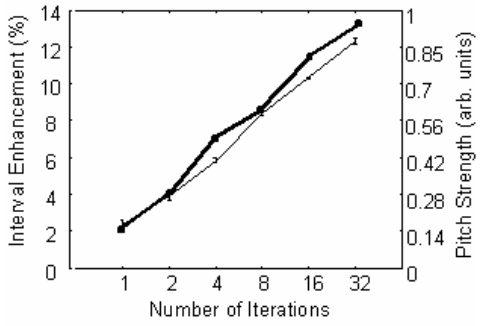


Fig. 3. Interval Enhancement (fine line and left Y-axis) and perceived pitch strength (thick line, after Yost, 1996b) of IRNS as a function of the number of iterations used to generate the IRNS. Note the linear relationship between Interval Enhancement and perceived pitch strength. Error bars represent standard errors across 3 simulations.

5.3 The pitch of harmonic complexes and the dominance region

One challenge for a temporal model of pitch such as the current one is the dominance region. The low harmonics contribute most to pitch strength but their periodicity corresponds to the period of the harmonic, not to the fundamental period of the complex. Wiegrebe and Winter (2001) and Winter et al. (2001) showed, however, that Chop-S units show interval enhancement when the integer multiple of the period of a harmonic is close to the CP. Thus, a unit spectrally tuned to the 2nd harmonic may enhance a period corresponding to the fundamental period when its CP corresponds to the fundamental period. Here a paradigm very similar to Ritsma (1967) is used to estimate the relative contribution of low and high harmonics to the overall Interval Enhancement in a population of Chop-S units. Two complexes are added where harmonics 1 to n have a lower f0 than harmonics n+1 to 38.

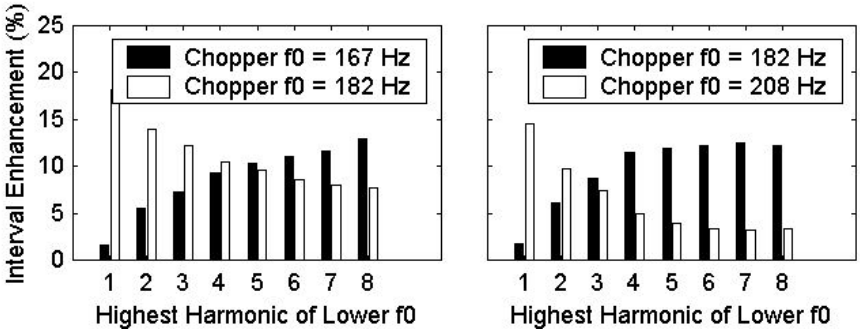


Fig. 4. Neural correlate of the dominance region in Chop-S units: With increasing n, the Interval Enhancement of the Chop-S population tuned to the higher f0 (open bars) decreases whereas the Interval Enhancement of the population tuned to the lower f0 (filled bars) increases. The crossing of the two population analyses should correspond to that n where the two f0s produce an equivalent pitch strength.

Simulated responses are obtained for two populations of Chop-S units; one with a CP equivalent to the reciprocal of the lower f0 and one population with a CP equivalent to the reciprocal of the higher f0. Interval Enhancement averaged across each population is shown in Fig. 4 as a function of n, i.e., the highest harmonic number of the lower f0. Data are shown for two different sets of f0s. Result show that a few low harmonics are sufficient to outweigh many high harmonics. Moreover, the harmonic number at the crossover point decreases with increasing f0 region. This is in line with psychophysical findings.

5.4 The pitch of inharmonic complexes

Chopper units are mostly associated with modulation tuning. Thus the pitches evoked by inharmonic complexes represent a special challenge for the chopper model because these pitches can only be explained through the interaction of both modulation and fine-structure information. Again, IRN is used as test stimuli because recent data reveal different pitches which depend not only on the inharmonic shift but also on the number of iterations. When the delayed noise is subtracted from the undelayed noise (IRN gain = -1), the spectral ripple is shifted by half the delay reciprocal, i.e., the stimuli are inharmonically shifted by 50%. Yost (1997) showed that with one iteration, IRN with a gain of –1 produces pitches of $\pm$ 10% of the delay reciprocal. With increasing number of iterations, however, listeners start to match a pitch equivalent to 1/2d, i.e., an octave below the pitch obtained with a positive gain.

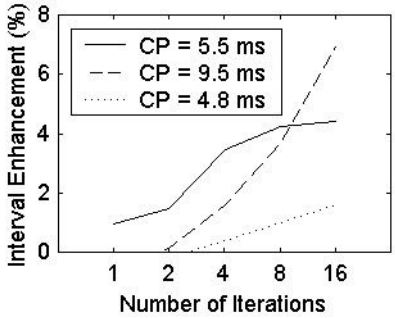


Fig. 5. The pitch of inharmonically shifted IRN. The stimulus was an IRNS(4.8,-1,n). For a low number of iterations, the predicted pitch is 182 Hz, i.e., about 10% below f0. With increasing number of iterations, Interval Enhancement of the Chop-S population tuned to 105 Hz (roughly an octave below f0) starts to dominate.

Figure 5 shows Interval Enhancement as a function of the number of iterations for an IRNS(4.8,-1,n). Simulated responses are obtained for three populations of Chop-S units, one with a CP equal to d, one with a CP near 1.1d and one with a CP near 2d. The results show that with one iteration, Interval Enhancement is strongest for the population tuned to 1.1d. With increasing n, Interval Enhancement of the population tuned to 2d starts to dominate. These simulation results are in good agreement with the perceptual data.

5.5 The effect of phase on pitch strength

Houtsma and Smurzynski (1990) showed that spectrally unresolved harmonic complexes added in Schroeder phase (resulting in a minimal envelope modulation)

produce weaker pitches than unresolved complexes added in sine phase. However, in line with experiments on the dominance region, both these pitches are much weaker than those produced by spectrally resolved complexes. Interval Enhancement for eight spectrally resolved harmonics added in sine or Schroeder phase and for eight unresolved harmonics also in sine and Schroeder phase is shown in Fig. 6. Note that while both the resolved complexes produce strong Interval Enhancement, the unresolved Schroeder-phase complex produces an even weaker Interval Enhancement than the unresolved sine-phase complex. These data are in good qualitative agreement with the experimental data of Houtsma and Smurzynski (1990).

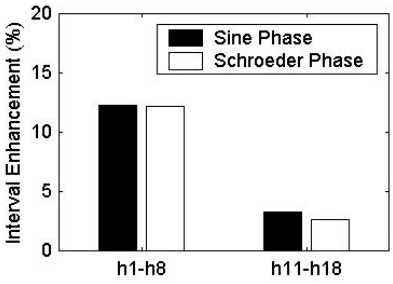


Fig. 6. Interval Enhancement for a group of 8 harmonics (h1-h8 or h11-h18) added either in sine phase (filled bars) or Schroeder phase (open bars). The low harmonics produce strong and phase-insensitive Interval Enhancement. High harmonics generally produce weaker Interval Enhancement but, in line with the psychophysics, sine-phase high harmonics produce stronger Interval Enhancement than Schroeder-phase high harmonics.

5.6 Sinusoidally amplitude modulated (SAM)-noise detection

SAM-noise detection deteriorates with increasing modulation frequency. Here an attempt is made to simulate these results in a population of Chop-S units.

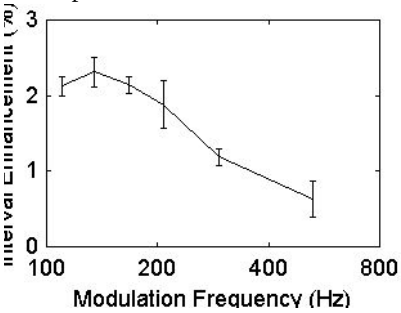


Fig. 7. Interval Enhancement of a population of Chop-S units with a CP equal to the modulation frequency in response to SAM noise. The decrease in Interval Enhancement is in qualitative agreement with the decreasing SAM sensitivity with increasing modulation frequency. Error bars represent standard errors across three simulations.

In general, Interval Enhancement in response to SAM noise is weak, in agreement with the weak SAM-noise pitch (Burns and Viemeister 1976). Moreover, Chop-S units may provide a physiological basis for modulation selectivity in the pitch range observed both perceptually and physiologically.

6 Summary and conclusions

Chop-S units may serve as a critical step in the conversion of the AN time code into a rate-place code of pitch as it is likely to exist in the central auditory system. Two Chop-S features are of special interest: first, Chop-S units exist with a range of CPs for each BF and second, Chop-S units can enhance an integer multiple of their input

periodicities. In the current model, the first feature is the basis for the generation of a temporal place code of pitch; the second feature is the basis for the temporal pitch extraction from spectrally resolved harmonics. For these model assumptions, the chopper model shows substantial parallels with a variety of critical psychophysical observations of pitch perception including the pitch and pitch strength of IRN, pitch shifts of inharmonic sounds, the effects of phase on pitch and pitch strength, and the pitch of SAM noise.

Acknowledgements

We thank Ian Winter for providing of the VCN recordings and many fruitful discussions. Supported by the Deutsche Forschungsgemeinschaft and the Medical Reasearch Council, UK.

References

Burns, E.M. and Viemeister, N.F. (1976) Nonspectral pitch. J.Acoust.Soc.Am. 60, 863-868.

Cariani, P.A. and Delgutte, B. (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J. Neurophysiol. 76, 1698-1716.

Cariani, P.A. and Delgutte, B. (1996b) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J. Neurophysiol. 76, 1717-1734.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892-2905.

Glasberg, B.R. and Moore, B.C.J. (2002) A model of loudness applicable to time-varying sounds. J. Audio Eng. Soc. 50, 331-342.

Hewitt, M.J., Meddis, R., and Shackleton, T.M. (1992) A computer model of a cochlear-nucleus stellate cell: responses to amplitude-modulated and pure-tone stimuli. J. Acoust. Soc. Am. 91, 2096-2109.

Houtsma, A.J.M. and Smurzynski, J. (1990) Pitch identification and discrimination for complex tones with many harmonics. J. Acoust. Soc. Am. 87, 304-310.

Lopez-Poveda, E.A. and Meddis, R. (2001) A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. 110, 3107-3118.

Meddis, R. and O'Mard, L. (1997) A unitary model of pitch perception. J. Acoust. Soc. Am. 102, 1811-1820.

Ritsma, R.J. (1967) Frequencies dominant in the perception of the pitch of complex sounds. J. Acoust. Soc. Am. 42, 191-198.

Sumner, C.J., Lopez-Poveda, E.A., O'Mard, L.P., and Meddis, R. (2002) A revised model of the inner-hair cell and auditory-nerve complex. J. Acoust. Soc. Am. 111, 2178-2188.

Wiegrebe, L. and Winter, I.M. (2001) Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus. J. Neurophysiol. 85, 1206-1219.

Winter, I.M., Wiegrebe, L., and Patterson, R.D. (2001) The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. J. Physiol. 537, 553-566.

Yost, W.A. (1996a) Pitch of iterated rippled noise. J. Acoust. Soc. Am. 100, 511-518.

Yost, W.A. (1996b) Pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 100, 3329-3335.

Yost, W.A. (1997) Pitch strength of iterated rippled noise when the pitch is ambiguous. J. Acoust. Soc. Am. 101, 1644-1648.

Pitch perception of complex tones within and across ears and frequency regions

Andrew J. Oxenham[1,2], Joshua G. Bernstein[1,2], and Christophe Micheyl[1,3]

[1] Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, {oxenham,cmicheyl,jgbern}@mit.edu
[2] Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts, USA
[3] Laboratoire Neurosciences et Systemes Sensoriels, UPRESA CNRS 5020, Lyon, France

1 Introduction

In judging the pitch of harmonic tone complexes, it is well known that lower-order harmonics (harmonic numbers less than about 10) produce a more salient (and accurate) pitch percept than higher-order harmonics, with a fairly sharp transition between the two regions (Houtsma and Smurzynski 1990). This difference has been explained in terms of whether or not the individual harmonics are peripherally resolved (e.g., Shackleton and Carlyon 1994). It is widely believed that the pitch of complexes containing resolved harmonics is derived from their individual frequencies, whereas the pitch of complexes containing only high harmonics is derived from the envelope repetition rate of the complex waveform produced when unresolved harmonics interact in the auditory periphery. While this is an appealing framework, which can be made to cover a wide range of pitch phenomena, some questions remain. The two questions addressed in this paper are: (1) Can harmonics that are not normally resolved contribute to the overall pitch if they are presented in a resolved manner, and (2) is there any evidence that the pitches produced by resolved and unresolved harmonics require some internal "translation" before they can be compared?

2 Experiment 1: Complex pitch perception with diotic and dichotic harmonic complexes

If poor pitch discrimination results from unresolved harmonics, can discrimination be improved by presenting alternate harmonics to opposite ears? If peripheral resolvability limits our ability to detect small changes in fundamental frequency (F0), then doubling the peripheral spacing between harmonics in each ear should lead to a doubling in the harmonic number at which F0 discrimination begins to

deteriorate. Earlier results have confirmed that the auditory system generally does combine information from both ears when deriving a pitch percept (Houtsma and Goldstein 1972; Darwin, Hukin, and al-Khatib 1995). However, results using two-tone complexes suggest that pitch perception is not improved by dichotic presentation. Houtsma and Goldstein (1972) found that the ability to perceive the F0 from two adjacent harmonics decreased with increasing harmonic number, regardless of whether the two harmonics were presented to the same or different ears. Similar conclusions have been reported in hearing-impaired listeners by Arehart and Burns (1999). Nevertheless, the pitch elicited by two components is generally very weak, leaving open the possibility that performance may be different with a larger number of components.

2.1 Methods

A method similar to that of Houtsma and Smurzinski (1990) was used. Stimuli consisted of twelve adjacent harmonics and F0 difference limens (DLF0s) were measured as a function of the lowest present harmonic number. The lowest harmonic number was roved across intervals by ±1 to reduce the effectiveness of spectral envelope cues. Listeners judged which of three intervals had the higher F0. The tones were presented in a background noise, which was uncorrelated in the two ears, had a spectrum level of 15 dB SPL below 600 Hz, and rolled off at 2 dB/oct. above 600 Hz, thereby producing masked thresholds that were nearly constant in terms of dB SPL between 200 and 4000 Hz. The equal-amplitude harmonic complexes were presented so that each component was at least 10 dB above its masked threshold in the noise. Each interval was 500 ms, including 30-ms raised-cosine ramps. A 2-down 1-up adaptive procedure tracked the 71% correct point. Fundamental frequencies of 100 and 200 Hz were tested, and components were added either in sine or in random phase. The components were either all presented to both ears (diotic condition) or were presented with alternate harmonics to each ear (dichotic condition). In the latter case, the distribution of harmonics across ears was randomized across intervals. Four listeners completed the sine-phase conditions. Four other listeners completed the random-phase conditions. All had normal audiometric thresholds.

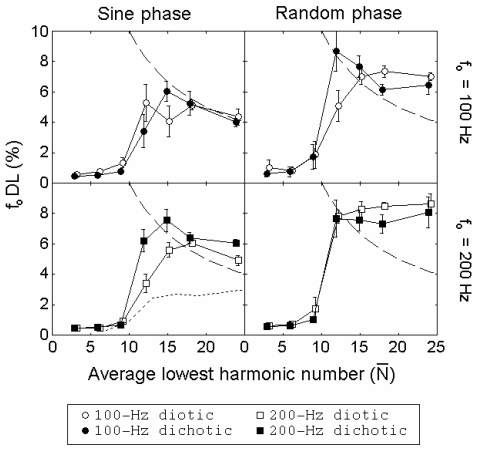


Fig. 1. Mean data from Experiment 1. Error bars denote ±1 s.e. of the mean. The dashed lines show the best possible performance based on the frequency of the lowest harmonic alone.

2.2 Results

The mean results are shown in Fig. 1. Consider first the results from the diotic conditions (open symbols). As expected based on earlier results,

performance deteriorates considerably as the lowest harmonic number is increased above about 9. The generally poorer performance of our subjects in conditions comparable to those of Houtsma and Smurzynski (1990) (dotted line in lower-left panel) is probably due to our lower signal-to-noise ratio (Hoekstra 1979). The fact that results from 100- and 200-Hz F0s (upper and lower panels, respectively) are very similar in terms of harmonic number and not absolute frequency is a confirmation that the deterioration in pitch perception with increasing harmonic number is not primarily an effect of absolute frequency due, for instance, to reduced phase locking. Consider next the results from the dichotic conditions (filled symbols). If peripheral resolvability *per se* were responsible for the deterioration in performance with increasing harmonic number, then the transition region between good and poor performance should have been shifted to the right by a factor of about two. In fact, results in the diotic and dichotic conditions are very similar. This implies that the transition between good and poor performance is determined at a level of processing higher than the auditory periphery. Finally, as expected (Houtsma and Smurzynski 1990), there is a trend for the random-phase conditions to produce higher thresholds than the sine-phase conditions at high harmonic numbers, consistent with idea that listeners use envelope cues in the presence of only high harmonics.

The main finding is that no benefit is obtained in F0 discrimination by presenting alternate harmonics to the two ears. In terms of models that incorporate harmonic templates, the results suggest that templates may only exist for harmonics that are normally resolved. This is consistent with the idea that such templates are learned from exposure to harmonic sounds (Terhardt 1974) or emerge from exposure to any broadband stimulation (Shamma and Klein 2000).

3 Experiment 2: Comparing pitches from resolved and unresolved harmonics

Differences in perception between resolved and unresolved harmonics have been used as evidence for two separate pitch mechanisms, in contrast to single-mechanism theories, such as that implemented by Meddis and O'Mard (1997). In perhaps the strongest version of a two-mechanism theory, Carlyon and Shackleton (1994) have suggested that pitch comparisons between two complexes are more difficult if one complex is resolved and the other is unresolved than if both are either resolved or unresolved, and have formalized this in terms of the involvement of a "translation" noise that is added when such across-mechanism comparisons of F0 are made. In their model, all stimuli are subject to an "encoding noise" (σ_e); comparisons across spectral region are subject to a further "comparison noise" (σ_c); and comparisons across resolvability are subject to the postulated translation noise (σ_t), where all noises are linear and additive with zero means.

The data supporting this proposal, however, are not strong. For instance, in deriving their noise estimates, Carlyon and Shackleton (1994) used F0DLs from *sequential* comparisons *within* a fixed spectral region to predict performance in *simultaneous* comparisons *across* spectral regions. The extent to which sequential

and simultaneous F0 comparisons reflect the same processes is not known; simultaneous differences may be detected by differences in perceived fusion regardless of perceived pitch, whereas sequential tasks must rely on a comparison of pitches and involves some form of memory. Thus, using one to predict the other may be misleading. Similarly, when resolved and unresolved harmonics are presented simultaneously, the resolved harmonics are likely to dominate the overall percept (e.g., Plomp 1967), perhaps making simultaneous differences in F0 across spectral regions less easy to detect. In other words, the pitch of the resolved harmonics may interfere with the pitch of the unresolved harmonics. Such interference may in fact be interpreted as evidence for a *single* mechanism, as interference implies interaction.

The purpose of the current experiment was to provide a stronger test for the presence of translation noise. All comparisons involved sequential presentations in an attempt to eliminate cues that were not directly related to perceived pitch differences. In a manner very similar to that used by Carlyon and Shackleton, two reference F0s were used in three spectral regions (LOW, MID, and HIGH, described below), such that complexes at both reference F0s contained resolved harmonics in the LOW region and contained only unresolved harmonics in the HIGH region. In the MID region, the complexes with the lower F0s contained only unresolved harmonics, while the complexes with the higher F0s contained resolved harmonics.

3.1 Methods

Sequential F0DLs were measured in six spectral configurations with two F0s (100 and 200 Hz) at two points on the psychometric function (71%- and 79%-correct) for a total of 24 conditions. The spectral configurations consisted of combinations of LOW, MID, and HIGH regions. The cutoff frequencies of the three regions were 600-1150 Hz, 1400-2500 Hz, and 3000-5250 Hz, respectively, with spectral slopes on either side of 48 dB/oct. The complexes were presented in a noise background with the same spectral characteristics as used in Experiment 1. The complexes were presented at a level of 45 dB SPL per component and the spectrum level of the noise in its flat spectral region, below 600 Hz, was 15 dB SPL. The complexes were added in sine phase and were 500 ms in duration, including 20-ms raised-cosine onset and offset ramps. Three spectral configurations involved within-region comparisons (LOW-LOW, MID-MID, and HIGH-HIGH) and three involved across-region comparisons (LOW-MID, LO-HIGH, MID-HIGH).

Thresholds were measured using an adaptive two-interval forced-choice method; listeners judged which of the two intervals contained the higher F0. Within each run, two independent tracks were randomly interleaved, with each track assigning the higher F0 to a different spectral region. For instance, in the LOW-MID condition, one track had the higher F0 assigned to the LOW region and the other track had the higher F0 assigned to the MID region. This allowed us to control for (and calculate) any response biases due to differences in spectral region. Of course, in within-region comparisons (e.g., LOW-LOW), there was no difference between the two tracks, but they were run in the same way for consistency. Six normal-hearing listeners took part. All listeners received at least 35 hours of

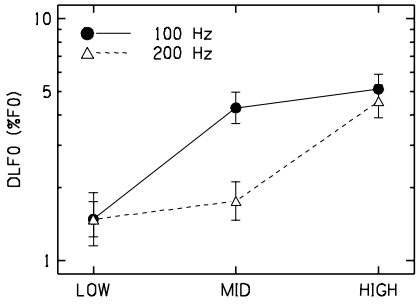


Fig. 2. Mean F0DLs for within-region comparisons.

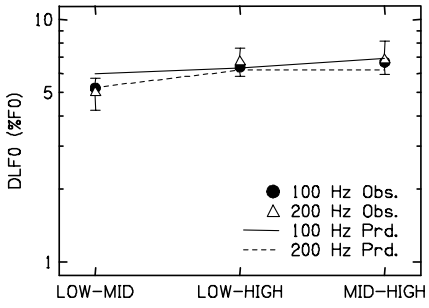


Fig. 3. Mean F0DLs for across-region comparisons. Symbols represent data; lines represent model predictions.

training before formal data collection began. Once thresholds had been collected, the "true" threshold and the bias were calculated from each interleaved pair of tracks, in a way similar to that described by Oxenham and Buus (2000). The last three such thresholds from each listener were geometrically averaged.

3.2 Results

Our measure of bias indicated no bias in the within-region conditions (as expected), and significant bias in only one of the across-region conditions, namely the 100-Hz MID-HIGH condition. This condition was the only one in which both intervals contained only unresolved harmonics; it is possible that because the pitch was weakest here, listeners were more prone to respond to spectral height (or timbre) in addition to pitch height.

The "true" thresholds from the 71% and 79% points on the psychometric function were generally different by a constant proportion, supporting the idea that sensitivity, as measured by d', is proportional to ΔF0 in Hz. For this reason, the results were collapsed across the two levels of sensitivity, so that the data shown in Figs. 2 and 3 represent performance at approximately 75% correct. The geometric mean results for the within-region comparisons are plotted in Fig. 2. Filled circles represent conditions with a 100-Hz F0 and open triangles represent conditions with a 200-Hz F0. The error bars represent the standard error of the geometric mean, after removing any overall differences in performance between listeners. As expected, conditions with resolved harmonics (100- and 200-Hz LOW and 200-Hz MID) produce substantially lower thresholds than conditions with only unresolved harmonics (100-Hz MID and HIGH and 200-Hz HIGH).

The thresholds in the across-region conditions are shown in Fig. 3. The symbols are the same as in Fig. 2. The lines represent predictions discussed in the following section. All the thresholds are higher than those in the within-region conditions. This suggests, consistent with many previous studies (e.g., Moore and Glasberg 1990), that F0 judgments across spectral regions are generally poorer than within-region judgments. In terms of the three additive noise sources outlined above, the results suggest that the comparison noise, σ_c, is significant compared to the

encoding noise, σ_e. A non-negligible translation noise would be reflected in elevated thresholds in conditions where one complex contained resolved harmonics and the other contained only unresolved harmonics (100-Hz LOW-MID and LOW-HIGH conditions; 200-Hz LOW-HIGH and MID-HIGH conditions). At first glance, this does not appear to be the case. This issue is pursued further below.

3.3 A simple model of additive noises

The results summarized in Fig. 2 were used to derive individual estimates of encoding noise, σ_e, for each combination of F0 and spectral region, using the following equation:

$$d' = \sqrt{2}k\Delta F0/\sigma_e, \quad (1)$$

where d' is the sensitivity, $\Delta F0$ is the difference in F0 in Hz, and k is a constant. The factor $\sqrt{2}$ is there because each stimulus is associated with the encoding noise and two stimuli are compared to make a judgment. The $\Delta F0$ is known, and the corresponding values of d' can be derived from the percent-correct levels (71% and 79%), giving us σ_e to within a constant. For simplicity, we assume k=1. A comparison of the results summarized in Figs. 2 and 3 was then made to derive estimates of the combined comparison noise and translation noise, $\sigma_c^2+\sigma_t^2$, using Eq 2. For illustration, this is done for a LOW-MID (LM) comparison:

$$\Delta F0_{LM} = \sqrt{(\sigma_{eL}^2 + \sigma_{eM}^2 + \sigma_{cLM}^2+\sigma_{tLM}^2)} \quad (2)$$

The terms $\Delta F0_{LM}$, σ_{eL}^2, and σ_{eM}^2 are all known, so it is straightforward to rearrange Eq 2 to determine $\sigma_{cLM}^2+\sigma_{tLM}^2$. The idea is that σ_c should be constant across all conditions; any conditions with a larger value of $(\sigma_{cLM}^2+\sigma_{tLM}^2)$ may be an indication of a translation noise. This formulation is essentially identical to that proposed by Carlyon and Shackleton (1994).

We derived values for $(\sigma_c^2+\sigma_t^2)$ for each subject in each of the twelve conditions (two F0s, three region comparisons, and two percent-correct levels), and then performed a repeated-measures analysis of variance. The dependent variable was the difference between the observed noise (right-hand side of Eq 2) and the predicted noise, assuming no comparison or translation noise. The factors were same-or-different resolvability, percent-correct level, and F0. Overall, the difference was significantly different from zero ($p<0.05$), confirming that the comparison noise was significant. However, there were no other main effects or interactions. This implies that the value of $(\sigma_c^2+\sigma_t^2)$ did not vary significantly across conditions. In other words, our assumption of a constant σ_c was valid and, furthermore, there was no evidence for an influence of σ_t, suggesting that it was either nonexistent or at least negligible. The predictions assuming a constant σ_c and no σ_t are shown as lines in Fig. 3. As can be seen, the fits are reasonably good. The best-fitting value of σ_c was about 14.8, which is greater than the encoding noise for the resolved harmonic regions (4.6), but not substantially larger than the encoding noise for the unresolved harmonic regions (13.7).

In summary, the results are not consistent with the idea that an additional translation noise is involved in comparisons between resolved and unresolved

harmonics. Instead it appears more likely that F0s from both resolved and unresolved harmonics are represented by the same neural code at the stage where sequential comparisons are made.

Acknowledgments

This work was supported by the National Institutes of Health (NIDCD grants R01 DC 05216 and T32 DC 00038).

References

Arehart, K.H. and Burns, E.M. (1999) A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss. J. Acoust. Soc. Am. 106, 993-997.

Carlyon, R.P. and Shackleton, T.M. (1994) Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? J. Acoust. Soc. Am. 95, 3541-3554.

Darwin, C.J., Hukin, R.W., and al-Khatib, B.Y. (1995) Grouping in pitch perception: Evidence for sequential constraints. J. Acoust. Soc. Am. 98, 880-885.

Hoekstra, A. 1979. Frequency discrimination and frequency analysis in hearing. Ph.D.,Institute of Audiology, University Hospital, Groningen, Netherlands.

Houtsma, A.J.M. and Goldstein, J.L. (1972) The central origin of the pitch of complex tones: Evidence from musical interval recognition. J. Acoust. Soc. Am. 51, 520-529.

Houtsma, A.J.M. and Smurzynski, J. (1990) Pitch identification and discrimination for complex tones with many harmonics. J. Acoust. Soc. Am. 87, 304-310.

Meddis, R. and O'Mard, L. (1997) A unitary model of pitch perception. J. Acoust. Soc. Am. 102, 1811-1820.

Moore, B.C.J. and Glasberg, B.R. (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. J. Acoust. Soc. Am. 87, 2163-2177.

Oxenham, A.J. and Buus, S. (2000) Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions. J. Acoust. Soc. Am. 107, 1605-1614.

Plomp, R. (1967) Pitch of complex tones. J. Acoust. Soc. Am. 41, 1526-1533.

Shackleton, T.M. and Carlyon, R.P. (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J. Acoust. Soc. Am. 95, 3529-3540.

Shamma, S. and Klein, D. (2000) The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. J. Acoust. Soc. Am. 107, 2631-2644.

Terhardt, E. (1974) Pitch, consonance, and harmony. J. Acoust. Soc. Am. 55, 1061-1069.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Internal noise and memory for pitch

Laurent Demany, Gaspard Montandon, and Catherine Semal

Laboratoire de Neurophysiologie, CNRS and Université Victor Segalen, Bordeaux, France
{laurent.demany,catherine.semal}@psyac.u-bordeaux2.fr

1 Introduction

Any sensation tends to be progressively "forgotten" with the passage of time. What does the forgetting process consist of and how does memory relate to the initial sensation? In this regard, a simple model based on signal detection theory (Green and Swets 1974) has been proposed by Kinchla and Smyzer (1967), who called it the "diffusion model of perceptual memory". This model was subsequently incorporated by Durlach and Braida (1969) into a more general conceptual framework, taking into account "context-coding" processes. According to the diffusion model, once a sensory trace x_0 is encoded in memory, it goes through a random walk process: the value stored in memory is randomly increased or decreased by some constant amount at a constant rate. After a time t, therefore, this value differs from x_0 by a quantity that can be considered as a Gaussian random variable with a mean of 0 and a variance proportional to t. When the task is to make a relative judgment on some attribute of two successive stimuli *S1* and *S2*, separated by an inter-stimulus interval (ISI) equal to t, performance will be limited by the sum of three internal noises: the two *sensory noises* inherent to the perception of the stimuli, plus the *memory noise* representing the imperfect retention of *S1*. In terms of d', the model states that:

$$d'(t) = \frac{2 \cdot \Delta S}{\sqrt{V_S + \Phi \cdot t}} \qquad (1)$$

in which ΔS is determined by the difference between *S1* and *S2*, V_S is the sum of the two sensory variances, and Φ represents the diffusion rate of the memory noise. As t increases, d' will decrease and it follows from Eq 1 that the relative decrease of d' will be:

$$\frac{d'(t)}{d'(0)} = \frac{1}{\sqrt{1 + (\Phi / V_S) \cdot t}} \qquad (2)$$

Thus, the relative speed at which d' decreases will be determined by Φ / V_S.

Kinchla and Smyzer (1967; see also Kinchla and Allan 1970) tested this model experimentally, with success. However, they manipulated only t and ΔS. In the

experiments reported here, by contrast, we tested the model by manipulating both V_S and t. In doing so, most importantly, we added to the model a basic assumption which was not made (explicitly) by Kinchla and Smyzer: we assumed that *the diffusion rate* Φ *of a given trace does not depend on the magnitude of the associated sensory noise.*

Two experiments will be reported. In each of them, listeners were required on each trial to compare the pitches of two sounds differing in frequency (or spectral centroid), and we measured d' as a function of two variables: (1) the ISI; (2) a physical parameter affecting pitch salience and thus V_S. In one condition, pitch salience was low, so that V_S was large. In another condition, pitch salience was high, so that V_S was small. Under the assumption about Φ stated above, the model predicted that, with an increase of the ISI (i.e., t), d' would decrease at a slower rate in the "large V_S" condition than in the "small V_S" condition. It was this prediction that we intended to test.

2 Experiment 1

2.1 Method

In experiment 1, all stimuli were bursts of sinusoids. In one condition, the fine structure of each burst consisted of only 6 sinusoidal cycles. In the other condition, the number of cycles was equal to 30. All stimuli had an amplitude envelope consisting of one cycle of a raised cosine function and were equated in energy. On each trial, the direction of the frequency shift from *S1* to *S2* was selected at random and the subject had to indicate which stimulus was higher in pitch. In order to prevent the subject from using a "context-coding" strategy (Durlach and Braida 1969), the frequency of *S1* was selected randomly (on a logarithmic frequency scale) from a wide range – namely between 400 and 2400 Hz. Visual feedback was provided on a screen following each response. Trials were grouped in blocks of 75, during which *S1* and *S2* contained a fixed number of sinusoidal cycles and were separated by a constant ISI: 200, 350, 500, 1000, 2000, or 4000 ms. Within each block, the delay separating a response from the onset of the next trial was equal to the ISI value incremented by 300 ms. When the ISI exceeded 1 s, a countdown was displayed on the screen, during both the ISI and the inter-trial delay.

In the experiment proper, the frequency shift from *S1* to *S2* ("ΔF_{rel}") had a fixed size on a log-frequency scale for each of the two stimulus conditions. Our aim was to obtain, for a 350-ms ISI, a d' value of about 2.0 in both conditions. For each subject, the appropriate values of ΔF_{rel} were estimated during a preliminary stage of the experiment that also served as a training phase. The experiment proper consisted of 10 sessions, run on different days. In each session, lasting about 1 h, 12 blocks of trials were run: one block for each of the six ISI values in each stimulus condition. From block to block, the "6 cycles" and "30 cycles" conditions were presented alternately and the ISI varied monotonically, from 200 ms to 4 s in half of the sessions and from 4 s to 200 ms in the other half. Four listeners with normal hearing, including two co-authors of the present paper, served as subjects.

2.2 Results and discussion

For a given subject, number of cycles and ISI, 750 trials had been run. They were treated as five successive subsets of 150 trials, from which five *d'* values were computed. The corresponding means and standard deviations of *d'* are displayed in Figure 1, together with the two ΔF_{rel} values which were used for each subject. The four subjects behaved similarly and the main findings can be summarized as follows: (1) In each stimulus condition, as the ISI increased, *d'* typically increased before decreasing; (2) the ISI for which *d'* reached its maximal value ("ISI_{opt}") was longer in the "30 cycles" condition than in the "6 cycles" condition; (3) above ISI_{opt}, *d'* did not appear to decrease more rapidly in the "30 cycles" condition than in the "6 cycles" condition; indeed, the overall trend was in the opposite direction.

The fact that *d'* was a non-monotonic function of the ISI could be anticipated from previous research (e.g., Massaro and Idson 1977; Tanner 1961). To our knowledge, however, there was no obvious reason to suspect that ISI_{opt} would differ in the two stimulus conditions and this comes as a surprise. Results such as those of Ronken (1972) seem to rule out the idea that, in either condition, the non-monotonicity could stem from a forward masking of *S2* by *S1* at the shortest ISI or ISIs. By itself, this non-monotonicity should not be taken as evidence against the model that we wished to test because one can argue that the forgetting process described by the model is liable to start some time after the end of the stimulus to be remembered. It is indeed reasonable to suppose that the *formation* of an optimally accurate pitch memory trace takes time and may require, if the stimulus is short, a duration exceeding by far the duration of the stimulus itself.

Let us adopt this point of view and admit that, in experiment 1, one should consider only the data obtained at and beyond ISI_{opt} in order to evaluate the model. From the data obtained at ISI_{opt}, we estimated – subject by subject, and under the assumption that *d'* was proportional to ΔF_{rel} (see in this respect Nelson and Freyman 1986) – the ratio R_V of the sensory variances involved in the two conditions [$R_V = V_S$ (*6 cycles*) / V_S (*30 cycles*)]. Besides, from the data obtained at and beyond ISI_{opt} in the "6 cycles" condition, we estimated for each subject the value of Φ / V_S in this condition; this was done by determining the best fit of Eq 2 according to a least-squares criterion. Under the assumption that Φ does not depend on V_S, the model implied that:

$$[\Phi / V_S]\ (30\ cycles) \ = \ R_V \cdot [\Phi / V_S]\ (6\ cycles) \qquad (3)$$

Using Eqs 2 and 3, we then calculated the *d'* values which were expected beyond ISI_{opt} in the "30 cycles" condition, given (i) R_V and (ii) $[\Phi / V_S]$ *(6 cycles).* In Figure 2, these expected values can be compared to those actually measured (replotted from Figure 1). For each subject, a steep drop in *d'* was expected beyond ISI_{opt}. In fact, the decrease of *d'* was much less abrupt. The discrepancy between expectations and measurements is so pronounced that it seems to provide rather strong evidence against the model.

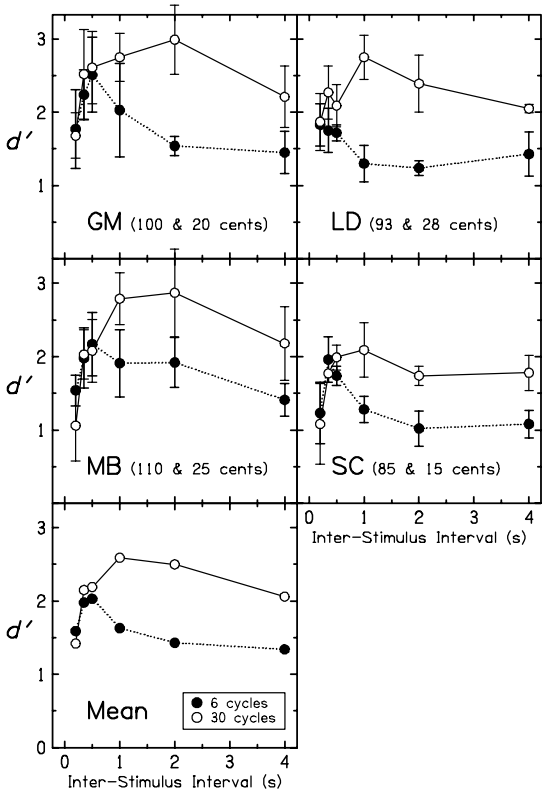


Fig. 1. Results of experiment 1. The four upper panels display the individual results, averaged in the bottom panel. *d'* represents the relative discriminability of ascending and descending frequency shifts (*S2* > *S1* versus *S2* < *S1*). The error bars have a total length of two standard deviations. For each subject, the ΔF_{rel} values used in the "6 cycles" and "30 cycles" conditions are indicated in musical cents (1 cent = 1/1200 octave).

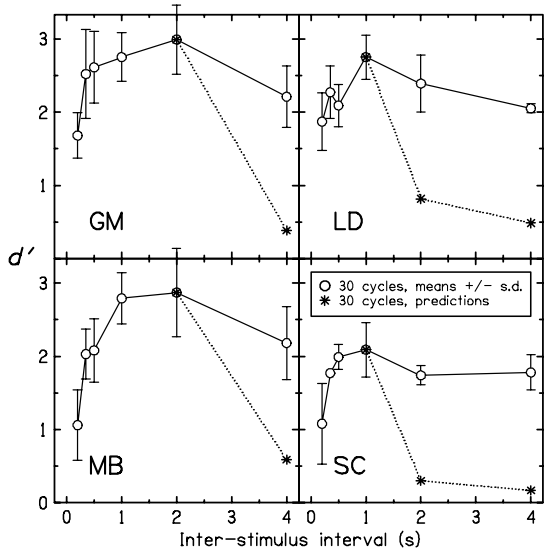


Fig. 2. Comparison between the results obtained in the "30 cycles" condition of experiment 1 (replotted from Fig. 1) and the results predicted by the model given (i) R_V and (ii)) $[\Phi/V_S]$ *(6 cycles)*.

3 Experiment 2

3.1 Method

In experiment 2, the "6 cycles" and "30 cycles" conditions of experiment 1 were replaced by conditions named, respectively, "Glide" and "Steady". The "Steady" condition was identical to the "30 cycles" condition, except that *S1* and *S2* had now exactly the same duration, corresponding to exactly 30 cycles of *S1* (but not *S2*). In the "Glide" condition, the stimuli were ascending frequency glides that covered 3 octaves with a straight trajectory on a log-frequency scale. On each trial in this condition, the central frequency of *S1* was selected randomly between 400 and 2400 Hz, like the frequency of *S1* in the "Steady" condition. The central frequency of *S2* was slightly lower or higher and the subject had of course to identify the direction of this shift. *S1* and *S2* had an identical duration, corresponding to that of a non-glided tone burst of 30 cycles at the central frequency of *S1*.

The procedure was essentially the same as that used previously. However, 8 different ISI values were used instead of 6, and 20 sessions were run following the preliminary stage. During the first 10 sessions, the ISI values were 100, 200, 350, 500, 1000, and 2000 ms. In the next 10 sessions, the ISI was equal to either 1500 or 4000 ms. Unfortunately, these two parts of the experiment were separated by a delay of several weeks. Among the four listeners who served as subjects, three had been previously tested in experiment 1.

3.2 Results and discussion

The data were analyzed exactly like those of experiment 1 and are displayed in Figure 3, together with the ΔF_{rel} values used in the two stimulus conditions. Again, as the ISI increased, *d'* initially increased and then decreased. This time, however, there was no significant difference between the ISI_{opt} values found in the two conditions. Overall, the decrease of *d'* beyond ISI_{opt} was not more rapid in the "Steady" condition than in the "Glide" condition, contrary to the model's prediction; only one of the four subjects (ED) produced a trend in the direction predicted by the model.

In Figure 4, the results obtained in the "Steady" condition can be compared to those predicted by the model (beyond ISI_{opt}), given (i) R_V and (ii) $[\Phi / V_S]$ *(Glide)*. The model's failure is not less spectacular than in experiment 1.

4 General discussion

The model that we tested, and apparently disproved, is the simplest model of perceptual memory that can be formulated within the framework of signal detection theory. In essence, it consists of the two following assumptions: (1) the accuracy of a perceptual comparison between two successive stimuli is limited by a

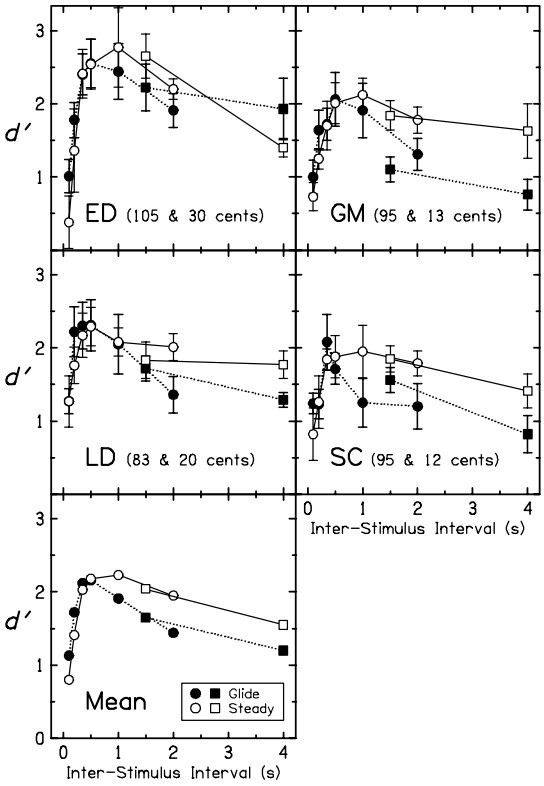


Fig. 3. Same as Fig. 1, but for experiment 2. The results obtained in the first and second parts of the experiment are respectively represented by circles and squares.

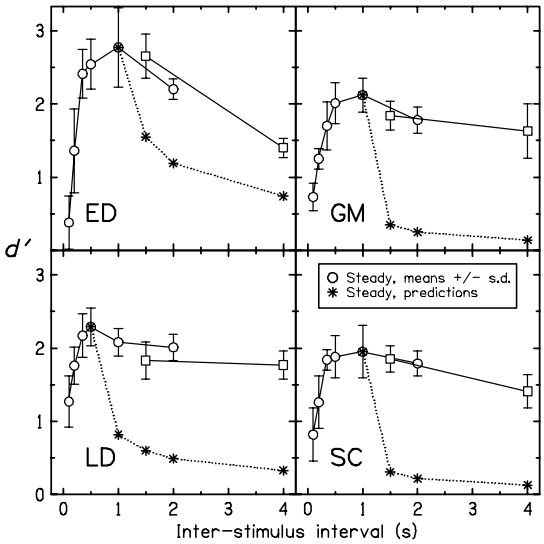


Fig. 4. Same as Fig. 2, but for the "Steady" condition of experiment 2.

combination of "sensory noise" and "memory noise" (the contribution of memory increasing with the ISI); (2) these two components of the overall internal noise are independent and added to each other. Assumption 1 is certainly valid, at least metaphorically. However, we suggest that assumption 2 is wrong.

It could be wrong for at least two reasons. First, the error may be limited to the postulate of addition. An alternative model positing that the two components of internal noise are independent but multiplied rather than added would make predictions more in line with the present data. And indeed, some previous investigators of perceptual memory have put forth empirical equations that amount to make this alternative assumption, in different terms (Laming and Scheiwiller 1985). From the physiological point of view, however, a multiplication of the two internal noises is more difficult to conceive of than an addition.

A second possibility is that the two sources of noise are not really independent or even separate. But this would contradict a common belief about memory: In the words of Magnussen (2001), "most concepts of memory assume a process that is distinct from on-line perceptual analysis in the sense that information is transferred to another location in the brain or transformed into a memory code, or both."

It is feasible to assess the relative merits of the multiplication hypothesis and the non-independence hypothesis in further experiments. Let us emphasize, however, that this further work should not be devoted exclusively to memory for pitch. Indeed, pitch may be memorized in a special way and quite differently from, for instance, loudness (Clément, Demany and Semal 1999).

References

Clément, S., Demany, L. and Semal, C. (1999) Memory for pitch versus memory for loudness. J. Acoust. Soc. Amer. 106, 2805-2811.

Durlach, N.I. and Braida, L.D. (1969) Intensity perception. I. Preliminary theory of intensity resolution. J. Acoust. Soc. Amer. 46, 372-383.

Green, D.M. and Swets, J.A. (1974) *Signal Detection Theory and Psychophysics*. Krieger, Huntington, New York.

Kinchla, R.A. and Allan, L.G. (1970) Visual movement perception: a comparison of sensitivity to vertical and horizontal movement. Percept. Psychophys. 8, 399-405.

Kinchla, R.A. and Smyzer, F. (1967) A diffusion model of perceptual memory. Percept. Psychophys. 2, 219-229.

Laming, D. and Scheiwiller, P. (1985) Retention in perceptual memory: a review of models and data. Percept. Psychophys. 37, 189-197.

Magnussen, S. (2001) Low-level memory processes in vision. Trends Neurosci. 23, 247-251.

Massaro, D.W. and Idson, W.L. (1977) Backward recognition masking in relative pitch judgments. Percept. Mot. Skills 45, 87-97.

Nelson, D.A. and Freyman, R.L. (1986) Psychometric functions for frequency discrimination from listeners with sensorineural hearing loss. J. Acoust. Soc. Amer. 79, 799-805.

Ronken, D.A. (1972) Changes in frequency discrimination caused by leading and trailing tones. J. Acoust. Soc. Amer. 51, 1947-1950.

Tanner, W.P. (1961) Physiological implications of psychophysical data. Ann. New York Acad. Sci. 89, 752-765.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Time constants in temporal pitch extraction: A comparison of psychophysical and neuromagnetic data

André Rupp[1], Stefan Uppenkamp[2,3], Jen Bailes[3], Alexander Gutschalk[1], and Roy D. Patterson[3]

[1] Sektion Biomagnetismus, Neurologische Klinik, Universität Heidelberg, andre.rupp@urz.uni-heidelberg.de
[2] Medizinische Physik, Fachbereich Physik, Universität Oldenburg, stefan.uppenkamp@uni-oldenburg.de
[3] CNBH, Department of Physiology, University of Cambridge, roy.patterson@mrc-cbu.cam.ac.uk

1 Introduction

Auditory models for the perception of pitch can be classified into two gross categories: (a) spectral pitch models where pitch is determined by peaks in the power spectrum of the sound, and (b) temporal pitch models where pitch is determined by time-intervals in the neural firing pattern produced by the sound. In the "Auditory Image Model" (Patterson, Handel, Yost, and Datta 1996) pitch corresponds to a stable pattern in the Auditory Image, which is created from the peripheral neural activity pattern by time-interval processing. There are several important time constants involved in temporal pitch extraction that determine the stability and persistence of the auditory image. In this paper, we compare psychophysical and neuromagnetic methods for estimating these time constants. Standard psychophysical tasks were used to determine the duration required to detect a short segment of regular-interval noise embedded in noise and vice versa. The results are compared with the neuromagnetic response to the stimuli in a subset of the psychophysical conditions.

Regular interval sounds (RIS) are created by delaying a copy of random noise and adding it back to the original. The resulting sound has some of the hiss of the original random noise, but it is also perceived to have a pitch with a frequency at the inverse of the delay time (Yost 1996). The pitch strength increases when the delay-and-add process is repeated. When the pitch is less than about 125 Hz (a delay time of 8 ms or more) and the stimuli are high-pass filtered at about 500 Hz, the RIS produces effectively the same average excitation as random noise in all frequency channels. There are no resolved spectral peaks in the excitation pattern on the basilar membrane. In this case, the perception of pitch is based on extracting time-intervals from the signal rather than spectral peaks. Figure 1 presents two stabilised auditory images created from simulated neural activity patterns in

response to random noise (left) and a RIS (right). Briefly, a histogram of the time intervals between peaks is created from the spike probability functions, one for each frequency channel. These time-interval histograms are stored in an image buffer, that decays with a half-life of 30 ms. RIS exhibit peaks in the summary auditory image at multiples of the delay in the delay-and-add process used to create the sound.

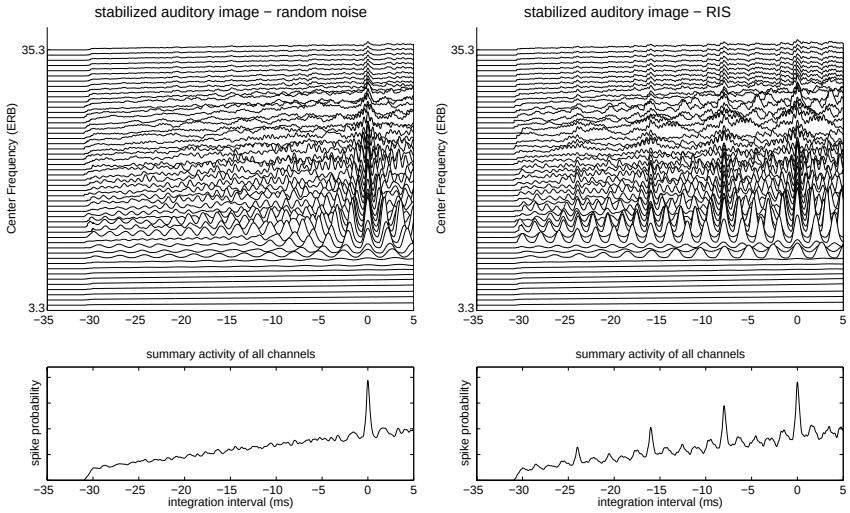


Fig. 1. Stabilised auditory images and summary images created from simulated neural activation in response to random noise (left) and a RIS (right) with 16 iterations and an 8 ms delay in the delay-and-add process.

2 Threshold for RIS in noise and vice versa

2.1 Methods

All of the RIS and noise were generated digitally and played through 16-bit D/A converters at a sampling rate of 20 kHz (TDT DD1). The stimuli were band-pass filtered from 500 Hz to 4 kHz (TDT PF1), and then attenuated to give an output level of 65 dB SPL (TDT PA4). They were presented diotically to the listener, via headphones with a diffuse-field response (AKG K240 DF). The listeners were seated in a double-walled, sound-attenuating booth, which contained a response box with four LEDs and corresponding buttons. Four normal-hearing listeners, aged 20 to 23 years, participated in the experiment.

Detection thresholds were measured using a two-interval, two-alternative, forced-choice adaptive procedure with feedback. Both intervals had durations of 500 ms (or 800 ms when the delay of the RIS was 20 or 24 ms). One interval was the standard, either a continuous random noise, or a RIS with a delay of 8, 12, 16, 20, or 24 ms, and with varying pitch strength determined by the number of itera-

tions in the delay-and-add process (1, 2, 4 or 16). The other interval had a short segment of either RIS or random noise embedded in the standard. The duration of this short test signal was varied with an adaptive procedure to find the minimum duration required to detect the interruption.

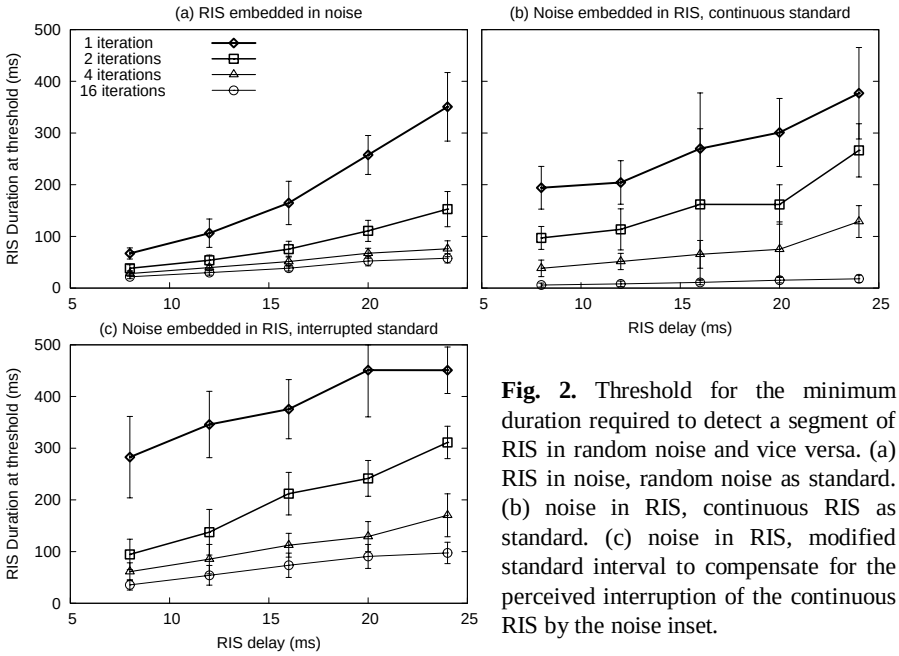


Fig. 2. Threshold for the minimum duration required to detect a segment of RIS in random noise and vice versa. (a) RIS in noise, random noise as standard. (b) noise in RIS, continuous RIS as standard. (c) noise in RIS, modified standard interval to compensate for the perceived interruption of the continuous RIS by the noise inset.

2.2 Results and discussion

Figure 2 shows the mean detection results for four listeners with standard deviations. In most stimulus conditions, the thresholds were similar across listeners. However, when the pitch of the RIS was weak (one iteration, low pitch) there were large differences between listeners as indicated by the standard deviations. In all tasks, detection of the RIS in noise and vice versa, thresholds were found to increase as pitch strength decreased, that is, as the RIS delay increased or the number of iterations of the delay-and-add process decreased. There was also an interaction between delay and pitch strength. The pitch strength had more effect at long delays than at short delays.

For strong pitch (16 iterations), detection of a short segment of noise in RIS (bottom line in Fig. 2b) appears to be much easier than a short segment of RIS in noise (Fig. 2a). Perceptually, the latter is very similar to simple detection of a tone in noise, while the former sounds more like gap detection. The *interruption* of the tone by the short noise appears to be the dominant cue, irrespective of whether this interruption is caused by silence, noise, or any other sound. It seems to interfere with the stable representation of the pitch of the RIS. When we compensated for

this interruption cue (Fig. 2c) by presenting a RIS in the RIS standard, listeners had to detect the presence of the noise on the basis of its timbre rather than the presence of an interruption, and the duration required to do this was considerably longer.

The most important finding from the psychophysical experiments is the strong asymmetry between the detection tasks, RIS in noise and noise in RIS. For most stimulus conditions, it was easier to detect a short tone in noise than vice versa. This indicates that the processing of tones and noises is different as suggested by time-interval models. We assume that there is a pitch-specific region in the auditory cortex, dealing with the temporal regularity of the sound. In previous brain imaging studies with functional MRI (Patterson, Uppenkamp, Johnsrude, and Griffiths 2002) and MEG (Gutschalk, Patterson, Rupp, Uppenkamp, and Scherg 2002), a region was located at the lateral end of Heschl's gyrus in both hemispheres that was activated differentially by RIS, while there was no centre differentially activated by noise. In the following section, the cortical response to RIS was explored with magnetoencephalography (MEG) in a recording paradigm that was derived from the psychophysical paradigm.

3 Neuromagnetic responses to RIS in noise and vice versa

Whole head MEG was employed to compare the perceptual data with the neuromagnetic responses of auditory cortex to RIS in continuous noise and noise in RIS. The perceptual data indicated that stimulus durations of 50 and 200 ms were in the range of threshold for strong and weak pitches, respectively and so the MEG experiments were limited to these durations. The RIS were created with delays of 8 and 20 ms, and 1 and 16 iterations.

3.1 Experimental procedures

The auditory evoked magnetic fields (AEFs) of ten normal hearing listeners were recorded with a Neuromag-122 whole head system during three separate sessions. In the first session RIS segments of 50 ms and 200 ms were embedded in continuous noise. In the second session noise bursts served as test signals and the background was continuous RIS. In the third session, continuous RIS was interrupted by a segment of another RIS having the identical delay and number of iterations, but based on a different noise sample. Each recording session lasted for 25 minutes resulting in about 320 averages per condition. Stimuli were presented diotically at a stimulus intensity level of 65 dB SPL via ER-3 earphones (Etymotic Research, Inc.) equipped with foam earpieces. The MEG data were recorded with a sampling rate of 1000 Hz (recording bandwidth: 0-330 Hz) and then filtered offline using a zero-phase filter ranging from 0.1 to 50 Hz. Spatio-temporal source analysis was performed using the BESA software (MEGIS Software GmbH). The time interval for the fit covered the N1m evoked by the RIS. There was one equivalent dipole in each hemisphere. For each subject the resulting BESA model of a condition with 16 iterations was held constant and used as a spatial filter to derive the source waveforms for each stimulus condition separately.

In the RIS-in-noise conditions the baseline preceding the N1m served as the reference. In the noise-in-RIS and RIS-in-RIS conditions, the magnitude of the N1m response was based on the peak-to-peak distance of the positive deflection with a maximum around 100 ms and the subsequent N1m elicited by RIS. The significance of the deflections was assessed using Student's *t*-intervals based on B=1000 bootstrap resamples.

3.2 Results and discussion

The grand average source waveforms for all stimulus conditions are shown in Fig. 3. A comparison of the morphology of the left and right hemisphere data revealed a very similar pattern for all stimulus conditions.

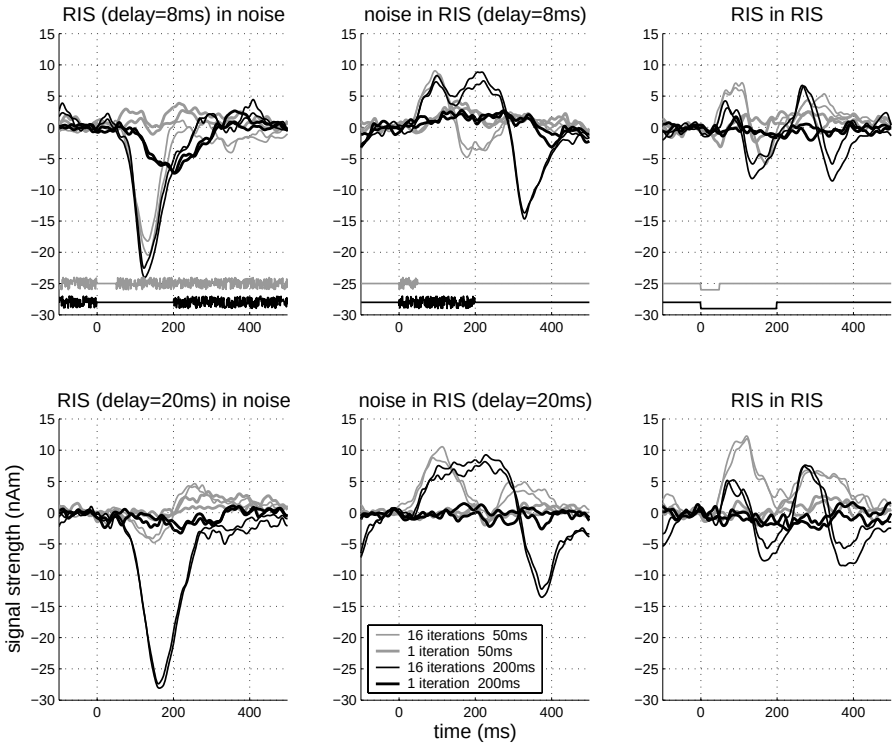


Fig. 3. Source waveforms from the two-dipole model based on a fit of the RIS-evoked N1m component. The additional lines in the top panels indicate the temporal position of the test tone in the continuous standard. Responses of both hemispheres are shown in the same line style. Note the asymmetric responses for RIS-in-noise, on the one hand, and both noise-in-RIS conditions, on the other hand.

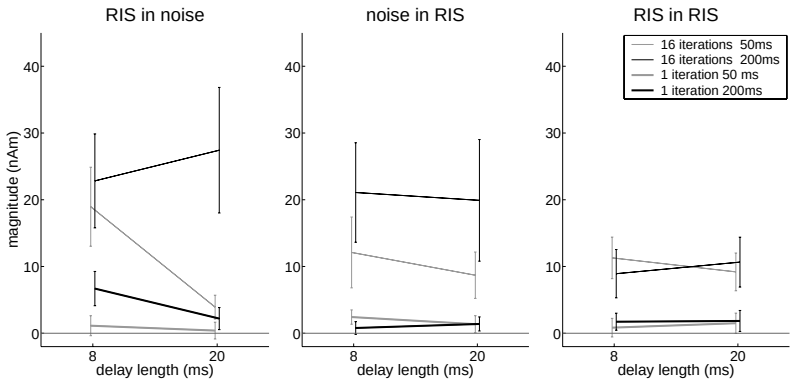


Fig. 4. Mean values of the N1m source strength for all stimulus conditions for the average data from the left and right hemispheres. The vertical bars give *t*-intervals.

RIS in noise evoked a prominent anterior N1m that increased in magnitude with the number of iterations, the length of the RIS segments and with decreasing delay length. An exception is the response derived from the RIS-in-noise condition with a delay of 20 ms based on 16 iterations, where the largest response was observed. The projection of the N1m on individual 3D-magnetic resonance images revealed that this component, referred to as the N100m' (Mäkelä, Hari, and Leinonen 1988) or the Pitch Onset Response (Krumbholz, Patterson Seither-Preisler, Lammertmann, and Lütkenhöner 2003), was located near the medial aspect of Heschl's gyrus in all subjects, close to the regularity-specific generator of the sustained field (Gutschalk *et al.* 2002). The latency of this component increased with delay (i.e., a drop in pitch), which agrees with the observations of Krumbholz *et al*. The N1m evoked by RIS with one iteration also peaked later than RIS with 16 iterations. One-sided *t*-tests (Fig. 4) indicated that no significant response was found for RIS inserts with 8 or 20 ms delay when the stimulus duration was 50 ms.

The source waves for the RIS-in-noise and RIS-in-RIS conditions had different morphologies. The interruption of the RIS standard evoked a positive deflection around 100 ms followed by the N1m elicited by the continuation of RIS. For long noise inserts in the conditions with 16 iterations, this positivity showed a plateau that was interrupted by the pitch specific response. Thus, it is unclear whether this late positive deflection represents an interruption of regularity-specific sustained fields (Gutschalk *et al.* 2002) or an excitation evoked by the noise burst onset. In these conditions, there was no big effect of RIS delay. For both, RIS-in-noise and RIS-in-RIS, neuromagnetic responses evoked by sounds based on one iteration were significantly smaller and very close to the critical *t*-value.

4 Conclusions

There is a strong asymmetry between the detection of RIS in noise, which is relatively easy, and the detection of noise in RIS, which is relatively difficult. A similar asymmetry is observed in the source analysis of the neuromagnetic responses to

these sounds in the auditory cortex. The magnitude of the RIS-evoked N1m depended strongly on the number of iterations of the delay-and-add process, that is, the perceived pitch strength. These observations are compatible with the suggestions that (i) the anterior N1m generator reflects a process closely related to pitch extraction based on regularity in the time interval patterns produced by the sound in the mid-brain, and (ii) there is a hierarchy in the processing of pitch and melody along the auditory pathway (Patterson *et al.* 2002).

There is no straightforward relationship between perceived pitch height and the minimum time required to detect a RIS in noise; a simple integration mechanism might have been expected to detect the relative excess of time-intervals at the RIS delay relatively quickly. The data indicate that there is probably an extra stage in the integration process that determines the perceived pitch strength from the number of iterations in the delay-and-add process. In another MEG study on the pitch of RIS (Krumbholz *et al.* 2003), it was demonstrated that the time constants of integration are affected by the frequency range of the stimuli as well, which parallels the reduction in pitch strength at higher frequencies observed psychophysically. There is good correspondence between the psychophysical data and neuromagnetic responses in the sense that RIS durations, that are subthreshold perceptually, do not result in significant MEG responses, while RIS stimuli with durations above the perceptual threshold give a clear N1m response. Moreover, the amplitude and latency of the neuromagnetic response are closely related to the pitch strength and duration of the RIS. In summary, our data indicate that the N1m responses evoked by regular-interval sounds might serve as an objective basis for estimating the time constants in temporal pitch processing.

Acknowledgement

These studies were supported by the UK MRC (G9901257) and DFG (Ru 652/3-1).

References

Gutschalk, A., Patterson R.D., Rupp, A., Uppenkamp, S. and Scherg, M. (2002) Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage 15, 207-216.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C. and Lütkenhöner, B. (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cerebral Cortex, *in press*.

Mäkelä, J.P., Hari, R. and Leinonen, L. (1988) Magnetic responses of the human auditory cortex to noise/square wave transitions. Electroenceph. Clin. Neurophysiol. 69, 423-430.

Patterson, R.D., Handel, S., Yost, W.A. and Datta, A.J. (1996) The relative strength of the tone and noise components in iterated rippled noise. J. Acoust. Soc. Am. 100, 3286-3294.

Patterson R.D., Uppenkamp, S., Johnsrude, I. and Griffiths, T.D. (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767-776.

Yost, W.A. (1996) Pitch of iterated rippled noise. J. Acoust. Soc. Am. 100, 511-518.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Auditory processing at the lower limit of pitch studied by magnetoencephalography

Bernd Lütkenhöner[1], Christian Borgmann[1], Katrin Krumbholz[2], Stefan Seither[1], and Annemarie Seither-Preisler[1]

[1] Institute of Experimental Audiology, Münster University Hospital, Münster, Germany, { lutkenh, christian.borgmann, seithers, preisler }@uni-muenster.de
[2] IME, Forschungszentrum Jülich, Jülich , Germany, k.krumbholz@fz-juelich.de

1 Introduction

A periodic click train produces a sensation of pitch when the repetition rate is larger than about 30 Hz (Krumbholz et al., 2000; Pressnitzer et al., 2001). However, when the rate is decreased below 30 Hz, the pitch fades away and the periodicity is perceived as roughness, flutter or pulsation. At rates below about 10 Hz, the individual clicks are heard as separate events (Warren and Bashford, 1981). The physiological substrate of the perceptual transition at the lower limit of pitch is largely unclear. Magnetoencephalography (MEG) could help to gain better insight, as it combines a reasonable spatial selectivity with an almost unlimited temporal resolution.

Most previous MEG studies on pitch perception focused on deflection N100m of the auditory evoked field (AEF), which occurs about 100 ms after the onset of an auditory stimulus. While earlier claims that N100m provides a window to a pitch-related functional map in the auditory cortex could not be confirmed (Lütkenhöner et al., 2001b; Lütkenhöner, 2003), it appears that pitch has a systematic effect on the N100m latency (Roberts and Poeppel, 1996; Roberts et al., 2000; Lütkenhöner et al., 2001a). Nevertheless, the question arises whether N100m is a good choice at all for the investigation of pitch extraction in the auditory cortex, because this response crucially depends on factors which have no psychoacoustic analogue, such as the time elapsed since the previous auditory stimulus (Imada et al., 1997). The occurrence of an N100m could even be disadvantageous: As this response often dominates the AEF, weaker responses, possibly being more relevant for the understanding of pitch extraction, might be masked.

Gutschalk et al. (2002) recently studied another AEF component, the sustained field (SF), which can be considered as a baseline shift roughly following the envelope of a sustained stimulus. They distinguished between an anterior source, located in lateral Heschl's gyrus (HG), and a posterior source, located in planum temporale (PT). A comparison between the responses to regular and irregular click trains suggested that the anterior source is particularly sensitive to regularity and largely

insensitive to sound level, and that the situation is reversed for the posterior source. The effect of regularity on the anterior source was most pronounced at short inter-click intervals (ICIs) and existed for ICIs as long as 25 ms.

The present study is related to the above one insofar as click trains were used. However, the main concern was not pitch perception *per se*, but the onset and offset of a pitch percept. In this respect, the present study is comparable to another recent MEG study (Krumbholz et al., 2003), in which transitions between regular-interval and random noises were considered.

2 Methods

Clicks were presented to the subject's right ear with cyclically varying ICI. Each stimulation cycle consisted of an acceleration stage (decreasing ICI, 600-ms duration), a periodic stage (constant ICI, 400-ms duration), and a deceleration stage (increasing ICI, 600-ms duration). Successive click train cycles were presented with silent intervals that were randomized between 300 and 350 ms. The click repetition rate during the periodic stage alternated, in random order, between 20, 30, 40, and 60 Hz. The longest ICI in both the acceleration and the deceleration stage was 100 ms. Figure 1a illustrates the four different cycles, and Fig. 1b shows a detail of the stimulation sequence as a whole. Stimulus intensity was adjusted to 50 dB above the behavioral threshold for single clicks. For convenience, the stimulation cycles will be named after their rate during the periodic stage.

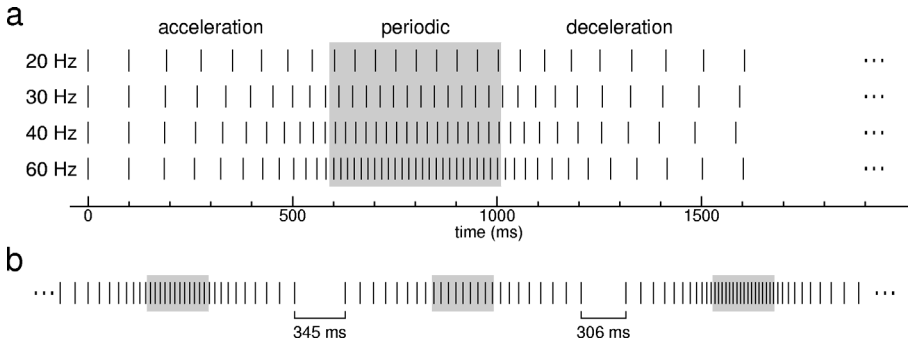


Fig. 1. a The four different stimulation cycles. Clicks are represented by vertical bars. The click rate during the periodic stage (gray highlight) is specified on the left. **b** Detail of the stimulation sequence as a whole (roughly three stimulation cycles visualized).

The magnetic field contralateral to the ear of presentation was recorded with a 37-channel neuromagnetometer (Biomagnetic Technologies, San Diego). Four subjects were studied, each in 8 independent sessions on four days (net measurement time almost 7 hours per subject). Epochs of 2.15-s duration were averaged separately for each type of stimulation cycle. The first click of a cycle served as the origin of the time axis; epochs began at –0.2 s. For a part of the analyses, the data were high-pass filtered (4^{th} order zero-phase Butterworth filter with ½ Hz edge

frequency), to reduce magnetic field fluctuations following the almost periodic presentation of stimulation cycles. This filtering was done before splitting up the data stream into epochs. Further methodological aspects, including details about dipole modeling (fixed dipole) and grand averaging of data from different sessions (in the same subject), are described in Lütkenhöner et al. (2003).

3 Results

The evoked magnetic field consists of two basic components: a slow response roughly following the stimulation cycles, and a fast response reflecting individual clicks. Figure 2 exemplifies, for each of the four cycles, the time course of the magnetic field at a single measurement location. The times of click presentation are indicated by dotted vertical lines. Since the cycles did not differ in the time points of the first two clicks (0 and 100 ms), the initial responses are basically identical. However, starting with the third click, which occurred at 186 ms in the 60-Hz cycle and 5 ms later in the 20-Hz cycle, the responses begin to diverge, although only marginally at first.

A closer look shows that the first click elicited two major positive deflections: one around 30 ms, denoted as P30m, and another around 80 ms. To facilitate an identification of corresponding deflections in the responses to subsequent clicks, the times 30 and 80 ms after a click are marked by an inverted triangle and a vertical bar, respectively. In the case of the 20-Hz cycle, both positive deflections are clearly recognizable in the responses to the first four clicks of the acceleration stage and the last four clicks of the deceleration stage. With decreasing ICI, however, the two deflections appear to merge, so only a single deflection remains during the periodic stage. Assuming that the 30- and 80-ms components in the responses to the first clicks do not become refractory at higher click rates, the low amplitude of the responses during the periodic stage of the 30-Hz cycle may mean that the two components are roughly in counter-phase and thus cancel each other at these click rates. Conversely, the large amplitude of the responses during the periodic stage of the 40-Hz cycle might be due to in-phase addition of the 30- and 80-ms components. During the periodic stage of the 60-Hz cycle the two presumed components would again be in phase (their latency difference corresponds to three times the ICI). The response amplitudes are, nevertheless, relatively small, possibly due to refractoriness caused by the relatively high stimulation rate. Refractoriness may also explain the fading of the fast response at the end of the periodic stage of the 60-Hz cycle.

The intricate dependence of the fast response on click repetition rate, largely resulting from refractoriness as well as interference of contributions from neighboring cortical areas, makes it unlikely that this response represents a direct correlate of pitch extraction. Fig. 2 suggests, however, that the *slow response* on which the fast response is “riding” might be such a correlate. In the case of the 20-Hz cycle, the slow response appears as an almost amorphous wave with a duration of about 2 s, corresponding to the *period* of a stimulation cycle, whereas at higher click rates it reflects the *structure* of the stimulation cycles as well. The response to the 60-Hz cycle shows two prominent negative deflections with latencies of about 750 and 1050 ms, respectively, which seem to be related to the transition from the accelera-

tion stage to the periodic stage and from the periodic stage to the deceleration stage. As these transitions correspond to the onset and offset of a pitch percept, the two deflections will be called pitch onset and offset responses, for the sake of convenience, although the basis for this terminology is just a working hypothesis.

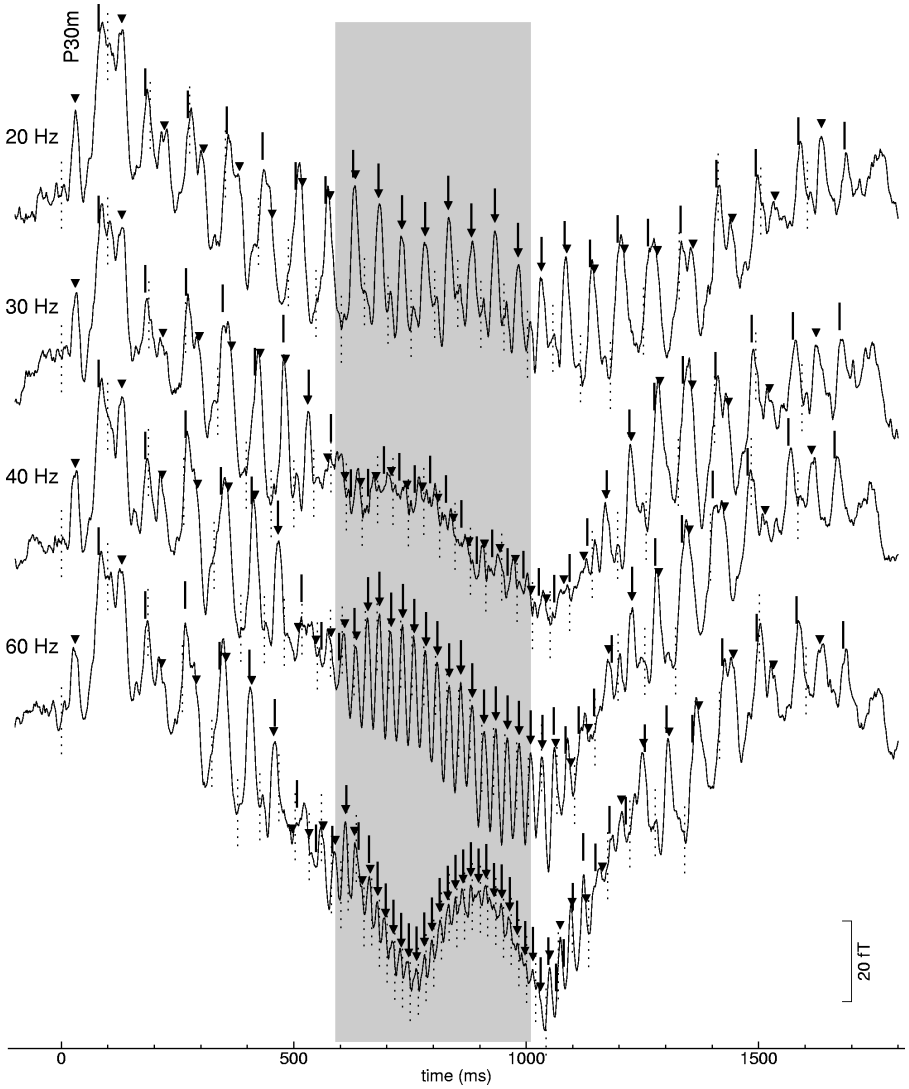


Fig. 2. Responses to the four different stimulation cycles (subject S1; magnetic field measured near its anterior inferior extremum; data low-pass filtered at 100 Hz). Click presentations are represented by dotted vertical lines. The times 30 and 80 ms after each click, corresponding to the two dominant positive deflections in response to the first click, are indicated by an inverted triangle and a vertical bar, respectively. Coinciding symbols form an arrow.

To focus on these responses, the data were bandpass filtered with edge frequencies of 0.5 and 10 Hz. Fig. 3 shows, for each subject and each cycle type, an overlay of the responses recorded at the 37 measurement locations. Except for subject S4, a distinct pitch onset response can be noticed only for the 60-Hz cycle, whereas an offset response can be observed also for the 40-Hz and 30-Hz cycles (even for the 20-Hz cycle in subject S1). Subject S1 also shows a strong response at about 100 ms; however, its positive polarity (cf. Fig. 2) is not compatible with the assumption that this is the well-known N100m.

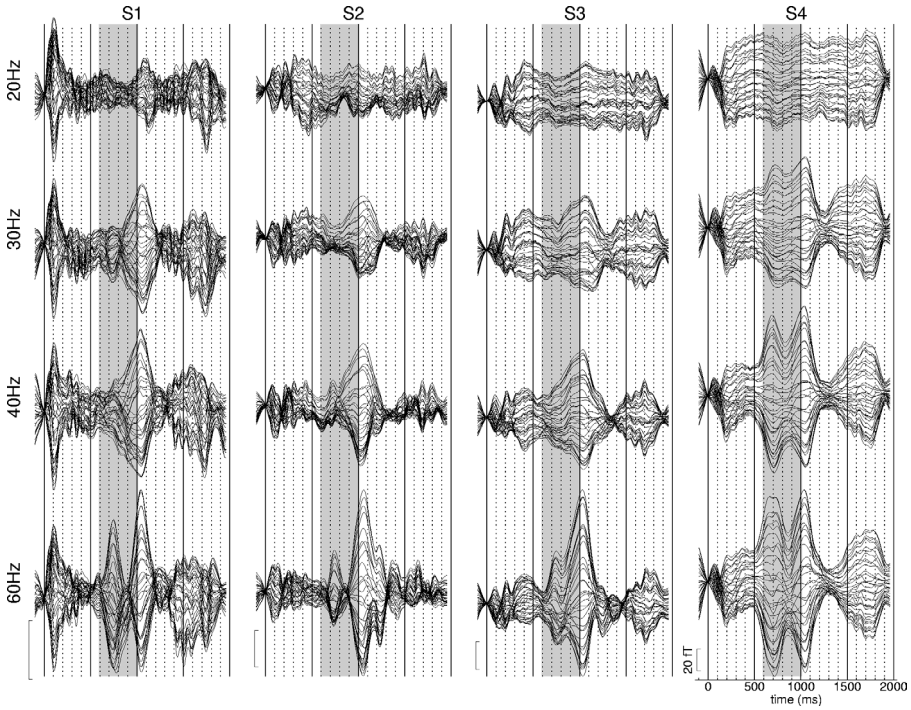


Fig. 3. Overlays of the responses at the 37 measurement locations. Each column represents one subject, each row one cycle type. Data bandpass filtered between 0.5 and 10 Hz.

Since a three-dimensional reconstruction of the auditory cortices was available for subject S1 (Krumbholz et al., 2003), it was tempting to perform a dipole source analysis in this subject, to get an idea of the activated cortical areas. The dipoles visualized in Fig. 4 appear to be lined up along Heschl's gyrus (HG). The most medial dipole was obtained for an early peak in the first derivative of the magnetic field, dP20m, which presumably reflects a fast activity increase in the primary auditory cortex (Lütkenhöner et al., 2003). Although dP20m is basically the point of steepest slope on peak P30m, the dipole obtained for P30m is located significantly more lateral, about halfway to the dipole representing the peak at 100 ms. All these dipoles were derived from the grand average of the responses to the four stimulation cycles. Only the 60-Hz cycle was considered when determining the dipoles for

the pitch onset and offset responses (black). Although the figure may suggest that the latter two dipoles have a similar location, close to the dipole obtained for the peak at 100 ms, a detailed analysis showed that the dipole explaining the pitch offset is not really a satisfactory model for the pitch onset, and vice versa.

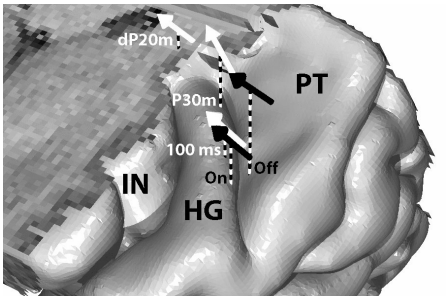


Fig. 4. Reconstruction of the auditory cortices with dipoles representing various components of the magnetic field (subject S1). The focus time range for the analysis with the fixed-dipole model was 17–22 ms for dP20m, 20–40 ms for P30m, 80–100 ms for the peak near 100 ms, 650–850 ms for the pitch onset response, and 900–1100 ms for the pitch offset response. The first three dipoles were derived from data lowpass filtered at 100 Hz, whereas the other two dipoles were derived from data band-pass filtered between 0.5 and 10 Hz. The dipoles were shifted upwards by 5 mm from their actual positions to prevent them from being partially hidden under the cortical surface. Each bar on the scales is 1 mm in height. Note that the dipoles do not mark exact source locations, but merely represent centers of activity. HG: Heschl's gyrus; PT: planum temporale; IN: insula.

4 Discussion

Krumbholz et al. (2003) observed a strong response to pitch onset, but almost no response to pitch offset. Thus, it may be surprising at first glance that, in the present study, the response to pitch offset was clearly dominating in three of four subjects. However, the two studies have little in common, apart from the fact that both dealt with the onset and offset of a pitch percept. Thus deviating results can probably be attributed to methodological differences.

Gutschalk et al. (2002) showed that the sustained field caused by a click train increases in amplitude with increasing click rate. Supposing that this relationship applies also to the more dynamic situation investigated in the present study, the slow response component following the almost periodic presentation of stimulation cycles (corresponding to about ½ Hz) could be explained, at least qualitatively. However, it is evidently not possible by this means to account for the emergence of a pitch onset or offset response, let alone an asymmetry between these two responses. Thus additional mechanisms have to be postulated.

Hypotheses about the nature of the observed pitch onset and offset responses are highly speculative for the time being, particularly as the data do not allow a unique interpretation. A serious problem is, for example, the definition of a baseline. In

Fig. 3, this problem was "solved" by high-pass filtering. However, such a data manipulation might lead one astray, as can be seen by considering the response to the 60-Hz cycle in subject S1. After high-pass filtering (Fig. 3), there seems to be no alternative to the assumption that a pronounced peak, representing the pitch-onset response, occurs at about 750 ms (150 ms after the beginning of the periodic phase). However, without high-pass filtering (Fig. 2) the data allow a completely different interpretation as well: A positive wave might start off at about 750 ms, reaching a maximum around 900 ms, and returning to baseline around 1050 ms. This wave might be related to a response component which, under different circumstances, would be denoted as P200m. Supposing that this interpretation is correct, the pitch offset response might be nothing else than the fast decrease of the sustained field after the end of the periodic phase. The responses to the other three cycles do not contradict this interpretation, although the positive wave postulated for the 60-Hz cycle is missing. Definite conclusions evidently require further experiments.

References

Gutschalk, A., Patterson, R.D., Rupp, A., Uppenkamp, S. and Scherg, M. (2002) Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. NeuroImage 15, 207-16.

Imada, T., Watanabe, M., Mashiko, T., Kawakatsu, M. and Kotani, M. (1997) The silent period between sounds has a stronger effect than the interstimulus interval on auditory evoked magnetic fields. Electroencephalogr Clin Neurophysiol 102, 37-45.

Krumbholz, K., Patterson, R.D. and Pressnitzer, D. (2000) The lower limit of pitch as determined by rate discrimination. J Acoust Soc Am 108, 1170-80.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C. and Lütkenhöner, B. (2003) Neuromagnetic evidence for a pitch processing centre in Heschl's gyrus. Cereb Cortex (in press).

Lütkenhöner, B., Lammertmann, C. and Knecht, S. (2001a) Latency of auditory evoked field deflection N100m ruled by pitch or spectrum? Audiol Neurootol 6, 263-78.

Lütkenhöner, B., Lammertmann, C., Ross, B. and Steinsträter, O. (2001b) Tonotopic organization of the human auditory cortex revisited: high precision neuromagnetic studies. In: Breebaart, D.J., Houtsma, A.J.M., Kohlrausch, A., Prijs, V.F., Schoonhoven, R. (Eds.), *Physiological and Psychophysical bases of auditory function*. Shaker, Maastricht, pp. 129-136.

Lütkenhöner, B. (2003) Single-dipole analyses of the N100m are not suitable for characterizing the cortical representation of pitch. Audiol Neurootol (in press).

Lütkenhöner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinsträter, O. and Patterson, R.D. (2003) Localization of primary auditory cortex in humans by magnetoencephalography. NeuroImage 18, 58-66.

Pressnitzer, D., Patterson, R.D. and Krumbholz, K. (2001) The lower limit of melodic pitch. J Acoust Soc Am 109, 2074-84.

Roberts, T.P. and Poeppel, D. (1996) Latency of auditory evoked M100 as a function of tone frequency. NeuroReport 7, 1138-40.

Roberts, T.P.L., Ferrari, P., Stufflebeam, S.M. and Poeppel, D. (2000) Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. J. Clin. Neurophysiol. 17, 114-129.

Warren, R.M. and Bashford, J.A., Jr. (1981) Perception of acoustic iterance: pitch and infrapitch. Percept Psychophys 29, 395-402.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Auditory maps in the midbrain: the inferior colliculus

Günter Ehret[1], Steffen R. Hage[2], Marina Egorova[3], and Birgit A. Müller[1]

[1] Department of Neurobiology, University of Ulm
guenter.ehret@biologie.uni-ulm.de
[2] Department of Neurobiology, German Primate Center Göttingen
shage@dpz.gwdg.de
[3] Laboratory of Comparative Physiology of Sensory Systems, I.M. Sechenov Institute of Evolutionary Physiology and Biochemistry, St. Petersburg
egorova@iephb.ru

1 Introduction

In the auditory midbrain inferior colliculus (IC) at least 19 major and minor ascending pathways from brainstem nuclei converge (e.g. Ehret 1997). Since the input projections show various patterns of spatial distributions (e.g. Kudo and Nakamura 1988; Oliver and Shneiderman 1991), including complex interactions of excitation and inhibition (e.g. Oliver, Winer, Beckius and Saint Marie 1994; LeBeau, Rees and Malmierca 1996; Ma, Kelly and Wu 2002), it is poorly understood how neural response patterns in the IC come about and how they may contribute to acoustical signal analysis and representation. A rather neglected aspect in the large number of publications on the IC is the study of spatial distributions of neural response properties. However, as in the visual system, orderly gradients and systematic representations of neural response properties over the available space (neural maps) may be the keys for understanding how sounds are transferred and transformed in the auditory system for perception and recognition.

Common to most auditory centers, from the cochlea up to the primary auditory cortex, is the well-known tonotopy. The central nucleus of the inferior colliculus (ICC) is tonotopically organized in so-called frequency-band laminae that are piled up like onion sheets through the three spatial dimensions of the ICC. Low frequencies are represented in the upper (dorsal) sheets, higher frequencies in the more ventral and ventromedial laminae. Further maps found in the whole ICC are those of tone-response threshold (Stiebler 1986), latency (Schreiner and Langner 1988a; Langner, Albert and Briede 2002), sharpness of frequency tuning (Stiebler 1987; Schreiner and Langner 1988a, b), and best-modulation frequency to amplitude-modulated tones in the cat and chinchilla (Schreiner and Langner 1988a, b; Langner et al. 2002). In recent studies on the mouse ICC, we have added maps of shapes of frequency tuning curves (Ehret, Egorova, Hage and Müller 2003; Hage and Ehret 2003), tone-response patterns and preferences for velocities and

directions of frequency sweeps (Hage and Ehret 2003). These maps and common features of all maps will be discussed here.

2 Methods

All mapping studies were done by extracellular recording of single-unit activity. The electrodes were advanced into the ICC stereotaxically with reference to the λ-point of the skull. Thus, every recorded neuron had a defined location with three coordinates, rostrocaudal and mediolateral of the λ-point and dorsoventral of the surface of the ICC. The recordings aimed at neurons with characteristic frequencies (CF) between 15 and 20 kHz which is the most sensitive frequency range of the mouse audiogram (Ehret 1974) and the area in the ICC with the largest spatial expansion of frequency-band laminae (Stiebler and Ehret 1985). All recording sites were projected to one horizontal plane corresponding to the horizontal projection of frequency-band laminae between 15 and 20 kHz.

3 Description of maps

3.1 Map of tuning-curve shapes

Neurons in the ICC have excitatory frequency tuning curves (FTCs) of various shapes. Four classes of FTCs can be discriminated mainly based on the steepness of slopes on the two sides of the FTCs (Egorova, Ehret, Vartanian and Esser 2001). Class I neurons had steep slopes (> 250 dB/octave) on the high-frequency side and shallow slopes (< 150 dB/octave) on the low-frequency side. Class II neurons had steep slopes on both sides (> 250 dB/octave, > 150 dB/octave, respectively), class III neurons had shallow slopes on both sides (< 250 dB/octave, < 150 dB/octave, respectively), and class IV neurons had more than one peak (two CFs). We found between 38% and 54% class I neurons and between 0% and 10% class IV neurons distributed all over a frequency-band lamina of the ICC. Class II neurons (24 – 27%) were concentrated in the center, class III neurons (24 – 25%) in the periphery. Thus, class II and III neurons showed significant concentric gradients of abundance over a frequency-band lamina (Fig. 1B,a) generating an average decrease of sharpness of tuning from the center to the periphery (Ehret et al. 2003; Hage and Ehret 2003).

3.2 Map of tone-response patterns

Numerous studies on the ICC (e.g. Ehret 1997) have shown a variety of tone-response patterns such as phasic-tonic, phasic, tonic, pauser, buildup (long latency) and chopper responses. If these patterns are measured at a certain level above threshold at the CF of a neuron, as is usually done, they do not unequivocally characterize the neuron's response in its excitatory receptive field, because of intensity and frequency dependence of the response patterns (Ehret and Moffat 1985; Ehret and Merzenich 1988). A more independent measure of a neuron's tone-

response pattern, is the whole-field peristimulus time histogram (PSTH) (Hage and Ehret 2003) summing up all responses within the excitatory receptive field (the area inside the FTC)

The mapping of whole-field PSTHs over frequency-band laminae led to about 58% neurons (N = 67) with phasic-tonic responses (including tonic, pauser and chopper responses) distributed all over frequency-band laminae, 22% neurons with purely phasic and 20% with buildup responses concentrated mainly in the periphery or center of a lamina, respectively. Thus, neurons with buildup and phasic patterns showed significant concentric gradients of abundance (Fig. 1B, b) generating at the same time a gradient of increasing precision of tone-onset coding from the center to the periphery of the ICC.

3.3 Maps of preference for velocity and direction of frequency sweeps

Neurons were stimulated with 10 repetitions of linear frequency sweeps, each at 30 dB and 50 dB above threshold at CF (Hage and Ehret 2003). We used upward and downward sweeps of the velocities 300 kHz/s, 600 kHz/s, and 3 MHz/s of 120 ms, 120 ms and 60 ms, respectively (always 5 ms rise and fall times included). All sweeps started and ended outside the FTCs of the neurons. Spike rates determined whether neurons responded to the sweep velocities at all and whether they preferred a sweep direction by producing at least twice the spike rate to one, compared to the other direction. We found (Hage and Ehret 2003) that 92.5% of the 67 neurons tested responded at least to one of the parameter combinations (velocity, direction, level) presented. Most neurons in the center of a frequency-band lamina did not respond or only weakly responded to the highest velocity, but responded well to the lowest. The responsiveness to higher velocities increased with a concentric gradient from the center to the periphery so that in the peripheral ICC the average response rate to all three velocities was very similar (Fig. 1B,c). Neurons in a strip at mediolateral coordinates of about 1.0 – 1.4 mm lateral to the λ-point (central part of the ICC) preferred upward sweeps. Neurons in more medial and lateral strips preferred downward sweeps. This is indicated by the arrows through the frequency-band laminae in Fig. 1A.

3.4 Other maps

The previously found maps of average sharpness of frequency tuning (Stiebler 1987; Schreiner and Langner 1988a, b) and average tone-response threshold (Stiebler 1986) had a concentric appearance with a gradient (sharpness decreasing, threshold increasing) from the center to the periphery of the frequency-band lamina (Fig. 1A, 10 kHz lamina). Two further maps of tone-response latency and best-modulation frequency to amplitude-modulated tones (Schreiner and Langer 1988a, b; Langer et al. 2002) are not concentric but have a gradient from medial (long latencies, low best-modulation frequencies) to lateral (short latencies, high best-modulation frequencies), as indicated by the shading on the 30 kHz frequency-band lamina in Fig. 1A.

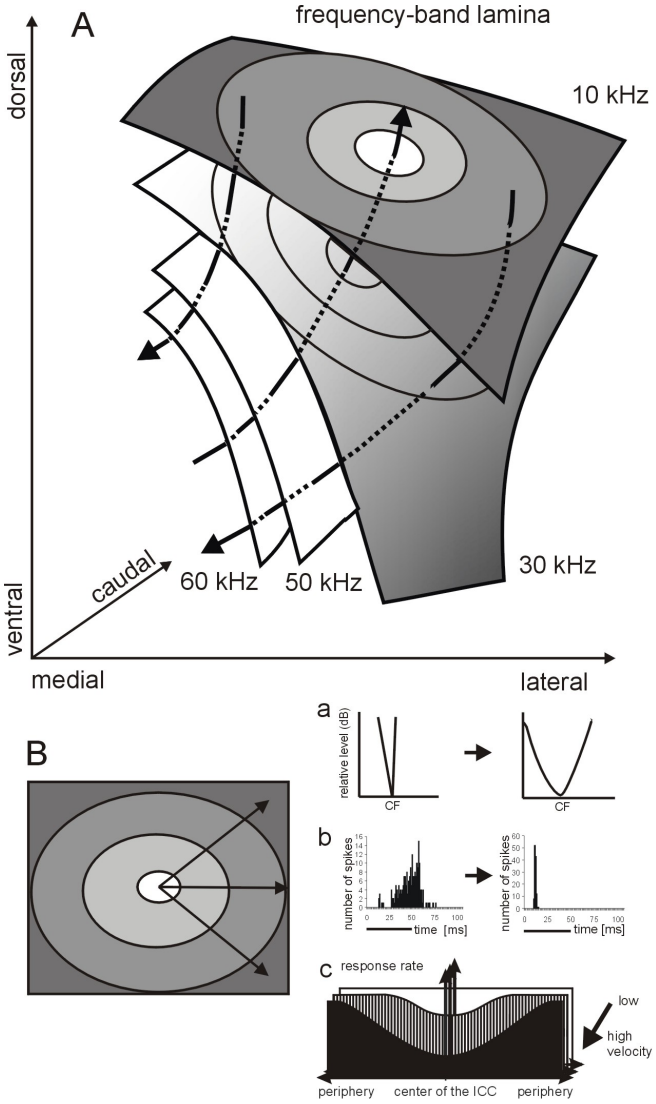


Fig. 1. A. Diagram of the 10, 30, 50, and 60 kHz frequency-band laminae in the three-dimensional space of the central nucleus of the inferior colliculus of the mouse (modified from Stiebler and Ehret 1985) indicating the tonotopic gradient. The shading across the 10 kHz lamina shows the gradient of five concentric collicular maps, the shading on the 30 kHz lamina, the dorsomedial to ventrolateral gradient of two maps (compare text). The arrows through the laminae indicate medial and lateral locations of neurons preferring frequency sweeps from low to high (upward) and central locations of preferences for high to low (downward sweeps). **B.** A schematic frequency-band lamina with changes from the center to the periphery of tuning-curve shape and sharpness of tuning (a), tone-response pattern from preference for buildup to phasic (b), responsiveness to low and high velocities of frequency sweeps (c) (modified from Hage and Ehret 2003).

4 Discussion of maps and mechanisms

On the background of the tonotopy, frequency-band laminae give room to further functional maps with basically three geometries: 1) concentric gradient; 2) gradients from dorsomedial to ventrolateral; 3) rostrocaudal strips at certain mediolateral positions. Concentric gradients relate to tone-response patterns and thresholds, composition of neurons of tuning-curve classes II and III, average sharpness of frequency tuning, and responsiveness to higher velocities of frequency sweeps. Tone-response latency and best-modulation frequency to amplitude-modulated tones follow the dorsomedial to ventrolateral gradient. Direction-selectivity for frequency sweeps is organized in three strips along the mediolateral coordinate.

We suggest that the concentric maps are based on at least two mechanisms showing the appropriate gradients, from the center to the periphery of frequency-band laminae: 1) decreasing inhibitory influences on neurons; 2) increasing numbers of intrinsically, phasically responding neurons. The arguments are that in the center of a frequency-band lamina class II neurons are sharply tuned due to strong sideband and whole-field inhibition (Egorova et al. 2001). Thus, they cannot respond to fast frequency sweeps. In the periphery, broadly tuned class III neurons with little inhibitory influence can respond to fast frequency sweeps, because the sweeping tone stays long enough and without inhibition in the excitatory response field. Neurons of class III tuning curves have, on average, higher tone-response thresholds compared to neurons of the other classes (Egorova et al. 2001). Further, the high incidence of the buildup response type in the center of a lamina indicates again the presence of strong and fast inhibition overriding the excitation there (see Rhode and Greenberg 1992; Young and Davis 2002). Finally, purely phasic responses seem to be due to intrinsic onset-spiking properties of ICC neurons. There are data from brain slices showing that onset-spiking neurons are mainly located in the periphery of the ICC just where the maximum number of neurons of the phasic response type are found (Reetz and Ehret 1999).

For the maps showing dorsomedial to ventrolateral gradients on frequency-band laminae, mechanisms are less clear. Recordings from neurons in brain slices of the ICC with electrical stimulation in the lateral lemniscus showed (Ehret and Reetz 1999) that the latency gradient found *in vivo* is present already *in vitro* and, thus, seem possibly to be an intrinsic property of the ICC connectivity. Further, projections from the cochlear nucleus seem to innervate the ventrolateral parts of frequency-band laminae more densely, compared to the dorsomedial parts (Casseday, Fremouw and Covey 2002), so that excitation may have stronger effects and produce faster responses in the ventrolateral, compared to the dorsomedial ICC.

Finally, the banded pattern of selectivity for the direction of frequency sweeps cannot be explained by the above-mentioned mechanisms. Fuzessery and Hall (1996) suggested that a main factor for the generation of direction-selectivity in neurons could be an asymmetric dendritic tree reaching for different distances into frequency-band laminae of higher and lower frequencies, thus collecting different amounts of excitation for upward and downward sweeps.

5 Conclusions

Sound processing in the auditory midbrain inferior colliculus is organized in functional maps. Besides tonotopy, eight such maps have been identified as orderly gradients on frequency-band laminae. The eight maps divide up into three basically different geometries suggesting three different complexes of mechanisms underlying the maps.

Acknowledgements

Supported by the VW-foundation (I/69589), the Deutsche Forschungsgemeinschaft (Eh 53/16,18), and the Russian Foundation for Basic Research (RFFJ.96–04–122).

References

Casseday, J.H., Fremouw, T. and Covey, E. (2002) The inferior colliculus: A hub for the central auditory systems. In: D.R. Oertel, .R.R. Fay, and A.N. Popper (Eds.), *Integrative Functions of the Mammalian Auditory Pathway*. Springer, New York, pp. 238-318.

Egorova, M., Ehret, G., Vartanian, I. and Esser, K.H. (2001) Frequency response areas of neurons in the mouse inferior colliculus. I. Threshold- and tuning- characteristics. Exp. Brain. Res. 140, 145-161.

Ehret, G. (1974) Age-depending hearing loss in normal hearing mice. Naturwissenschaften 61, 506.

Ehret, G. (1997) The auditory midbrain, a "shunting yard" of acoustical information processing. In: G. Ehret and R. Romand (Eds.), *The Central Auditory System*. Oxford University Press, New York, pp. 259-316.

Ehret, G., Egorova, M., Hage, S.R. and Müller, B. (2003) Spatial maps of frequency tuning-curve shapes in the mouse inferior colliculus. NeuroReport (in press)

Ehret, G. and Merzenich, M.M. (1988) Complex sound analysis (frequency resolution, filtering and spectral integration) by single units of the inferior colliculus of the cat. Brain Res. Revs. 13, 139-163.

Ehret, G. and Moffat, A.J.M. (1985) Inferior colliculus of the house mouse II. Single unit responses to tones, noise and tone-noise combinations as a function of sound intensity. J. Comp. Physiol. A 156, 619-635.

Fuzessery, Z.M. and Hall J.C. (1996) Role of GABA in shaping frequency modulation in the rat auditory cortex. Eur. J. Neurosci. 7, 438-450.

Hage, S.R. and Ehret G. (2003) Mapping responses to frequency sweeps and tones in the inferior colliculus of house mice (submitted).

Kudo, M. and Nakamura, Y. (1988) Organization of the lateral lemniscal fibers converging onto the inferior colliculus in the cat: an anatomical review. In: J. Syka, and R.B. Masterton (Eds.), *Auditory Pathway: Structure and Function*. Plenum Press, New York, pp. 171-183.

Langner, G., Albert, M. and Briede, T. (2002) Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*Chinchilla laniger*). Hear. Res. 168, 110-130.

LeBeau, F.E., Rees, A. and Malmierca, M.S. (1996) Contribution of GABA- and glycine-mediated inhibition to the monaural temporal response properties of neurons in the inferior colliculus. J. Neurophysiol. 75, 902-919.

Ma, C.L., Kelly, J.B. and Wu, S.H. (2002a) Presynaptic modulation of GABAergic inhibition by GABA(B) receptors in the rat's inferior colliculus. Neuroscience 114, 207-215.

Oliver, D.L. and Shneiderman, A. (1991) The anatomy of the inferior colliculus: a cellular basis for integration of monaural and binaural information. In: R.A. Altschuler, R.B. Bobbin, B.M. Clopton and D.W. Hoffman (Eds.), *Neurobiology of Hearing: The Central Auditory System*. Raven Press, New York, pp. 195-222.

Oliver, D.L., Winer, J.A., Beckius, G.E. and Saint Marie, R.L. (1994) Morphology of GABAergic neurons in the inferior colliculus of the cat. J. Comp. Neurol., 340, 27-42.

Reetz, G. and Ehret, G. (1999) Inputs from three brainstem sources to identified neurons of the mouse inferior colliculus slice. Brain Res., 816 527-543.

Rhode, W.S. and Greenberg, S. (1992) Physiology of the cochlear nuclei. In: A.N. Popper, and R.R. Fay (Eds.), *The Mammalian Auditory Pathway: Neurophysiology*. Springer-Verlag, New York, pp. 94-152.

Schreiner, C.E. and Langner, G. (1988a) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J. Neurophysiol. 60, 1823-1840.

Schreiner, C.E and Langner, G. (1988b) Coding of temporal patterns in the central auditory nervous system. In: W.M. Edelman, W.E. Gall, and W.M. Cowan (Eds.), *Auditory Function. Neurobiological Bases of Hearing.* Wiley & Sons, New York, pp. 337-361.

Stiebler, I. (1986) Tone threshold mapping in the inferior colliculus of the house mouse. Neurosci. Lett. 65, 336-340.

Stiebler, I. (1987) Frequenzrepresentation und Schallempfindlichkeit im Colliculus inferior und auditorischen Cortex der Hausmaus (*Mus musculus*). *Konstanzer Dissertationen*, Vol.173, Hartung-Gorre, Konstanz.

Stiebler, I. and Ehret, G. (1985) Inferior colliculus of the house mouse. I. A quantitative study of tonotopic organization, frequency representation, and tone-threshold distribution. J. Comp. Neurol. 238, 65-75.

Young, E.D. and Davis, K.A. (2002) Circuitry and function of the dorsal cochlear nucleus. In: D. Oertel., R.R. Fay, and A.N. Popper (Eds.), *Integrative Functions in the Mammalian Auditory Pathway*. Springer, New York, pp. 160-206.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Representation of frequency modulation in the primary auditory cortex of New World monkeys

Craig Atencio[1,2,3], Fabrizio Strata[2,3], David Blake[2,3], Ben Bonham[2,3], Benoit Godey[2,3] , Michael Merzenich[2,3], Christoph Schreiner[1,2,3], and Steven Cheung[2,3]

[1] UCSF/UCB Bioengineering Graduate Group
[2] Coleman Memorial Laboratory, UCSF
[3] Keck Center for Integrative Neuroscience, UCSF, {craig, fab, dblake, ben, begodey, merz, chris, cheung}@phy.ucsf.edu

1 Introduction

Vocalizations are the primary means of communication among non-human primates. Indeed, among New World monkeys, such as the squirrel monkey and owl monkey, vocalizations may take many forms, such as cackles, growls, trills, twitters, etc. (Winter et al. 1966; Jurgens 1986; Wang et al. 1995; Jurgens 1998). Thus each species has many stereotyped communication calls in its vocal repertoire, each communicating some unique information to other members of its social environment.

In this study we quantitatively described the responses of the primary auditory cortex (AI) of the anesthetized squirrel monkey and awake owl monkey to tone bursts, random tone pips, and frequency modulated sinusoids (FM sweeps) - sounds whose frequency content is stationary or whose frequency content changes smoothly over time from one frequency to another, respectively. FM sweeps serve as simplifications of the frequency transitions that are present in many natural monkey calls. Discerning how these sweeps are represented across the auditory cortex allows us to see if there is a systematic and orderly representation of this parameter in AI.

This study is unique in that it employs highly dense mapping of neurons throughout the extent of AI in the squirrel monkey and single unit recording in the awake owl monkey. While we previously showed that there is a precise and systematic map of CF and bandwidth across AI of the squirrel monkey (Cheung et al. 2001) the picture for FM sweep processing is less clear. We discuss the spatial organization of FM sweep responses as well as the population responses for both species.

2 Methods

We report here the results from experiments involving three young adult male squirrel monkeys and two young adult male owl monkeys. All procedures were approved by the institutional animal welfare committee at UCSF and were consistent with national and state animal welfare guidelines. Our surgical procedures and anesthetic protocol were described in our previous studies (deCharms et al. 1998; deCharms et al. 1999; Cheung et al. 2001). The main difference between the two types of experimental preparations was the anesthetic state. The squirrel monkeys were anesthetized with pentobarbital sodium for the duration of each experiment, usually 3-4 days. By contrast, the owl monkeys were implanted with chronic recording electrodes allowing for many recording sessions in the awake, non-behaving state. The stimulation protocols were quite similar. For both species we recorded responses to 16 logarithmic FM sweeps at speeds of 10, 17, 30, 52, and 90 octaves per second over the frequency range 50-21000 Hz in both the up up and down sweep directions. In the squirrel monkey pure tone tuning curves were also computed while in the owl monkey spectro-temporal receptive fields (STRFs) were constructed using random tone pip stimuli. Tone pip densities were the same as in a previous study, and all later calculations were from STRFs obtained using a tone pip density of one tone pip per octave per 64 milliseconds (Blake and Merzenich 2002).

For the FM sweep responses for both monkey types we created post-stimulus time histograms (PSTHs). We then fit Gaussian curves to each PSTH, thereby calculating the latency, width, and amplitude of each response. We defined the response to a particular sweep speed as the area beneath the best fitting Gaussian. From these responses we calculated directional selectivity indices as well as centroid (weighted best speed) measures, as in previous studies (Shamma et al. 1993; Nelken and Versnel 2000).

3 Results

The results in these studies are from 507 penetration sites in 3 squirrel monkeys (SM01, SM64, SM82) and from 128 sites in two owl monkeys (OM1, OM2). AI was physiologically located. For the squirrel monkey recordings AI was also verified anatomically.

3.1 Single site examples

The responses of a typical squirrel monkey AI cell to these FM sweep stimuli are shown in Fig. 1, where the left column shows the responses to upward sweeps and the right column the responses to downward sweeps. Each individual plot represents the response of the cell to sixteen presentations of the same FM sweep stimulus. The bar under each histogram represents the duration of the FM sweep. This cell most vigorously responded to low sweep speeds, similar to the FM sweep values present in squirrel monkey twitter calls.

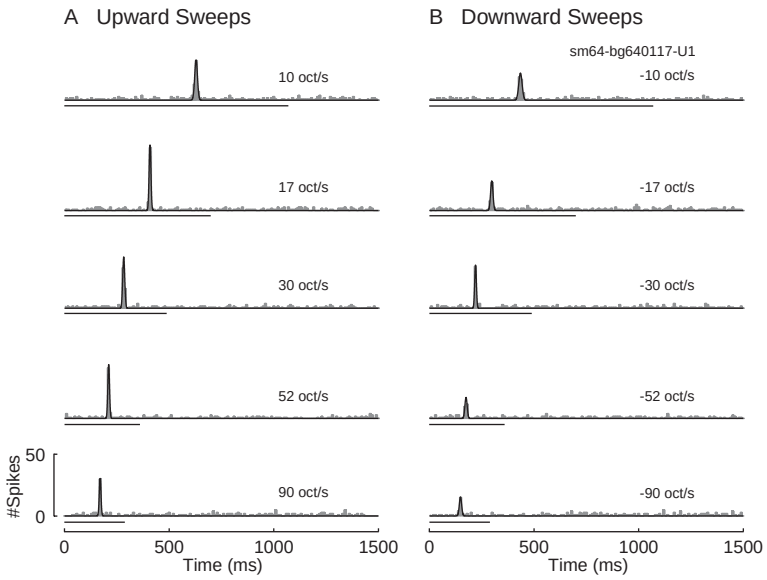


Fig. 1. Example PSTH from squirrel monkey of FM sweep responses in the up (A) and down (B) directions. Gaussian curve fits superimposed on each PSTH. Bar beneath each PSTH indicates FM sweep stimulus duration.

To quantify the nature of the response we fit Gaussian curves to each of the ten PSTHs obtained for each FM sweep stimulus. In Fig. 1 these are seen as the black outlines superimposed on each PSTH.

In Fig. 2A we plot the latency of the responses to FM sweeps versus the inverse of FM sweep speed. From Fig. 2A we see that each directions gives a different straight line fit (Up: r=0.994, p<1e-3; Down: r=0.997, p<1e-3). Using the slopes of these lines we computed the frequency at which the cell first responds to the FM sweep, termed the trigger frequency (Heil and Irvine 1998). The trigger frequency for upward sweeps is given by $F_{UP} = 50 * 2^{S}$ *Hz*; for downward sweeps by $F_{DN} = 21000 * 2^{-S}$ *Hz*. For the cell in Fig. 2 the frequency calculated from the up direction sweeps is 8340 Hz and from the down direction sweeps this value is 9280 Hz. The pure tone CF of the unit was 9780 Hz.

In Fig. 2B we also show plots of normalized area versus sweep speed. The normalized areas are calculated by dividing all responses, in both sweep directions, by the largest area value, as determined from both directions. These curves of response magnitude show the selectivity for sweep speed in either the up or down directions. The example in Fig. 2B preferred both upward and downward direction sweeps at low sweep speeds. Note the similar tuning to sweep speed for both up and down directions. This strong correlation was evident for the majority of our sites.

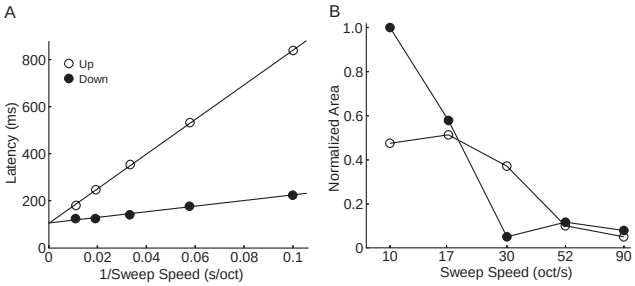


Fig. 2. Analysis of data if Fig. 1. A: Latency vs. inverse sweep speed for up and down FM sweep direction. B: Normalized area vs. sweep speed.

3.2 CF and BW predictions from FM sweep responses

In Fig. 3 we show predictions of CF obtained using the average of the two trigger frequencies versus the CF values obtained from pure tone tuning curves and STRFs. It is clear that CF is well predicted from the FM sweep responses. Similar plots for tuning curve bandwidth resulted in very scattered, non-significant correlations, as might be expected from contributions by tuning asymmetry, inhibitory sidebands, and neural nonlinearities.

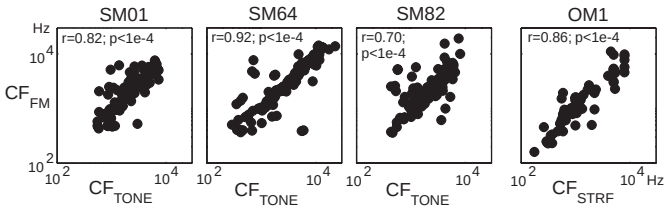


Fig. 3. CF predictions from FM sweep responses versus CFs from tuning curves or STRFs.

3.3 Population responses

In Fig. 4 we show summary data for squirrel and owl monkeys. The top row shows the directional selectivity index (DSI) for each penetration site in the monkeys. The DSI is bound between -1 and 1, where positive DSI values indicate sites that showed a preference for upward FM sweeps, and negative values indicate the converse. The striking result is that the median of the indices is located at 0 DSI and that there appears to be little difference between the anesthetized and awake preparations.

In the bottom row we show the tallied centroid values, thus obtaining a measure of best speed for each site. We see that the peak in each histogram is centered between 20-40 oct/s, which corresponds to actual FM sweep speeds in natural monkey vocalizations.

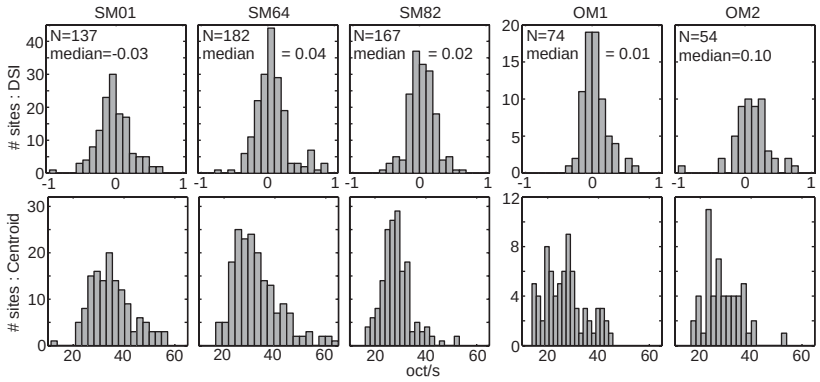


Fig. 4. Population distributions for directional selectivity indices and centroid (best speed - oct/s) measures (5 animals).

3.4 Topographic maps

We surmised that there may be spatial clustering of cells in AI with similar FM sweep preferences. The top row of Fig. 5 shows the spatial distribution of FM sweep DSI values. Each column represents one squirrel monkey.

A clear clustering of upward selective neurons is present in the more ventral, low frequency areas of AI. Each site in the cluster represents directional selectivity indices greater than 0.15, implying an upward sweep selective sites. Regression analysis (lower panels in Fig. 5) confirm the inverse correlation between CF and DSI for all three monkeys. Thus it appears that there is an orderly spatial representation of DSI in the cortex of squirrel monkeys.

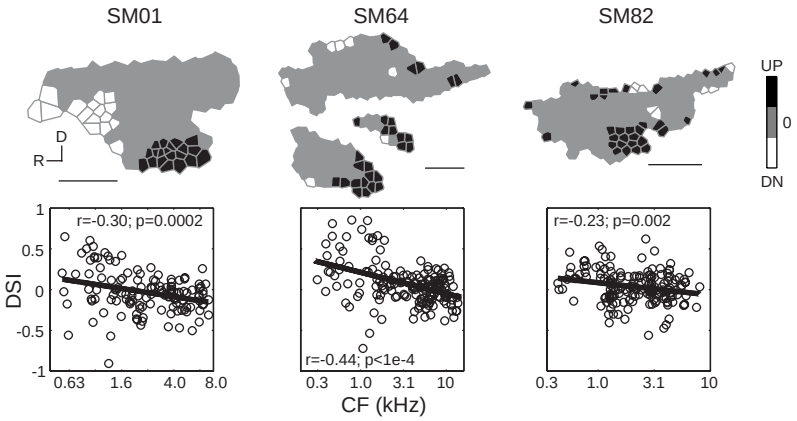


Fig. 5. Spatial maps of DSI values (top row) and DSI versus CF plots (bottom row) for each squirrel monkey. DSI values are classified for FM sweep direction preferences as upward (UP), downward (DN), or non-selective (0). Scale bar = 1mm. D = dorsal, R = rostral.

3.5 FM sweep predictions from STRFs

For one owl monkey we computed STRFs. From these STRFs we derived estimates of FM sweep directional selectivity and best speed values, following a method outlined by DeAngelis and colleagues (DeAngelis et al. 1999). Briefly, we compute the magnitude of the 2D FFT of the STRF, termed the ripple transfer function (Depireux et al. 2001), transforming the STRF to a representation where the axes have units of cycles/octave and cycles/second (spectral and temporal modulation, respectively). By selecting the greatest values in the 1st and 2nd quadrants, V_1 and V_2, of the transfer function, we estimate directional selectivity via $DSI = (V2-V1)/(V2+V1)$. Also, finding the peak in the transfer function, and locating the corresponding maximum spectral modulation (SM_{max}) and temporal modulation values (TM_{max}), we estimate best sweep speed via $BS = TM_{max}/SM_{max}$ (oct/sec). These calculations are valid if the STRF is an adequate descriptor of the neuron under study and if the neuron behaves in a linear manner.

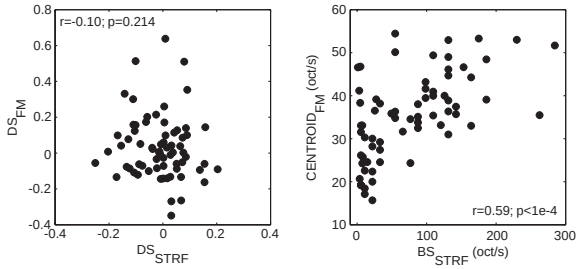


Fig. 6. Left: DSI values from FM sweeps versus those from STRFs. Right: Centroid best speed values from FM sweeps versus best speed calculations from STRFs.

Figure 6 shows these results, where it is evident that the we cannot adequately predict directional selectivity from STRFs. However, we do see a high correlation between best speed of FM sweeps and STRFs.

4 Discussion

Our results reveal several intriguing findings. First, we see that in most cases neurons are not highly selective for speed and direction, as shown by similar area-best speed plots. This is borne out by the DSI population distributions which are centered about zero. Second, we do not see a strong difference between the anesthetized and awake preparations, suggesting that it is adequate to draw conclusions from anesthetized monkey studies for stimulus conditions that result in infrequent and phasic responses. Third, a clear frequency dependent spatial distribution of DSI values is seen in the squirrel monkey, identifying another possible organizing parameter in AI. Finally, we found that our STRF measures do

not accurately predict FM direction selectivity due to either response nonlinearities or the inadequacy of our STRF methods. However, STRF predicted preferred sweep speeds correlated significantly with the actual best FM speeds.

References

Blake, D. T. and M. M. Merzenich (2002). "Changes of AI receptive fields with sound density." J Neurophysiol 88(6): 3409-20.

Cheung, S. W., P. H. Bedenbaugh, S. S. Nagarajan and C. E. Schreiner (2001). "Functional organization of squirrel monkey primary auditory cortex: responses to pure tones." J Neurophysiol 85(4): 1732-49.

DeAngelis, G. C., G. M. Ghose, I. Ohzawa and R. D. Freeman (1999). "Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons." J Neurosci 19(10): 4046-64.

deCharms, R. C., D. T. Blake and M. M. Merzenich (1998). "Optimizing sound features for cortical neurons." Science 280(5368): 1439-43.

deCharms, R. C., D. T. Blake and M. M. Merzenich (1999). "A multielectrode implant device for the cerebral cortex." J Neurosci Methods 93(1): 27-35.

Depireux, D. A., J. Z. Simon, D. J. Klein and S. A. Shamma (2001). "Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex." J Neurophysiol 85(3): 1220-34.

Heil, P. and D. R. Irvine (1998). "Functional specialization in auditory cortex: responses to frequency-modulated stimuli in the cat's posterior auditory field." J Neurophysiol 79(6): 3041-59.

Jurgens, U. (1986). "The squirrel monkey as an experimental model in the study of cerebral organization of emotional vocal utterances." Eur Arch Psychiatry Neurol Sci 236(1): 40-3.

Jurgens, U. (1998). "Neuronal control of mammalian vocalization, with special reference to the squirrel monkey." Naturwissenschaften 85(8): 376-88.

Nelken, I. and H. Versnel (2000). "Responses to linear and logarithmic frequency-modulated sweeps in ferret primary auditory cortex." Eur J Neurosci 12(2): 549-62.

Shamma, S. A., J. W. Fleshman, P. R. Wiser and H. Versnel (1993). "Organization of response areas in ferret primary auditory cortex." J Neurophysiol 69(2): 367-83.

Wang, X., M. M. Merzenich, R. Beitel and C. E. Schreiner (1995). "Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics." J Neurophysiol 74(6): 2685-706.

Winter, P., D. Ploog and J. Latta (1966). "Vocal repertoire of the squirrel monkey (Saimiri sciureus), its analysis and significance." Exp Brain Res 1(4): 359-84.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Frequency change velocity detector: A bird or a red herring?

Pierre L. Divenyi

Veterans Affairs Northern California Health Care System and East Bay Institute for Research and Education, Martinez CA, USA, pdivenyi@ebire.org

1 Introduction

Our acoustic environment is dynamic: although at different rates, every feature of the sounds we hear continuously changes. The importance of determining sensitivity to frequency modulation (FM) was recognized several decades ago (see e.g., Shower and Biddulph, 1931) because FM perception was regarded as a convenient window that opens into the area of dynamic pitch processing. Interestingly however, the auditory mechanisms responsible for the analysis of frequency modulation have remained elusive despite several notable attempts to model them. Most theories proposed to date, such as those by Hartmann (Hartmann and Klein, 1980) and by de Cheveigné (2000), as well as earlier models summarized in their articles, are based on peripheral frequency analysis supplemented by some particular time-based analysis by an autocorrelation mechanism. Although, on the whole, the models are able to account, at least qualitatively, for a number of features of FM perception, on the one hand, they are conceptually quite similar to one another, while, on the other hand, they all rest on a number of assumptions which, as an ensemble, make them less realistic. One particular possibility that has not received significant attention, except by a few investigators (Pollack, 1968; Kluender and Jenison, 1992), is that the auditory system may characterize the spectrum of the signals it receives not only in the form of a neural spectrogram but also as successive time derivatives thereof — especially the first two of these, i.e., the velocity and acceleration of frequency changes. Such a representation may be particularly useful for a system required to analyze unidirectional frequency changes abundant in speech as well as in environmental sounds.

The present study is an attempt to measure sensitivity to the dynamics of pure tone glides, in order to determine whether the perception of such signals can be explained by sensitivity to duration and/or frequency change (Pollack, 1968; van Wieringen and Pols, 1995) or, alternatively, whether one can perceive such changes in the absence of frequency and duration cues — i.e., whether it is necessary to postulate the existence of a velocity and/or an acceleration detector.

2 Experiment 1: Discrimination of sinusoidal glide velocity

The first experiment was designed to estimate d'=1.0 thresholds for the discrimination of two tone glides of different rates of frequency change. The stimuli in the two observation intervals were designed such that the only valid cue for their discrimination was the velocity of frequency change. In other words, neither the starting or ending frequencies of the glides nor their frequency range, nor their duration was correlated with the velocity of frequency change. Five young normal-hearing listeners with at least two months of experience in psychoacoustic tasks served as subjects; one of them (indicated as "best subject" in the figures) had 18 months of training. The thresholds were estimated from responses obtained in a multiple-level fixed 2AFC paradigm in which the velocity of a "standard" glide was compared with one of four "variable" velocities, as illustrated in Fig. 1. Threshold estimates were based on psychometric functions obtained from data of at least four blocks of trials in each condition. In each condition, both glides were either ascending or descending; the range of their starting frequencies extended from 400 to 1300 Hz. The frequency change encompassed a range of 1.25 to 10 basilar membrane (BM) mm/s, using Greenwood's (1961) classic formula.

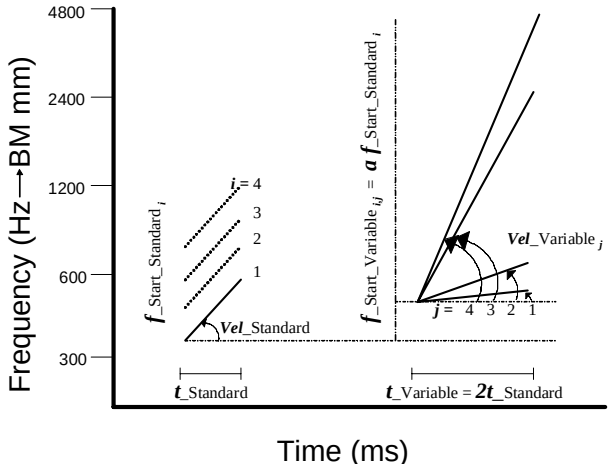


Fig. 1. Schematic time-frequency diagram of the stimulus used in experiment 1 in which a "standard" and a "variable" pure tone glide had to be discriminated. The standard's frequency change velocity (in basilar-membrane distance per s) and duration were constant but it could take up four different starting (and ending) frequencies. Each of the four standard glides could be paired with four different variable glides having a duration twice that of the standard, and starting frequencies *a* times that of the standard. Two of the variables had velocities lower and two higher than the standard.

The first question addressed in this experiment was whether discriminability of the rate of frequency change was dependent on the base velocity of the glide. Fig. 2 illustrates the results for standard glide durations of 25, 50, and 100 ms. The first observation one has to make is that the just noticeable velocity differences (jnd's)

were rather large: between about 50 and 75 percent on the average for the five subjects combined and between 20 and 60 percent for the best listener. Nevertheless, these glide velocity jnd's, obtained with randomized starting glide frequencies, are not overly large in view of the 20 to about 85 percent jnd's obtained for frequency modulation depth of randomized sinusoidal carriers (Plack and Carlyon, 1994). The second observation is that the velocity jnd's appear to be independent of the base glide velocity, with a weak trend of inverse proportionality at standard durations below 100 ms, and mainly for the best listener. Across glide durations, however, the velocity thresholds change only in minor ways.

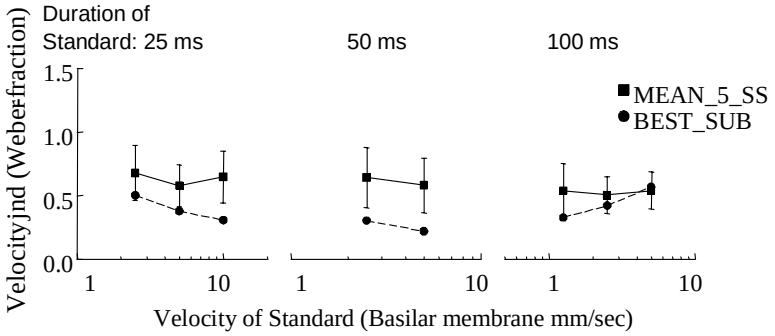


Fig. 2. Results of Experiment 1: estimated just-noticeable frequency change velocity difference (expressed as Weber fractions) as a function of standard glide velocity, shown at three standard glide durations. Results of the average of five listeners and of the best listener.

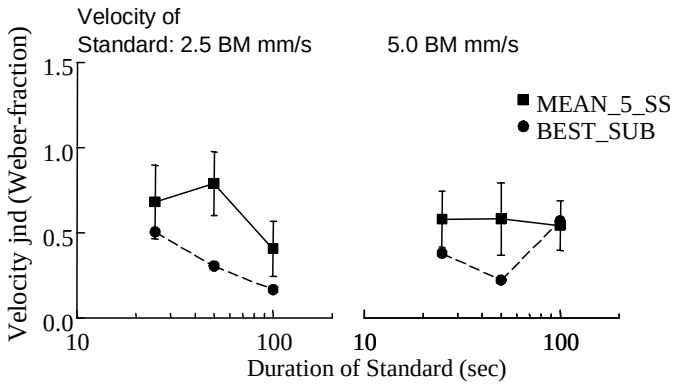


Fig. 3. Results of Experiment 1: estimated just-noticeable frequency change velocity difference as a function of standard glide duration shown at two standard glide velocities.

Duration dependence of velocity discrimination can be better seen in Fig. 3 that illustrates glide jnd's as a function of standard glide duration for two standard glide velocities: 2.5 and 5.0 BM mm/s. For the slower standard, the jnd's appear to be inversely related to glide duration, suggesting that a longer observation period helps velocity discrimination as long as the rate of change is low.

The effect of glide direction is shown in Fig. 4 in which glide jnd is represented as a function of standard glide velocity for 50- and 100-ms standard glide durations. It appears that the jnd of downward glide velocity tends to be less than that of upward glide velocity. This finding comes somewhat as a surprise in light of data demonstrating the superiority of peak frequency for the discrimination of FM frequency excursion (Demany and Clement, 1997).

In summary, Experiment 1 demonstrated that discrimination of sinusoidal glides is possible when neither duration nor frequency excursion can be reliably used, i.e., when the only remaining cue is the slope — the velocity — of the glide. Under these circumstances, glide velocity jnd's are by-and-large constant over the range of velocities and durations investigated, with a weak inverse relation between the Weber fraction for velocity and glide duration.

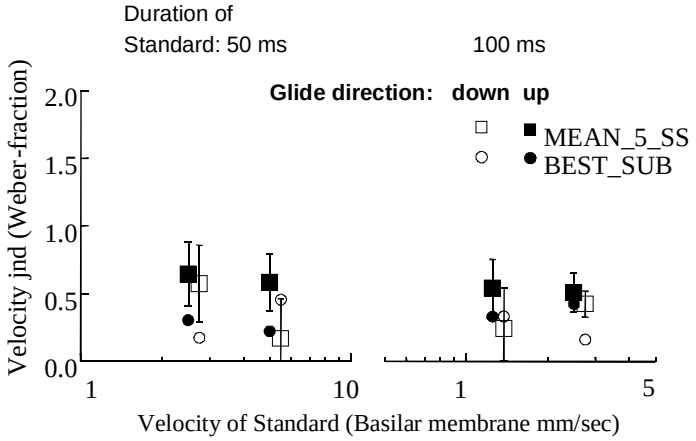


Fig. 4. Results of Experiment 1: effect of glide direction on estimated just-noticeable frequency change velocity difference, shown as a function of standard glide velocity at two standard glide durations.

3 Experiment 2: Discrimination of sinusoidal glide acceleration

Glides going up or down at a constant rate are rare in our acoustic environment. More often, frequency modulation is, itself, modulated: the rate of change increases and decreases in time. Can listeners discriminate changes in the rate of frequency change in sinusoidal carriers? To fully answer this question, a series of conditions with a range of glide velocities, durations, and velocity changes should be examined. Experiment 2 addressed only the very basic question of whether positive or negative acceleration of frequency change can be discriminated at all.

The stimulus, schematically illustrated in the diagram of Fig. 5, had the listener compare in a multiple-level 2AFC paradigm a standard consisting of a steady-state tone followed by a 1-BM mm/s glide (positive acceleration) or a 1-BM mm/s glide followed by a steady-state tone (negative acceleration) with a variable consisting of

a two-segment glide with the first segment having a velocity either lower than the second (positive acceleration) or higher than the second (negative acceleration). The velocity change always occurred in the temporal middle of the stimulus. The duration of the standard was 200 or 400 ms, whereas that of the variable was opposite: 400 or 200 ms. The particular standard glide pattern as well as the starting frequency of each of the two standard glide patterns, 400 or 700 Hz, were randomly chosen at each trial. Also randomized at each trial were the variable glide pattern (the second segment of the glide having a velocity different from that of the first by ±0.5 or ±1.0 BM mm/s) as well as the variable starting frequency (two for each variable glide pattern). The same five listeners served as subjects.

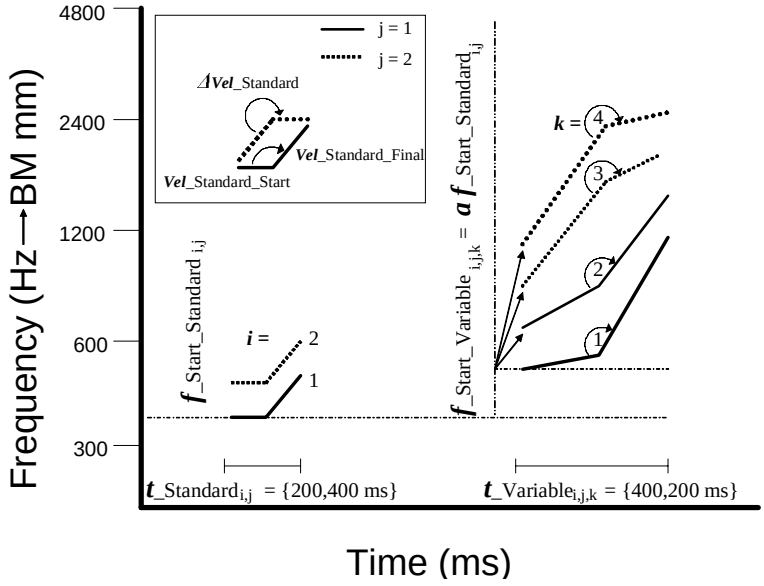


Fig. 5. Schematic time-frequency diagram of the stimulus used in Experiment 2. The change in the velocity of the standard glide had to be compared with four different changes in the velocity of the variable glide.

Results are displayed in Fig. 6 as the jnd for velocity change as a function of the duration of the standard, separately for the condition in which the velocity change in the standard was positive, and the one in which it was negative. The most apparent finding is that positive and negative frequency acceleration patterns are more discriminable from each other than a positive from another positive and a negative from another negative. This being said, velocity changes of the same sign are also discriminable but not very well (Weber fractions between 0.9 and 1.9). Duration, at least in the inordinately long range investigated, does not seem to have an effect.

In summary, whereas different directions of velocity change appear to be discriminable at about the same level (Weber fractions between 0.2 and 0.5) as velocity differences in Experiment 1, telling whether a glide accelerated more or

less than another glide having the same direction of velocity change is very difficult.

4 Discussion and conclusions

The results reported above are clearly pilot in nature. In essence, they are in agreement with Pollack's classic study (1968) — velocity of sinusoidal glides can be discriminated — but they failed to confirm the two-tiered, all-or-none dependence of velocity discrimination found by Pollack (i.e., total reliance on either frequency excursion cues or duration cues). While it is possible that the failure of replicating Pollack's neatly aligned discrimination functions could be due to the narrow parametric range investigated in the present study, it is equally likely that our randomization of both the frequency range and the frequency excursions, together with the 2:1 ratio of the standard and variable durations in each trial, made the duration and frequency excursion cues practically useless. If, indeed, those cues were not available to the listeners, one has to assume that they based their responses on comparing glide velocities, i.e., different rates of frequency change.

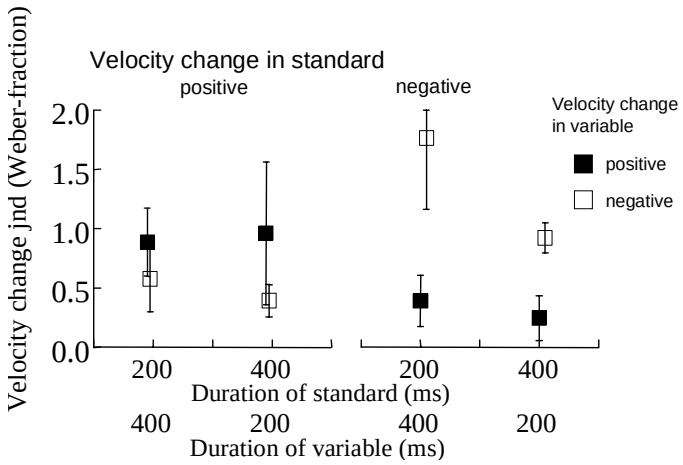


Fig. 6. Results of Experiment 2: Discrimination of velocity changes in sinusoidal FM signals displayed as Weber fraction of the velocity change, for pairs of standard and variable of different durations and directions of velocity change. Averaged results of five subjects.

But what mechanism does the auditory system use to track frequency changes in the 25- to 200-ms range investigated? At present, no known model deals explicitly with the phenomenon reported in our two experiments, although one can imagine that, with a limited number of added assumptions and/or processing stage, some previously proposed sensory-based models — such as the by-now classic sampling-and-correlation model by Hartmann (Hartmann and Klein, 1980) or the elegant delayed-correlation model by de Cheveigné (2000) — could account for the results. An alternative model, unexplored by hearing science but adopted by automatic

speech recognition technology (Aikawa, Singer, Kawahara, and Tohkura, 1996) postulates the existence of a frequency change detector having a time window appropriate for tracking formant transitions in speech, i.e., of a duration exceeding auditory peripheral time constants but approaching cortical integration time (Schreiner and Urbas, 1988). The advantages offered by such a detector for speech processing, especially in a "cocktail-party" situation, would be significant.

Acknowledgments

The author thanks his colleague Brian Gygi for many protracted and useful discussions on the experiments reported, and Alex Brandmeyer for his technical help. The research has been supported by a grant from National Institute on Aging and by the Veterans Affairs Medical Research.

References

Aikawa, K., Singer, H., Kawahara, H., and Tohkura, Y. (1996) Cepstral representation of speech motivated by time-frequency masking: an application to speech recognition. J. Acoust. Soc. Am., 100(1), 603-614.

de Cheveigné, A. (2000) A model of the perceptual asymmetry between peaks and troughs of frequency modulation. J. Acoust. Soc. Am., 107(5 Pt 1), 2645-2656.

Demany, L., and Clement, S. (1997) The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform. J. Acoust. Soc. Am., 102(5 Pt 1), 2935-2944.

Greenwood, D. D. (1961) Auditory masking and the combination band. J. Acoust. Soc. Am., 33, 484-502.

Hartmann, W. M., and Klein, M. A. (1980) Theory of frequency modulation detection for low modulation frequencies. J Acoust Soc Am, 67(3), 935-946.

Kluender, K. R., and Jenison, R. L. (1992) Effects of glide slope, noise intensity, and noise duration on the extrapolation of FM glides through noise. Percept. Psychophys, 51(3), 231-238.

Plack, C. J., and Carlyon, R. P. (1994) The detection of differences in the depth of frequency modulation. J. Acoust. Soc. Am., 96(1), 115-125.

Pollack, I. (1968) Detection of rate of change of auditory frequency. J. Exp. Psychol. 77(4), 535-541.

Schreiner, C. E., and Urbas, J. V. (1988) Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hear. Res., 32(1), 49-63.

Shower, E. G., and Biddulph, R. (1931) Differential pitch sensitivity of the ear. J. Acoust. Soc. Am. 3(2), 275-287.

van Wieringen, A., and Pols, L. C. W. (1995) Discrimination of single and complex consonant–vowel- and vowel–consonant-like formant transitions. J. Acoust. Soc. Am. 98(3), 1304-1312.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Coding of FM and the continuity illusion

Robert P. Carlyon[1], Christophe Micheyl[1,2], and John Deeks[1]

[1] MRC Cognition & Brain Sciences Unit, {bob.carlyon,john.deeks}@mrc-cbu.cam.ac.uk
[2] MIT RLE cmicheyl@mit.edu

1 Introduction

Here we investigate the encoding of frequency modulation (FM) by the auditory system. A simple class of explanation, which can account for the detection of FM in many situations, is that the central auditory system compares successive samples of the instantaneous frequency of a modulated tone (e.g. Demany and Semal, 1989). If snapshots taken at different times reveal different frequencies, the listener reports the presence of FM, in much the same way as one would detect a difference in frequency between two steady tones. However, there is also evidence for a different, or additional form of processing, which may reflect neural units, or "feature detectors", that specialise in the processing of *dynamic* changes in frequency. Physiological evidence for such units was provided by Rees and Moller (1983), who found neurons in the rat inferior colliculus (IC) whose firing rates were tuned to the frequency of FM.

Recently, Cusack and Carlyon (in press) have provided psychophysical evidence for a "perceptual asymmetry", roughly analogous to that used as evidence for feature detectors in vision. They required subjects to detect a single FM tone in a mixture of up to 32 steady tones, distributed quasi-randomly in frequency and time. Performance was considerably better than in the converse case, where a steady tone had to be detected from a mixture of FM tones. It has also been argued (Carlyon, 2000) that the phenomenon of FM Detection Interference (FMDI) (Wilson *et al.*, 1990; Moore *et al.*, 1991; Carlyon, 1994) reflects a similar perceptual asymmetry. FMDI occurs when subjects have to identify which of two sequentially presented tones is modulated, and in which a second tone, having a very different carrier frequency (fc), is presented simultaneously with both of them. The asymmetry arises because the interfering tone impairs performance when it is modulated but not when it is steady. Finally, Plomp (1982) made an informal observation of particular relevance to the present article: When a portion of an FM tone is replaced by a burst of noise, subjects not only hear the tone continue through the noise (the "continuity illusion"), but also heard the *modulation* continue. This is consistent with a feature detector signalling the presence of FM, and continuing to do so throughout the noise burst. The experiments described here explore Plomp's observation more

formally, and investigate whether the phase as well as the presence of FM is preserved during the illusion.

There are some reasons to suspect that the presence and rate of FM can be encoded in a way that is not sensitive to FM phase. Physiologically, the fact that the IC neurons described by Rees and Moller encoded changes in FM rate by changes in firing rate, rather than by some temporal aspect of their response, makes it unlikely that they would encode FM phase. Psychophysically, subjects cannot discriminate between in-phase and out-of-phase FM applied to two inharmonically related carriers, provided that cues such as beats and auditory distortion products are controlled for (Carlyon, 1991; Carlyon, 1994); under such circumstances, FMDI is also insensitive to relative FM phase. Note, though, that the psychophysical evidence comes from experiments in which subjects made comparisons between two simultaneous tones. It is possible that FM phase is encoded when a single tone is present, but that somehow this information is lost when making an across-frequency comparison. In contrast, the experiments described here use the continuity illusion to investigate whether (and how) FM phase is encoded within a single tone. In doing so, they rely not only on subjective reports, but also on objective measures of performance using a forced-choice task.

2 Experiment 1

Experiment 1 measured the extent to which the percept of FM is preserved during the continuity illusion. Although Plomp informally reported that the modulation percept persisted, it is not known whether the perceived depth of the FM was the same as if the modulated tone had been physically continuous.

The first stimulus presented in each trial consisted of two 200-ms 1-kHz sinusoids frequency modulated at a depth of ± 10% and at a rate of 5 Hz. The tones were turned on and off with 5-ms raised-cosine ramps and separated by a 200-ms gap (zero-voltage points). This gap was filled by a digitally-generated 800-1200-Hz 220-ms bandpass noise that was turned on 5 ms before the start of the first tone's offset ramp, and ended 5 ms after the second tone had reached full amplitude. The level of the tone was 60 dB SPL and the spectrum level of the noise was 52 dB SPL. The second stimulus was presented 500 ms later and consisted of a 600-ms 5-Hz-FM tone, whose modulation depth during the first and last 200 ms was also 10%. During the middle 200 ms the FM depth was, on separate trials, either 4, 6, 8, 10, 12, 14, 16, or 20%.

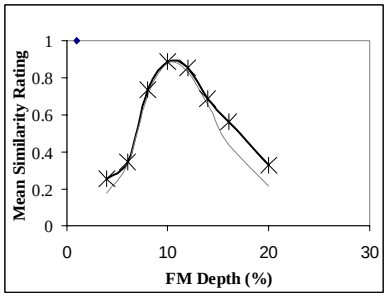


Fig. 1. Solid line: Results of experiment 1. Faint line: Supplementary experiment described in section 6

Four normally hearing subjects were instructed to ignore the noise during the first stimulus and to focus on the middle portion of the tone. They then adjusted a

horizontal slider to indicate how similar this sounded to the second stimulus. These readings were digitised, and normalised to the maximum value used by each subject. The mean data are plotted by the heavy solid curve and crosses in Fig. 1, as a function of the FM depth in the central portion of the second sound. The function peaked when this FM depth was 10%, indicating that during the noise burst subjects heard the modulation continue to the same extent as was physically present before and after the noise.

3 Experiment 2

Experiment 1 showed that subjects do indeed hear the modulation continue unabated during the continuity illusion. We reasoned that if the phase of the FM is also preserved, then subjects should be able to tell whether the tone burst presented after the noise continues in the phase that it "would have" had if it had been physically continuous. This prediction rests on the assumption that subjects can detect a *physical* FM reversal. Experiment 2 tested this assumption.

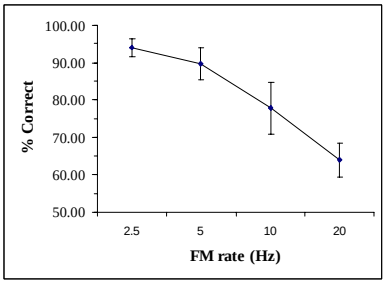


Fig. 2. Results of expt. 2

Each interval of every 2IFC trial contained an 800-ms 1-kHz pure tone, with an FM depth of ± 10%. The FM phase reversed after 400 ms in the signal but not in the standard interval (see Fig. 3a). The initial modulator phase was randomly selected from 0° and 180° in each interval. The FM rate for each block of 25 trials was fixed at either 2.5, 5, 10, 15, or 20 Hz, and blocks were run in a counterbalanced order until 200 trials had occurred for each condition. Eight normally hearing subjects took part

The percentage of trials on which subjects performed correctly is shown as a function of modulator rate (fm) in Fig. 2. Average performance was at least 90% for fm=2.5 and 5 Hz, but dropped markedly at the higher rates. Therefore, a modulator rate of 5 Hz was selected for subsequent experiments.

4 Experiment 3

4.1 Method

The main condition of experiment 3 is shown in Fig. 3b. Subjects heard two 1-kHz tones, modulated at a rate of 5 Hz and a depth of ± 10%. The tones were separated by a 200-ms gap that was filled by the same noise as in experiment 1. The nominal duration of each tone was 400 ms (see later). In the standard interval of each 2IFC trial the modulation of the second tone started in "preserved phase" –i.e. the phase that the first tone would have had at that point if it had physically continued throughout the noise. In the signal interval it started 180° out of phase. Experiments

1 and 2 showed that the percept of FM is maintained throughout the continuity illusion, and that subjects can detect a physical reversal of FM phase. If modulator phase information is also preserved during the illusion then subjects should be able to perform this discrimination.

The "gap" condition (Fig. 3c) was identical except that the noise was absent. One might expect chance performance in this condition, but inspection of Fig 3c reveals an alternative cue; the two tone bursts in each interval start in the same phase during the standard interval, but in opposite phases during the signal interval. To prevent this, the durations of the first and second burst in each interval were increased at random (and independently of each other) by 0, 100, or 200 ms from presentation to presentation (see example in Fig. 3d). This was done in all conditions. The "continuous" condition (Fig. 3a) was otherwise identical to the 5-Hz physical-reversal condition of experiment 2. Finally, the "steady" condition (Fig. 3e) was like the gap condition, except that the silent gap was filled by an unmodulated tone. Five normally hearing subjects ran counterbalanced blocks of 25 trials per condition, until each had completed 200 trials/condition.

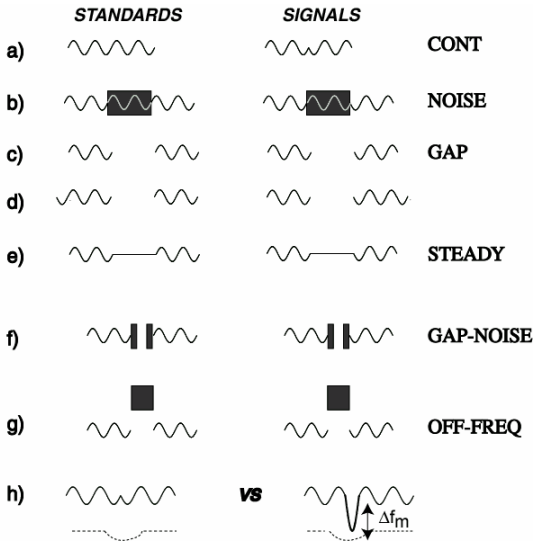


Fig. 3. Schematic spectrograms of trials for various conditions of experiments 2, 3, and 4. The FM tone was absent during all noise bursts, but is indicated in white in the NOISE condition (part b) to illustrate the illusory phase reversal in the signal interval.

4.2 Results

The percentage of trials correct, averaged across subjects, is plotted for each condition in Fig. 4. The "continuous" condition is essentially a replication of the 5-Hz condition of experiment 2, and the results confirm that subjects can detect a physical reversal in FM phase. In contrast, performance

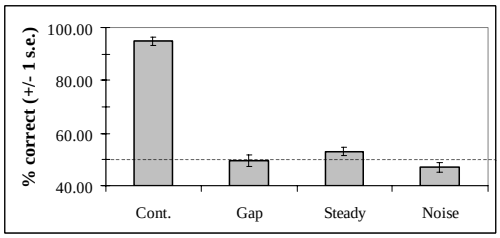


Fig. 4. Results of expt 3.

is at chance in the noise condition. Hence, although they perceive the modulation continue unabated through the noise (experiment 1), subjects cannot detect a phase reversal that would have been easily detectable had the FM tone really been continuous. Unsurprisingly, subjects could not detect whether or not a modulation continued in preserved phase after a 200-ms silent gap. They could not do so either when that gap was filled with a steady tone. This is consistent with subjects' reports that they did not hear the modulation continue during that steady portion.

5 Experiment 4

5.1 Method

Experiment 4 addresses the same issue as experiment 3 but using a paradigm in which the continuity illusion is predicted to *impair* performance unless information on FM phase is retained. Four of the conditions – noise, gap, continuous and steady (Figs. 3a,b,c,e) – were the same as in experiment 3, but with one important change. This was that the randomization of tone duration was removed. We reasoned that this would allow subjects to perform the task in the gap condition (Fig. 3c) by comparing the starting phases of the two FM bursts in each interval (see section 4.1). We also reasoned that if the FM is perceived as continuous then subjects should not be able to use this cue, as they would hear only one (long) tone in each interval. Hence, performance in the noise condition should be at chance unless information on FM phase is explicitly preserved throughout the illusion.

Because we predicted that the noise should impair performance, by virtue of inducing the continuity illusion, two further conditions controlled for other ways in which such an impairment might occur. To control for general distracting effects, the "off-frequency" condition replaced the noise with one filtered between 2000-2400Hz (Fig. 3g). This noise had the same level as the original and should still distract the subject, but not induce the continuity illusion. Second, because performance in the gap condition was expected to be dependent on a comparison of the beginning the two tone bursts in each interval, we wished to control for any effect of the noise masking these onsets. To do so, we replaced the central 100 ms of the noise with silence (Fig. 3f); hence the end of the first burst and start of the second burst in each interval were masked, but the silent gap would be expected to eliminate the illusion. Subjects and procedure were as in experiment 3

5.2 Results

The percentage of trials correct, averaged across subjects, is plotted for each condition in Fig. 5. Unlike the results of experiment 3, performance is very good (91%) in the "gap" condition, because subjects could now compare the FM phase at the start of the two bursts in each interval. In contrast, when the continuity illusion was induced in the noise condition, performance was close to chance. Performance was substantially better when we presented an off-frequency (82%) or gapped (80%) noise. Interestingly, performance was also close to chance in the steady condition. Although subjects could hear the modulation stop and start in this condition (cf.

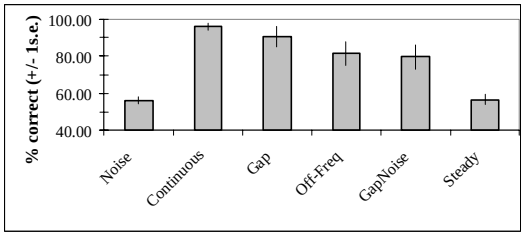


Fig. 5. Results of experiment 4

experiment 3), they apparently could not compare the FM phase at the start of the tone to that when the FM was re-introduced after the steady portion (compare Figs. 3c, 3e). The fact that they could do so when the two FM portions were separated by a silent gap is consistent with them being able to make phase comparisons only between the start of two separate sounds, and not between two portions of the same sound.

The trends described above were evaluated statistically by performing a one-way ANOVA, which showed a highly significant effect of condition ($F(5,20)=25.7$, $p<0.001$). Uncorrected pairwise comparisons revealed that performance in the noise condition did not differ significantly from that in the steady condition, but was significantly worse than in the off-frequency ($p<0.01$), gapnoise ($p<0.05$), gapped, ($p<0.005$) and continuous ($p<0.001$) conditions.

6 Discussion

Our results suggest that, at least for the modulation rates and depths studied, information on the phase of periodic FM is not preserved during the continuity illusion. Specifically, they show that a) subjects hear the modulation continuing through the noise at the same FM depth as it would have had it been uninterrupted, b) they can detect physical reversals of FM phase, but c) they cannot detect an "illusory" phase reversal. Several aspects of these findings merit further comment.

First, the present results reflect a paradoxical percept, in that subjects perceive a modulation as continuous, but do not notice an FM phase change that would have been obvious had the modulation really been uninterrupted. Another paradoxical percept has been obtained by replacing the band of noise with a 73-dB-SPL 1000-Hz tone frequency modulated by ± 10% at 13.6 Hz. Under these circumstances one hears a tone modulated at a rate 5 Hz continue behind another tone modulated at 13.6 Hz. As both tones have the same carrier frequency there is no physical stimulus that would give rise to the percept of these two modulation rates occurring concurrently. Nevertheless, the 5-Hz tone is heard to continue to be modulated at its full depth of 10% - as shown by the faint line in Fig. 1.

Second, combined with evidence for listeners' insensitivity to differences in the phase of FM applied to different carriers of a complex tone (Carlyon, 1991; Moore *et al.*, 1991; Carlyon, 1994), the data point to a lack of explicit encoding of FM phase by the auditory system. It is important to restrict this conclusion to the periodic form of FM used here. Clearly, subjects can discriminate between a linear frequency glide that is falling and one that is rising. Similarly, one would expect that, when a tone is frequency modulated at a very slow rate and over a wide range, listeners could consciously follow the up-and-down pattern of that modulation. However, it appears that there is a range of modulation rates and depths at which

subjects can hear the *presence* of modulation, but where the direction in which frequency is changing is not encoded on a moment-by-moment basis.

One way in which FM could be encoded is suggested by the results of experiment 2, where the detection of an FM phase reversal deteriorated as modulation rate increased (Fig. 2). This is consistent with instantaneous frequency being averaged over some "sliding pitch integrator" having a finite duration (dotted traces under left panel of Fig. 3h). As that window passes over the phase reversal, its output will rise or fall (depending on the phase at which the reversal occurs). Furthermore, the size of this change will decrease as modulation rate increases, because several complete cycles of FM will fall within the window. This could account for the observed reduction in performance with increasing rate. It could also account for the ability to hear the direction of linear glides, for which the window output will change monotonically, and for slow rates and wide extents of periodic FM, where the window should follow the modulations. If so, then subjects should find it difficult to discriminate between a stimulus containing a phase reversal and one in which a half-cycle of the modulation has been increased in depth (Fig. 3h). We have recently presented evidence for this (Carlyon *et al.*, 2002), and plan to explore the parameters of the proposed window in a future publication.

References

Carlyon, R. P. (1991). Discriminating between coherent and incoherent frequency modulation of complex tones. J. Acoust. Soc. Am. 89, 329-340.

Carlyon, R. P. (1994). Further evidence against an across-frequency mechanism specific to the detection of FM incoherence between resolved frequency components. J. Acoust. Soc. Am. 95, 949-961.

Carlyon, R. P. (2000). Detecting coherent and incoherent frequency modulation. Hearing Research 140, 173-188.

Carlyon, R. P., Micheyl, C., Deeks, J. M. and Moore, B. C. J. (2002). FM phase and the continuity illusion. J. Acoust. Soc. Am. 111, 2468.

Cusack, R. and Carlyon, R. P. (in press). Auditory pop-out: Perceptual asymmetries in sequences of sounds. Journal of Experimental Psychology: Human Perception and Performance

Demany, L. and Semal, C. (1989). Detection thresholds for sinusoidal frequency modulation. J. Acoust. Soc. Am. 85, 1295-1301.

Moore, B. C. J., Glasberg, B. R., Gaunt, T. and Child, T. (1991). Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals. Q. J. Exp. Psychol. 43A, 327-348.

Plomp, R. (1982). Continuity effects in the perception of sounds, Psychoacoustics of music; Jablonna, Poland.

Rees, A. and Moller, A. R. (1983). Responses of neurons in the inferior colliculus of the rat to AM and FM tones. Hearing Research 10, 301-310.

Wilson, A. S., Hall, J. W. and Grose, J. H. (1990). Detection of frequency modulation (FM) in the presence of a second FM tone. J. Acoust. Soc. Amer. 88, 1333-1338.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

The role of spectral change detectors in sequential grouping of tones

Makio Kashino[1] and Minae Okada[1, 2]

[1] NTT Communication Science Laboratories, NTT Corporation, kashino@avg.brl.ntt.co.jp
[2] Graduate School of Integrated Arts and Social Sciences, Japan Women's University, mokada@st.jwu.ac.jp

1 Introduction

The perceptual organization of sequential spectral components is a crucial feature of auditory scene analysis. The sequential organization can be demonstrated by presenting a sequence of high- and low-frequency tones (H and L) in an alternating pattern (HLHL...). When the tone presentation rate is slow or the frequency separation (Δf) between the tones is small, listeners tend to perceive the sequence as a connected series of tones ordered in time (called *temporal coherence*). When the tone presentation rate is fast or the Δf is large, however, the alternating sequence perceptually splits into two parallel unrelated sequences, or *streams*, one high and one low in pitch (called *fission* or *stream segregation*) (Bregman and Campbell 1971; van Noorden 1975). The neural mechanism underlying this perceptual effect has until now been unclear.

Here we propose that the sequential grouping of spectral components with different frequencies is mediated by neural units that are sensitive to the direction of spectral changes. Neurons responding selectively to upward or downward frequency changes have been found at various neural sites from the cochlear nucleus to the auditory cortex (For a review, see Palmer 1995). We hypothesize that the activation of such neural units, or *spectral change detectors*, links up separate spectral components into a coherent stream. When the presentation rate of successive components is too fast or the Δf is too large to activate spectral change detectors, the components are not linked, resulting in stream segregation.

In the present study, we evaluated the hypothesis using the selective adaptation paradigm, which has been used to examine the selective processing of spectral change direction (Gardner and Wilson 1979; Kayahara 1998; Shu, Swindale and Cynader 1993). If the judgment of spectral change directions is affected by adaptation to an alternating sequence of high and low frequency tones, it would indicate that the same population of neural units is involved in both the sequential grouping of tones and the detection of spectral change direction.

2 Methods

The test sound was a 100-ms linear frequency glide (including 10-ms raised-cosine onset and offset ramps) centered at 1000 Hz (Fig. 1). The adaptor was an alternating sequence of high- and low-frequency tones (H and L) that were repeated 20 times. The center frequency of H and L tones was fixed at 1000 Hz, and the Δf was either 1/6, 1/3, or two octaves. The initial tone of the adaptor was either high (HLHL…HL, referred to as a *high-first* adaptor) or low (LHLH…LH, a *low-first* adaptor). The duration of each tone was 40 ms (including 10-ms raised-cosine onset and offset ramps), and that of a cycle of two tones (HL or LH) was 200 ms including silent intervals between the tones. Onset asynchrony between the first and second tones in a cycle was either 60, 100, or 140 ms. Perceptual organization of the adaptor sequence changed depending on stimulus conditions. At 100-ms onset asynchrony, where H and L tones were isochronous, listeners perceived the adaptor as a coherent stream (LHLH… or HLHL…) when Δf was 1/6 octave. At 60-ms and 140-ms onset asynchronies, temporally closer tones were grouped together (For example, for the low-first adaptor, the organization was LH-LH-…LH at 60-ms onset asynchrony and L-HL-HL…-H at 140-ms onset asynchrony) when Δf was 1/6 octave (L: 944 Hz, H: 1060 Hz). With a Δf of two octaves (L: 500 Hz, H: 2000 Hz), H and L tones were perceptually segregated at all onset asynchronies. Perception was somewhat intermediate when Δf was 1/3 octave (L: 891 Hz, H: 1123 Hz). The interval between the adaptor and the test sound was 100 ms.

In all experiments, the method of constant stimuli was used. In each trial, an adaptor sequence was presented first, followed by the test sound. Subjects were asked to judge the direction of pitch change of the test sound, as downward or upward. For each trial, the frequency change rate of the test sound was selected

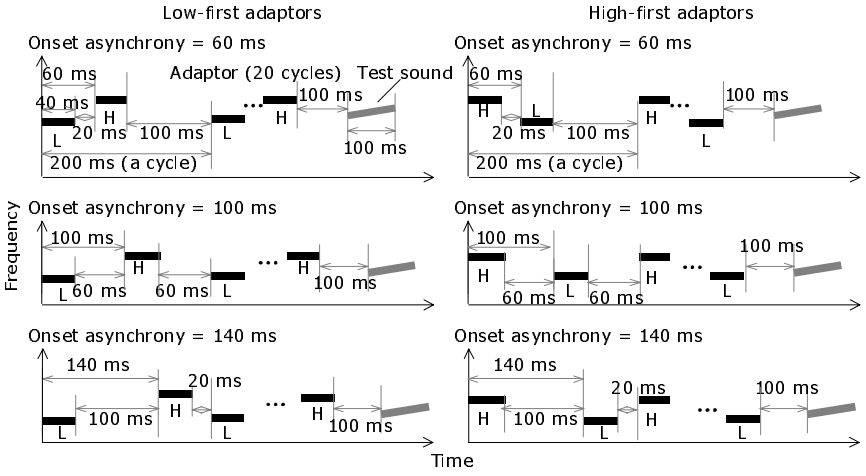


Fig. 1. Schematic representation of the stimuli used in the experiments. Six types of adaptor sequences were used. For details, see text.

randomly from nine values determined based on each subject's ability to judge the frequency change direction measured in preliminary sessions. The initial tone (L or H), onset asynchrony (60, 100 or 140 ms), and Δf (1/6, 1/3 or two octaves) of the adaptor were fixed throughout a session. Each value of the frequency change rate was tested at least 20 times. In the control condition, only the test sound was presented. The time required for a session was 10 - 20 min. The stimuli were presented diotically through headphones (Sennheiser HDA200) at 60 dB SPL.

Three young adults with normal hearing participated in the experiments with informed consent. All subjects received several hours of training in the judgment of pitch change direction in advance of the experimental sessions.

3 Results

The proportion of responses in which subjects judged that the pitch change of the test sound was upward was calculated as a function of the frequency change rate for each subject in each adaptor condition. Then psychometric functions were estimated using the maximum likelihood method. A frequency-change-rate value corresponding to a 50% point on a psychometric function is defined as the point of subjective constancy in pitch.

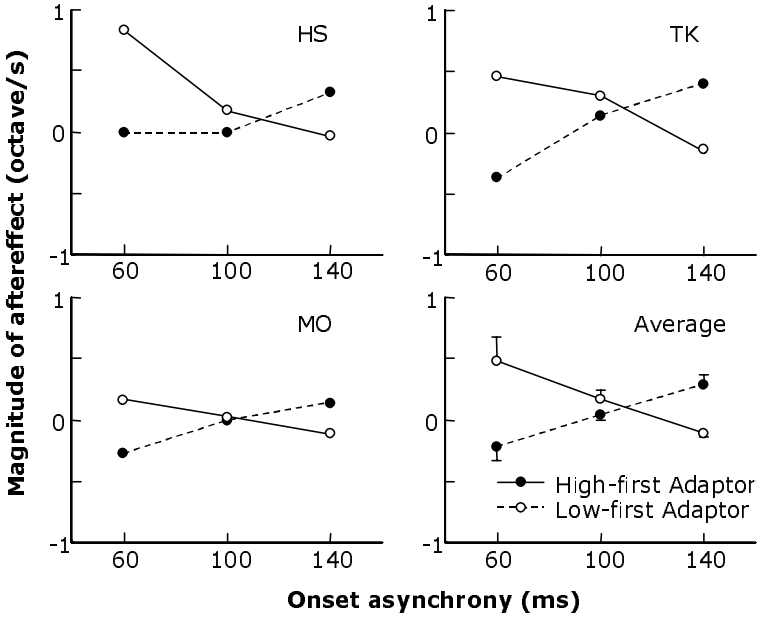


Fig. 2. The magnitude of the aftereffect for the high-first and low-first adaptors with a Δf of 1/6 octave as a function of onset asynchrony between two adapting tones in a cycle. For further explanation on how the magnitude was estimated, see text. Individual data for three subjects and the average of them are shown. Error bars indicate standard errors across subjects.

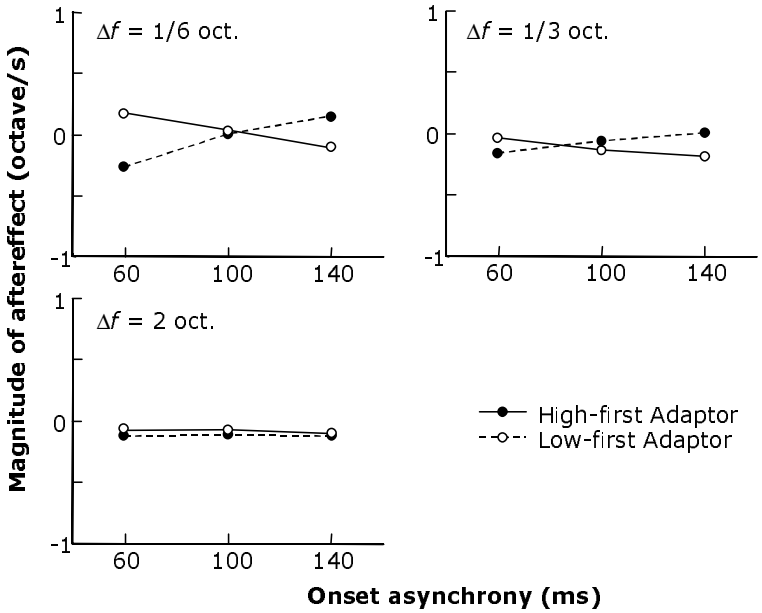


Fig. 3. The magnitude of the aftereffect for the high-first and low-first adaptors with Δ*f*s of 1/6, 1/3, and two octaves as a function of onset asynchrony between two adapting tones in a cycle. Data for subject MO is shown.

We define the magnitude of the aftereffect as the difference between the points of subjective constancy in each adaptor condition and in the no-adaptor control condition. When Δ*f* was 1/6 octave, the magnitude of the aftereffect was negative following the high-first adaptor with 60-ms onset asynchrony or the low-first adaptor with 140-ms onset asynchrony, and positive following the high-first adaptor with 140-ms onset asynchrony or the low-first adaptor with 60-ms onset asynchrony (Fig. 2). This means that a test sound with a constant frequency was perceived as if its pitch changed upward following adaptation to tone sequences in which every pair of high and low tones were perceptually grouped (HL-HL…-HL or L-HL-HL…-H) and as if its pitch changed downward following adaptation to tone sequences in which every pair of low and high tones were perceptually grouped (H-LH-LH…-L or LH-LH…-LH). The magnitude of the aftereffect decreased significantly when onset asynchrony was 100 ms both in the high-first and low-first adaptors. The absolute magnitude of the aftereffect varied across subjects, but the overall tendency was common for all subjects.

Figure 3 shows the effect of Δ*f* on the magnitude of the aftereffect for one subject. The magnitude significantly decreased when Δ*f* was 1/3 octave, and virtually disappeared when it was two octaves. This means that the magnitude of the aftereffect depends on the strength of the grouping of low and high or high and low tones in the adaptor sequence.

4 Discussion

We found that, in certain conditions, adaptation to an alternating sequence of high- and low-frequency tones did affect the subsequent judgment of spectral change directions depending on the perceptual grouping of the adaptor sequence. This indicates that the same population of neurons is involved both in the sequential grouping of tones and the detection of spectral change direction.

The observed shift of the point of subjective constancy in pitch following adaptation could be explained as follows: According to our hypothesis, a sequence of low and high tones close in frequency is assumed to activate neural units sensitive to upward spectral change, and a sequence of high and low tones activates those sensitive to downward spectral change. When the two tones are isochronously alternated, the former and the latter are activated to the same extent, producing no bias in the subsequent judgment of frequency change direction. When high and low tones are temporally closer than low and high tones in the alternating-tone sequence, the neurons sensitive to downward spectral change are activated more strongly than those sensitive to upward change. Prolonged exposure to this pattern would reduce the sensitivity of the former more than it reduces that of the latter. Consequently, the latter is activated more strongly than the former for the subsequent constant frequency sound. This pattern of neural activity is equivalent to that produced by an upward frequency glide without adaptation. As a result, the constant frequency sound is perceived as if its frequency were changing upward. This prediction is consistent with the results of the present experiments.

We also found that the aftereffect was decreased as the Δf of adaptor became larger, and virtually disappeared at a Δf of two octaves, where high and low tones were perceptually segregated into two streams (HHH…H and LLL…L). This could be explained as follows: A discontinuous frequency change of two octaves in 100 ms is too large or too rapid to activate the spectral change detectors, producing no effect on the subsequent judgment of spectral change direction.

Thus, the observed aftereffect is consistent with the hypothesis we have put forward. A possible role of spectral change detector in sequential grouping has been suggested by van Noorden (1975), but the idea has not been examined experimentally. Recent models of sequential grouping do not incorporate mechanisms that explicitly serve to link sequential spectral components (Beauvois and Meddis 1996; McCabe and Denham 1997). The present study provides psychophysical evidence for the existence of neural units linking separate spectral components presented sequentially.

The present theory is applicable to various aspects of the sequential organization of sounds, other than the effect of Δf and presentation rate. First, our hypothesis could explain why temporal order judgment of components is accurate within a stream but poor across different streams (Warren, Obusek, Farmer, and Warren 1969; Bregman and Campbell 1971); the temporal order of components linked into a stream by spectral change detectors can be represented as a direction of spectral change (Okada and Kashino 2003), whereas there is no such explicit neural representation for the order of unlinked components. Second, our hypothesis could also explain the tendency that the probability of hearing temporal coherence

decreases steadily over time in an alternating two-tone sequence (Anstis and Saida 1985); prolonged exposure to the alternating sequence would reduce the activity of spectral change detectors, resulting in a weakened link between spectral components in different frequency regions. Third, our hypothesis is consistent with the finding that, when high and low tones in an alternating sequence are connected by frequency glides, the tendency for the sequence to split into high and low streams is reduced and order perception of high and low tones becomes easier (Bregman and Dannenbring 1973; Dorman, Cutting, and Raphael 1975), because the frequency glides would activate the spectral change detectors more strongly.

A promising avenue of research would be to examine the relevance of the present theory to these and other aspects of the sequential organization of sounds.

Acknowledgments

This study was supported in part by a grant to RCAST of Doshisha University from the Ministry of Education, Culture, Sports, Science and Technology.

References

Anstis, S. and Saida, S. (1985) Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol.: Human Percept. Perform. 11, 257-271.

Beauvois, M. W. and Meddis, R. (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. J. Acoust. Soc. Am. 99, 2270-2280.

Bregman, A. S. and Campbell, J. (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J. Exp. Psychol, 89, 244-249.

Bregman, A. S. and Dannenbring, G. L. (1973) The effect of continuity on auditory stream segregation. Percept. Psychophys. 13, 308-312.

Dorman, M. F., Cutting, J. E. and Raphael, L. J. (1975) Perception of temporal order in vowel sequences with and without formant transitions. J. Exp. Psychol.: Human Percept. Perform. 104, 121-129.

Gardner, R. B. and Wilson, J. P. (1979) Evidence for direction-specific channels in the processing of frequency modulation. J. Acoust. Soc. Am. 66, 704-709.

Kayahara, T. (1998) Changing frequency after-effect of a linear frequency glide. J. Acoust. Soc. Am. 103, 3020.

McCabe, S. L. and Denham, M. J. (1997) A model of auditory streaming. J. Acoust. Soc. Am. 101, 1611-1621.

Okada, M. and Kashino, M. (2003). The role of spectral change detectors in temporal judgment of tones. NeuroReport, 14, 261-264.

Palmer, A. R. (1995) Neural signal processing. In: B. C. J. Moore (ed.), *Hearing*. Academic Press, San Diego, pp.75-121.

Shu, Z. J., Swindale, N. V. and Cynader, M. S. (1993) Spectral motion produces an auditory after-effect. Nature, 364, 721-723.

van Noorden, L. P. A. S. (1975) Temporal coherence in the perception of tone sequences. Unpublished doctoral dissertation. Eindhoven University of Technology, Netherlands.

Warren, R. M., Obusek, C. J., Farmer, R. M. and Warren, R. P. (1969) Auditory sequence: Confusion of patterns other than speech or music. Science, 164, 586-587.

Performance measures of auditory organization

Christophe Micheyl[1,2], Robert P. Carlyon[2], Rhodri Cusack[2], and Brian C.J. Moore[3]

[1] MIT-RLE, Cambridge, MA, USA, cmicheyl@mit.edu
[2] MRC-Cognition and Brain Sciences Unit, Cambridge, UK, {bob.carlyon, rhodri.cusack}@.mrc-cbu.cam.ac.uk
[3] Department of Experimental Psychology, University of Cambridge, Cambridge, UK, bcjm@cus.cam.ac.uk

1 Introduction

Many studies have been devoted to characterizing the stimulus-related factors that govern the perceptual organization of sound sequences, and more specifically, how the phenomenon of streaming depends upon different spectral or temporal characteristics of the stimuli (e.g., van Noorden 1975; Bregman 1990; Grimault, Bacon, and Micheyl, this volume). In contrast, relatively few studies have explored how auditory streaming influences or correlates with the ability to detect, recognize, or discriminate sounds (or sound sequences) embedded in sequences of (other) sounds. Yet, this question is of both theoretical and practical interest: it improves our understanding of the role of streaming in hearing, and allows us to devise indirect but accurate methods for the measurement of streaming, which are not dependent on purely subjective reports. Schematically, earlier studies on the perceptual consequences or correlates of streaming suggest that primitive stream segregation processes can impose strong limitations on tasks that require the combination of information across streams. For example, listeners lose to a large extent the ability to compare the timing of sounds that fall into different streams (e.g., Bregman and Campbell 1971). At the same time, stream segregation appears to facilitate the comparison or combination of information between sounds within a stream by "shielding" this information from potentially interfering influences by the other stream(s). For example, listeners can better recognize or discriminate pitch sequences interleaved temporally with extraneous sounds when the target tones differ widely in frequency from the targets, and are thus likely to be perceptually segregated from them (Dowling 1973; Micheyl and Carlyon 1998; Bey and McAdams, 2002).

The experiments described here further explore and quantify factors that are known to affect streaming and shows that these factors also affect the recognition, detection, or discrimination of sounds or sound features embedded in temporal

sequences. A key feature is that the experiments include tasks both where streaming would be expected to improve and where it should impair performance, using very similar stimuli and the same group of subjects. The results indicate dramatic changes in performance in the detection of local and global features, as well as in frequency discrimination, with variations in stimulus parameters that are known to affect the way in which the sequences are perceptually organized. Besides providing new information on the potential role of streaming in auditory perception, the experiments described here suggest new ways of testing for streaming indirectly, through performance measures.

2 Streaming and the detection of global and local features

As mentioned above, results in the literature suggest two main ways in which stream segregation may influence perceptual performance. On the one hand, stream segregation may limit the combination of elementary features, carried locally by consecutive sounds, into global features. On the other hand, stream segregation may facilitate the extraction of local features carried by individual sounds in sequences by preventing involuntary perceptual merging of features across adjacent sounds. We are not aware of a published study in which these two forms of influence of streaming on perceptual performance have been tested using comparable tasks, similar stimuli, and in the same listeners. Here, the same listeners successively performed two tasks: In the first, they had to detect a local feature consisting of an amplitude modulation (AM) applied to the middle (B) tone of two ABA triplets (where A and B stand for tones of (usually) different frequencies), embedded in a sequence of 14 such triplets. As the A tones in each triplet were either left steady or had an AM independently of the corresponding B tone, subjects had to focus on the B tones in order to do the task. Based on the hypothesis that stream segregation facilitates the detection of a feature within a given stream by shielding it from interfering influences by the other stream, we predicted that performance would *improve* with perceptual segregation between the A and B tones in this task. In the second task, listeners had to detect a global feature consisting of the AM of both A and B tones within a given ABA triplet. Once again, only two out of the 14 triplets contained in the sequence carried this global feature, and their position within the sequence was varied across trials. In this second task, listeners had to combine information across A and B tones. Based on the hypothesis that stream segregation limits the comparison or combination of information across streams, we predicted that performance would *worsen* with increasing perceptual segregation of the A and B tones in this task. For simplicity, we will refer to the two above-described tasks as “segregated easy” and “integrated easy”, respectively. The degree of stream segregation was controlled via two parameters: the frequency separation between the A and B tones (dF_{AB}), and the inter-triplet interval (ITI). These two parameters have been shown to be the most dominant factors influencing streaming with ABA sequences (Bregman, Ahad, Crum and O’Reilly 2000). The influence of these two parameters on the degree of stream segregation was checked in a preliminary experiment where listeners simply indicated whether they heard the sequences as integrated or segregated.

2.1 Methods

The stimuli were sequences of ABA triplets, where A and B represent 125-ms tone pips (including 25-ms $\cos^2$ ramps). The triplets were separated by 100, 200, or 300 ms ITIs. Within a triplet, the inter-tone interval was zero. Each sequence contained 14 triplets, for a total duration of 9.45 sec. The frequency of the A tones was always 500 Hz. That of the B tones was constant within each sequence but could vary across sequences, with possible values of 0, 1, 3, 6, or 9 semitones (ST) above that of A. Within a given triplet, the two A tones and/or the B tone were either left steady or modulated sinusoidally in amplitude (at a 40-Hz rate), leading to four types of triplets: MMM, SSS, MSM, or SMS, where S stands for steady, and M for modulated. The subject's task was to detect two (target) triplets of a given type interspersed randomly among more frequent triplets of the other two categories within each sequence. In the "segregated easy" task, the SMS triplets were used as targets and SSS and MSM as non-targets (Fig. 1). In the "integrated easy" task, MMM triplets were used as targets and MSM and SMS triplets as non-targets. In addition, we performed a rating experiment to measure the proportion of "2 stream"' vs. "1 stream" judgments as a function of dF_{AB} using the same sequences and the same subjects. Stimuli were played at 44.1 kHz through a 16-bit audio card, and delivered diotically via Sennheiser HD250 earphones. Four young normal-hearing subjects took part. They were tested in an IAC booth.

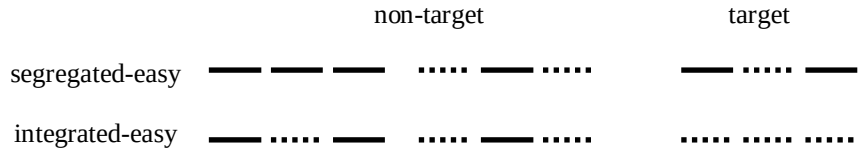


Fig. 1. Schema of the stimuli used in the "segregated easy" and "integrated easy" conditions of the local- and global-feature detection experiment (solid lines denote steady tones; dashed lines denote AM tones).

2.2 Results

Figure 2 shows the results of the stream-rating experiment. The percentage of "two stream" answers increased with dF_{AB} [$F(1,5)=580.46$, $p< 0.0005$]; it also increased significantly with decreasing ITI but only for intermediate dF_{AB}s [$F(1,5)=10.04$, $p<.05$].

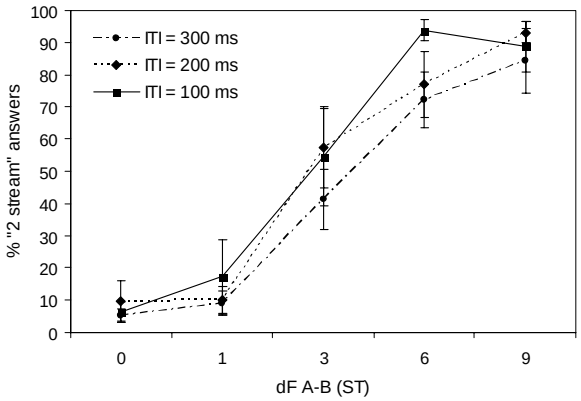


Fig. 2. Proportion of "2 stream" answers as a function of the frequency separation between the A and B tones. The error bars show standard errors of the mean across subjects.

The results of the detection experiment are shown in Fig. 3. The left-hand panel shows the results of the "segregated easy" condition; the right-hand panel, those of the "integrated easy" condition. In the "segregated easy" condition, performance increased as dF_{AB}, the frequency separation between the A and B tones, increased [$F(1,3)=308.24$, $p<0.0005$]. It also increased slightly with decreases of the ITI from 300 to 100 ms [$F(1,3)=13.55$, $p<.05$]. Opposite effects were found in "integrated easy" condition [dF: $F(1,3)=39.95$, $p<0.01$; ITI:$F(1,3)=28.01$, $p<0.05$].

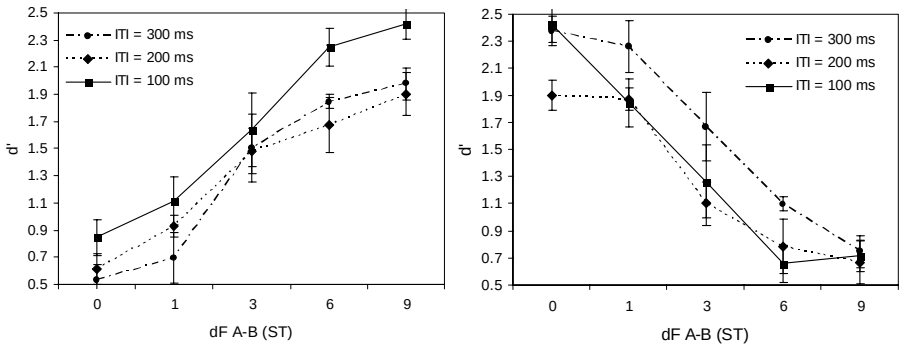


Fig. 3. Detection performance as a function of dF_{AB}, with ITI as parameter, in the "segregated easy" (left) and "integrated easy" (right) conditions. The error bars show standard errors of the mean across subjects.

2.3 Discussion

The results are consistent with the general hypothesis that, in conditions where consecutive sounds are perceptually segregated into different streams, the voluntary or involuntary combination of features between these temporally neighboring sounds is strongly constrained. Thus, in tasks where the combination of information between consecutive sounds is beneficial, performance decreases as stream segregation increases; conversely, in tasks where the combination of information between consecutive sounds is detrimental, because it produces confusion or distraction, performance increases with stream segregation.

3 Streaming and frequency discrimination

Earlier results (Micheyl and Carlyon, 1998; Gockel, Carlyon and Micheyl 1999) suggest that in certain conditions, thresholds for pitch discrimination between two sequential tones (targets) are dramatically elevated by the presentation of other tones before and after each target. This occurs particularly when the target and non-target tones fall in the same spectral region, and have a similar pitch and perceived laterality. Interestingly, these conditions are also those in which the target and non-target sounds are likely to fall in the same stream, suggesting that the temporal

interference effects in pitch discrimination depend upon the integration of consecutive target and non-target tones into the same stream.

To test this hypothesis, we measured frequency-discrimination thresholds (FDTs) for B tones presented in sequences of repeating ABA triplets as a function of two parameters: the frequency separation between A and B (dF_{AB}) and the number (N) of reference triplets preceding the comparison triplet that contained a frequency-shifted B-tone. We reasoned that as dF_{AB} increased, stream segregation too would increase, and FDTs should decrease. Furthermore, at intermediate dF_{AB}s stream segregation should increase with N; therefore, at these intermediate dF_{AB}s, a decrease in FDTs should be observed with increasing N. In addition, we used two ITIs, and predicted that, as stream segregation should be stronger at the shorter of the two, FDTs should decrease with increasing ITI.

3.1 Methods

The stimuli were sequences of ABA tone triplets, as illustrated schematically in figure 4. The A and B tone pips were 200 and 100 ms long (including 10 ms $\cos^2$ ramps), respectively. The triplets were separated by silent gaps of 200 ms (slow condition) or 100 ms (fast condition). The first N triplets in the sequence were identical. In the last triplet (N+1), the frequency of B was shifted up or down by an amount ($dF_{BB'}$) relative to its value in previous triplets; to avoid confusion, we refer to the shifted B frequency as B' hereafter. The subject's task was to indicate the direction of the shift. The value of $dF_{BB'}$ was decreased (by a factor of 2 until the fourth reversal, $\sqrt{2}$ thereafter) after three consecutive correct responses, and increased after each incorrect response. The three-down, one-up adaptive tracking procedure stopped automatically after 12 reversals and the frequency discrimination threshold (FDT) for the B tones was computed as the geometric mean of $dF_{BB'}$ over the last 8 reversals. N was set to 1, 2, 5, or 10. The frequency of B was set to 0, 1, 4, 7, or 10 ST above that of A. To prevent subjects from relying on across-trial comparisons, the frequency of A was chosen randomly in a +/-12% interval around 1 kHz on each trial; the frequencies of B and B' were adjusted accordingly. The tones were sampled at 40 kHz and played out through a 16-bit D/A (CED1401plus). They were presented at a level of 56 SPL in pink noise (15 dB SPL/Hz at 1 kHz) to the subject's right ear via Sennheiser HD250 earphones. Four young normal-hearing subjects were tested in an IAC booth.

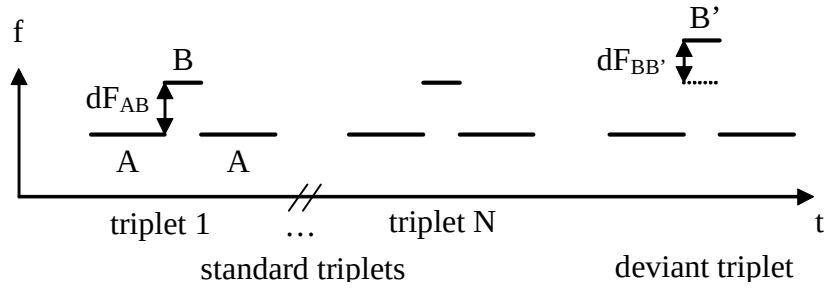


Fig. 4. Schematic illustration of the stimuli used to investigate the relation between streaming and frequency-discrimination performance (see text for details).

3.2 Results

Figure 5 shows the average FDTs as a function of dF_{AB} (the frequency separation between the A and B tones), for the different values of N (number of standard triplets in the sequence) at the 200-ms ITI. FDTs decreased dramatically as dF_{AB} increased above 1 ST [$F(1,3)=15.23$, $p<.05$]. A decrease in FDTs was also observed with increasing N, but this was significant for the 4-ST dF_{AB} only [$F(1,3)=13.25$, $p<.05$]. Finally, FDTs were found to decrease significantly as the ITI was decreased from 200 to 100 ms [$F(1,3)=24.76$, $p<.05$] (not shown here).

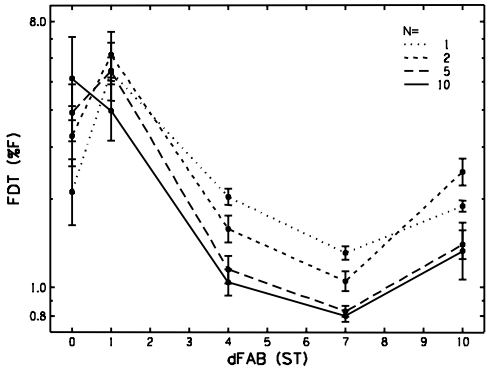


Fig. 5. FDTs as a function of the frequency separation between the A and B tones (dF_{AB}), for different numbers of standard triplets in the sequence (N).

3.3 Discussion

Improvements in FDTs between non-consecutive B tones in ABA triplets with increasing A-B frequency separation are consistent with the hypothesis that stream segregation prevents the involuntary combination of frequency information between consecutive tones that has been described in earlier studies as temporal over-integration of pitch information (Micheyl and Carlyon 1998). However, this result alone does not provide evidence that frequency-discrimination performance depends upon stream segregation. A second argument for the idea that the FDL is linked to stream segregation comes from the finding of significant improvements in FDTs with increases in N at intermediate dFs. This improvement is consistent with the build up of stream segregation over time (Anstis and Saida, 1985). If this point can be confirmed in later studies, the effect demonstrated here could provide the first performance measure of the build-up of stream segregation. However, at this point, one cannot exclude the possibility that other perceptual mechanisms may explain this effect. We are currently investigating whether a multiple-looks mechanism combined with temporal integration of frequency information can account for the observed dependence of FDTs on the number of standard triplets without resorting to an explanation in terms of stream segregation. Whichever conclusion is reached on this point, yet another argument for the notion that the performance in the task used here is related to streaming comes from the finding that FDTs decreased with increasing ITI. It is difficult to conceive of a reason for this, except that, all other things being equal, the reduction in ITI contributed to increase the amount of stream segregation. Thus there appears to be converging evidence for the notion that frequency-discrimination performance in the type of

task used here reflects streaming, and as such, may be used as an "objective" measure of the phenomenon.

4 Conclusions

The above-described experimental results demonstrate that performance in certain recognition, detection, or discrimination tasks with sound sequences depends heavily on factors that also affect the way in which these sequences are perceptually organized. Although these results do not prove that streaming determined performance, they certainly contribute to the converging evidence for this hypothesis. The results presented here are consistent with the hypothesis that stream segregation limits or prevents voluntary or involuntary combination of information across temporally adjacent sounds, so that when such combination is beneficial, performance decreases with increasing stream segregation, and when it is detrimental, performance increases with stream segregation. From a more practical point of view, the experiments described here indicate new methods for assessing streaming indirectly through performance measures, which may prove helpful in future psychophysical studies on streaming in humans or behavioral studies on perceptual auditory organization in animals.

Acknowledgments

The experiments were performed whilst the first author was visiting at the MRC-CBU, with funding from GR/N64861/01, Experimental Psychology Department, University of Cambridge, UK, and CNRS, UMR 5020, Lyon, France.

References

Anstis, S. and Saida S. (1985) Adaptation to auditory streaming of frequency-modulated tones. J. Exp. Psychol. Hum. Percept. Perf. 11, 257-271.

Bey, C. and McAdams, S. (2002) Schema-based processing in a auditory-scene analysis. Percept. Psychophys. 64, 844-854.

Bregman, A.S. (1990) *Auditory Scene Analysis*. MIT Press, Cambridge Massachusetts.

Bregman, A.S., Ahad, P.A., Crum, P.A. and O'Reilly, J. (2000) Effects of time intervals and tone durations on auditory stream segregation. Percept. Psychophys. 62, 626-636.

Bregman, A.S., and Campbell, J. (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J. Exptl. Psychol. 89, 244-249.

Dowling, W.J. (1973) The perception of interleaved melodies. Cog. Psychol. 5, 322-337.

Gockel, H., Carlyon, R.P., and Micheyl, C. (1999) Context dependence of fundamental-frequency discrimination: lateralized temporal fringes. J. Acoust. Soc. Am. 106, 3553-3563.

Grimault N., Bacon S.P., Micheyl C. (2003) Auditory streaming without spectral cues in hearing impaired subjects. In: D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Eds.), *this volume*.

Micheyl, C., and Carlyon R.P. (1998) Effects of temporal fringes on fundamental-frequency discrimination. J. Acoust. Soc. Am. 104, 3006-3018.

Van Noorden, L.P.A.S. (1975) *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Auditory streaming without spectral cues in hearing-impaired subjects

Nicolas Grimault[1,2], Sid P Bacon[1], and Christophe Micheyl[2,3]

1 Psychoacoustic Laboratory, Department of Speech and Hearing Science, Arizona State University spb@asu.edu
2 Département Neuroscience et Systèmes sensoriels, Université Claude Bernard Lyon 1, nicolas.grimault@olfac.univ-lyon1.fr
3 Now at: MIT – Research Laboratory of electronic, Cambridge, MA, USA. cmicheyl@mit.edu

1 Introduction

1.1 The importance of spectral cues for streaming

Sounds that rapidly follow each other in time tend to be organized into perceptual streams by the auditory system, based on similarity in frequency, timbre or pitch among other possible cues (Bregman 1990). Van Noorden (1975) showed that an alternating sequence of pure tones A and B (ABA-ABA-... where – represents a silence) can be perceived either as a single stream (a gallop) or as two streams when varying the frequency gap between A and B and the duration of the silence.

The channeling theory (Hartmann and Johnson 1991) explains numerous streaming data. The basic idea of this theory is that frequency, timbre and pitch similarities are all closely related to the similarity of the peripheral auditory excitation patterns evoked by the stimuli. This theory can partially account for the larger pitch difference required to perceptually segregate unresolved than resolved complex tones (Vliegen and Oxenham 1999; Vliegen, Moore and Oxenham 1999; Grimault, Micheyl, Carlyon, Arthaud and Collet 2000). A slight pitch shift leads to a larger variation of the excitation pattern for resolved complex tones than for unresolved complex tones.

Hearing-impaired and elderly people generally experience listening difficulties in environments in which several acoustic sources are present at the same time (Cherry 1953). Grimault, Micheyl, Carlyon, Arthaud and Collet (2001) have proposed that part of these difficulties might be explained by a reduced ability to segregate concurrent auditory streams due to reduced frequency selectivity. Indeed, sensorineural hearing loss is frequently associated with enlarged auditory filters (Moore 1985). This could reduce spectral cues that are typically used for sequential stream segregation. Figure 1 depicts the pattern of results from Grimault *et al.* (2001). In this study, the authors presented repetitions of sequences of alternating

complex tones (pattern ABA-...). The streaming performance is expressed in terms of d' as a function of the F0 difference (in octave) between the pitch of A and B complex tones. Based on data in the literature on auditory filter bandwidths in normal-hearing and hearing-impaired listeners, Grimault *et al.* (2001) estimated that, in the lower panel, A and B complex tones were all resolved for normal-hearing subjects and marginally resolved for hearing-impaired subjects. Based on their earlier work showing that, in normal-hearing listeners, less stream segregation was obtained with unresolved than with resolved harmonics (Grimault *et al.* 2000), they suggested that the reduced frequency resolution in hearing-impaired listeners could explain the reduced stream segregation in this group. Alternatively, in the upper panel where the performance overlaps for the two groups, the A and B complex tones were presumably unresolved for all subjects.

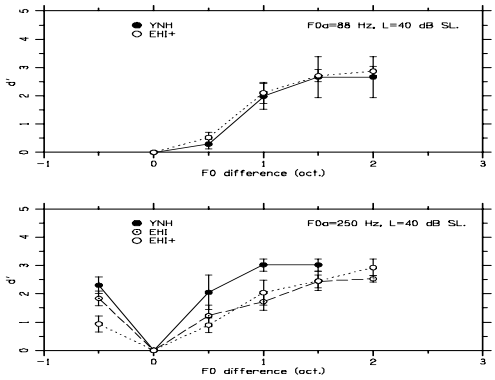


Fig. 1. Streaming performance for harmonic complex tones as a function of the F0 difference between the complexes replotted from Grimault *et al.* (2001). All stimuli were presented at 40 dB SL. Three groups of subjects with different amounts of hearing loss were tested. YNH: young normal-hearing listeners; EHI: elderly hearing-impaired listeners having normal hearing for their age (Davis, 1995); EHI+: elderly hearing-impaired listeners with an additional hearing loss. A constant-stimuli procedure was used. d' is computed from the percent of two-stream responses using a false-alarm rate equal to the percent of two-stream responses in the 0 F0-difference condition.

1.2 The use of temporal cues for streaming

The important role of spectral cues for stream segregation is tempered by recent findings showing that segregation can occur in conditions where mostly or only temporal cues are available for streaming (Vliegen and Oxenham 1999; Vliegen *et al.* 1999; Grimault *et al.* 2000). In the above-cited study by Grimault *et al.* (2001), for example, listeners reported hearing two streams at large F0 separations even

when the complex tones were filtered in a spectral region high enough for none of their harmonics to be individually resolved by the auditory periphery.

Strong evidence for streaming without spectral cues for normal-hearing listeners has been provided recently by Grimault, Bacon and Micheyl (2002). For A and B, the authors used bursts of modulated broadband noises that differed in modulation rate. The peripheral excitation patterns of the A and B stimuli were presumably identical, and thus differences in the temporal envelope between A and B were the only available cues for streaming. Figure 2 summarizes the results. As can be seen, most sequences were segregated into two streams as soon as a difference of about one octave was introduced between A and B modulation rates.

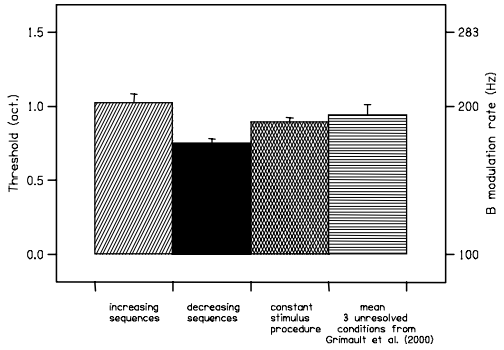


Fig. 2. Streaming performance of normal-hearing subjects. The first three bars summarize the streaming thresholds obtained with bursts of amplitude-modulated broadband noises at different rates (Grimault et al. 2002). Three different procedures (increasing or decreasing sequences and constant-stimuli procedure) were used. These are described in detail in section 2.2. The far-right bar summarizes the streaming thresholds obtained with unresolved complex tones in Grimault et al. (2000).

While these recent results demonstrate that stream segregation can occur on the basis of temporal envelope cues in normal-hearing listeners, they also suggest the possibility that hearing-impaired listeners in whom reduced frequency selectivity leads to weak or absent spectral cues, may take advantage of temporal cues for streaming. The temporal envelope is apparently preserved in hearing-impaired subjects (e.g. Bacon and Gleitman 1992) as long as the stimulus is audible throughout its spectrum (as would be the case for subjects with flat hearing losses listening at comfortable presentation levels). However, there could be some specific, and as yet unknown, factors limiting the use of temporal cues for the purpose of stream segregation in hearing-impaired listeners. The study described here was designed to test the extent to which hearing-impaired subjects can use temporal envelope cues for stream segregation.

2 Experiment

2.1 Subjects

Three sensorineural hearing-impaired subjects with ages from 59 to 80 years participated. All had a mild hearing loss from 125 Hz (mean 30 dB HL; SE 2.89 dB) to 8000 Hz (mean 43 dB HL; SE 6.67 dB). The ability of the hearing-impaired subjects to detect and discriminate amplitude modulation was, on average, similar to that of the normal-hearing subjects tested in Grimault *et al.* (2002).

2.2 Material and methods

All stimuli, materials and procedures were identical to those from Grimault *et al.* (2002). They consisted of temporal sequences of sinusoidally amplitude-modulated bursts of broadband noise. The stimulus sequences were formed by a repeating ABA- pattern, where the A and B correspond to fully modulated bursts having different modulation rates, and - represents a 20-ms silent interval. Each burst was 100 ms in duration, including rise and fall cosine ramps of 10 ms each. The stimuli were presented at a spectrum level of 50 dB SPL.

The streaming test procedure was the "timing procedure" described in Grimault *et al.* (2002). It involved the use of repeating ABA- sequences wherein the modulation rate of A was constant at 100 Hz, and the modulation rate of B varied either from 100 to 800 Hz (blocks of "increasing" sequences) or from 800 to 100 Hz (blocks of “decreasing” sequences). Ten different variation rates were used, leading to 10 different sequence durations, ranging from 10.88 s to 33.6 s. The different variation rates were produced by varying, simultaneously, the number of consecutive ABA- triplets in which the amplitude modulation rate of the B noises remained constant from 1 to 5, and the variation step from 0.03 to 0.3 oct. A and B were independent samples of noise. Moreover, new samples were used each time the modulation rate of the B noise varied. Each of the 10 sequences was presented 3 times, in one block, in a pseudo-randomized order. The use of different sequence durations prevented the subject from knowing *a priori* the length of the sequence. Any substantial effect of sequence duration would then indicate a response bias. This bias has been tested using the index **I**, the calculation of which is detailed in section 2.3. On the basis of this index, the results in conditions using “decreasing” sequences have been invalidated. Subjects were required to press a button on a response box (TDT) as many times as they wanted during the presentation of the sound sequence to indicate whether they heard one stream or two. The distribution of 1-stream responses minus the distribution of 2-stream responses across the 10 variation-rate conditions and 3 repetitions distributed in 0.1-octave wide bins was then plotted and fitted to compute a streaming threshold.

All stimuli were played through a 16-bit digital-to-analog converter (TDT DD1) at a sampling rate of 44.1 kHz and passed through an anti-aliasing filter (TDT FT6-2; -60 dB at 1.15 times the corner frequency) with a corner frequency set to 8 kHz. A continuous background noise was added in order to mask the perception of distortion products which could have been elicited by the amplitude-modulated noise stimuli (Wiegrebe and Patterson 1999). This masker was produced by

successively feeding the output of a white noise generator (TDT WG1) to an anti-aliasing filter (TDT FT6-2) with a corner frequency of 8 kHz and to a low-pass filter (TDT PF1; -25 dB at 1.15 times the corner frequency) with a corner frequency of 900 Hz. The spectrum level of the masker (within the pass band) was also 50 dB SPL. The signals and the background noise were independently led to two separate programmable attenuators (TDT PA4). The outputs of the attenuators were summed (TDT SM3) and led to the right earpiece of Sony MDR 7506 headphones via a pre-amplifier (TDT HBC) and a resistor box aimed to prevent electrical cross talk.

2.3 Results and discussion

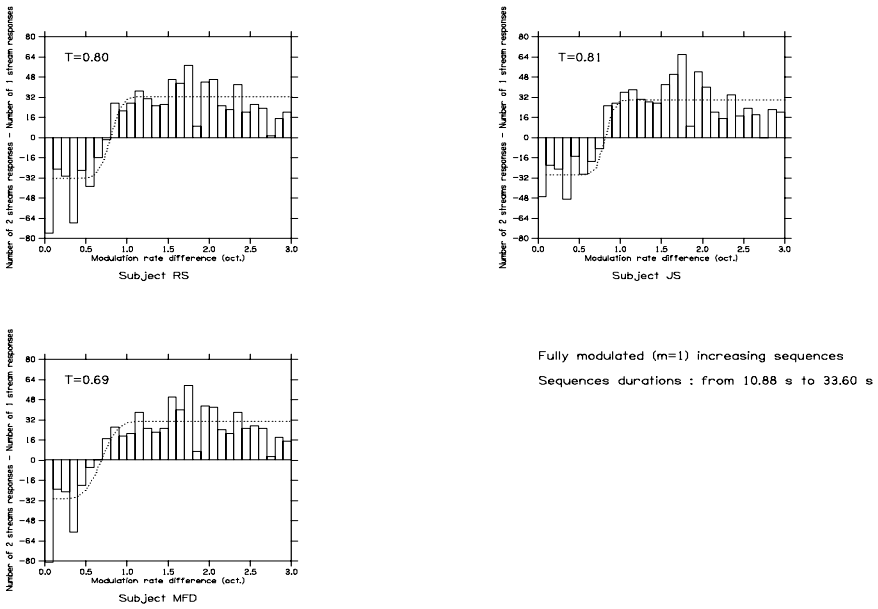


Fig. 3. Difference between the cumulated (across variation rates) distribution of 2-streams responses and the cumulated distribution of 1-stream responses as a function of the modulation rate difference between A and B noise bursts (in octaves). A cumulative gaussian curve (dotted line) has been used for the fitting that led to a streaming threshold **T** indicated on each panel.

The individual results for the “increasing” sequence conditions are plotted in Fig. 3. As indicated above, the validity of these results has been verified by computing an index **I**. For the results to be considered valid, **I** must be small. In order to compute this index, the distribution difference in each 10 variation rate conditions (i.e. sequence duration conditions) have been plotted for each individual (Fig. 4). **I** has then been defined as the sum of the euclidian distances between all the data points

(previously normed) that fell in the hatched areas (see Fig. 4) defined by the horizontal line at 0 and the vertical line at **T**.

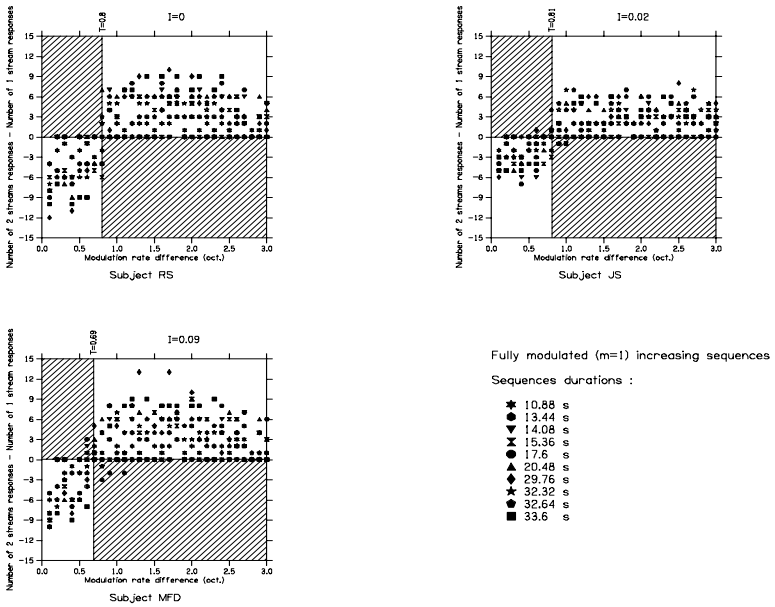


Fig. 4. Difference for each sequence duration condition between the distribution of 2-streams responses and the distribution of 1-stream responses as a function of the modulation rate difference between A and B noise bursts (in octaves). The individual thresholds (**T**) and index (**I**) values are given.

Low values of index **I** (0, 0.02 and 0.09) were computed indicating valid measurements with "increasing" sequences. Unfortunately, the values of the index for "decreasing" sequences were high (0.12, 0.75 and 0.29), indicating a large bias in the data that prevents them from being reported here.

Altogether, the individual streaming thresholds for the three hearing-impaired subjects are similar to those for normal-hearing subjects (Fig. 2, first bar: **T**=1.02 oct. on average). In fact, the thresholds for the hearing-impaired subjects even tend to be slightly lower than those for the normal-hearing subjects. The difference in performance (if any) is unlikely to be due to the lower loudness of the stimuli for hearing-impaired subjects. Indeed, Grimault *et al.* (2001) reported no significant effect of signal presentation. Further testing is required to explore this trend.

2.4 Conclusions

The following can be concluded from the present results:

(1) The hearing-impaired subjects tested here could use temporal envelope cues for streaming;
(2) Indeed, these subjects may have been slightly better at using these cues than previously tested normal-hearing subjects;
(3) Although the results from the increasing sequences were considered valid, the results from the decreasing sequences were not. It is unclear whether this increased variability or response bias reflects the hearing loss or the advanced age of the hearing-impaired subjects.

Acknowledgments

This work was supported by a post-doctoral fellowship from the Fondation pour la Recherche Médicale (FRM) awarded to the first author and a grant from NIDCD (DC01376) awarded to the second author.

References

Bacon, S.P. and Gleitman, R.M. (1992) Modulation detection in subjects with relatively flat hearing losses. J. Speech Hear. Res. 35, 642-653.

Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound.* MIT Press, Cambridge.

Cherry, E.C. (1953) Some experiments on the recognition of speech with one or two ears. J. Acoust. Soc. Am. 25, 975-979.

Davis, A. (1995) *Hearing in adults.* Whurr Publisher, London.

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P. and Collet, L. (2000) Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. J. Acoust. Soc. Am. 108, 263-271.

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P. and Collet, L. (2001) Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. Brit. J. Audiol. 35, 173-182.

Grimault, N., Bacon, S.P. and Micheyl, C. (2002) Auditory stream segregation on the basis of amplitude-modulation rate. J. Acoust. Soc. Am. 111, 1340-1348.

Hartmann, W.M. and Johnson, D. (1991) Stream segregation and peripheral channeling. Mus. Perc. 9, 155-184.

Moore, B.C.J. (1985) Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. Brit. J. Audiol. 19, 189-201.

van Noorden L.P.A.S. (1975) *Temporal coherence in the perception of tone sequences*. unpublished doctoral dissertation, Technische Hogeschool Eindhovern, Eindhoven, The Netherlands.

Vliegen, J. and Oxenham, A.J. (1999) Sequential stream segregation in the absence of spectral cues. J. Acoust. Soc. Am. 105, 339-346.

Vliegen, J., Moore, B.C.J. and Oxenham, A.J. (1999) The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. J. Acoust. Soc. Am. 106, 938-945.

Wiegrebe, L. and Patterson, R.D. (1999) Quantifying the distortion products generated by amplitude-modulated noise. J. Acoust. Soc. Am. 106, 2709-2718.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

The role of temporal structure in envelope processing

Neal F. Viemeister, Mark A. Stellmack, and Andrew J. Byrne

Department of Psychology, University of Minnesota, {nfv, stell006, byrne031}@umn.edu

1 Introduction

The ability to extract information from relatively slow amplitude changes, the envelope, appears to be a crucial aspect of auditory communication. The general question we address is how can this ability best be understood and described. One approach, the "modulation filterbank" (MF) model (Dau, Kollmeier, and Kohlrausch 1997; Ewert and Dau 2000) postulates that there are filters within the auditory system that are selectively tuned to envelope frequency. This, essentially, is a translation of the well-documented notion of auditory spectral filtering into the domain of modulation frequency. The general concept of modulation filters is based on data from experiments on modulation masking that indicate "tuning" for modulation frequency (see Ewert, Verhey, and Dau 2002). It is not surprising, therefore, that the general MF model can account for these data. The model, however, has considerable predictive power and can account for envelope effects, notably the differences seen in detection of sinusoidal amplitude modulation (SAM) for different types of carriers (Dau, Verhey, and Kohlrausch 1999; Kohlrausch, Fassel, and Dau 2000).

Unfortunately, direct experimental evaluation of the critical aspect of the MF models, modulation filters, has proved elusive and inconclusive. Several experiments that attempt such evaluation have used complex modulators or maskers (Lorenzi and Berthommier 1999; Moore, Sek, and Glasberg 1999; Sheft and Yost 2001). There recently have been examinations of phase effects that are not easily explained by the MF models, at least in their current forms (Strickland and Viemeister 1996; Moore and Sek 2000; Sek and Moore 2003). Part of the problem, one common to most models of envelope processing, is uncertainty as to that aspect of the "internal" envelope (the decision variable) that is important in detection and, more generally, in envelope-based perception.

This paper does not attempt a direct evaluation of modulation filters or, more generally, of approaches based on the modulation spectrum. Rather, it is an initial examination of detection in stimulus situations for which the local temporal structure of the envelope is manipulated systematically. The general question is to what extent does local temporal structure determine envelope perception. The

emphasis is on the time-domain representation of the envelope rather than on the modulation frequency domain representation.

2 Methods

In all conditions, a three-interval, three-alternative forced-choice task was used. The stimulus in each interval, $x(t)$, was defined as follows:

$$x(t) = \{[1 + m(t)] + [1 + m_s \cos(\omega_s t)]\}\, n(t) \quad (1)$$

where $m(t)$ is the masker modulator and m_s and ω_s are the modulation index and modulation frequency of the signal modulator. The carrier, $n(t)$, was a broadband Gaussian noise low-pass filtered at 10 kHz. The noise carrier was generated randomly from trial to trial but the same sample of noise was used in all three intervals of each trial. The starting phase of the signal modulator was always zero. Each interval was 1000 ms in duration with 500 ms of silence between the intervals. Correct-answer feedback was provided after each response.

The modulation index of the signal modulator (m_s) in the signal interval was varied adaptively in a 3-down-1-up procedure that estimated the 79.4 percent correct point on the psychometric function (Levitt 1971). In the non-signal intervals of each trial, m_s was set to zero. The initial step size for m_s was set to 2 dB (in units of 20 log m_s) and was reduced to 1 dB after four reversals. A block of trials was terminated after a total of 12 reversals, and threshold was estimated as the mean modulation depth of the final eight reversals. Four blocks of trials were run in each condition and the four threshold estimates were averaged to obtain the final threshold estimate for that condition. A block was terminated and no threshold was estimated for that block if the addition of the signal would have produced overmodulation.

Modulation-detection thresholds were measured separately for the signal modulator in the presence of several different masker modulators. In the "No Masker" condition, $m(t)$ was set to zero. In the "Sine Masker" condition: $m(t) = \cos(\omega_m t - \pi/2)$. In the "Pulsed Masker" conditions the masker modulator was a periodic rectangular pulse wave with duty factor (ratio of pulse width to signal period) ranging from 0.15 to 0.50. The fundamental frequency was equal to that of the signal (4, 8, 24, 48, and 64 Hz). The pulsed maskers were generated with the restriction that the amplitude and phase of the fundamental were always equal to those of the sine masker. Because of the manner in which $m(t)$ was generated, the DC of the pulse train with a duty factor of 0.25 was -2.8 dB and the DC for a duty factor of 0.15 was -8.1 dB relative to those of the other maskers. In a subset of conditions the overall DC of the masker was adjusted to be constant across duty factors. In all conditions, the waveform for the signal interval was scaled such that its rms was equal to that of the non-signal intervals.

In the Pulsed Masker conditions, modulation detection thresholds were measured when either the complete signal modulator was present or the signal

modulator was present only in the troughs of the masker modulator. Examples of the waveforms used in these conditions are shown in Fig. 1.

Stimuli were generated and presented via Matlab (Math Works) on a PC equipped with a high-quality, 24-bit sound card (Echo Audio Gina) at a sampling rate of 44.1 kHz. Stimuli were presented monaurally via Sony MDR-V6 stereo headphones to listeners seated in a sound-attenuating chamber. The listeners were four undergraduate students (one male and two female) from the University of Minnesota who were paid to participate in the study. All listeners had normal hearing according to lab audiometric standards. Listeners were allowed to practice in a variety of modulation masking conditions until their thresholds stabilized (about 2 weeks).

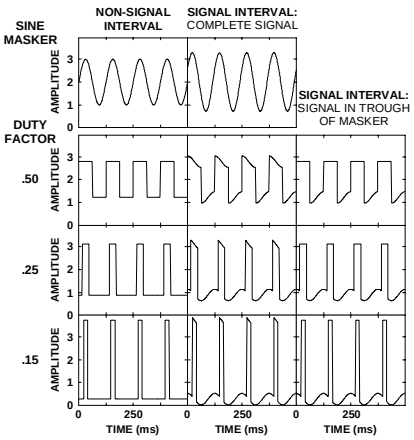


Fig. 1. Envelopes in the 8-Hz conditions. The amplitude of the fundamental but not the DC is constant across the conditions shown.

3 Results

Figure 2 shows data for two of the six modulation frequencies examined. The important trends are similar across modulation frequency and because the thresholds are similar across listeners, averaged thresholds are shown. Two reference conditions are shown by the symbols on the left-hand side of each panel. Also shown are thresholds for a condition in which the DC component of the masker for duty factors of 0.15 and 0.25 (inverted triangles) was adjusted to be equal to that in the Sine Masker and 0.50-duty factor conditions. The thresholds in the No Masker and Sine Masker conditions are in good agreement with those reported in the literature (e.g. Viemeister 1979; Wakefield and Viemeister 1990; Ozimek and Sek 1988).

One important aspect of the data shown in Fig. 2 is the general decrease in modulation threshold with decreasing duty factor. Similar decreases were observed for the other modulation frequencies. Another noteworthy

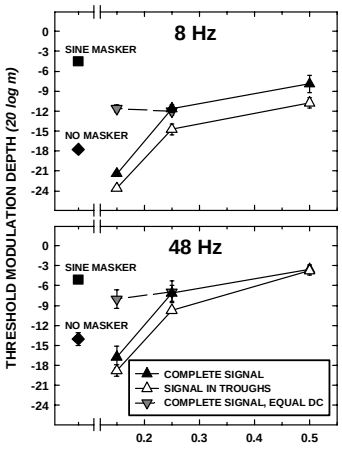


Fig. 2. Data from the 8- and 48-Hz conditions. The masker modulator, when present, was pulsed (triangles) or sinusoidal (squares). The fundamentals of all maskers were equal in amplitude.

result is that the thresholds for "signal in troughs" (open symbols) follow the same general trend as those for the complete signal but usually are slightly lower than those for the complete signal. A possible explanation for the lower thresholds is that truncating the signal introduces modulation energy in the envelope spectrum at harmonics of the signal frequency and that somehow this "splatter" can be used to aid detection of the signal. At this point we view these slight differences (the largest between the black and white symbols shown in Fig. 2 is 3.1 dB) as a secondary concern. The more important point is that the thresholds for the troughs-only condition are comparable to those for the complete signal.

The data for the "equal DC" condition make it clear that the large decrease in thresholds for a duty factor of 0.15 in the other conditions is due in part to the fact that the DC was allowed to vary across duty factor, although the change in DC had no effect for a duty factor of 0.25. Despite the obvious effects of varying the DC, the important comparison in Fig. 2 is between the thresholds in the Pulsed Masker conditions (triangles) and those for the Sine Masker (squares). For these conditions, addition of the signal causes an increment in the fundamental of the masker, which is equal across all of these conditions. The substantial changes in threshold across these conditions suggest that such a frequency-domain approach is not a useful way to interpret the data.

An alternative way to interpret the data shown in Fig. 2 is to consider that signal detectability in the Pulsed Masker conditions may be determined by the *local* modulation depth in the troughs of the masker. This notion is illustrated in Fig. 1 where it is clear that because the DC in the trough decreases with decreasing duty factor, the modulation depth of the portion of the signal in the trough is increased. The modulation depth in the trough of the masker can be computed as

$$m_t = (A_{max} - A_{\min})/(A_{max} + A_{min}) \quad (2)$$

where A_{max} and A_{min} are the maximum and minimum amplitudes of the signal plus masker in the trough of the masker. For the fixed amplitude signal used in Fig. 1, 20 log m_t is -13.7, -11.0, and -0.6 dB for duty cycles of 0.50, 0.25, and 0.15 respectively.

This notion is pursued in Fig. 3. The symbols connected by lines are the data from Fig. 2 expressed as modulation depths in the troughs, 20 log m_t, as defined above. The important point illustrated in Fig. 3 is that local modulation depth, specifically m_t, appears to eliminate the strong dependence on duty factor shown in Fig. 2. This clearly is true for the unequal DC conditions (solid symbols).

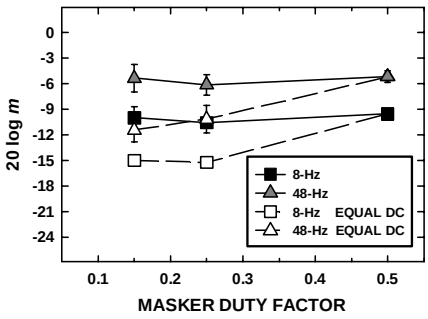


Fig. 3. A portion of the data from Fig. 2 expressed as the modulation depth in the troughs of the pulsed masker.

There are other aspects of the data that are notable. For all duty factors, thresholds are higher for 48 Hz than for 8 Hz and the increases in thresholds are in agreement with those for simple detection (the No Masker condition). However, the thresholds for the Sine Masker shown in Fig. 2 are the same for 8 and 48 Hz. This condition is modulation depth discrimination and, because it is frequency independent, appears to be a different process from that involved in the Pulsed Masker conditions. An explanation is that the Pulsed Masker conditions involve detection of the signal in the troughs of the masker rather than discrimination of a change in the amplitude of the fundamental of the envelope spectrum. This explanation is consistent with the notion that local modulation depth determines signal threshold.

It is not clear why the thresholds in Fig. 3 for the equal DC condition are lower than those for the other conditions. In the equal DC condition the overall level is higher but this should not affect the detectability of modulation (Viemeister 1979). Despite these uncertainties, Fig. 3 illustrates the main point of this paper, namely that local modulation depth may determine detectability with complex maskers.

In an experiment related to the present experiment a wider range of masker duty factors was examined using maskers for which the amplitudes of both the DC and fundamental were held constant across duty factor. To avoid overmodulation, a potential problem using the adaptive procedure, performance (percent correct) was measured for fixed signal depths (m_s). An important result is that there were large differences in performance for the complementary pairs of duty factors examined, specifically (0.15, 0.85) and (0.25, 0.75). For such pairs the amplitude spectra of the envelopes are identical. Thus, this result further underscores the importance of phase and temporal structure.

4 Discussion

The general question addressed in this paper is the nature of the decision variable used in envelope processing. This is a fundamental question and is one that has received attention over the past few decades (for a review see Strickland and Viemeister 1996). It appears that the decision variables that have been proposed can, at best, describe the data only for a limited subset of conditions, namely detection of SAM. None appears to be able to account for the present data.

Our qualitative, black-box account is essentially an extension of the notion of "multiple looks" (Viemeister and Wakefield 1991) in which information from brief samples or looks at some central representation is selectively and intelligently combined to make decisions about the stimulus. In the present situation this central representation is the result of relatively early stages of envelope processing. An example of such processing is lowpass filtering preceded by a nonlinearity such as half-wave rectification. Another is modulation filtering possibly including some type of recombination of the filter outputs. At this point in model development the details of this processing seem less important than in describing how the output is processed to arrive at a decision variable.

A decision variable that appears promising is one based on local modulation depth at the signal frequency. This is closely related to m_t in Eq 2 which provided a good, but incomplete, account of the general trends of the data from this experiment. Thus, after initial envelope processing the running modulation depth of the waveform at the signal frequency is computed and those temporal regions in which the local depth is largest is given the most weight. To be more concrete, one such statistic is based on

$$D(t) = [e'(t) - e'(t + T/2)]/[e'(t) + e'(t + T/2)] \quad (3)$$

where $e'(t)$ is the "internal envelope" after initial processing and T is the period of the signal. A simple decision variable would be the maximum value of $D(t)$. There are, of course, other ways to compute local modulation depth including local rms and running cross correlation. The key element, however, appears to be that the statistic is computed over some relatively short time window and is allowed to run over the course of the modulation. This permits extraction of temporal structure and distinguishes this type of decision statistic from those that have been previously examined.

The duration of the window probably should be related to the period of the signal rather than fixed. This conjecture is based on our observations in experiments similar to those reported here but using maskers and signals that differ in their periods. A window related to the signal captures the notion that listeners have some knowledge of the signal (and masker) periodicity and offers the appealing possibility that the frequency selectivity shown in envelope processing, namely modulation rate discrimination and tuning in modulation masking, can be explained on a purely temporal basis.

5 Summary and conclusion

The results from this experiment strongly suggest a potentially important role for local temporal structure in envelope processing. The data indicate that the detectability of SAM in the presence of pulsatile AM maskers is approximately determined by the local modulation depth in the trough of the masker (Fig. 3). The data may not be inconsistent with spectrally-based models, however, especially if the decision variable somehow incorporates phase information. In our opinion, a more promising approach emphasizes a time-domain analysis. We have outlined the rudiments of a multiple-looks type of model based on local modulation depth that captures, at least qualitatively, the general characteristics of the data. Such a model may provide the basis for a purely temporally-based account of this fundamentally important aspect of auditory perception.

Acknowledgments

This work was supported by Research Grant No. R01 DC 00683 from the National Institute on Deafness and Communication Disorders, National Institutes of Health.

References

Dau, T., Verhey, J. and Kohlrausch, A. (1999) Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. J. Acoust. Soc. Am. 106, 2752-2760.

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892-2905.

Ewert, S.D. and Dau, T. (2000) Characterizing frequency selectivity for envelope fluctuations. J. Acoust. Soc. Am. 108, 1181-1196.

Ewert, S.D., Verhey, J.L. and Dau, T. (2002) Spectro-temporal processing in the envelope-frequency domain. J. Acoust. Soc. Am. 112, 2921-2931.

Kohlrausch, A., Fassel, R. and Dau, T. (2000) The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J. Acoust. Soc. Am. 108, 723-734.

Levitt, H. (1971) Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467-477.

Lorenzi, C. and Berthommier, F. (1999) Discrimination of amplitude-modulation phase spectrum. J. Acoust. Soc. Am. 105, 2987-2990.

Moore, B.C.J., Sek, A. and Glasberg, B.R. (1999) Modulation masking produced by beating modulators. J. Acoust. Soc. Am. 106, 908-918.

Moore, B.C.J. and Sek, A. (2000) Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation. J. Acoust. Soc. Am. 108, 2337-2344.

Ozimek, E. and Sek, A. (1988) AM difference limens for noise bands. Acustica 66, 153-160.

Sek, A. and Moore, B.C.J. (2003) Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators. J. Acoust. Soc. Am. (in press)

Sheft, S. and Yost, W.A. (2001) AM detection with interrupted modulation. In: D.J. Breebart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs and R. Schoonhoven (Eds.), *Physiological and Psychophysical Bases of Auditory Function.* Shaker, Maastricht, pp. 290-297.

Strickland, E.A. and Viemeister, N.F. (1996) Cues for discrimination of envelopes. J. Acoust. Soc. Am. 99, 3638-3646.

Viemeister, N.F. (1979) Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Am. 66, 1364-1380.

Viemeister, N.F. and Wakefield, G.H. (1991) Temporal integration and multiple looks. J. Acoust. Soc. Am. 90, 858-865.

Wakefield, G.H. and Viemeister, N.F. (1990) Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise. J. Acoust. Soc. Am. 88, 1367-1373.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Detecting changes in amplitude-modulation frequency: A test of the concept of excitation pattern in the temporal-envelope domain

Christian Füllgrabe[1], Laurent Demany[2], and Christian Lorenzi[1,3]

[1] Laboratoire de Psychologie Expérimentale - UMR CNRS 8581, Université René Descartes Paris 5, {christian.fullgrabe,christian.lorenzi}@psycho.univ-paris5.fr
[2] Laboratoire de Neurophysiologie - UMR CNRS 5543, Université Victor Segalen Bordeaux 2, laurent.demany@psyac.u-bordeaux2.fr
[3] Institut Universitaire de France

1 Introduction

In the audio-frequency domain, the concept of auditory filter has been successfully used to account for frequency selectivity. Based on this notion, the classic "Excitation Pattern (EP) model" (Zwicker 1956, 1970; Maiwald 1967; Florentine and Buus 1981) postulates that a sound of a given frequency evokes a so-called EP which is psychoacoustically defined as the short-term output of the auditory filters as a function of their center frequency. Physiologically, the EP along the tonotopic scale corresponds to a pattern of neural activity (discharge rate) as a function of the characteristic frequency of the excited neurons. As a consequence, any change in frequency results in a change in the distribution of excitation evoked by the sound. Frequency changes may therefore be detected on the basis of a lateral movement of the EP or a local variation in excitation level, by monitoring the output of a single or several auditory filters.

Since changes in level also influence the amount of excitation, the perception of changes in frequency and changes in level depends on the same underlying mechanism according to the EP model. If this is true, one can expect that the ability to detect changes in frequency will be degraded by the co-occurrence of random variations in level: In the presence of such random variations, the variation in excitation level of a single filter will no longer be a reliable cue. Psychophysical experiments confirmed that, under conditions where no temporal coding of frequency information was possible, frequency discrimination thresholds (DLFs) and frequency modulation detection thresholds (FMDLs) increased when random changes in level were added to the stimuli (Moore and Glasberg 1989).

In the temporal-envelope domain, the concept of modulation filter has been used to account for frequency selectivity in amplitude-modulation (AM) masking (e.g.,

Houtgast 1989; Bacon and Grantham 1989). According to the modulation filterbank (MFB) model (e.g., Dau, Kollmeier and Kohlrausch 1997; Ewert, Verhey and Dau 2002), an array of broad overlapping modulation filters, each tuned to a different AM frequency, decomposes the temporal envelope of sounds into its spectral components. As in the case of auditory filters and tonotopy, physiological data suggest that AM frequency information present in the stimulus envelope may be mapped onto a "periodotopically" organized space (e.g., Langner 1992; Schulze, Hess, Ohl and Scheich 2002). Moreover, according to a simplified version of the MFB model, the auditory system monitors long-term changes in the output of the modulation filters activated by the stimuli in order to detect temporal envelope fluctuations (Ewert and Dau 2000).

The present study aimed at investigating whether or not the concept of EP as put forward by Zwicker (1956) and Maiwald (1967) could be transposed to the temporal-envelope domain in order to account for the ability to perceive changes in AM frequency. According to this concept, it should be expected that listeners compare excitation levels of the EP along a periodotopic scale when discriminating AM frequencies or detecting continuous variations in AM frequency (*i.e.* FM in the temporal-envelope domain). If this is true, random fluctuations in AM depth of the stimuli are liable to have a deleterious effect on the listeners' performance in the two tasks. However, this may not be the case if the perception of changes in AM frequency is based on the timing information related to the stimulus envelope (for instance, time intervals between successive peaks in the stimulus envelope).

In the first experiment, AM frequency discrimination thresholds with fixed or randomly varying AM depth were measured. AM depth varied either between the two stimuli of a given trial (within-trial randomization) or from one trial to the next (between-trial randomization). In the second experiment, detection thresholds for FM imposed on a sinusoidally amplitude modulated (SAM) stimulus were obtained for fixed AM depth and in the presence of an additional, "2^{nd}-order" AM, *i.e.* a sinusoidal modulation of the depth of the SAM stimulus (for a more detailed description of 2^{nd}-order AM, see Lorenzi, Soares and Vonner 2001).

2 Discrimination of AM frequency

2.1 Method

Five listeners (aged 27 to 48 years) with audiometrically normal hearing sensitivity participated in this experiment. All listeners had previous experience with psychoacoustical tasks and were tested individually in a sound-attenuating booth.

A two-interval, two-alternative forced-choice (2I, 2AFC) procedure with feedback was used to assess frequency difference limens (DLFs) for SAM imposed on white noise. On each trial, a standard stimulus with a fixed (reference) modulation frequency (f_m) of either 4, 16, or 64 Hz, and a target stimulus with a variable modulation frequency ($f_m+\Delta f_m$) were presented successively in random order. The listeners' task was to indicate the interval containing the higher modulation frequency. Over a given experimental block, AM depth, *m*, was either (i) held constant at 0.2, 0.6, or 1 in three "fixed-depth" conditions, or (ii) varied

randomly from 0.2 to 1 in two "randomized-depth" conditions. In the "between-trial randomization" condition, *m* varied between successive trials, but was identical in the two observation intervals (standard and target) of a given trial. By contrast, *m* also varied within trials in the "within-trial randomization" condition. Each time, the variation of depth was at least 0.2. All stimuli were equated in rms and presented at 75 dB SPL. The starting phase of SAM was always set to zero. Each stimulus had a duration which was chosen at random between 2 and 2.5 s, and was gated on and off with 50-ms raised-cosine ramps. Randomization of the duration was used to prevent listeners from simply counting the number of modulation cycles. The inter-stimulus interval was fixed at 1 s.

The AM frequency of the target stimulus was varied using a two-down one-up stepping rule that estimates the difference limen, Δf_m (in Hz), necessary for 70.7 % correct discrimination. The step size of the variation corresponded initially to a factor of 1.26, and was reduced to 1.12 after the first two reversals. The mean of the values of Δf_m at the last 10 reversals in a series of 16 reversals was taken as the threshold estimate for that block. For each listener and f_m, thresholds are based upon the three lowest estimates out of four.

2.2 Results and discussion

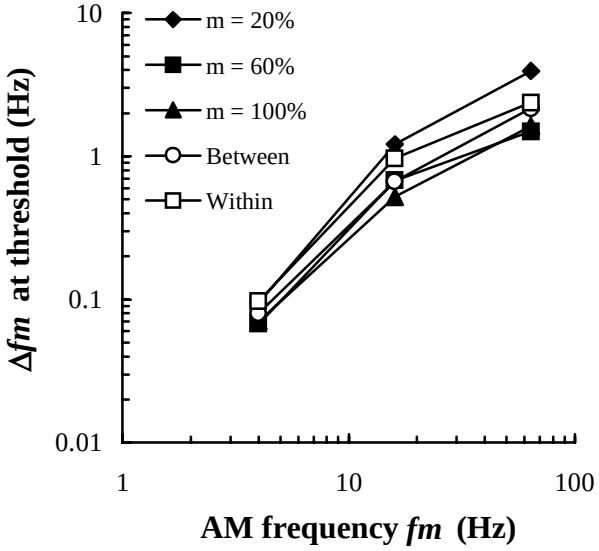


Fig. 1. Average DLFs as a function of reference AM frequency, f_m. The parameter was AM depth, *m*, fixed at 0.2 (filled diamonds), 0.6 (filled squares), or 1 (filled triangles), or randomly varying from 0.2 to 1 between trials (open circles) or within trials (open squares).

In agreement with previous studies (e.g., Formby 1985; Hanna 1992), it was found that AM frequency discrimination thresholds (in Hz) increased with reference AM frequency (see Fig. 1). The data obtained in the fixed-depth conditions showed that AM depth only influenced frequency discrimination at f_m=16 and 64 Hz. This may be explained by the fact that an AM depth of 0.2 is getting closer to AM detection threshold when the reference AM frequency increases from 4 to 64 Hz (e.g., Lorenzi *et al.* 2001). The data also showed that DLFs for random AM depth fell within the range of DLFs obtained in the fixed AM-depth conditions. In addition, DLFs for the two randomized-depth conditions did not differ significantly from each other (although at f_m=16 Hz thresholds were slightly better when m was held constant within trials).

Clearly, these findings cannot be accounted for by a simple (*i.e.* a single-band) EP model. Zwicker (1956) hypothesized that the displacement of an EP produced by a frequency shift is detected if, and as soon as, there is a detectable change in the excitation level of a single filter - the most sensitive one. Here, a similar model would predict larger DLFs in the within-trial randomization condition than in the other conditions. However, a more elaborate EP model may explain why this was not observed. It has to be assumed that the auditory system is able to combine excitation level information across different parts of an EP in order to estimate the "centroid" or the peak of this EP. If this were the case, then frequency shifts could be detected independently of shifts in level (in the audio-frequency domain), or AM depth (in the temporal-envelope domain).

3 Detection of frequency modulation imposed on AM

3.1 Method

The same five listeners participated in this experiment. Their task was to detect the presence of a sinusoidal FM imposed on SAM. On each trial, a standard (without FM) and a target stimulus (with FM) were presented successively in random order. In one condition, FMDLs were measured using a fixed AM depth (m=0.5). Equation 1 describes the standard, $S(t)$, consisting of a white-noise carrier, $n(t)$, amplitude-modulated at f_m=4, 16, or 64 Hz:

$$S(t) = [1 + m \sin(2\pi f_m t)]\, n(t) \quad (1)$$

The target, $T(t)$, given in Eq 2, was identical to the standard, except that f_m was modulated with a frequency F_m equal to 1 or 4 Hz:

$$T(t) = [1 + m \sin(2\pi f_m t + (\Delta f_m / F_m) \sin(2\pi F_m t + \phi)]\, n(t) \quad (2)$$

Δf_m (in Hz) corresponds to one half of the total frequency swing, and ϕ corresponds to the FM starting phase, which was randomized on each trial. In a second condition, FMDLs were measured with sinusoidally varying AM depth (*i.e.* 2nd-order AM) within each interval. For this condition, Eq 3 describes the depth m of standard and target stimuli as a function of time:

$$m(t) = [0.5 + 0.2 \sin(2\pi F_m t)] \quad (3)$$

Thus, in each interval, AM depth varied cyclically between 0.3 and 0.7, at a frequency (F_m) which was equal to the FM frequency in the target stimulus (Eq 2). In the two conditions, each stimulus had a total duration of 2 s and was gated on and off with 50-ms raised cosine ramps. All stimuli were equated in rms and presented at 75 dB SPL. The inter-stimulus interval was set to 1 s.

FM detection thresholds were obtained using a two-down one-up, 2I, 2AFC adaptive procedure with feedback. This estimated the value of Δf_m necessary for 70.7 % correct detection. The step size of Δf_m variation corresponded initially to a factor of 1.59; it was reduced to 1.26 after the first two reversals. The mean of the values of Δf_m at the last 10 reversals in a block of 16 reversals was taken as the threshold estimate for that block. Thresholds correspond to the four lowest estimates out of five.

3.2 Results and discussion

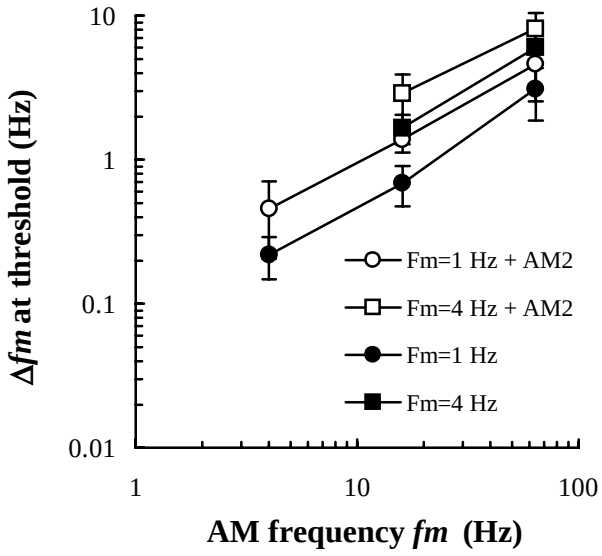


Fig. 2. Average FMDLs as a function of AM frequency, f_m. The sinusoidal FM of frequency F_m equal to 1 Hz (circles) or 4 Hz (squares) had to be detected alone (filled symbols) or in the presence of a 2nd-order AM (open symbols) with a frequency always equal to the FM frequency, F_m. Error bars show ± one standard deviation about the mean across listeners.

Overall, the measured FMDLs increased as AM frequency (f_m) increased from 4 to 64 Hz (Fig. 2). It is worthy to note that this increase was somewhat different from that observed in the first experiment between 4 and 16 Hz. FMDLs also increased, by a factor of 2, when the FM frequency F_m increased from 1 to 4 Hz.

If the concept of EP holds in the envelope domain, then FM should have induced a beat of frequency F_m at the output of the modulation filters tuned near the

AM frequency, in the same way as 2nd-order AM did. Hence, FMDLs and 2nd-order AM detection thresholds should show a similar dependence on F_m. Consistent with the FM data, previous results on 2nd-order AM sensitivity (Lorenzi *et al.* 2001) indicate that detection thresholds obtained for f_m=16 Hz double indeed when the 2nd-order frequency (F_m) increases from 1 to 4 Hz (no data available for f_m=64 Hz).

As predicted by a simple EP model, adding 2nd-order AM to the standard and target stimuli had a deleterious effect on FMDLs. However, the observed effect was rather small in size. The mean FMDLs increased by a factor ranging from 1.4 (for f_m=64 Hz and F_m=4 Hz) to 2.1 (for f_m=4 Hz and F_m=1 Hz). Overall, the effect was highly significant (p<0.001); but post-hoc tests (Tukey HSD) indicate that it was locally significant (with p<0.05) only for the two lowest AM frequencies, f_m=4 and 16 Hz.

4 Conclusions

Taken together, the psychophysical data reported here do not permit a simple answer to the question the present study was designed to address, namely: Can one account for the detectability of changes in AM frequency by an EP model derived from the MFB concept ?

In the first experiment, it was found that random variations in AM depth did not impair the discrimination of *discrete* changes in AM frequency. This result strongly argues against a single-band EP model in which the output of a single modulation filter is monitored, but may still be explained by a multi-band version according to which the auditory system computes the centroid or peak of the EP produced by SAM in the MFB.

By contrast, the results of the second experiment provide some support to the idea that *continuous* changes in AM frequency are detected on the basis of short-term variations of excitation level at the output of modulation filters: As predicted by a single-band EP model, FMDLs worsened when a 2nd-order AM was added to the stimuli. Since this increase was not dramatic (on average a factor of 1.73), excitation level information may again have been combined, to some extent, across different parts of the EP. However, the fact that FMDLs increased with 2nd order AM would then imply that the combination process is rather “sluggish” in nature, *i.e.*, not optimally efficient when the EP is moving continuously, even at such a low FM frequency as 1 Hz.

On the other hand, the results of the first experiment are also compatible with the notion of a purely temporal mechanism based on the measurement of time intervals between successive peaks in the stimulus envelope. For the detection of FM in the presence of 2nd-order AM, this mechanism will not operate optimally if the cyclic variation in the magnitude of AM peaks disrupts phase-locking to the (1st-order) AM, or shifts attention away from the FM.

It is conceivable that completely different mechanisms (such as a purely temporal mechanism and a multi-band process) underlie frequency discrimination and FM detection in the temporal-envelope domain, as already hypothesized in the audio-frequency domain (e.g., Moore and Glasberg 1989; Moore and Skrodzka 2002). Further studies will have to investigate the different hypotheses listed above.

Acknowledgments

This research was supported by a MENRT grant to C. Füllgrabe, and a grant from the Institut Universitaire de France to C. Lorenzi.

References

Bacon, S.P. and Grantham, D.W. (1989) Modulation masking patterns: Effects of modulation frequency, depth, and phase. J. Acoust. Soc. Am. 85, 2575-2580.

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation: I. Modulation detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892-2905.

Ewert, S.D. and Dau, T. (2000) Characterizing frequency selectivity for envelope fluctuations. J. Acoust. Soc. Am. 108, 1181-1196.

Ewert, S.D., Verhey, J.L. and Dau, T. (2002) Spectro-temporal processing in the envelope-frequency domain. J. Acoust. Soc. Am. 112, 2921-2931.

Florentine, M. and Buus, S. (1981) An excitation-pattern model for intensity discrimination. J. Acoust. Soc. Am. 70, 1646-1654.

Formby, C. (1985) Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners. J. Acoust. Soc. Am. 78, 70-77.

Hanna, T.E. (1992) Discrimination and identification of modulation rate using noise carrier. J. Acoust. Soc. Am. 91, 2122-2128.

Houtgast, T. (1989) Frequency selectivity in amplitude-modulation detection. J. Acoust. Soc. Am. 85, 1676-1680.

Langner, G. (1992) Periodicity coding in the auditory system. Hear. Res. 60, 115-142.

Lorenzi, C., Soares, C. and Vonner, T. (2001) Second-order temporal modulation transfer functions. J. Acoust. Soc. Am. 110, 1030-1038.

Maiwald, D. (1967) Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas. Acustica 18, 193-207.

Moore, B.C.J. and Glasberg, B.R. (1989) Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. J. Acoust. Soc. Am. 86, 1722-1732.

Moore, B.C.J. and Skrodzka, E. (2002) Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation. J. Acoust. Soc. Am. 111, 327-335.

Schulze, H., Hess, A., Ohl, F.W. and Scheich, H. (2002) Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil. Eur. J. Neurosci. 15, 1077-1084.

Zwicker, E. (1956) Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs. Acustica 6, 356-381.

Zwicker, E. (1970) Masking and psychological excitation as consequences of the ear's frequency analysis. In: R. Plomp and G.F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing.* Sijthoff, Leiden.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Modeling the role of duration in intensity increment detection

Erick Gallun, Ervin R. Hafter, and Anne-Marie Bonnel

Dept. of Psychology, University of California, Berkeley,
{gallun,hafter,ambonnel}@socrates.berkeley.edu

1 Introduction

A tonal signal added in-phase to an ongoing tonal pedestal of the same frequency produces changes in both the energy at the frequency of the tone and the shape of the amplitude envelope of the pedestal (see Fig. 1). A variety of models based on combinations of these two cues were used here to predict performance 1) in a detection task where signals of various durations were presented in a quiet background, and 2) in a detection task where various amplitude-modulated maskers were presented as well.

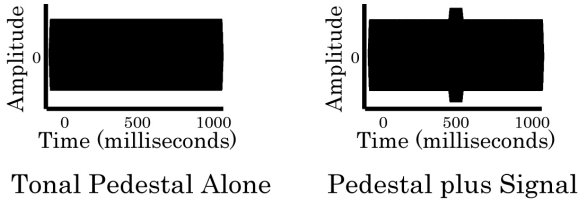


Fig. 1. A tonal pedestal might or might not have a signal added to it. Signals are shown larger than in the actual experiment to facilitate viewing.

2 Modeling

Previously, modeling the detection of added signals has focused on either the change in energy (Green, Birdsall and Tanner 1957) or the change in envelope (Oxenham 1997; 1998). This study tested an energy model, an envelope model and three "mixed" models, each of which is sensitive to both changes in energy and changes in envelope.

(1) The "energy" model described in Green and Swets (1966, p. 211) performs a temporal integration on the power in a critical band filter centered on the frequency of the pedestal. A two-hundred millisecond time constant was added in order to capture the maximum integration time suggested by Green et al. (1957).

(2) The envelope model is similar to the off-frequency listening model of Leshowitz and Wightman (1971) except that the "unsigned transient" model used here is sensitive to off-frequency energy in the spectrum of the envelope rather than

the spectrum of the carrier. Oxenham (1998) has suggested that when all of the energy is confined to a single critical band, detection is based on changes in the frequency spectrum of the envelope between 70 and 150 Hz without regard to sign. That is, with equal to sensitivity to changes in the envelope whether they are due to increases or decreases in level.

(3) In the "equal weighting" model, on-frequency and off-frequency energy are treated as independent sources of information. Performance is predicted to be the square-root of the mean of the squared predictions ("RMS") of the first two models.

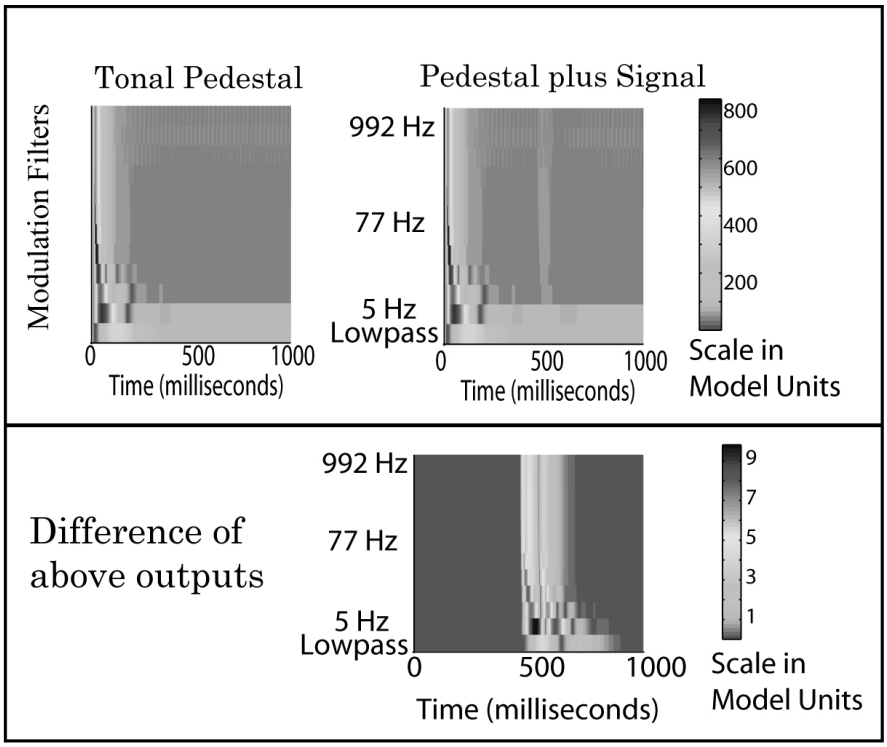


Fig. 2. The stimuli in Fig. 1 as processed by the Dau modulation filterbank model.

(4) The Dau "modulation filterbank" model (Dau, Kohlrausch and Kollmeier 1997a; 1997b) has seldom been applied to increment detection, but it proves to be a good predictor. In this model, the envelope is extracted and then passed through a bank of twelve "modulation" filters with center frequencies distributed equally (on a log scale) between 2.5 and 997 Hz. The difference is used to predict listener performance. The output of the filters is greater whether the duration is longer or the envelope change is more rapid. See Fig. 2.

(5) The fifth model is the Shamma modulation filterbank model. As described in Chi, Gao, Guyton, Ru, and Shamma (1999), this model consists of five modulation filter-banks each with a characteristic spectral bandwidth. The model,

initially proposed by Shamma, Chadwick, Wilbur, Morrish, and Rinzel (1986) was inspired by the spectro-temporal representation of sound that best characterizes a variety of mammalian auditory-cortical receptive fields.

Model predictions were generated by a least-squares fit to the average data of human listeners. These adjustments give the models the best chance to predict the data and are necessary because the output of the models is in model units rather than performance units.

3 Experiment 1: Detection of signals with various durations

Detection thresholds were derived from five-point psychometric functions obtained for five levels of two types of signals. The first signal type ("single-burst") always began and ended with 8 ms cosine-squared ramps. Total durations varied between 17 ms and 601 ms. The second signal type ("multiple-burst") was composed of one, two, four, eight or sixteen repetitions of a 20-ms tone-burst with 10-ms onsets and offsets, resulting in total signal durations that varied between 20 and 320 ms. All signals were 477-Hz tones added in-phase to 477-Hz tonal pedestals of one second total duration presented at 60 dB SPL. Examples of the two types appear in Fig. 3. Data from an experiment using 8-ms ramps were used in order to allow longer durations to be compared. Performance was the same as with 10-ms ramps for durations up to 85 ms.

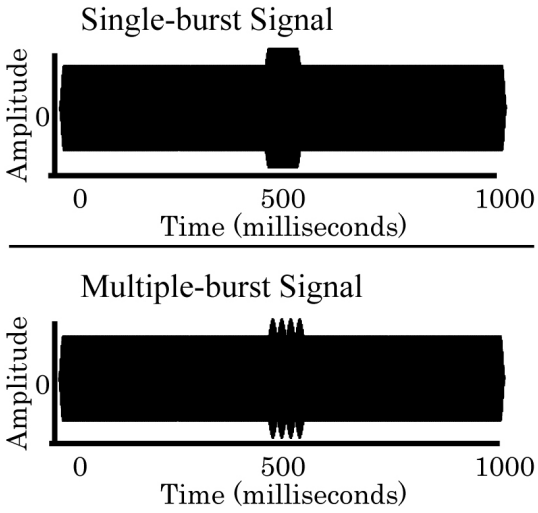


Fig. 3. Examples of the signal plus pedestal stimuli presented in Experiment 1.

The reason for comparing single and multiple-burst signals is that the "unsigned transient" model, based as it is on high-frequency energy in the envelope, predicts equal detectability for the single-burst signals but increasing detectability with duration for the multiple-burst. The modulation filterbank models predict similar performance for the two signal types.

Not shown here are additional data with asymmetrical signals that demonstrate the equivalent detectability of fast-onsets and fast-offsets. It was on the basis of that data set that the high-frequency envelope spectrum model (Oxenham 1998) was chosen.

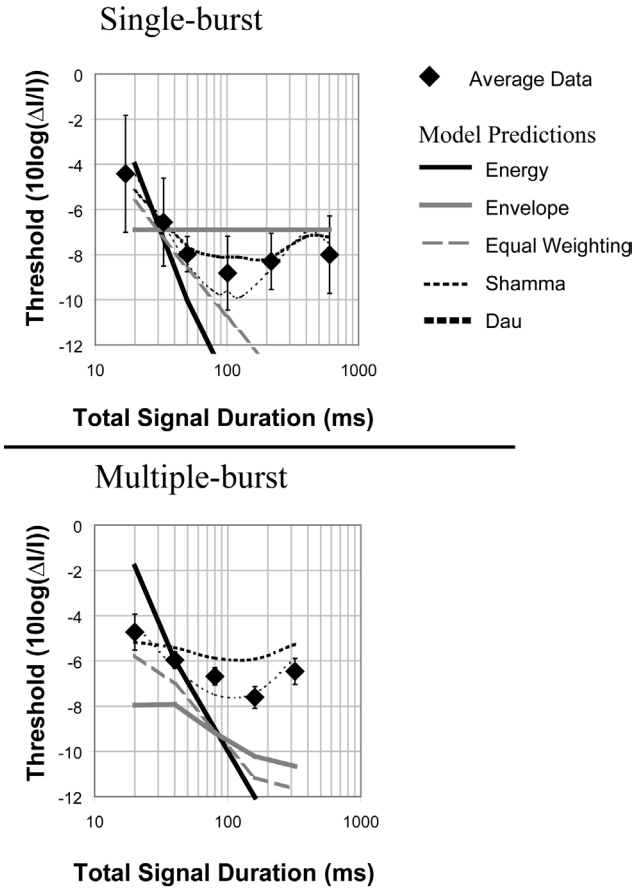


Fig. 4. Thresholds and model predictions are graphed for two types of duration-varying signals. Of the models tested, the modulation filterbanks give the closest fit to the data.

4 Results and modeling

Figure 4 shows the results and the predictions of the five models for the single-burst and multiple-burst signals. The ordinate represents thresholds (d′=1) in units of the Weber fraction (10logΔI/I). Clearly, the improvement with duration is less than that predicted by the energy model with a 200-ms temporal integration window. A 100-ms window would improve the fit for the single-burst signals but not the multiple-burst. Conversely, for the single-burst signals, the envelope model fails by predicting no change in performance, while for the multiple-burst it fails by predicting too much improvement with additional bursts. Consequently, the equal weighting model fails at long durations. Only the two modulation filterbank

models predict the appropriate functions. Note that thresholds in the multiple burst condition are somewhat higher due to the reduction in total energy for a given duration.

5 Experiment 2: Detection in the presence of AM maskers

The modeling from Experiment 1 suggested that listeners were using a mechanism that could readily be modeled as a modulation filterbank. As a further test of this hypothesis, detection was measured again, but this time in the presence of amplitude-modulated maskers. Performance was compared to a condition where the masker was an unmodulated tone whose peak level was the same as for the modulated maskers. Since the unmodulated masker had more energy than the modulated maskers, the energy model predicts that an unmodulated masker would cause the worst performance. The envelope spectrum model predicts worst performance for the condition where the masker is a tone modulated at a high frequency. Both modulation filterbank models predict the greatest masking for low-frequency modulated maskers. This is illustrated in Fig. 2 where it can be seen that the greatest activity in the model is in the low-frequency modulation filters. Additionally, both models predict interference even if the modulated tone has a carrier frequency that is in a different critical band than the pedestal carrier, similar to studies of modulation masking using a modulation detection task (Yost and Sheft, 1989). Because the Dau model is based on psychophysical results and the Shamma model on neurophysiology, they differ in the manner in which carrier frequencies are combined. This and other differences lead to divergent predictions of the pattern of interference that should be found as modulation rate is varied.

The pedestal was, again, a 477-Hz tone of one second duration. The added signal was a 477-Hz tone added in-phase with the pedestal. Signals were 85 ms in total duration with 40 ms raised-cosine onsets and offsets. Their peak levels were a fixed -1.09 dB and -3.85 dB with the 477-Hz and 2013-Hz maskers, respectively. Maskers were tones either unmodulated or amplitude-modulated to a depth of 80% with carrier frequencies of either 477 Hz or 2013 Hz. Both pedestal and maskers were presented at a peak level of 60 dB SPL. Signal levels were set to produce detection scores of about $d' = 2$ in the presence of the unmodulated masker. For the 477-Hz masker, the modulation rates were 4, 48 and 96 Hz; for the 2013-Hz masker, 4, 8, 12, 16, 24, 32, 48, 64 and 96 Hz.

6 Results and modeling

Figure 5 shows d' values with the 477-Hz and 2013-Hz maskers, as well as predictions from the modulation filterbank models. Note that with a signal level of -1.09 dB and no masker, detection was unplottably high (*). Modulation masking is defined as a decrease in performance compared to with an unmodulated masker. Peak levels of 60 dB in the modulated maskers ensured that they had less energy than the unmodulated masker. The energy model predicts that none of the modulated maskers should cause more interference than the unmodulated masker.

This was never true. It is clear, then, that an energy model cannot explain these data. With modulated maskers, both on-frequency and off, the most effective modulation was 4 Hz. We also see that for the 2013-Hz masker, there was no masking with an unmodulated masker, showing the traditional result with critical bands, where on-frequency masking is more effective than a masker at a remote frequency. What is new is that when the task is to detect a tonal signal added to a tonal pedestal, a modulated masker is effective whether within the critical band of the signal or not, albeit less effective when the carrier is remote.

Unlike the first experiment, predictions of the modulation filterbank models are accurate only in that they predict the modulation masking. What is more, even a

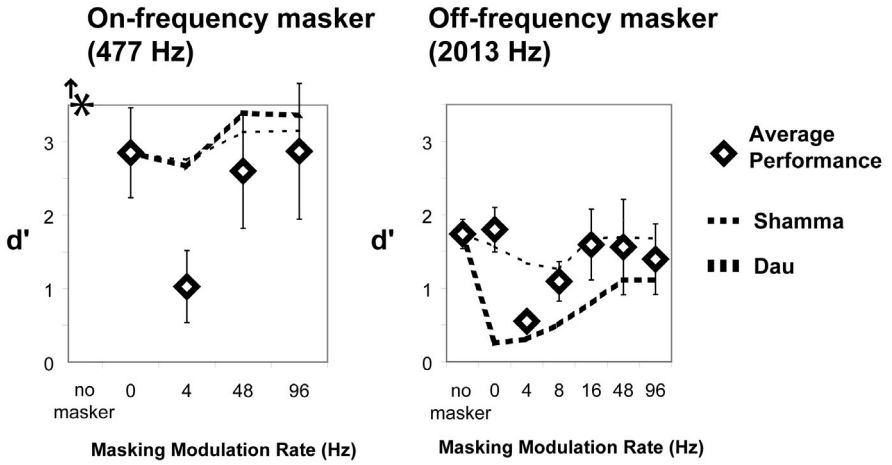


Fig. 5. The effects of masking with 80% amplitude-modulated tones. Models are from Dau and Verhey (1999) and Chi et al. (1999). See text for details. Signal levels were $10\log\Delta I/I$ = -1.09 dB for the 477 Hz and -3.85 dB for the 2013 Hz. (The * refers to perfect performance.)

qualitative description of the off-frequency masking in the Dau model requires expansion as suggested in Dau and Verhey (1999) to include two filterbanks, one for each critical band, whose outputs are summed. With this prediction, the pattern of modulation masking is predicted well but the unmodulated masker is predicted to be just as effective when it is at 2013 Hz as when it is at 477 Hz. The Shamma model escapes this faulty prediction, but does so by sacrificing precision in its modulation masking predictions. Despite these failings, the central insight of these models is preserved: detection does seem to rely on a mechanism especially tuned to low-rate fluctuations in energy over time. Additionally, it is clear from the results with the 2013-Hz masker that whatever this mechanism is, it must be sensitive to modulation across critical bands while still being immune to the masking effects of an unmodulated masker at a distant frequency.

7 Summary

When listeners are asked to detect an increment in the intensity of a tone, their performance can be predicted by assuming that they are sensitive to slow changes in the energy of the envelope. The use of this model was validated by an experiment in which the detection of increments was impaired by the presentation of amplitude-modulated maskers. Though the modulation filterbank models are not able to predict the data perfectly, they capture the central findings. Future work aspires to improve the value of modulation filterbank models for this paradigm by measuring bandwidths and shapes of the putative modulation filters.

Acknowledgments

This research was supported by a grant from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, DCD00087.

References

Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. J. Acoust. Soc. Am., 106(5), 2719-2732.

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997a). Modeling auditory processing of amplitude modulation: I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am., 102(5), 2892-2905.

Dau, T., Kollmeier, B. and Kohlrausch, A.(1997b). Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration. J. Acoust. Soc. Am., 102(5), 2906-2919.

Dau, T. and Verhey, J. L. (1999). Modeling across-frequency processing of amplitude modulation. In: T. Dau, V. Hohmann, and B. Kollmeier (eds.), *Psychoacoustics, Physiology and Models of Hearing.*World Scientific, Singapore, pp. 229-234.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: Wiley.

Green, D. M., Birdsall, T. G., & Tanner, W. P. J. (1957). Signal detection as a function of signal intensity and duration. J. Acoust. Soc. Am., 29, 523-531.

Leshowitz, B., & Wightman, F. L. (1971). On-frequency masking with continuous sinusoids. J. Acoust. Soc. Am.,, 49(4), 1180-1190.

Oxenham, A. J. (1997). Increment and decrement detection in sinusoids as a measure of temporal resolution. J. Acoust. Soc. Am., 102(3), 1779-1790.

Oxenham, A. J. (1998). Temporal integration at 6 kHz as a function of masker bandwidth. J. Acoust. Soc. Am., 103(2), 1033-1042.

Shamma, S., Chadwick, R., Wilbur, J., Morrish, K. and Rinzel, J. (1986). A biophysical model of cochlear processing: Intensity dependence of pure tone responses. J. Acoust. Soc. Am., 80(1), 133-145.

Yost, W.A. and Sheft, S. (1989). Across critical band processing of amplitude modulated tones. J. Acoust. Soc. Am. 85, pp. 848-857.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Minimum integration times for processing of amplitude modulation

Stanley Sheft and William A. Yost

Parmly Hearing Institute, Loyola University Chicago {ssheft,wyost}@luc.edu

1 Introduction

In the spectral domain, the ability to discriminate short-term spectral variation has been used to estimate a minimum integration time for auditory processing (Green 1973). One estimation approach involves discrimination between transient stimuli with identical energy but differing phase spectra (Patterson and Green 1970; Green 1973). For these stimuli, discrimination ability decreases with stimulus duration. The duration at which discrimination is no longer possible suggests integration within a single time window, preventing resolution of stimulus dynamics. Though shorter values have been obtained (see Henning and Gaskill 1981), measures of minimum integration time are generally consistent with estimates of temporal resolution derived from the temporal modulation transfer function (TMTF; Viemeister 1979; Sheft and Yost 1990). Direct estimation of temporal resolution from the TMTF associates thresholds for detecting amplitude modulation (AM) with the time constant of a single lowpass filter. More recent models for AM processing replace the lowpass filter with a modulation filterbank (Dau, Kollmeier, and Kohlrausch 1997; Sheft and Yost 2001; Ewert, Verhey, and Dau 2002). For filterbank models, a single time constant or minimum integration time is an incomplete descriptor.

The present study utilized a modulation-domain analog of the procedure used to measure minimum integration time in the spectral domain. The current procedure measured as a function of duration the ability to discriminate stimuli whose modulators differed only in terms of their phase spectra. Patterson and Green (1970; Green 1973) demonstrated that measures of minimum integration time are not established by the constraints of peripheral auditory filtering. The present work attempted to evaluate potential constraints set by filtering in the modulation domain. By varying modulator bandwidth, the intent was to determine if results are consistent with the multiple time constants of modulation-filterbank models.

2 Method

A practical motivation for the present work was evaluation of the role of the modulation phase spectrum in signal discrimination for subsequent application to automated signal recognition. Therefore, a class of minimum-phase modulators was chosen for the signal standard. A common method of generating signals that differ only in phase is through time reversal. Performance was thus measured for discriminating whether two wideband-noise (WBN) carriers had been modulated by the same function or if the modulator of one had been time reversed. Additional conditions were also run in which maskers were added to the modulators.

Modulators were minimum-phase reconstructions of samples of lowpass Gaussian noise. Modulator bandwidth varied from 4 to 2048 Hz with duration ranging from 15.625 to 500 ms. Reconstruction was based on reverse-cepstral analysis (Oppenheim and Shafer, 1975). As an analog to autocorrelation in log space, cepstral analysis of stochastic modulators concentrates energy near signal onset with the dominance of the onset increasing with modulator bandwidth (see Fig. 1). In the masking conditions, a second noise band was added to the minimum-phase modulator with masker bandwidth and lower cutoff frequency as separate parameters. Independent masker bands were generated for each stimulus presentation.

Modulators were normalized through several stages. First, the dc level was removed and rms amplitude was normalized for the signal band, and also masker band if present. The modulator was then dc shifted for a minimum amplitude of zero. The carrier for each stimulus presentation was an independent sample of Gaussian noise. Following carrier modulation, stimuli were shaped with 1-ms cosinusoidal rise/fall times and lowpass filtered at 12 kHz. Overall stimulus level was fixed at 70 dB SPL.

Performance was measured with a cued single-interval procedure. The cue was a modulated sample of WBN. The task was to determine whether the stimulus presented during the observation interval repeated the cue modulator or if the modulator was reversed in time. A different cue preceded each trial. Data were collected from six listeners with performance levels based on 400 trials per condition for each listener.

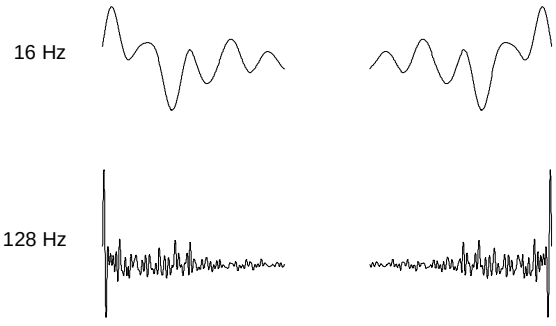


Fig. 1. Illustration of two minimum-phase modulators paired with their time reversal. Modulator bandwidth was 16 Hz for the top row and 128 Hz for the bottom.

3 Results

Figure 2 shows mean discrimination ability as a function of stimulus duration. Results indicate an interaction between modulator bandwidth and duration. For bandwidths of 512 Hz and less, performance was best at intermediate durations. With larger bandwiths, discrimination improved monotonically with duration.

The decline in discrimination ability with decreasing duration over some range is consistent with the notion of minimum integration times. However, results do not follow constraints anticipated by parameters used in modulation-filterbank models. Derived from masking data, the models assume that modulation-filter bandwidth increases with center frequency. If performance were limited by integration set by filter time constants, discrimination ability would improve with increasing bandwidth. In that this trend was not obtained, results do not indicate minimum integration times associated with filterbank parameters. The goal then

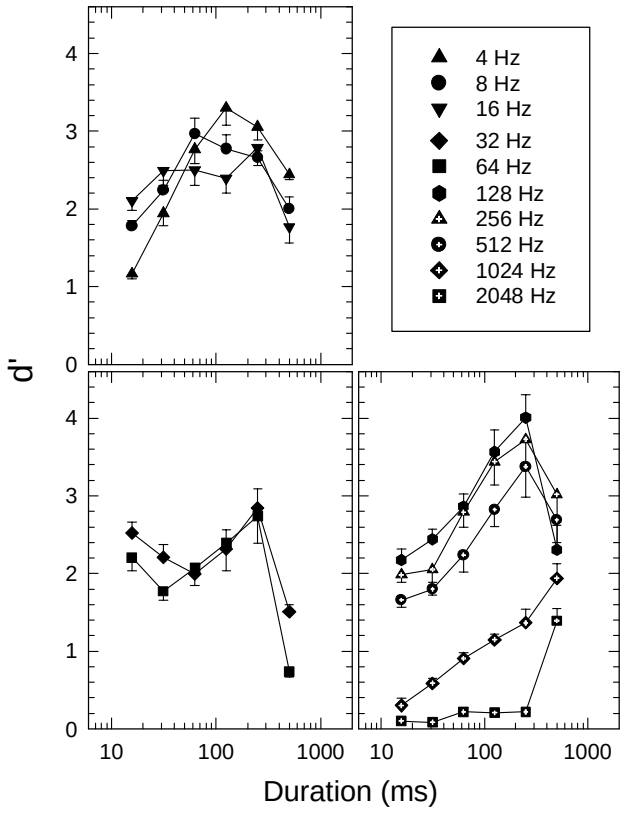


Fig. 2. Mean discrimination ability as a function of stimulus duration. The parameter is modulator bandwidth. The separate panels display results obtained with a different range of bandwidth values. Error bars represent the standard error of mean performance levels.

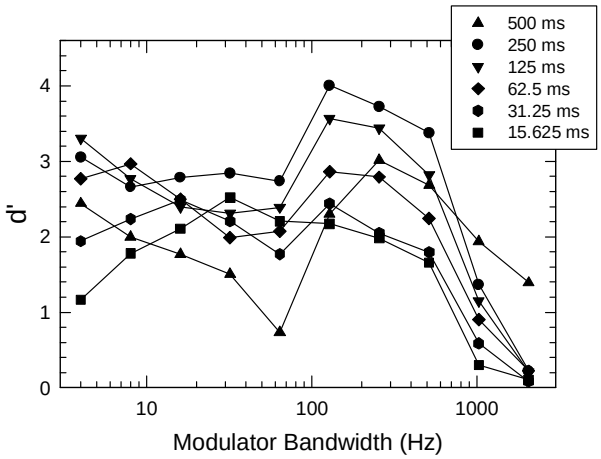


Fig. 3. Data from Fig. 2 replotted as a function of modulator bandwidth with duration the parameter.

becomes estimating aspects of modulation processing consistent with the data set.

Across duration, performance was generally best at intermediate values of modulator bandwidth with discrimination ability falling sharply once bandwidth exceeded 512 Hz (Fig. 3). The nonmonotonic psychometric functions of Fig. 3 follow two general shapes. Except at the two shortest durations, performance initially declined with increasing bandwidth then improved with further increment. The initial lowpass segment is consistent with results obtained in a previous study evaluating the discriminability of phase randomization of lowpass-noise modulators (Sheft and Yost 2002). With a decision statistic based on cross-correlation, modulation-filterbank simulations were able to account for the lowpass result. As discussed in Sec. 2, the prominence of the onset of minimum-phase modulators increases with bandwidth. The improvement in discrimination ability at moderate modulator bandwidths suggests involvement of a second cue derived from onset/offset characteristics. Discrimination based on an interaction between two cues precludes direct estimation of minimum integration times from the current results.

The short-duration results of Fig. 3 show performance initially improving then declining as a function of modulator bandwidth. At short durations and narrower bandwidths, the correlation between the ac component of a modulator and its time reversal is inversely related to bandwidth. For a correlation receiver, d' is inversely related to the correlation between signals. The highpass segment of the short-duration functions thus follows the predictions derived from a decision statistic based on cross correlation. The lowpass segment of the short-duration results may then reflect the same process that leads to the drop in performance at all durations with the broader modulator bandwidths.

In the masking conditions, a second noise band was added to the modulators above the lowpass minimum-phase signal bands. Parameters were signal and masker bandwidth, frequency separation between the bands, and stimulus duration. Masker ac-rms amplitude was constant across change in bandwidth. Significant masking was obtained primarily with a narrow signal bandwidth. Initial conditions

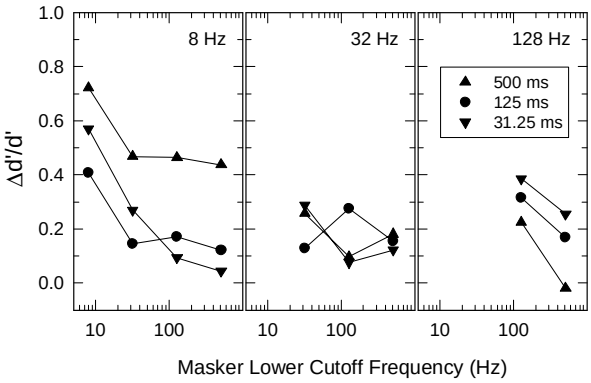


Fig. 4. Relative change in d' due to masking as a function of the lower cutoff frequency of the masker band. The parameter is stimulus duration. Masker bandwidth was one octave. Signal bandwidth is shown at the top of each panel.

used a 500-ms stimulus duration. With an 8-Hz signal bandwidth, the amount of masking decreased with increasing frequency separation between the signal and masker bands, but was independent of masker bandwidth. Small amounts of masking were obtained with a broader signal bandwidth, but only when the signal and masker bands were contiguous in frequency.

Stimulus duration affected masking. Figure 4 shows the relative change in performance due to masking at three durations with an octave masker bandwidth. With an 8-Hz signal bandwidth (left panel), the amount of masking dropped with reduction of stimulus duration from 500 ms. This same pattern was obtained in additional conditions in which instead of an octave bandwidth, masker bandwidth was fixed at 8 Hz across frequency location. As in the initial masking conditions, masking was greatest when the masker was contiguous in frequency with the 8-Hz signal band. At signal bandwidths of 32- and 128-Hz (center and right panels of Fig. 4, respectively), d' changed at most by a factor of 0.4 due to the addition of a masking band. With a 32-Hz signal bandwidth, there was no consistent effect of duration on masked performance levels. Though the extent of masking was less, the effect of duration with a 128-Hz signal bandwidth exhibited a trend opposite that obtained at 8 Hz.

The nonmonotonic effect of modulator bandwidth in conditions without maskers present was taken to suggest use of multiple discrimination cues by observers. At narrow bandwidths, the temporal pattern of the envelope distinguishes signals, while at broader bandwidths, envelope onset/offset characteristics cue discrimination. The masking results also show an effect of bandwidth. Across conditions, masking was generally greatest with a narrow signal bandwidth where temporal pattern is the discrimination cue. The prevalence of masking at a narrow signal bandwidth may then indicate greater lability of the decision statistic used to extract envelope pattern than the one that processes onset/offset characteristics. Without maskers present, there was a nonmonotonic effect of duration at narrow modulator bandwidths with discrimination best at intermediate durations. That with a narrow signal bandwidth masking was greatest at the longest stimulus duration is consistent with the relative performance decrement at this duration without maskers present.

4 Simulations

In conditions with spectral cues unavailable, models of AM processing generally assume uniformity of decision statistic across modulation rate. The present results are interpreted in terms of utilization of two bandwidth-dependent discrimination cues. Recent physiology offers evidence of two coding schemes for AM by the auditory system with slower fluctuations coded by temporal response pattern and faster ones by discharge rate (Schulze and Langer 1997; Lu, Liang, and Wang 2001a, 2001b). For discrimination based on envelope temporal pattern at modulator bandwidths of 64 Hz and less, simulations cross-correlated processed signal pairs. The nonmonotonic effect of duration at these bandwidths was assumed related to a memory constraint modeled through an exponential decay. At greater modulator bandwidths, simulations based discrimination on envelope peak values weighted by their temporal location. Lu *et al.* (2001a) reported separate populations of cortical neurons responsive to either ramped or damped envelope peaks. Current simulations captured this asymmetry through temporal weighting of peak values defined relative to channel variance. Stages of modulation-filterbank processing, based on the work of Dau *et al*. (1997), have been described elsewhere (Sheft and Yost 2001). Relevant model parameters in the present application included compression, loss of envelope-phase information and logarithmic spacing for filters centered above 10 Hz, and a constant internal noise level.

Simulation results followed most major data trends (Fig. 5). In comparison to subject data (Fig. 3), the largest discrepancies between actual and simulated performance levels relate to the effect of duration at narrower modulator bandwidths. Modeling memory limitation by an exponential decay is a rough approximation. More importantly, it is unlikely that the auditory system cross-correlates stimulus representations on individual point (or spike) coding. More likely, the system keys off of distinctive features of stimuli for cross comparison. Running, and perhaps smoothed, cross-correlation triggered by envelope peaks might offer a more accurate descriptor of observer performance.

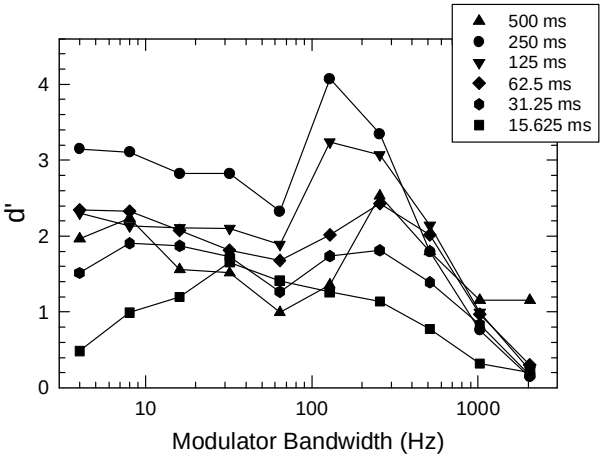


Fig. 5. Model results for the data of Fig. 3.

5 Conclusions

Discrimination of WBN modulated by lowpass minimum-phase signals showed nonmonotonic effects of modulator bandwidth and duration. Major data trends were successfully modeled with processing by a modulation filterbank that extracted decision statistics based on cross correlation and peak envelope values weighted by their temporal location. Better approximation to observer performance may be obtained with modification, based on stimulus features, of the correlation algorithm used at the narrower modulator bandwidths.

Acknowledgments

This work was supported by contract no. F30602-01-C-0039 with the U.S. AFRL.

References

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892-2905.

Ewert, S.D., Verhey, J.L. and Dau, T. (2002) Spectro-temporal processing in the envelope-frequency domain. J. Acoust. Soc. Am. 112, 2921-2931.

Green, D.M. (1973) Minimum integration time. In: A.R. Møller (Ed.), *Basic Mechanisms in Hearing.* Academic Press, New York, pp. 829-843.

Henning, G.B. and Gaskell, H. (1981) Monaural phase sensitivity with Ronken's paradigm. J. Acoust. Soc. Am. 70, 1669-1673.

Lu, T., Liang, L. and Wang, X. (2001a) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. J. Neurophysiol. 85, 2364-2380.

Lu, T., Liang, L. and Wang, X. (2001b) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nature Neurosci. 4, 1131-1138.

Oppenheim, A.V. and Schafer, R.W. (1975*) Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, New Jersey.

Patterson, J.H. and Green, D.M. (1970) Discrimination of transient signals having identical energy spectra. J. Acoust. Soc. Am. 48, 894-905.

Schulze, H. and Langer, G. (1997) Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm? J. Comp. Physiol A 181, 651-663.

Sheft, S. and Yost, W.A. (1990) Temporal integration in amplitude modulation detection. J. Acoust. Soc. Am. 88, 796-805.

Sheft, S. and Yost, W.A. (2001) AM detection with interrupted modulation. In: D.J. Breehaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs and R. Schoonhoven (Eds.), *Physiological and Psychophysical Bases of Auditory Function.* Shaker Publishing, Maastricht, pp. 290-297.

Sheft, S. and Yost, W.A. (2002) Envelope phase-spectrum discrimination. AFRL Prog. Rep. 2, contract no. SPO700-98-D-4002.

Viemeister, N.F. (1979) Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Am. 66, 1364-1380.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Neural mechanisms for analyzing temporal patterns in echolocating bats

Ellen Covey[1] and Paul A. Faure[2]

Department of Psychology, University of Washington, Seattle, WA 98195-1525, USA
([1]ecovey@u.washington.edu; [2]paul4@u.washington.edu)

1 Introduction

Echolocation in bats is a specialized auditory behavior. Nevertheless, the perceptual processes and underlying neural mechanisms are probably similar to those used by all mammals to analyze complex sound sequences. These processes include linking neural representations of sounds that occur at different times to generate predictions of what sounds should occur next, or to reinterpret a specific sound based on subsequent information. Backward and forward masking are simple forms of interaction between sounds that occur at different times. Temporal masking has been studied extensively in psychophysical experiments, but little is known about the underlying neural mechanisms. We have used a neural analog of temporal masking to study the characteristics of neural inhibition and temporal processing in a specialized population of midbrain auditory neurons in the big brown bat, *Eptesicus fuscus.*

The mammalian auditory brainstem comprises multiple, parallel pathways that converge at the midbrain, in the inferior colliculus (IC). The IC also receives descending input from the auditory cortex and other brain regions, and transmits information to motor control systems of the cerebellum and superior colliculus. Hence, the IC is a complex integrative center for auditory information processing (for a review, see Casseday, Fremouw and Covey 2002). Because the time to evoke a neural response to a given sound is longer in some pathways than in others (*e.g.* Haplea, Covey and Casseday 1994), it is possible to create neural circuits in the IC that compare different sounds or sound elements through coincidence detector mechanisms. Many IC neurons respond selectively to specific temporal features of sound such as the time between two signals, the rate, depth and direction of a frequency modulated (FM) sweep, or the duration of a sound. Previous studies have shown that some forms of response selectivity are created through the convergence of excitatory and inhibitory synaptic potentials (*e.g.* Casseday, Ehrlich and Covey 1994; Fuzessery and Hall 1996, 1999). These inputs produce complex temporal sequences of neural excitation and inhibition that may peak at the same or different times (Covey, Kauer and Casseday 1996). A good example of such temporal interactions can be seen in the responses of neurons tuned to signal duration.

2 Mechanisms underlying duration tuning

The midbrain is the first stage in the central auditory pathway where duration tuned neurons have been found, suggesting that they are created there. Blocking the inhibitory neurotransmitters GABA or glycine alters the spike count and latency of many IC cells (*e.g.* Park and Pollak 1993), and abolishes duration tuning (Casseday, Ehrlich and Covey 1994; 2000). This indicates that inhibition is necessary for duration tuning, and suggests that the inhibition has a different time course and latency from that of the excitation. Whole-cell patch clamp recordings support this idea and have led to a conceptual model of how duration selective neurons are created in the IC (Fig. 1).

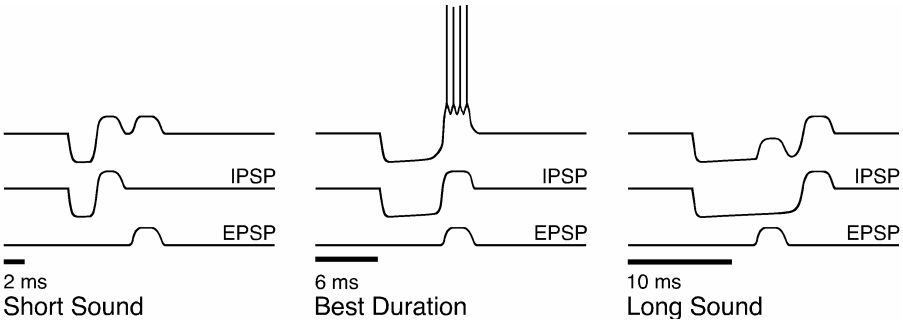


Fig. 1. Conceptual model for the creation of bandpass duration tuning. Three traces are shown above each stimulus (solid bars). The middle and bottom traces represent hypothetical synaptic inputs to an IC neuron (IPSP and EPSP, respectively); the top trace illustrates the resulting change in the cell's membrane potential and, if suprathreshold, its spike output. The model illustrates how the coincidence of two excitatory events—onset-evoked EPSP plus post-inhibitory rebound—interact to push the membrane potential of the cell above spike threshold for a sound at best duration (6 ms), but fail to provide the necessary coincidence of subthreshold excitation at shorter (2 ms) and longer (10 ms) durations. The model also works for shortpass duration tuning. Modified from Faure, Fremouw, Casseday and Covey (2003).

The model has three components, each of which likely represents the effects of summed, multiple synaptic inputs to a neuron: (1) a transient, onset-evoked, subthreshold excitatory post-synaptic potential (EPSP); (2) a sustained, onset-evoked, inhibitory post-synaptic potential (IPSP) with a latency no longer than that of the EPSP; and (3) transient, offset-evoked excitation, which may be due to rebound from the sustained inhibition. A neuron fires action potentials whenever the signal duration is such that two excitatory events (1 and 3) coincide. The cell fails to spike when the sound duration is so short that the offset-evoked excitation occurs before the onset-evoked EPSP arrives, or when it is so long that the sustained IPSP overrides the EPSP. The model predicts that a cell's best duration, range of duration selectivity, and duration filter characteristic are controlled, in part, by the amount of time by which inhibition precedes excitation. For details on this model, see Ehrlich, Casseday and Covey (1997) and Faure, Fremouw, Casseday and Covey (2003).

3 Paired-tone stimulation characterizes the inhibition

Duration tuned neurons provide an excellent opportunity to study the effect of one stimulus upon another when the two occur at different times. The number and timing of action potentials in response to a best duration (BD) stimulus can be used as a probe to measure the effective strength and time-course of the inhibition evoked by a second stimulus set to a non-excitatory (NE) duration. Figure 2 summarizes how such paired-tone stimulation can be used to characterize aspects of neural inhibition in duration tuned neurons. Details are given in Faure *et al*. (2003).

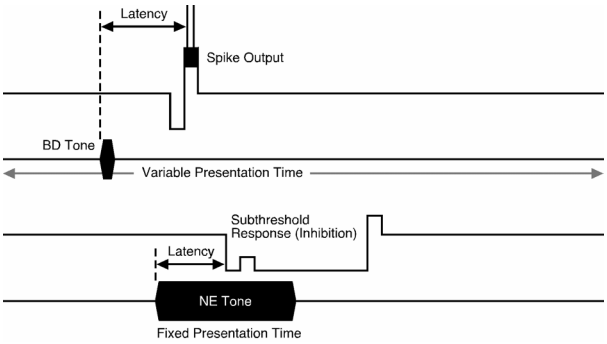


Fig. 2. Suprathreshold excitation (Spike Output) evoked by a best excitatory frequency tone set to a neuron's best duration (BD), was used as a probe to measure the strength and time-course of the inhibition (Subthreshold Response (Inhibition)) evoked by a second tone of the same frequency, but set to a non-excitatory (NE) duration. The presentation time of the NE tone was fixed between trials, whereas the relative timing of the BD tone was randomly varied (gray arrow). Excitatory and inhibitory changes in the cell's membrane potential (derived from Fig. 1) are illustrated as square waves to emphasize the temporal sequence of EPSP and IPSP inputs, and do not reflect the actual amplitudes or the time constants of the subthreshold potentials. Adapted from Faure *et al.* (2003).

The hypothesis was that, whenever the time relationship between the paired stimuli was such that the suprathreshold excitation evoked by the BD tone coincided in time with the inhibition evoked by the NE tone, spikes would be reduced in number and/or altered in their latency. In effect, the NE tone acted as a masker of the responses evoked by the BD tone, while avoiding the complications inherent in presenting a masking stimulus that itself elicits an excitatory response.

4 Characteristics of inhibition in duration tuned cells

Figure 3 shows example results from a duration tuned neuron tested with paired-tone stimulation. When the BD and NE tones were presented monaurally to the ear contralateral to the recording IC (Fig. 3A), spikes were progressively eliminated from the latter part of the response as the offset of the BD tone approached the onset

of the NE tone. The pattern and timing of the spike loss indicates that the inhibition causing the suppression was evoked by the onset of the NE tone. The fact that the suppression began while there was still a gap between the leading BD tone and the lagging NE tone demonstrates that the inhibition evoked by the NE tone had a shorter latency than the excitation evoked by the BD tone. Spikes were suppressed throughout and beyond the duration of the composite stimulus, indicating that the inhibition was sustained, and that it persisted longer than the NE tone evoking it.

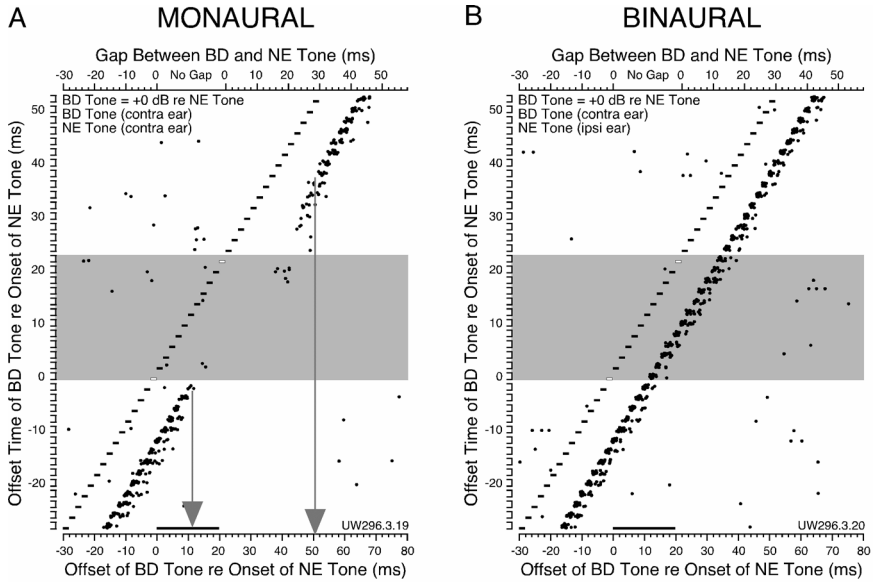


Fig. 3. Dot raster responses of a shortpass neuron to paired-tone stimulation. (A) Monaural stimulation. The BD and NE tones were presented to the contralateral ear. When the two tones temporally overlapped they summed to form a single, composite tone with an amplitude pedestal. Gray arrows indicate the calculated start and end times of the inhibition evoked by the NE tone. (B) Binaural stimulation. The BD tone was presented to the contralateral ear; the NE tone was presented to the ipsilateral ear. (A and B) The BD and NE tones were matched in frequency, amplitude and phase. The presentation time of the BD tone was randomized; the onset time of the NE tone was fixed. The NE tone is shown with a long solid bar above the x-axis; the BD tone is shown as a series of vertically stacked short solid bars. BD bars with a white fill indicate intervals when the BD tone was contiguous with but did not overlap the NE tone. Gray box indicates the range of times over which the BD tone was contiguous with or overlapped the NE tone. BD tone duration, 2 ms; NE tone duration, 20 ms; BD and NE tone frequency, 27 kHz; BD and NE tone amplitude, 31 dB SPL; BD tone threshold, 21 dB SPL; 15 trials per stimulus. Data from Faure, Covey and Casseday (unpublished).

The baseline first spike latency of the cell, measured from trials when the BD tone preceded the NE tone and the two tones were well separated in time, was 14.4 ms. For this neuron, a 50% change in baseline spike count was used as the indicator to quantify the effective start and end times of the spike suppression (gray arrows).

The start of the inhibition, and hence its latency, was calculated to occur 11.4 ms after the onset of the NE tone. Thus, inhibition preceded the cell's excitatory first spike latency by 3.0 ms. The end of the inhibition was calculated to occur 50.4 ms after the onset of the NE tone, lasting for a total duration of 39.0 ms. Therefore, inhibition persisted 19.0 ms longer than the duration of the 20 ms NE tone.

We used these methods to quantify the properties of NE tone-evoked inhibition in a population of duration tuned neurons. During monaural stimulation, when the BD and NE tones were equal in amplitude the latency of inhibition was shorter than that of excitation in 32 of 34 neurons tested, with the average latency difference being 3.9 ms (range: -3.3 to 11.9 ms). By testing with NE tones of different duration, we were able to demonstrate that the inhibition evoked by the NE tone was sustained for at least the duration of the stimulus, and usually beyond. Indeed, when the BD and NE tones were equal in amplitude, all 34 neurons showed evidence of persistent inhibition, lasting an average of 14.1 ms (range: 1.7 to 43.7 ms). By testing with BD tones at different relative amplitudes, we were able to show that the excitation evoked by the BD tone could more readily overcome the inhibition evoked by the NE tone when it occurred late in the inhibitory period. This suggests that the inhibition is quite strong at stimulus onset and decays slowly thereafter. For complete details of these results, see Faure *et al.* (2003).

Figure 3B shows paired-tone stimulation results from the same cell in Fig. 3A, only now the NE tone was presented to the ipsilateral ear and the BD tone was presented to the contralateral ear. During binaural paired-tone stimulation, the NE tone was no longer able to suppresses spikes evoked by the BD tone when the two tones were equal in amplitude (Fig. 3B). This suggests that the inhibition responsible for creating duration tuning in the IC is monaural in nature. However, when the level of the ipsilateral NE tone was increased by +20 dB re the BD tone, a reduction in spike number and an increase in first spike latency was observed for some responses evoked by the BD tone during the period of temporal overlap (Faure, Covey and Casseday unpublished). The timing of the suppression and the latency change suggest the hypothesis that binaural or crossed inhibition is different from the monaural inhibition that creates duration selectivity. This hypothesis is currently being addressed with additional paired-tone experiments.

5 Neural inhibition and temporal masking

Using paired-tone stimulation, complex temporal interactions between excitatory and inhibitory inputs to duration tuned neurons were readily observed and quantified (Faure *et al.* 2003). The suppression of sound-evoked responses can be thought of as a neural analog of temporal masking in psychophysics. From the point of view of a single neuron, the BD tone (or pedestal) served as the probe and the NE tone as the masker. We observed neurophysiological equivalents of backward, simultaneous, and forward masking, and obtained evidence that monaural neural inhibition could produce all three phenomena.

Spike suppression when the BD tone preceded the NE tone is equivalent to backward masking and could be attributed to the relatively short latency of the "leading" inhibition evoked by the onset of the NE tone. In backward masking, a

probe signal may be undetectable when it precedes a masker signal by up to 50 ms (Zwicker and Fastl 1990). Leading inhibition has been observed in IC neurons, whether or not they are duration tuned (*e.g.* Carney and Yin 1989), and can last up to 25 ms (*e.g.* Covey *et al.* 1996; Park and Pollak 1993).

Spike suppression when the BD and NE tones overlapped in time to form a single composite stimulus with an amplitude pedestal is equivalent to simultaneous masking and could be attributed to the sustained inhibition evoked by the NE tone.

Spike suppression when the BD tone followed the NE tone is equivalent to forward masking and could be attributed to the persistent inhibition evoked by the NE tone. In forward masking, a probe signal may be undetectable when it follows a masker signal for up to 150 ms (Zwicker and Fastl 1990). Persistent inhibition has been observed in IC neurons, whether or not they are duration tuned (*e.g.* Faingold, Anderson and Caspary 1991; Kuwada *et al*. 1997; Pollak and Park 1993), and can last up to 100 ms (*e.g.* Covey *et al.* 1996; Litovsky and Delgutte 2002; Yin 1994).

In human psychophysics, the time-courses of backward and forward masking are asymmetrical, the time frame of backward masking being approximately one-third that of forward masking. The results of paired-tone stimulation experiments on duration tuned neurons in the mammalian IC reveal that leading and persistent inhibition are also asymmetrical, with the duration of leading inhibition being approximately one-third that of persistent inhibition. The importance of these results with respect to the mechanisms of temporal masking is that leading, sustained, and persistent inhibition suggest a potential mechanism that could produce temporal masking phenomena in the central nervous system.

Finally, it is well established that temporal masking is strongest when the masker and probe signals are presented to the same ear (monaural masking), rather than when the masker and probe are presented to opposite ears (binaural or central masking; Deatherage and Evans 1969; Ingham 1957; Wegel and Lane 1924; Zwislocki, Damianopoulos, Buining and Glantz 1967). Likewise, spike suppression during paired-tone stimulation was strongest when the BD and NE tones were presented to the contralateral ear, whereas binaural stimulation with the NE tone in the ipsilateral ear and the BD tone in the contralateral ear resulted in less suppression (Faure, Covey and Casseday unpublished). Although our interpretation of these results must be considered as preliminary until more neurons have been tested, in general, the difference in spike suppression observed during monaural and binaural paired-tone stimulation parallels the difference in magnitude between monaural and binaural temporal masking in human psychophysics.

Acknowledgements

We thank John H. Casseday and Thane Fremouw for help with data collection and analysis, and Appalachia Martine, Kimberly Miller and Brandon Warren for technical support. Research funded by National Institute of Health Research Grants DC-00607, DC-00287, and Research Core Center Grant DC-04661 from the National Institute on Deafness and Other Communication Disorders.

References

Carney, L.H., Yin, T.C.T. (1989). Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: excitatory and inhibitory components. J Neurophysiol. 62, 144-161.

Casseday, J.H., Ehrlich, D., Covey, E. (1994). Neural tuning for sound duration: role of inhibitory mechanisms in the inferior colliculus. Science 264, 847-850.

Casseday, J.H., Ehrlich, D., Covey, E. (2000). Neural measurement of sound duration: control by excitatory-inhibitory interactions in the inferior colliculus. J. Neurophysiol. 84, 1475-1487.

Casseday, J.H., Fremouw, T., Covey, E. (2002). The inferior colliculus: a hub for the central auditory system. In: D. Oertel, R.R. Fay and A.N. Popper (Eds.), *Integrative Functions in the Mammalian Auditory Pathway*. Springer, New York. pp. 238-318.

Covey, E., Kauer, J.A., Casseday, J.H. (1996). Whole-cell patch-clamp recording reveals subthreshold sound-evoked postsynaptic currents in the inferior colliculus of awake bats. J. Neurosci. 16, 3009-3018.

Deatherage, B.H., Evans, T.R. (1969). Binaural masking: backward, forward, and simultaneous effects. J. Acoust. Soc. Am. 46, 362-371.

Ehrlich, D., Casseday, J.H., Covey, E. (1997). Neural tuning to sound duration in the inferior colliculus of the big brown bat, *Eptesicus fuscus*. J. Neurophysiol. 77, 2360-2372.

Faingold, C.L., Anderson, C.A.B., Caspary, D.M. (1991). Involvement of GABA in acoustically-evoked inhibition in inferior colliculus neurons. Hear. Res. 52, 201-216.

Faure, P.A., Fremouw, T., Casseday, J.H., Covey, E. (2003). Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus. J. Neurosci. 23, 3052-3065.

Fuzessery, Z.M., Hall, J.C. (1996). Role of GABA in shaping frequency tuning and creating FM sweep selectivity in the inferior colliculus. J. Neurophysiol. 76, 1059-1073.

Fuzessery, Z.M., Hall, J.C. (1999). Sound duration selectivity in the pallid bat inferior colliculus. Hear. Res. 137, 137-154.

Haplea, S., Covey, E., Casseday, J.H. (1994). Frequency tuning and response latencies at three levels in the brainstem of the echolocating bat, *Eptesicus fuscus*. J. Comp. Physiol. A. 174, 671-683.

Ingham, J.G. (1957). The effect upon monaural sensitivity of continuous stimulation of the opposite ear. Quart. J. Exp. Psychol. 9, 52-60.

Kuwada, S., Batra, R., Yin, T.C.T., Oliver, D.L., Haberly, L.B., Stanford, T.R. (1997). Intracellular recordings in response to monaural and binaural stimulation of neurons in the inferior colliculus of the cat. J. Neurosci. 17, 7565-7581.

Litovsky, R.Y., Delgutte, B. (2002). Neural correlates of the precedence effect in the inferior colliculus: effect of localization cues. J. Neurophysiol. 87, 976-994.

Park, T.J., Pollak, G.D. (1993). GABA shapes a topographic organization of response latency in the mustache bat's inferior colliculus. J. Neurosci. 13, 5172-5187.

Pollak, G.D., Park, T.J. (1993). The effects of GABAergic inhibition on monaural response properties of neurons in the mustache bat's inferior colliculus. Hear. Res. 65, 99-117.

Wegel, R.L., Lane, C.E. (1924). The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev. 23, 266-285.

Yin, T.C.T. (1994). Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. J. Neurosci. 14, 5170-5186.

Zwicker, E., Fastl, H. (1990). *Psychoacoustics. Facts and Models.* Springer-Verlag, Berlin.

Zwislocki, J.J., Damianopoulos, E.N., Buining, E., Glantz, J. (1967). Central masking: some steady-state and transient effects. Percept. Psychophys. 2, 59-64.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Time-critical frequency integration of complex communication sounds in the auditory cortex of the mouse

Diana B. Geissler and Günter Ehret

Department of Neurobiology, University of Ulm, diana.geissler@biologie.uni-ulm.de

1 Introduction

When the auditory system is stimulated by a stream of acoustic patterns, it has not only to analyze the patterns but also to group together spectral and temporal elements according to their coherence and synchrony in order to detect and recognize acoustic objects and images of potential importance. Psychophysical experiments mainly in humans on auditory streaming and scene analysis (e.g. Bregman 1990; Hirsh and Watson 1996) have shown the importance of spectro-temporal properties of sounds for the discrimination of acoustical images. The neurobiological bases for acoustical object and Gestalt perception are, however, poorly understood. Studies on the primate (including human) auditory cortex indicate that spectrally rich and in the time-domain complex sounds, such as communication sounds and speech, especially activate areas outside the primary auditory fields (e.g. Rauschecker and Tian 2000; Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman and Possing 2000). Hence, activation of these, not of the primary fields, may signal the biological relevance of a sound. On the other hand, the representation of vowels by their formant interaction has been shown for the primary auditory cortex (Ohl and Scheich 1997). These examples demonstrate the need for further studies. Here we show that the biological significance of sounds is differentially represented in higher-order fields of the mouse auditory cortex.

2 Acoustic communication and labeling of auditory cortex

Mice communicate acoustically only with few different call types. One of these are the wriggling calls of pups produced when pushing for the mother's nipples during nursing sessions. Most often, the vocalizing pup emits series of several calls, and only such call series reliably release maternal behavior in their mothers (Ehret 1975; Ehret and Bernecker 1986; Ehret and Riecke 2002). Wriggling calls consist of three (or more) harmonics, often close to 4, 8, and 12 kHz. This formant

spectrum is also decisive for the release of maternal behavior and, thus, together with the rhythm of the repetition, must carry the meaning of the sounds (Ehret and Riecke 2002). The meaning or biological significance of a call series is lost, if the formants are not heard in synchrony, i.e. if, for example, the first formant is started with a lead time of more than 30 ms compared to the higher formants (Geissler and Ehret 2002). This suggests that mother mice must hear a single stream of calls (adequate objects) to recognize them as biologically significant. As far as we know, call recognition is instinctive, so that the mice do not need training for that.

Here, we present series of two synthesized call models to the mothers, one that releases maternal behavior (Fig. 2a, call model A) and one that does not (Fig. 2a, call model G) (Geissler and Ehret 2002). Thus, we stimulate the auditory cortex with two identical patterns in the spectral domain differing only in one time parameter, the lead time of the first formant. The mothers, listening to the call models, are free to show maternal behavior which we record. We then evaluate the auditory cortex for correlates of call perception and recognition.

The division of the auditory cortex of the mouse (Mus musculus) into several fields (Fig. 1a) has been studied by electrophysiological mapping (Stiebler, Neulist, Fichtel and Ehret 1997). Here, we use c-Fos immunocytochemistry to visualize local patterns of neural activation, which has become a widely applied method of neural activity mapping with cellular resolution over large spatial dimensions (e.g. Sagar, Sharp and Curran 1988; Sheng and Greenberg 1990; Ehret and Fischer 1991; Friauf 1992; Calamandrei and Keverne 1994; Chaudhuri 1997). The immunocytochemical protocol used here is similar to that of Reimer (1993) with the modification that the secondary antibody is conjugated with horse radish peroxidase so that a PAP reaction leads to the labeling of Fos-positive cells.

3 Auditory cortical activation: perception vs. recognition

Fos-immunoreactive cells occurred in all auditory cortical fields (Fig. 1a) after stimulation with both call models. In the primary auditory fields AI and AAF of both hemispheres three sectors with a significant increase of Fos-positive cells could be distinguished from the amount of Fos-positive cells in the rest, termed background-labeling in sectors B1, B2, B3 and B4 (Fig. 1b). The sectors of increased activation corresponded very closely to the position of the stimulation frequencies 3.8, 7.6 and 11.4 kHz on the tonotopies of AI and AAF (Stiebler et al. 1997). There was no difference in the average number of Fos-positive cells in AI and AAF (the latter data are not shown in the figure) between the behaviorally relevant (call model A) and the irrelevant (call model G) stimulus (Fig. 2c). That is, we do not see a time-critical frequency integration based on synchronous onsets of frequency components neither in the activation of AI nor of AAF. In other words, the meaning of communication sounds seems not to be explicitly expressed in the general activation of primary auditory cortical fields.

Within the higher-order auditory fields AII and DP, Fos-positive cells appeared in relative constant numbers all over the fields. A quantitative evaluation of labeled cells from 15 sections of the center of both fields, separately for the left and right hemisphere, displayed statistically significant differences between calls that were

perceived without a response and calls that were perceived and responded to with maternal behavior (Fig. 2b). Mothers stimulated with the adequate call model A showed significantly less labeled cells in AII than mothers stimulated with the inadequate call G. In both groups, hemisphere differences in labeling did not occur. In DP, however, call model A led to a significantly higher amount of labeling in the left cortical hemisphere compared to the right side and to both hemispheres of mothers stimulated with the inadequate call G (Fig. 2b). Obviously, the difference in the biological significance of the two call models is seen in the activation of both higher-order cortical fields AII and DP with a left-hemisphere dominance associated with call recognition in DP. The difference in significance between the two call models is signaled here only by one time-critical parameter, the onset synchrony (or asynchrony) of the three frequency components. Since this parameter determines recognition of the spectrum as a relevant call (Geissler and Ehret 2002), time-critical frequency integration takes place in higher-order fields of the auditory cortex.

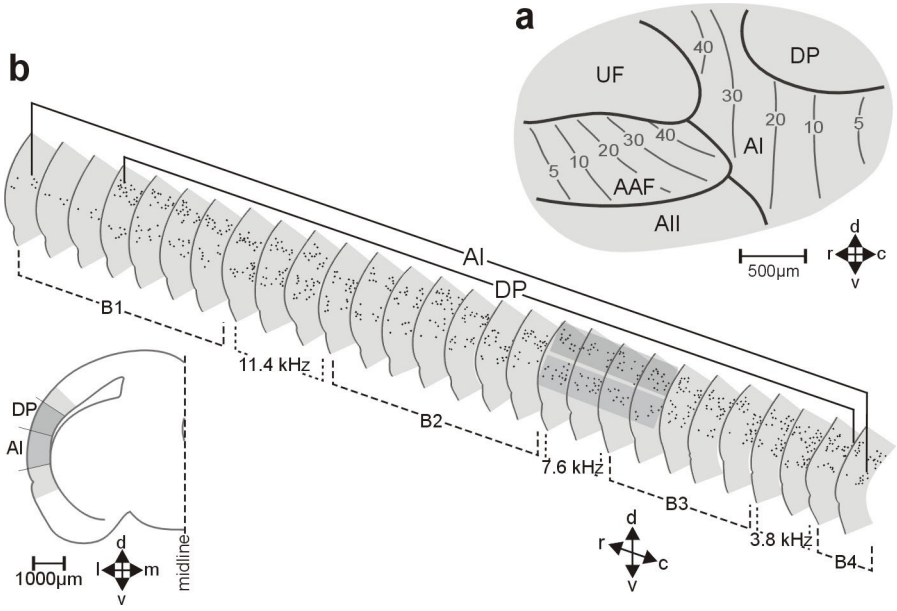


Fig. 1. a: Lateral view of the left-side auditory cortex of the mouse with its field arrangement and the tonotopy in the primary auditory fields AI and AAF (modified from Stiebler et al. 1997). Numbers indicate characteristic frequencies in kHz. **b:** Frontal sections through the left-side neocortex in the area of AI and DP shown as a serial reconstruction. The brain is from a mother stimulated with call model A in response to which she had shown maternal behavior. Fos-positive cells are indicated by dots. AI is divided in seven sectors on the basis of the amount of Fos-positive cells. In B1-4, lower numbers of labeled cells were found compared to the other three sectors in which the stimulation frequencies 3.8, 7.6 and 11.4 kHz are represented. AI, primary auditory field; AII, second auditory field; AAF, anterior auditory field; DP, dorsoposterior field; UF, ultrasonic field; c, caudal; d, dorsal; l, lateral; m, medial; r, rostral; v, ventral.

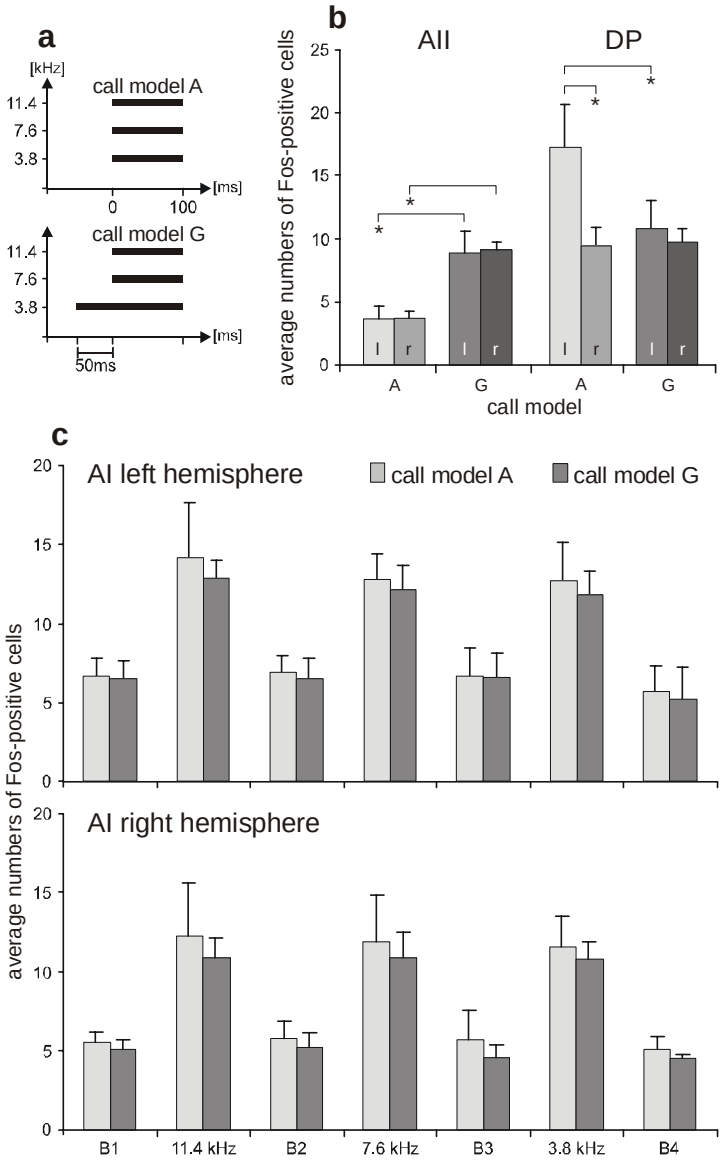


Fig. 2. a: Frequency and time structure of call models A and G (Geissler and Ehret 2002). In call model A, the frequencies 3.8, 7.6 and 11.4 were simultaneously present with a duration of 100 ms. In call model G, the fundamental frequency started 50 ms earlier than the higher harmonics. **b:** Average numbers of Fos-positive cells with standard deviations in AII and DP of both hemispheres (l, left; r, right) of 15 mothers stimulated with call model A or G. Statistically significant differences between the amounts of labeled cells are indicated with *p < 0,001. **c:** Average numbers of Fos-positive cells within the different sectors of AI as defined in Fig. 1b. Labeling is shown for the left and right hemisphere and for call models A and G. No statistically significant differences were found.

4 Discussion

In the primary (AI) and the anterior auditory field (AAF), neither the labeling pattern nor the quantitative evaluation of the number of labeled cells presented evidence that they are involved in wriggling call recognition. Hence, spectro-temporal integration of a stream of acoustic objects (single calls) to an acoustic Gestalt (call series of biological significance) seems not to be associated with the primary fields of the auditory cortex. This is different from the changes of activation patterns in the mouse auditory cortex when mothers recognize ultrasonic calls of their pups compared to virgin females, who do not recognize them (Fichtel and Ehret 1999). In this study, call recognition led to a decrease in the number of Fos-positive cells in the primary cortex (the ultrasonic field UF, in this case) and to an increase in AII compared to perception without recognition. Our present data do not show changes in primary cortex and a *decrease* of labeling in AII associated with recognition (Fig. 2b). The difference between ultrasound and wriggling call recognition and cortical activation may be due to the complexity of the call structure. Ultrasonic calls are recognized by one spectral component only (Ehret and Haack 1982) whereas wriggling calls are recognized by three formants which have to occur simultaneously in a series (spectro-temporal integration). It seems that the fields of the auditory cortex differ in their tasks: primary fields may represent time-varying spectra and the timing of the spectral components, higher-order fields may do the time-critical comparisons between and integrations of the spectral components, i.e. they perform the synthesis of the analyzed sound parameters into a Gestalt.

These ideas are compatible with the functional specializations beautifully demonstrated for the various fields of the mustached bat auditory cortex (e.g. Suga 1988). Neurons in specialized higher-order fields show, for example, response preferences to the distance and relative velocity between the bats and their prey by integrating spectral information of formants and formant transitions over critical time intervals between the echolocation calls and their echoes. Such integrations are not seen in the AI of the bat. Further, our results are compatible with Fos-labeling of auditory cortical fields of the rat when stimulated with a battery of 30 familiar and novel sounds (Wan, Warburton, Kuśmierek, Aggleton, Kowalska and Brown 2001). Labeling in the primary field was similar for novel and familiar sounds. In a higher-order field corresponding to AII of the mouse cortex, novel sounds elicited significantly more labeling than familiar sounds. This is exactly the situation of the mother mice in our tests when they heard the unfamiliar, non-relevant call model G which led to more labeling in AII than the adequate, expected call model A. Finally, our results fit well with results on primates showing that, in general, neurons in higher-order auditory cortical fields prefer complex spectro-temporal acoustic patterns against simple sounds while the reverse is seen in primary fields (e.g. Rauschecker and Tian 2000).

Last but not least, the highly interesting hemisphere difference in labeling in the higher-order field DP only in call-recognizing mothers needs a comment. In most humans, the left hemisphere is specialized for processing and recognition of speech sounds (e.g. Berlin 1977; Benson 1986). The left-hemisphere dominance for the

perception of species-specific communication sounds seems to be a general feature of the mammalian brain (Ehret 2003), present in monkeys (e.g. Beecher, Petersen, Zoloth, Moody and Stebbins 1979), rats (Fitch, Brown, O'Connor and Tallal 1993), and mice (Ehret 1987). Here we show, that the left-hemisphere advantage of sound processing occurs first in a higher-order field of the auditory cortex (DP, Fig. 2b). This is a very convincing demonstration that the left-hemisphere dominance of communication-sound perception is not a property of the primary auditory cortex.

5 Conclusions

A discriminative response (including a left-hemisphere dominance) to communication calls that can be recognized only after time-critical frequency integration is seen in the auditory cortex only in high-order fields, not in the primary fields AI and AAF.

Acknowledgements

Work supported by the Deutsche Forschungsgemeinschaft, Eh 53/8 and 17, 1-3

References

Beecher, M.D., Peterson, M.R., Zoloth, S.R., Moody, D.B. and Stebbins, W.C. (1979) Perception of conspecific vocalizations by Japanese macaques. Evidence for selective attention and neural lateralization. Brain Behav. Evol. 16, 443-460.

Benson, D.F. (1986) Aphasia and the lateralization of language. Cortex 22, 71-86.

Berlin, C.I. (1977) Hemispheric asymmetry in auditory tasks. In: S. Harnard, R.W. Doty, L. Goldstein, J. Jaynes and G. Krauthamer (Eds.) Lateralization in the Nervous System. Academic Press, New York, pp. 303-323.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N. and Possing, E.T. (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512-528.

Bregman, A.S. (1990) Auditory Scene Analysis. MIT-Press, Cambridge, MA.

Calamandrei, C. and Keverne, E.B. (1994) Differential expression of fos-protein in the brain of female mice dependent on pup sensory cues and maternal experience. Behav. Neurosci. 108, 113-120.

Chaudhuri, A. (1997) Neural activity mapping with inducible transcription factors. NeuroReport 8, 3-8.

Ehret, G. (1975) Schallsignale der Hausmaus (Mus domesticus). Behaviour 52, 38-56.

Ehret, G. (1987) Left hemisphere advantage in the mouse brain for ultrasound recognition in a communicative context. Nature 325, 249-251.

Ehret, G. (2003) Hemisphere dominance of brain function – which functions are lateralized and why? In: L. van Hemmen, T.J. Sejnowski (Eds.) 23 Problems on Systems Neuroscience. Oxford University Press, New York, in press.

Ehret, G. and Bernecker, C. (1986) Low-frequency sound communication by mouse pups (Mus musculus): wriggling calls release maternal behaviour. Anim. Behav. 34, 821-830.

Ehret, G. and Fischer, R. (1991) Neuronal activation and tonotopy in the auditory system visualized by c-fos gene expression. Brain Res. 567, 350-354.

Ehret, G., Haack, B. (1982) Ultrasound recognition in house mice: Key-stimulus configuration and recognition mechanism. J. Comp. Physiol. 148, 245-251.

Ehret, G. and Riecke, S. (2002) Mice and humans perceive multiharmonic communication sounds in the same way. Proc. Natl. Acad. Sci. USA 99, 479-482.

Fichtel, I. and Ehret, G. (1999) Perception and recognition discriminated in the mouse auditory cortex by c-Fos labeling. NeuroReport 10, 2341-2345.

Fitch, R.H., Brown, C.P., O'Connor, K. and Tallal, P. (1993) Functional lateralization for auditory temporal processing in male and female rats. Behav. Neurosci. 107, 844-850.

Friauf, E. (1992) Tonotopic order in the adult and developing auditory system of the rat as shown by c-fos immunocytochemistry. Eur. J. Neurosci. 4, 798-812.

Geissler, D.B. and Ehret, G. (2002) Time-critical integration of formants for perception of communication calls in mice. Proc. Natl. Acad. Sci. USA 99, 9021-9025.

Hirsh, I.J. and Watson, C.S. (1996) Auditory psychophysics and perception. Annu. Rev. Psychol. 47, 461-484.

Ohl, F.W. and Scheich, H. (1997) Orderly cortical representation of vowels based on formant interaction. Proc. Natl. Acad. Sci. USA 94, 9440-9444.

Rauschecker, J.P. and Tian, B. (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. USA 97, 11800-11806.

Reimer, K. (1993) Simultaneous demonstration of Fos-like immunoreaktivity and 2-deoxy-glucose uptake in the inferior colliculus of the mouse. Brain Res. 616, 339-343.

Sagar, S.M., Sharp, F.R. and Curran, T. (1988) Expression of c-fos protein in brain: metabolic mapping at the cellular level. Science 240, 1328-1331.

Sheng, N. and Greenberg, M.E. (1990) The regulation and function of c-for and other immediate early genes in the nervous system. Neuron 4, 477-485.

Stiebler, I., Neulist, R., Fichtel, I. and Ehret, G. (1997) The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation. J. Com. Physiol. A 181, 559-571.

Suga, N. (1988) Auditory neuroethology and speech processing: complex-sound processing by combination-sensitive neurons. In: G.M. Edelman, W.E. Gall, W.M. Cowan (Eds.) Auditory Function: Neurobiological Bases of Hearing. John Wiley, New York, pp. 679-720.

Wan, H., Warburton, E.C., Kuśmierek, P., Aggleton, J.P., Kowalska, D.M., Brown, M.W. (2001) Fos imaging reveals differential neuronal activation of areas of rat temporal cortex by novel and familiar sounds. Europ. J. Neurosci. 14, 118-124.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Transformation of stimulus representations in the ascending auditory system

Israel Nelken[1,2], Nachum Ulanovsky[1], Liora Las[1], Omer Bar-Yosef[1], Michael Anderson[4],Gal Chechik[3], Naftali Tishby[2,3], and Eric Young[4]

[1] Dept. Of Physiology, Hebrew University – Hadassah Medical School, Jerusalem, Israel, {Israel, nachumu, lioraa, omerbary}@md.huji.ac.il
[2] Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel
[3] Institute for Computer Science, Hebrew University, Israel, {ggal, tishby}@cs.huji.ac.il
[4] Dept. of Biomedical engineering, Johns Hopkins University, Baltimore MD, USA, {anderson, eyoung}@bme.jhu.edu

1 Introduction

The auditory system has the most extensive subcortical component of all sensory systems. Starting from the cochlear nucleus, multiple information streams can be identified by their anatomical sources and targets, by the cellular morphology of the participating neurons, and by their physiological properties (Smith and Spirou 2002). Some of this extensive subcortical processing can be attributed to specific processing requirements of the auditory system, such as the computations of binaural disparities in the superior olive. However, compared with the visual system, the shortest path from the sensory receptors (hair cells) to the cortex has one additional synapse, and most of the parallel pathways starting at the cochlear nucleus are even longer. Furthermore, even in the cochlear nucleus there are neurons with highly complex response properties (Nelken and Young 1994; Spirou and Young 1991). In consequence, any response property described in the auditory cortex could be generated subcortically.

Very few studies have compared responses to complex sounds in auditory cortex and in subcortical stations (Ehret and Merzenich 1988; Ehret and Schreiner 1997; Fitzpatrick, Kuwada, Kim, Parham, and Batra 1999; Miller, Escabi, Read, and Schreiner 2002; Ulanovsky, Las, and Nelken 2003). The purpose of the study presented here was to identify a set of interesting cortical response properties, and study their development along the ascending auditory system. Three stations were chosen: the inferior colliculus (IC), since all lower processing streams converge there; the medial geniculate body (MGB), which is the main thalamic auditory station; and primary auditory cortex (A1).

All the results described here were collected in gas-anesthetized cats. Detailed description of the methods is found elsewhere (Bar-Yosef, Rotman, and Nelken 2002).

2 Results

2.1 Redundancy reduction in the ascending auditory system

Bird vocalizations that consisted mainly of a frequency- and amplitude-modulated tones were extracted from natural recordings (Bar-Yosef *et al.* 2002). These stimuli are relatively simple, consisting of a dominant tonal component and a noise component that is 15-20 dB weaker. The noise component itself contains echoes of the tonal component, which occupy the same frequency band as the tonal component, and a wideband component that is much weaker. The stimuli were decomposed into their basic components, and the responses of neurons to these components and their combinations were tested.

The responses of neurons in cortex and IC to one of the natural sound segments, to its main tonal component, and to the noise component (natural sound minus the tonal component) are shown in Fig. 1. The two pairs had largely overlapping frequency response areas (FRAs, Figs. 1C and G), which included the frequency range of the chirps. In the IC, the total spike count evoked by these stimuli had similar variation in the two neurons: the natural sound and the clean tonal component (Main) evoked similar, large spike counts, whereas the noise component evoked a smaller number of spikes. On the other hand, the cortical neurons responded differently to the different versions: in Fig. 1A, the neuron responds robustly to the natural sound and to the noise component, with very weak responses to the noise. In Fig. 1B, the responses are reversed. The full response profile of the two neurons to a large set of such sounds are shown in Fig. 1D (for the cortical neurons) and Fig. 1H (for the IC neurons). The two profiles are similar in the IC, but are very different in the cortex.

The similarity between the responses of pairs of neurons in IC, MGB and A1 to this set of stimuli was quantified by using the mutual information (MI) between the distributions of the responses. The MI is defined as

$$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \qquad (1)$$

where x and y are the possible values of the two variables, $p(x,y)$ is the joint distribution of the two, and $p(x)$, $p(y)$ are the marginal distributions of X and Y. Here, these variables were taken to be the total spike counts evoked by each stimulus in the two neurons, although we tested a large number of other choices (for example first spike latency) with similar results.

The MI of the joint spike-count distributions can be seen as a measure of informational redundancy between the responses of the two neurons. The higher $I(X,Y)$, the more redundant the neurons are, in the sense that knowing the response of one of them strongly constrains the responses of the other one. The MI does not assume any model for the relationships between the responses of the two neurons, as is implicitly assumed when e.g. computing the linear correlation coefficient. It is similar to some other measures of association between random variables. For example, for two independent variables, it has asymptotically a χ^2 distribution (Cover and Thomas 1991). However, when the two variables are correlated, its

value gives additional information about the strength of the association between the two, information that is difficult to interpret when using e.g. the χ^2 statistic.

Since most neurons were not recorded simultaneously, it was necessary to couple them together in some way in order to estimate redundancy. We chose to couple the neurons by assuming that given the stimulus, the numbers of spikes evoked in the two neurons were independent of each other. Therefore, knowing that one neuron responded to a stimulus with a somewhat higher spike count than its average response to that stimulus does not influence whether the other neuron would simultaneously respond with somewhat higher or somewhat lower spike count relative to the average response to the same stimulus. We checked this assumption in the small number of cases in which we had simultaneous recordings of pairs of neurons, and found that it generally held reasonably well.

In Fig. 1, the spike counts evoked in the two IC neurons were highly dependent: if one neuron responded with a large spike count, the other one was also expected to respond with a large spike count. This happened, in spite of the conditional independence assumption, because the spike counts depended on the stimulus: a large spike count in one neuron occurred when the stimulus efficiently activated this neuron. Since the frequency response areas of the two neurons were largely overlapping, the stimulus had high chances of being also highly effective for the other neuron. Under these circumstances, the MI between the spike counts of the two neurons was relatively high: 0.56 bits.

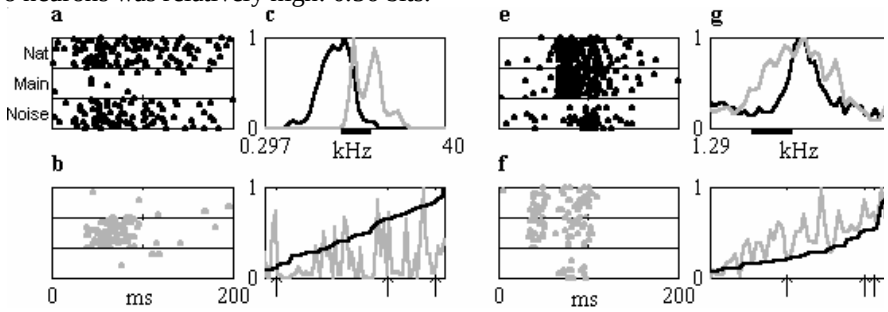


Fig. 1. a and b. Responses of two neurons in A1 to a natural bird chirp (Nat), the clean bird chirp (Main) and the difference between the two (Noise). c. Normalized responses as a function of frequency of the two neurons at the sound level of Main. Black and gray lines correspond to a and b respectively. The thick bar marks the frequency extent of the Main chirp. d. Response profile (normalized mean spike count in response to 64 natural stimuli) of the two neurons, ordered by the rank of the responses of the neuron in a. The arrows indicate the 3 stimuli whose responses are presented in a and b. e-h. Responses of two neurons in IC to the same stimuli. Same conventions as in a-d.

In contrast, in spite of the overlap in frequency response areas of the two cortical neurons in Fig. 1, their responses to the same stimuli could be very different from each other. As a result, knowing that one neuron had a large spike count did not help in guessing how many spikes the other neuron fired. As a result, the MI between the responses of the two was low: 0.014 bits.

The redundancy turned out to be rather large in IC, but small in MGB and A1. The mean redundancy in IC, after normalizing by the size of the mutual information

between neurons and stimuli, was 0.45, whereas in MGB it was 0.04 and in A1 it was 0.07. Furthermore, the redundancy in IC was related to the frequency sensitivity of the neurons: neurons with closer best frequencies tended to be more redundant (regression slope=0.07/oct, n=16, $p<0.01$); this did not occur in MGB and A1 (MGB: slope=0.0008/oct, n=36, n.s.; AI slope=0.009/oct, n=45, n.s.). Thus, neurons in IC with similar BFs were more redundant than neurons in MGB or in A1 with the same BFs, when tested with this set of bird chirps.

2.2 Sensitivity to weak acoustic components

We have previously shown that neurons in A1 are highly sensitive to weak acoustic components. For example, the cortical neurons in Fig. 1 responded to the full natural chirp in the same way that they responded to the noise, although the natural sound included a much stronger acoustical component, the main chirp, that entered their FRA. In fact, even in the population of A1 neurons whose FRA intersected the chirp frequency range, the correlation coefficient between the spike counts evoked by the full natural sound and the spike counts in response to the clean chirp was small, though significant ($r=0.26$, $p<0.01$).

In a substantial number of cases, the similarity between the responses to the noise component and to the natural sound was greater than the similarity between the responses to the main chirp and to the full natural sound. To quantify the similarity in temporal patterns of the responses to the two stimuli, we computed the χ^2 statistic for the difference between the poststimulus time histograms (PSTHs) of the responses to the two stimuli. The histograms were generated with non-uniform time bins, such that the sum of the spike counts evoked by the two stimuli was constant in each bin. The χ^2 statistic was divided by the number of degrees of freedom, resulting in a dissimilarity index (DI). DI values that were much larger than 1 indicated significantly different temporal response patterns. In A1, the DIs between the responses to the clean chirp and to the natural stimulus were more often larger than the DIs between the responses to the noise component and to the natural stimulus (54% of all comparisons; the DIs were not significantly different: $t=1.2$, $df=278$, n.s.). In the IC, the responses to the cleaned chirp were much more similar to the responses to the full natural sound, as long as the neuronal FRA intersected with the chirp frequency range ($r=0.68$, $p<0.001$). Only a minority of the DIs between the responses to the clean chirp and the natural stimulus were larger than the DIs between the responses to the noise component and to the natural stimulus (39% of all comparisons; the DIs were significantly different: $t=2.3$, $df=45$, $p<0.05$).

One possible explanation for this finding is that neurons in MGB and A1 are highly sensitive to noise stimuli, and do not respond well to tonal stimuli. To refute this explanation, we show here the results of a different set of experiments. We studied the masking of low-level tones by strong fluctuating noise. The basic results are illustrated in Fig. 2. When tested with fluctuating noise, the MGB neuron locked to the stimulus envelope (Fig. 2A, continuous gray line – the response peaks occur at each rise of the envelope). The tone did not evoke much response by itself (Fig. 2A, dotted line). However, the response to the sum of the tone and the noise showed a suppression of the envelope locking response to the noise alone (Fig. 2A,

continuous black line). The signal to noise ratio was –28 dB. A special signature of this type of suppression is the fact that the first noise cycle after tone onset was not suppressed: only the responses to the 2nd and following noise cycles were suppressed. The same findings can be demonstrated in A1.

In contrast, the responses in the IC were strikingly different. The neuron in Fig. 2B showed suppression of the envelope locking, but the suppression was identical at all noise cycles following tone onset. Furthermore, the signal to noise ratio at which this suppression occurred was higher: -17 dB. Thus, the highly sensitive suppression of envelope locking with its distinctive temporal pattern first appears in MGB. The long latency of this suppression may indicate the involvement of feedback circuits between MGB and cortex in the generation of this phenomenon.

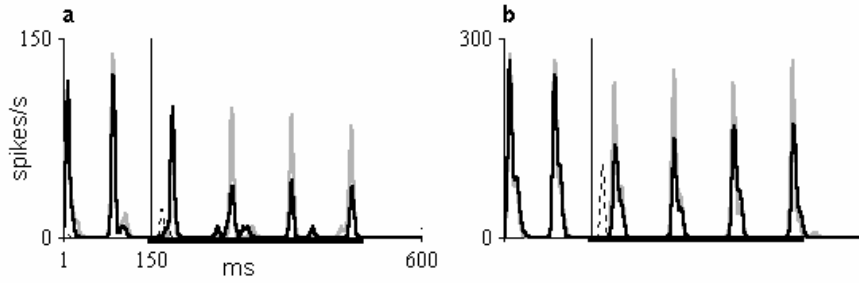


Fig. 2. Responses of an MGB neuron (a) and an IC neuron (b) to a tone (dashed line), fluctuating noise (continuous gray line) and the sum of the two (continuous black line). The vertical line marks tone onset (150 ms after noise onset). The thick black line marks the tone duration. MGB: BF 13.5 kHz, BW 13.5 kHz; IC: BF 22.1 kHz, BW 22.1 kHz.

In conclusion, low-level tones are as efficient at modifying the responses of MGB and A1 neurons to high-level noise, as low-level noise is efficient at modifying their responses to high-level tonal stimuli. Thus, neurons in MGB and cortex are especially sensitive to low-level acoustic components, whether they are narrowband or wideband.

2.3 Sensitivity to rare sounds

Neurons in A1 are not only sensitive to low-level acoustic components, but also show sensitivity to rare sounds (Ulanovsky *et al.* 2003). We tested neurons with pairs of tones of slightly different frequencies. The two tones appeared in a block in which one of the tones was common and the other one rare (usually 90%/10%), and in another block where the roles of the two tones were reversed. The number of spikes evoked by a tone tended to be higher when it was the rare stimulus in the block, than when it was the common stimulus in the block (Fig. 3). Thus, the probability of a stimulus can also influence the responses of A1 neurons. Such sensitivity to probability was not found in MGB, however.

The sensitivity to rare sounds demonstrated by A1 neurons share many characteristics with a component of auditory evoked potentials called mismatch negativity (MMN). MMN is evoked by rare sounds but is preattentive, being present under anesthesia (Naatanen, Tervaniemi, Sussman, Paavilainen, and

Winkler 2001). Thus, it is tempting to speculate that the type of stimulus-specific adaptation illustrated in Fig. 3 is the single neuron correlate of MMN (Ulanovsky *et al.* 2003).

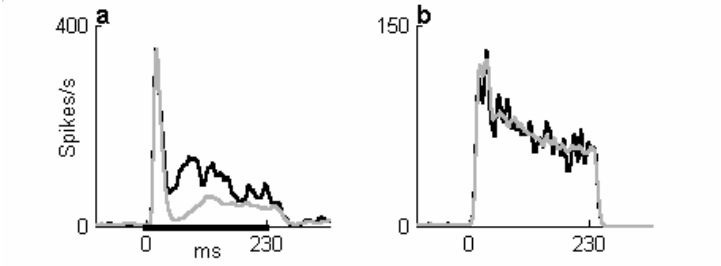


Fig. 3. Responses of an A1 neuron (a) and an MGB neuron (b) to the same sounds when common (gray lines) and rare (black lines). Two frequencies (f2/f1=1.1) centered on BF were used. They were presented in two blocks or 400 stimulus presentations. In one of the blocks f1 appeared 10% of the time and f2 90% of the time, and in the other block the roles were reversed. The PSTHs are averages of the responses to the two tone frequencies at each probability condition.

3. Auditory scene analysis in A1?

The data presented above suggest that dramatic changes in stimulus representation occur as information flows from IC through MGB to A1. Responses in IC are reasonably well described, to a first approximation, by the tuning properties of the neurons to pure tones. Our results are consistent with the idea that the dominant processing step in generating IC responses is the linear filtering of sound energy through the neuronal tuning curve. This claim has already been made before, in the context of other stimuli. For example, Ehret and Merzenich (Ehret *et al.* 1988) argued that neurons in IC show correlates of the perceptual critical band, which is described as a linear filter followed by a pure energy detector.

Neurons in IC can process the filtered stimulus in different ways: some IC neurons roughly respond to the short-term energy of the filtered stimulus (or to its envelope, we cannot separate between these two possibilities), whereas others can be sensitive to more sophisticated features. Nevertheless, neurons in IC with the same BF essentially process the same signal, and are therefore highly redundant.

The situation in A1 is much more involved. First, neurons in A1 are extremely sensitive to low-level acoustic components, even in the presence of much stronger sounds. This sensitivity is idiosyncratic, with different neurons showing very different response profiles to the same set of sounds. However, the most important aspect of our findings is the fact that the sensitivity of A1 neurons to low-level acoustic components is not completely arbitrary. The responses to a mixture were often more similar to the responses to the weak component when presented alone than to the responses to the strong component when presented alone. Neurons in MGB were more similar to A1 neurons than to IC neurons in this context.

Thus, a first transformation that occurs on the way from IC to A1 is the enhancement of the representation of low-level acoustic components in mixtures.

This transformation is to a large extent complete already at the level of the MGB, at least for the two types of mixtures that we used here.

Neurons in A1 further show effects of context: responses to a sound, when rare, are stronger than the responses to the same sound, when common. The time scale of this adaptation is in the range of a few seconds. Neurons in MGB do not show this adaptation under conditions in which it is very strong in cortex. Thus, the effect of temporal context appears to be substantially stronger in A1 than in MGB.

We would like to interpret these findings in the context of auditory scene analysis. It has been previously argued (de Cheveigne 2001) that a major problem in auditory scene analysis is that of splitting simultaneous auditory events that occur within the same critical band. The results presented here suggest that whereas IC neurons do not perform such splitting in the context of the sounds we used, neurons in MGB and cortex do. Furthermore, neurons in A1 also split streams of auditory events, for example based on the probability of sounds. Thus, we would like to suggest that our results are manifestations of auditory scene analysis, specifically source segregation, occurring in MGB and even more strongly in A1.

References

Bar-Yosef, O., Rotman, Y. and Nelken, I. (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J Neurosci 22, 8619-32.

Cover, T. and Thomas, J. (1991) *Elements of information theory* Wiley and Sons, NY.

de Cheveigne, A. (2001) The auditory system as a "separation machine". In: R. Schoonhoven, (Ed.), *Physiological and Psychophysical Bases of Auditory Function.* Shaker Publishing, Maastricht. pp. 453-460.

Ehret, G. and Merzenich, M.M. (1988) Complex sound analysis (frequency resolution, filtering and spectral integration) by single units of the inferior colliculus of the cat. Brain Res 472, 139-63.

Ehret, G. and Schreiner, C.E. (1997) Frequency resolution and spectral integration (critical band analysis) in single units of the cat primary auditory cortex. J Comp Physiol A 181, 635-50.

Fitzpatrick, D.C., Kuwada, S., Kim, D.O., Parham, K. and Batra, R. (1999) Responses of neurons to click-pairs as simulated echoes: auditory nerve to auditory cortex. J Acoust Soc Am 106, 3460-72.

Miller, L.M., Escabi, M.A., Read, H.L. and Schreiner, C.E. (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J Neurophysiol 87, 516-27.

Naatanen, R., Tervaniemi, M., Sussman, E., Paavilainen, P. and Winkler, I. (2001) "Primitive intelligence" in the auditory cortex. Trends Neurosci 24, 283-8.

Nelken, I. and Young, E.D. (1994) Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrow-band and wide-band stimuli. J. Neurophysiol. 71, 2446-2462.

Smith, P.H. and Spirou, G.A. (2002) From the Cochlea to the Cortex and Back. In: D. Oertel, R.R. Fay, and A.N. Popper, (Eds.), *Integrative Functions in the Mammalian Auditory Pathway*. Springer, New York. pp. 6-71.

Spirou, G.A. and Young, E.D. (1991) Organization of dorsal cochlear nucleus type IV unit response maps and their relationship to activation by bandlimited noise. J. Neurophysiol. 66, 1750-68.

Ulanovsky, N., Las, L. and Nelken, I. (2003) Processing of low-probability sounds by cortical neurons. Nat. Neurosci. 6, 391-398.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

AM and FM coherence sensitivity in the auditory cortex as a potential neural mechanism for sound segregation

Dennis L. Barbour and Xiaoqin Wang

Johns Hopkins University School of Medicine, Department of Biomedical Engineering, {dbarbour,xwang}@bme.jhu.edu

1 Introduction

The distinct modulation patterns of two simultaneous acoustic stimuli have been recognized as potential cues for their perceptual segregation (Bregman 1990; Cohen and Chen 1992). In particular, the coherence between simultaneous modulated stimuli has been altered in experiments to elicit psychophysical phenomena such as comodulation masking release (CMR) and modulation detection interference (MDI). (Hall, Haggard and Fernandes 1984; Yost, Sheft and Opie 1989). We have shown recently that auditory cortex neurons respond in systematic fashion to variations in temporal coherence between simultaneous AM or FM tones (Barbour and Wang 2002). Neurons were found to be suppressed either for coherent or for incoherent tones, implying that the excitatory and inhibitory inputs to the neurons studied were synchronized to modulation envelope.

AM and FM detection in the auditory system has been proposed to capitalize upon the same machinery (Saberi and Hafter 1995; Zwicker 1962), but FM has been shown to elicit less prominent grouping behavior than does AM (Carlyon 1992; Carlyon 1994; Carlyon 2000). We explore further in this article the behavior of auditory cortex neurons in response to two simultaneous AM or FM tones and compare the response characteristics between the two types of modulations. This experimental protocol can be expanded through modulation of different carrier stimuli, such as filtered noise or synthetic vocalizations.

2 Methods

Single-unit responses were collected from the bilateral primary auditory cortices (A1) of three awake marmoset monkeys (*Callithrix jacchus*). Detailed methods have been described previously (Barbour and Wang 2002). Briefly, each unit was characterized first by stimuli presented in isolation (tones if not otherwise indicated, but also bandpass noise or marmoset trill calls), modulated in amplitude or frequency, and then by two simultaneous modulated stimulus components. One (the f_1

component) always had its primary energy content fixed at the characteristic frequency (CF) of the unit; the carrier frequency of the other component (f_2) was varied systematically. Two stimulus conditions were studied: "coherent" (modulations of both components in phase at 0° offset) or "incoherent" (modulations out of phase at 180° offset). Sound level and other parameters of the f_1 component were chosen such that high rates were elicited; level of the f_2 component was chosen to be at the lowest level at which flanking inhibition could be elicited. Attenuation of 0 dB for a pure tone at 1 kHz is equivalent to 93 dB SPL.

Coherence sensitivity was quantified by the coherence sensitivity index (CSI):

$$\mathrm{CSI} = \frac{\mathrm{TMTFRF}(f_2,0°) - \mathrm{TMTFRF}(f_2,180°)}{\mathrm{TMTFRF}(\mathrm{CF},0°)}$$

TMTFRF is the two-modulated-tone frequency response function, which is the response as a function of f_2 carrier frequency while the f_1 component remains fixed at CF. This index reflects the coherent response minus the incoherent response normalized by the coherent response at CF. Upon visual inspection of the CSI functions, four general categories of response could be discerned:

- **Type 0**: CSI of 0 for all frequencies: no difference detected between coherent and incoherent stimuli.
- **Type 1a**: CSI of 0 near CF; negative CSI at flanking frequencies (can be asymmetric): flanking release from inhibition for incoherent stimuli.
- **Type 1b**: Positive CSI near CF; CSI of 0 for flanking frequencies: suppression near CF for incoherent stimuli.
- **Type 2**: CSI of 0 or negative near CF; positive CSI at flanking frequencies (can be asymmetric): flanking release from inhibition for coherent stimuli? (appears to be a heterogeneous response class)

CSI analysis of the entire population of units studied with AM and FM showed the greatest effect of coherence at CF±1/4 octave for AM and CF±1/8 octave for FM. Consequently, relative CSI values between CF and these frequencies were used to separate Types 1 and 2 for comparison. The CSI at CF minus the CSI at the appropriate flanking frequency is positive for Type 1, negative for Type 2.

3 Results

3.1 CSI comparison between AM and FM

The coherence sensitivity observed with FM appeared qualitatively similar to that observed with AM. Fig. 1A,B show the rate response and CSI as a function of f_2 carrier frequency of a Type 1a unit tested with AM. Fig. 1C shows the CSI as a function of f_2 carrier frequency for all neurons studied with AM (filled symbols) and FM (open symbols). The AM population CSI peaks at CF and dips around ±1/4 octave, indicating that coherence reaches maximal suppression approximately one critical band above and below CF. The FM population, on the other hand, tends to reach a local minimum around ±1/8 octave.

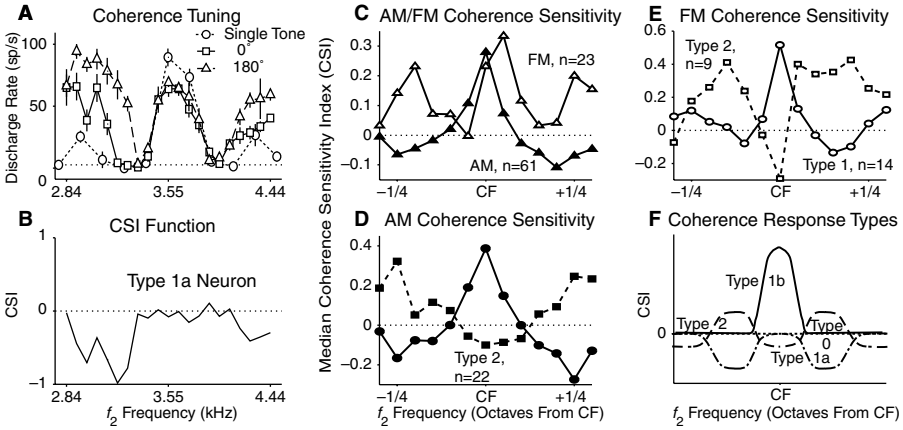


Fig 1. (**A**) For one neuron, tuning of a single tone as a function of carrier frequency (circles), and of two coherent (squares) or incoherent 16 Hz AM tones (triangles) as a function of f_2 frequency. (**B**) Coherence sensitivity index (CSI) as a function of f_2 frequency. (**C**) Median CSI as a function of f_2 frequency for AM (open triangles) and FM (solid triangles). (**D**) CSI for AM subdivided by the sign of CSI at CF minus CSI at ±1/4 octave. Type 1 response (circles); Type 2 response (squares). (**E**) CSI for FM subdivided by the sign of CSI at CF minus CSI at ±1/8 octave. (**F**) Four types of coherence sensitivity based upon CSI measures.

If the AM population is subdivided on the basis of relative CSI between CF and ±1/4 octave (see Methods), the CSI plot in Fig. 1D results. Approximately 2/3 of the neurons (Type 1) account for the population summary seen in Fig. 1C. These neurons have maximal CSI values at CF and minimal values near ±1/4 octave. Coherent f_2 tones therefore excite these neurons relative to incoherent tones near CF but suppress the neurons at flanking frequencies. The remaining 1/3 of the neurons (Type 2) show a complementary response profile. Coherent f_2 tones suppress these neurons if located near CF but excite them if located at flanking frequencies.

The FM population can be similarly subdivided by the relative CSI between CF and ±1/8 octave, yielding the profiles shown in Fig. 1E. Approximately 3/5 of these neurons show near-CF coherent excitation similar to the Type 1 AM population. The remaining 2/5 show a complementary Type 2 response, with the exception of a relative asymmetry in the flanking frequencies of coherent excitation, which contributes to the upward shift in the overall FM population summary in Fig. 1C.

Closer inspection of individual Type 1 responses reveals two distinct populations. Type 1a shows a flanking release from inhibition by incoherent stimuli while Type 1b shows a near-CF suppression in response to coherent stimuli. Type 2 responses overall tend to be a more heterogeneous group that has yet to be accurately assessed. The response types are summarized in Fig. 1F.

3.2 Level distribution for AM and FM

Psychophysical measures of coherence-induced masking typically test responses near detection threshold of the probe stimulus in the presence of a masking noise in

a "signal-vs.-noise" task. Physiological measures can also be probed for shifts in threshold (Nelken, Rotman and Bar Yosef 1999; Pressnitzer, Meddis, Delahaye and Winter 2001), but most real-world problems of interest involve segregating two or more competing signals well above threshold in a "signal-vs.-signal" task. For this reason neurons were tested in these experiments over a wide range of sound levels for coherence sensitivity. Figure 2A shows the distribution of the f_1 or CF tones for both AM and FM. Coherence sensitivity was found at all sound levels.

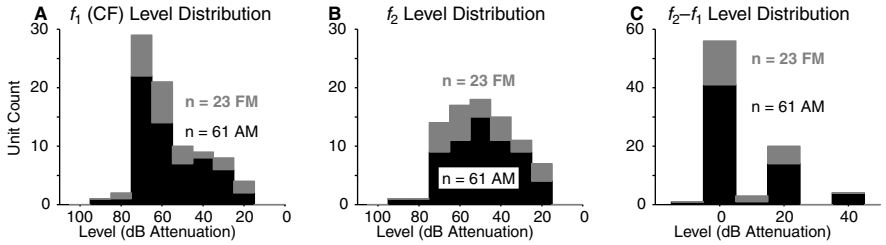


Fig 2. (**A**) Distribution of attenuation values for the f_1 or CF tones. (**B**) Distribution of attenuations for the f_2 tones. (**C**) Distribution of relative attenuations between f_1 and f_2 tones.

Distribution of f_2 tone levels tends toward higher values, seen in Fig. 2B and in the relative level distribution in Fig. 2C. The f_2 tones can modulate the inhibitory sidebands for most neurons at the same sound level as the tone at CF. A few require somewhat more intense tones for the effects to be seen.

3.3 Simultaneous synthetic trills with variable-coherence FM

Marmoset trills contain prominent AM and FM. Figure 3A shows a single-trill modulation transfer function, where the frequency of the call's sinusoidal FM component was varied over a physiologic range. Figure 3B shows the neuron's tuning to carrier frequency of a single call's fundamental. Also shown on the same axes are the responses to two simultaneous calls with identical parameters except for coherent (0°) or incoherent (180°) FM phase. The incoherent signals eliminate all response in the excitatory frequency range of the neuron, corresponding to a Type 1b response. This neuron's typical Type 1b CSI function can be seen in Fig. 3C.

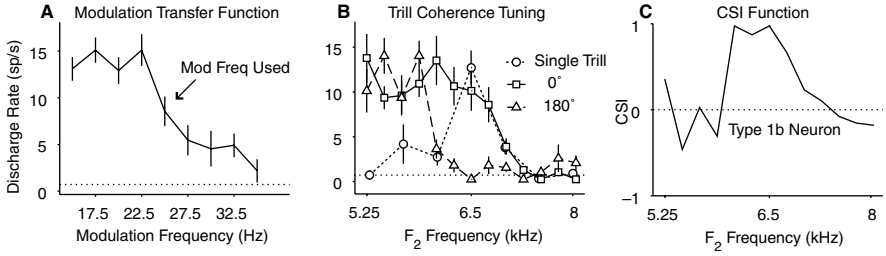


Fig 3. (**A**) For one neuron, modulation transfer function of a single synthetic marmoset trill call. (**B**) Tuning of one trill as a function of carrier frequency (circles), and of two coherent (squares) and or incoherent trills (triangles) as a function of f_2 frequency. (**C**) CSI function.

3.4 AM noise/tone

Many neurons in auditory cortex do not respond to tones but will respond to bandpass noise. Figure 4A shows the tuning of one of these neurons to a single bandpass AM noise stimulus. When stimulated by two coherent AM noise bands, it reveals flanking inhibition. When the noise bands are incoherently modulated, however, the neuron shows a bilateral flanking release from inhibition consistent with a Type 1a response. This neuron does not respond to single tones, as shown in Fig. 4B. If an AM noise band is delivered for the f_1 stimulus while a coherent AM tone is used for the f_2 stimulus, flanking inhibition is revealed. If the tone is then made incoherent, a release from flanking inhibition similar to that shown in Fig. 4A can be seen.

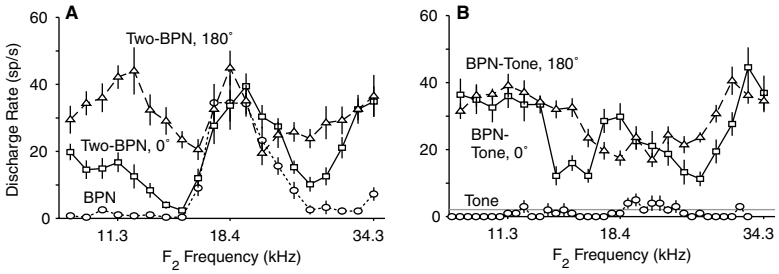


Fig 4. (**A**) For one neuron, tuning of a single 0.25-octave AM noise band as a function of carrier frequency (circles), and of a 0.5-octave AM f_1 noise band with a 0.25-octave AM f_2 noise band coherently (squares) or incoherently modulated (triangles) as a function of f_2 frequency. (**B**) Tuning of a single AM tone (circles) and of a 0.5-octave AM f_1 noise band with an AM f_2 tone coherently (squares) or incoherently modulated (triangles) as a function of f_2 frequency. Noise from (A) and tones from (B) both reveal a symmetric Type 1a CSI.

4 Neuronal model

Convergence of subcortical input neurons with a variety of temporal response properties and a range of characteristic frequencies can create the relative temporal sensitivity observed in auditory cortex. Subsequent projections of these cortical neurons onto neurons in higher cortical areas could be used to signal the presence of incoherent stimuli, i.e., stimuli that need to be segregated perceptually. Example circuitry that could be used for this task is shown in Fig. 5. Neurons with differently tuned flanking coherence sensitivity need only combine with a single neuron showing the effect at CF and a single neuron showing little effect to create an array of output neurons responsive only under particular conditions of stimulus coherence over a narrow range of carrier frequencies. The existence of such coherence-specific neurons elsewhere in the auditory cortex is still speculative. Finding such neurons, if they do exist, should be relatively straightforward: they should be relatively abundant, located in an area to which primary auditory cortex projects and only respond to temporally complex sounds—possibly only to multiple simultaneous sounds.

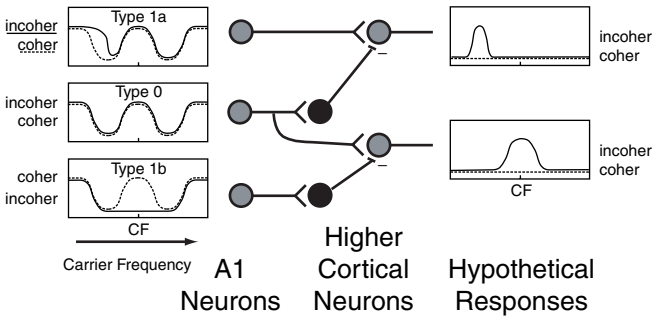


Fig 5. Circuitry model demonstrating how the observed neuronal responses might be used to detect incoherence. Such detection could be useful for signaling that energy close in frequency belongs to two sources.

The usefulness of such neurons would probably lie in their feedback projections to earlier auditory centers. Activity in these pathways could indicate which auditory filters (spectral or temporal) to emphasize or suppress in order to sharpen the distinction between easily confusable simultaneous sounds. Alternatively, they could project to higher cognitive centers and play a role in stimulus feature binding.

5 Conclusions

Auditory cortex neurons respond to the relative temporal structure of a complex stimulus in one of several ways. Type 0 responses show little sensitivity to relative signal coherence. Type 1a responses show incoherent flanking release from inhibition. Type 1b responses show coherent suppression of response near CF. Instead of unmodulated stimuli, these neurons seem to prefer modulated stimuli at intermediate modulation frequencies. This "bandpass" modulation transfer function feature does not apply strictly to the temporal structure of a single carrier, however; instead, the behavior can be attributed to the neuron's stimulus integration throughout its entire excitatory frequency range. Type 2 responses show flanking release from inhibition for coherent stimuli and may represent a heterogeneous group.

Type 1a responses probably account for previous observations in auditory cortex of lower thresholds for modulated than for unmodulated masking noise (Nelken, Rotman et al. 1999), while Types 1b and 2 appear to exhibit the opposite threshold shifts. Altogether, these responses may represent a coding scheme for segregating two competing acoustic signals based upon temporal coherence.

Neurons not driven by modulated tones can be studied using a similar coherence/incoherence protocol with more complex stimuli such as synthetic vocalizations or noise bands. The responses observed from these experiments fall into the same categories as those observed for modulated tones. Neurons that do not respond to tones at CF seem to be inhibited by flanking tones, implying that the inhibitory sidebands are less selective than are the excitatory frequency ranges.

AM and FM both show the same qualitative types of responses to coherence. The main difference appears to be the carrier frequency range of coherence sensitivity: neurons tested with FM seem to be sensitive to coherence only within one critical band centered on CF while neurons tested with AM seem to be sensitive

over a range of two critical bands. In both cases this phenomenon appears to be rather local in carrier frequency. FM coherence has been found to contribute little as a grouping cue independent of harmonicity (Carlyon 1992; Carlyon 1994; Carlyon 2000). The compact nature of the phenomenon reported here implies that it might be more useful as a local segregation than as an across-frequency grouping cue. In other words, these neurons might signal that incoherent energy falling within a particular narrow frequency range should be attributed to more than one percept. This type of processing seems intuitively to be more fundamental than the grouping of multiple already separated components into a single percept (de Cheveigné 2001).

Acknowledgments

Supported by NIH NIDCD Grant R01 DC–03180.

References

Barbour, D. L. and X. Wang (2002). "Temporal coherence sensitivity in auditory cortex." *J Neurophysiol*, 88, 2684-2699.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sounds*. Cambridge, Massachusetts, The MIT Press.

Carlyon, R. P. (1992). "The psychophysics of concurrent sound segregation." *Philos Trans R Soc Lond B Biol Sci*, 336, 347-355.

Carlyon, R. P. (1994). "Further evidence against an across-frequency mechanism specific to the detection of frequency modulation (FM) incoherence between resolved frequency components." *J Acoust Soc Am*, 95, 949-961.

Carlyon, R. P. (2000). "Detecting coherent and incoherent frequency modulation." *Hear Res*, 140, 173-188.

Cohen, M. F. and X. Chen (1992). "Dynamic frequency change among stimulus components: effects of coherence on detectability." *J Acoust Soc Am*, 92, 766-772.

de Cheveigné, A. (2001). "The Auditory System as a 'Separation Machine'." *Physiological and Psychophysical Bases of Auditory Function: Proceedings of the 12th International Symposium on Hearing*. A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs and R. Schoonhoven. Maastricht, Shaker Publishing.

Hall, J. W., M. P. Haggard and M. A. Fernandes (1984). "Detection in noise by spectro-temporal pattern analysis." *J Acoust Soc Am*, 76, 50-56.

Nelken, I., Y. Rotman and O. Bar Yosef (1999). "Responses of auditory-cortex neurons to structural features of natural sounds." *Nature*, 397, 154-157.

Pressnitzer, D., R. Meddis, R. Delahaye and I. M. Winter (2001). "Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus." *J Neurosci*, 21, 6377-6386.

Saberi, K. and E. R. Hafter (1995). "A common neural code for frequency- and amplitude-modulated sounds." *Nature*, 374, 537-539.

Yost, W. A., S. Sheft and J. Opie (1989). "Modulation interference in detection and discrimination of amplitude modulation." *J Acoust Soc Am*, 86, 2138-2147.

Zwicker, E. (1962). "Direct comparisons between the sensations produced by frequency modulation and amplitude modulation." *J Acoust Soc Am*, 34, 1425-1430.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Auditory perception with slowly-varying amplitude and frequency modulations

Fan-Gang Zeng[1,2], Kaibao Nie[1], Ginger Stickney[1], and Ying-Yee Kong[2]

[1] Department of Otolaryngology – Head and Neck Surgery, University of California, Irvine, CA 92697-1275, USA, fzeng@uci.edu, knie@uci.edu, stickney@uci.edu
[2] Department of Cognitive Sciences, University of California, Irvine, ykong@uci.edu

1 Introduction

Amplitude modulation (AM) and frequency modulation (FM) are abundant in natural stimuli, including speech, music, and animal communication sounds. Although amplitude and frequency modulations have been extensively studied physiologically and psychophysically (e.g., Riesz 1928; Grinnell 1963; Suga 1964; Gordon and O'Neill 1998), it is still unclear whether and how the auditory system extracts and uses these cues. For example, there is an ongoing debate on whether amplitude modulation is processed via envelope extraction in the temporal domain (Viemeister 1979) or a second filtering process in the spectral domain (Dau, Kollmeier, and Kohlrausch 1997). It is also unsettled whether frequency modulation is processed independently of amplitude modulation via specialized "FM channels" in the auditory system (Kay and Matthews 1972; Regan and Tansley 1979; Moore and Sek 1996), by a common mechanism (Moore and Sek 1995; Saberi and Hafter 1995).

Regardless of the underlying processing mechanisms of amplitude and frequency modulations, both cues have been shown to contribute to speech recognition in quiet laboratory conditions. Remez et al. (1981, 1990) demonstrated that speech could be reliably recognized with three sinusoids that tracked the formant movement, namely frequency modulation. On the other hand, Shannon et al. (1995) demonstrated that speech could also be reliably recognized with primarily temporal envelope cues, namely amplitude modulation. These results have been traditionally taken as an indication of the redundancy of multiple cues in natural speech sounds.

Motivated by how to deliver the fine structure cue to cochlear implants, recent studies have implicated possible independent contributions of amplitude and frequency modulations to auditory perception (e.g., Smith, Delgutte, and Oxenham 2002). We have developed a signal-processing strategy that extracts slowly-varying amplitude and frequency modulations from the traditionally defined temporal envelope and fine structure cues, i.e., Hilbert transform. This novel strategy also

provides a platform to test systematically the independent contribution of amplitude and frequency modulations to auditory perception. Our results suggest that, while amplitude modulation provides essential information for speech recognition in quiet, frequency modulation is needed for speech recognition with competing talkers and music perception.

2 Methods and materials

2.1 Subjects

A total of 26 normal-hearing listeners participated in the study. Five of them participated in the phoneme recognition experiment, 15 subjects consisting of 3 groups of 5 each participated in the sentence recognition experiment, and 6 additional subjects participated in the melody perception experiment. Local IRB approval and informed consent were obtained.

2.2 Stimuli

Phoneme stimuli included 12 /hvd/ vowels spoken by 3 male, 3 female, and 3 girl talkers (Hillenbrand, Getty, Clark, and Wheeler 1995) and 20 /aCa/ consonants by 2 male and 2 female talkers (Shannon, Jensvold, Padilla, Robert, Wang 1999). The noise was speech-spectrum-shaped and was presented at 0 and –5 dB signal-to-noise ratios. Sentence stimuli were 60 IEEE sentences spoken by a male talker. The noise was a competing sentence spoken by a different male talker. Both sentences had the same onset, but the competing sentence was always longer. Melody stimuli included two sets of 12 familiar songs with one set containing the rhythmic cue and the other containing no rhythmic cue. The rhythmic cue was removed by forcing all notes to be 350 ms in duration with a silent period of 150 ms between notes.

2.3 Processing

Figure 1 displays the basic structure of the novel signal processing strategy that analyzes and synthesizes a stimulus according to its AM and FM components. The original stimulus was first filtered into *N* narrow-bands (*N* ranging from 1 to 64 in octave steps). Each narrow-band signal was then subjected to separate AM and FM extraction pathways. The AM was derived by half-wave rectification followed by a low-pass filter. The low-pass filter controlled the amplitude modulation rate, which was set at 500 Hz in the present study. Similar to earlier work on phase vocoders (Flanagan and Golden 1966), the FM was derived by phase-orthogonal demodulators to remove the center frequency of the narrow-band signal. Two independent low-pass filters were used to control the FM depth and rate. In this study, the FM depth was set at 500 Hz or the critical bandwidth, whichever was narrower, while the FM rate was set at 400 Hz. The delay difference was compensated between the AM and FM pathways before recovering the center frequency to re-synthesize the original stimulus.

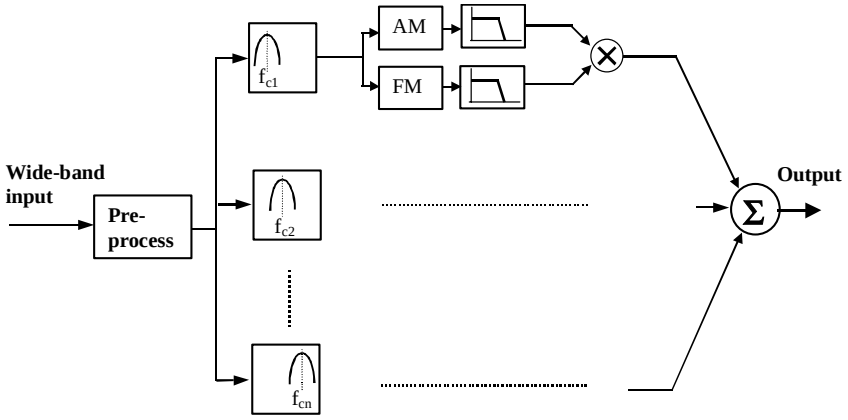


Fig. 1. Signal processing in sound analysis and synthesis using AM and FM cues.

To appreciate the novelty of the proposed processing, a synthetic token /bai/ was used in an 8-band processor to contain AM only and AM+FM components. Figure 2 shows the spectrogram of the original token (left panel), the AM only token (middle panel), and the AM+FM token (right panel). Although neither the AM only nor AM+FM token contained the detailed harmonic structure as in the original token, the AM+FM token clearly preserved formant transition information (see the initial formant transitions from /b/ to /a/ in the first 40 msec of the stimulus as well as the much longer transitions from /a/ to /i/ for the last 300 msecs of the stimulus).

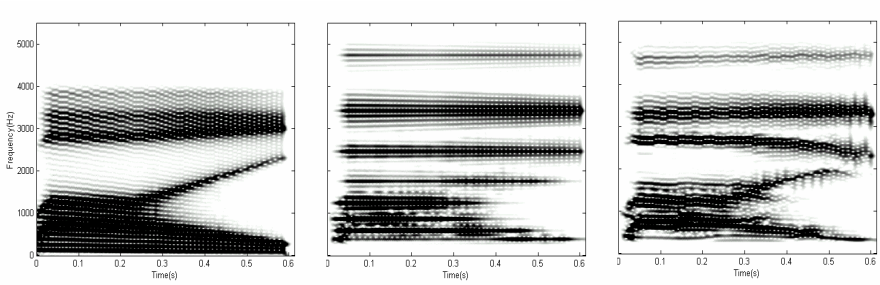


Fig. 2. Spectrograms of the original token /bai/ (left panel), the 8-band AM only token (middle panel), and the 8-band AM+FM token (right panel).

2.4 Procedures

In the phoneme recognition experiment, the subject was asked to identify the randomly presented phoneme by clicking on the GUI that contained all possible phonemes. Trial-by-trial feedback was provided. In the sentence recognition experiment, the subject heard 60 target sentences both in quiet and in the presence of a single competing sentence at different signal-to-noise ratios. The subject was then asked to type in the sentence via a keyboard. No feedback of any form was given. The keywords correctly identified were computed and reported as percent

correct. In the melody recognition experiment, the subject heard a melody and had to choose from 1 of the 12 melodies whose names were displayed on a computer screen. Trial-by-trial feedback was provided. A practice run was always given before formal data collection. All stimuli were presented monaurally through a Sennheiser headphone at 65 dB SPL. The subject performed these experiments in a double-walled, sound-attenuated chamber.

3 Results

3.1 Phoneme recognition

Figure 3 shows vowel (left panel) and consonant (right panel) recognition scores as a function of signal to noise ratios in the 8-band condition. The vowel recognition was generally at a high level between 70 and 90% correct for all conditions. A repeated measures ANOVA revealed no statistical difference between the AM and the AM+FM conditions [$F(1,4)= 4.748$, $p=0.095$] but a significant difference between the noise and quiet conditions [$F(2,8)=38.663$, $p<0.001$]. On the other hand, a significant difference in consonant recognition was observed for both the modulation [$F(1,4)= 35.911$, $p=0.004$] and noise [$F(2,8)= 114.534$, $p<0.001$] factors. A significant interaction was also observed between the modulation and noise factors, with the AM+FM stimuli producing better performance than the AM-only stimuli in noise but essentially no difference in quiet.

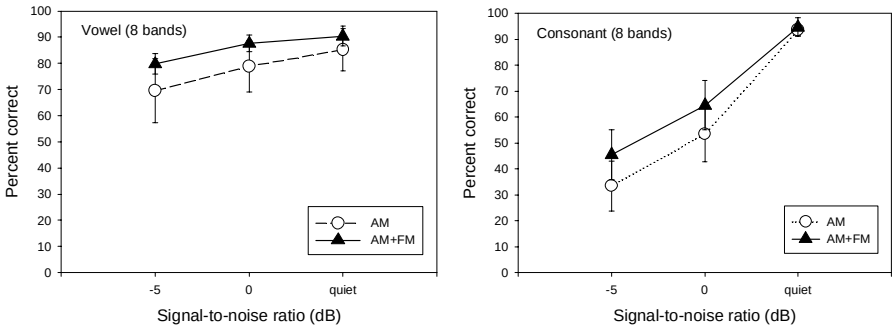


Fig. 3. Vowel (left panel) and consonant (right panel) recognition as a function of signal-to-noise ratios in an 8-band processor. The open circles represent data collected with the AM-only condition while the filled triangles represent the AM+FM condition.

3.2 Sentence recognition

Figure 4 shows sentence recognition scores as a function of signal-to-noise ratios in the presence of a competing talker. Different from the modest difference in phoneme recognition, the additional FM cue produced significantly better results

than the AM only condition [F(2,24) = 452.08, p<.001], particularly at low signal-to-noise ratios where the improvement was as much as 70 percentage points.

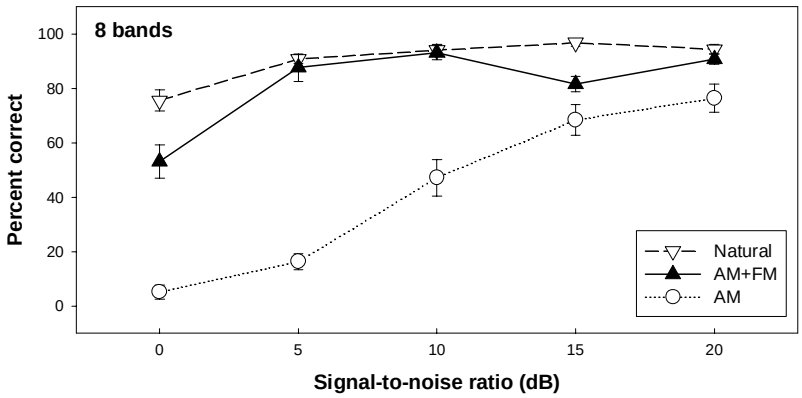


Fig. 4. Sentence recognition as a function of signal-to-noise ratios in an 8-band processor. The noise was a sentence from another talker.

3.3 Music perception

Figure 5 shows melody recognition as a function of the number of frequency bands with (left panel) and without (right panel) the rhythmic cue. Clearly the rhythmic cue contributed to a relatively high level of performance independent of both the number of bands and the addition of the FM cue. However, when the rhythmic cue

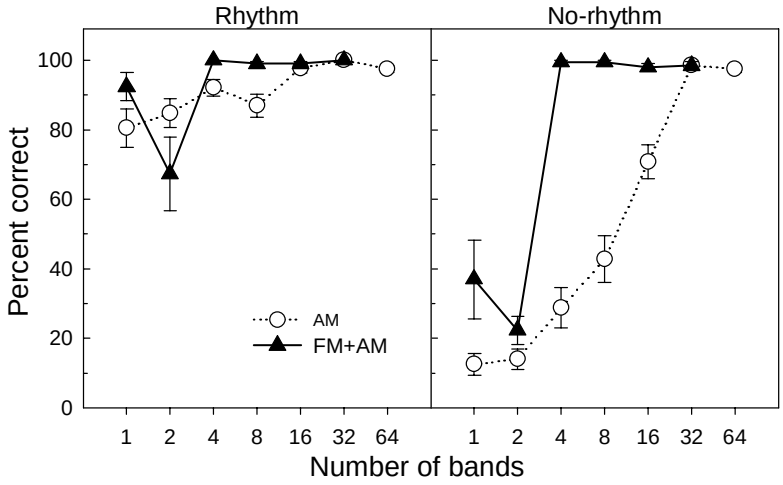


Fig. 5. Melody recognition as a function of the number of bands in the presence (left panel) and absence (right panel) of the rhythmic cue.

was removed, the AM condition needed 32 bands of spectral information to achieve perfect melody recognition while the AM+FM condition only required 4 bands. Between 4 and 16 bands, the AM+FM condition produced significantly better performance (t-tests, $p<0.01$) than the AM condition. The dip in performance with the 2-band FM condition was possibly due to an inappropriate FM representation of the original melody information.

4 Discussion

Together with previous studies, the present data show that, while AM information is sufficient for speech recognition in quiet, FM information is required for speech recognition in noise and for melody recognition without rhythmic cues. The largest improvement was observed for sentence recognition with a competing talker, emphasizing the importance of the FM cue in speech perception under realistic listening environments, e.g., at a cocktail party. We have collected preliminary data suggesting that the FM cue might have allowed the listener to tell one talker (signal) apart from the other (noise). In other words, there appears to be an independent contribution of the AM and FM cues to speech recognition: the AM mostly contributes to "what is said" whereas the FM mostly contributes to "who says what".

Different from previous studies in which FM might consist of rapid changes across multiple critical bands (Remez, Rubin, Pisoni, and Carrell 1981), the present study only extracts the slowly-varying FM components around the center frequency of a frequency band. With both the modulation depth and rate limited to a few hundred Hertz, this slowly-varying FM cue might be used by cochlear implant users as an efficient means for encoding fine-structure information.

The basic principle underlying the AM and FM cues may also be applied to low-rate, high-quality audio coding and processing. For example, for the 5000-Hz sub-band, there will be no need to transmit the 5000-Hz information, rather an FM signal with a bandwidth of 500-Hz or less is needed for transmission.

5 Summary

We have developed a signal processing strategy that can independently extract slowly-varying amplitude and frequency modulations within a frequency band with the number of bands as an independent variable. While the AM provides sufficient information for speech recognition in quiet, the additional FM significantly improves speech recognition in noise and music perception. The FM may be used as an efficient means to convey the fine-structure information in cochlear implants and audio coding.

Acknowledgments

We thank Ackland Jones, Michael Vongphoe, Elsa Del Rio, and Sheetal Desai for technical support. This research was supported by grants from NIH (R01-DC02267 and F32-DC-5900) and Chinese NSF (30000041).

References

Dau, T., Kollmeier, B. and Kohlrausch A. (1997) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892-2905.

Flanagan, J.L. and Golden, R.M. (1966) Phase Vocoder. Bell Sys. Tech. J. 45, 1493-1509.

Gordon, M and O'Neill, W.E. (1998) Temporal processing across frequency channels by FM selective auditory neurons can account for FM rate selectivity. Hear. Res. 122, 97-108.

Grinnell, A.D. (1963) The neurophysiology of audition in bats: Intensity and frequency parameters. J. Physiol. 167, 38-66.

Hillenbrand, J., Getty, L.A., Clark, M.J. and Wheeler, K. (1995) Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97, 3099-3111.

Kay, R.H. and Matthews, D.R. (1972) On the existence in the human auditory pathway of channels selectively tuned to the modulation present in frequency-modulated tones. J. Physiol. 225, 657-677.

Moore, B.C.J. and Sek, A. (1995) Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation. J. Acoust. Soc. Am. 97, 2468-2478.

Moore, B.C.J. and Sek, A. (1996) Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. J. Acoust. Soc. Am. 100, 2320-2331.

Regan, D and Tansley, B.W. (1979) Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selectively sensitive to frequency ranges. J. Acoust. Soc. Am. 65, 1249-1257.

Remez, R.E., Rubin, P.E., Pisoni, D.B. and Carrell, T.D. (1981) Speech perception without traditional speech cues. Science 212, 947-949.

Remez, R. and Rubin, P.E. (1990) On the perception of speech from time-varying acoustic information: contributions of amplitude variation. Percept. Psychophys. 48, 313-325.

Riesz, R.R. (1928) Differential intensity sensitivity of the ear for pure tones. Phys. Rev. 31, 867-875.

Saberi, K. and Hafter, E.R. (1995) A common neural code for frequency- and amplitude-modulated sounds. Nature 374, 537-539.

Shannon, R.V., Jensvold, A., Padilla, M., Robert, M.E. and Wang, X. (1999) Consonant recordings for speech testing. J. Acoust. Soc. Am. 106, L71-L74.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J. and Ekelid, M. (1995) Speech recognition with primarily temporal cues Science 270, 303-304.

Smith, Z.M., Delgutte, B., and Oxenham, A.J. (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416, 87-90.

Suga, N. (1964) Recovery cycles and responses to frequency modulated tone pulses in auditory neurons of echo-locating bats. J. Physiol. 175, 50-80.

Viemeister, N.F. (1979) Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Am. 66, 1364-1380.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

The role of auditory-vocal interaction in hearing

Steven J. Eliades and Xiaoqin Wang

Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering,
Johns Hopkins University
xwang@bme.jhu.edu

1 Introduction

Self-produced sounds, such as speech, are sensory inputs to the auditory system that have an important behavioral role. Humans continuously monitor their vocal output to correct perturbation in any one of a number of parameters. Alteration in the spectral pattern of speech feedback, for example, will result in compensatory adjustments in produced fundamental and formant frequencies (Burnett, Freedland, Larson, and Hain 1998; Houde and Jordan 1998). Several animal species show similar vocal control behavior, including frequency and temporal patterning in birdsong (Leonardo and Konishi 1999; Osmanski Dooling, and Venkatachalam 2003) and amplitude in primate vocalizations (Sinnott, Stebbins, and Moody 1975). However, the neural mechanisms underlying this sensory-motor control and the function of the auditory system during vocalization remain largely unknown.

Auditory-vocal interaction in humans and non-human primates is found primarily in the auditory cortex. Megnetoencephalogram studies have revealed a reduction in auditory cortical responses during speech production compared to passive listening (Houde, Nagrajan, Sekihara, and Merzenich 2002). These same experiments have shown that this reduction is largely absent from auditory-brainstem recordings. PET imaging studies have also shown reduced cortical activation during speech (Paus, Perry, Zatorre, Worsley, and Evans 1996). Similar patterns of reduced activation during electrically-stimulated vocalizations have been recorded from the primate auditory cortex (Müller-Preuss and Ploog 1981). We have investigated the role of sensory-motor interaction in the auditory cortex of a vocal primate, the common marmoset (*Callithrix jacchus*), using single-unit recordings during self-initiated vocalizations. This article summarizes and extends our recent study on this subject (Eliades and Wang 2003).

2 Vocalization-induced modulations of the auditory cortex

We have observed modulations of both spontaneous and stimulus driven activities of cortical neurons during voluntarily produced vocalizations. The predominant response observed was a suppression of cortical neurons (Fig 1A). Vocalization-induced suppression began several hundred milliseconds *prior to* the onset of vocal production (median 220 ms, Fig 1B). This pre-vocal onset of the suppression is suggestive of a neurally-mediated inhibition caused by signals from vocal control centers. The median reduction in firing rate by this suppression was 71%. A second, smaller, group of neurons displayed excitation during vocalization (Fig 1C). This excitation began *after* the start of vocal production and is likely an auditory response to feedback of the produced sound (Fig 1D).

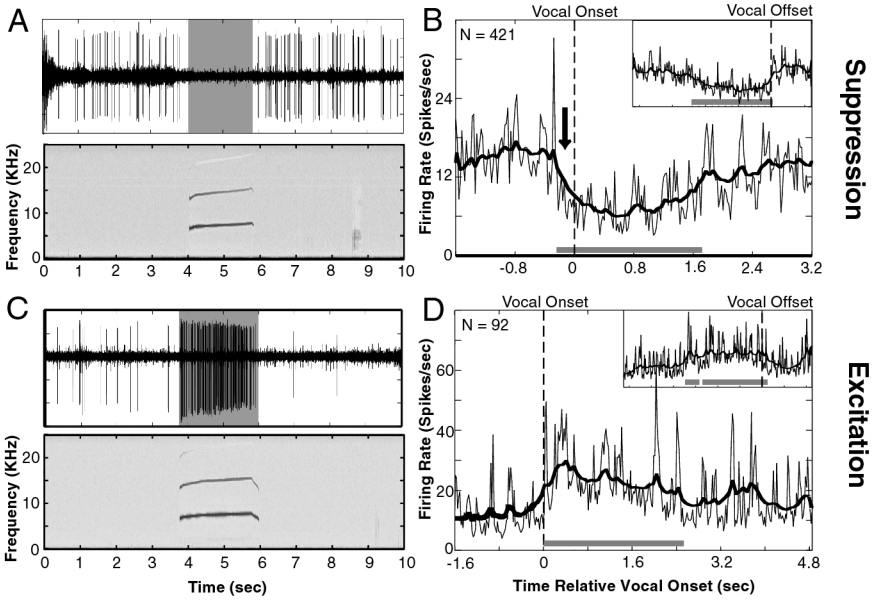


Fig. 1. Vocalization-induced suppression and excitation. A,C: Representative examples are shown for both response types. B,D: Response histograms were constructed from vocal onset-aligned data to observe the magnitude and timing of modulations. The insets show response histograms aligned by vocal offset. Thick bars along the horizontal axis indicated statistically significant ($p<0.01$) rate changes. Adapted from Eliades and Wang (2003).

3 Mechanisms of auditory-vocal interaction in the cortex

The onset of suppression before vocalization suggests that this modulation arises through inhibition triggered by signals from vocal production centers. One likely site for this inhibition is within the auditory cortex itself. While previous work has shown reduced brainstem activity during vocal production (Suga and Shimozawa

1974), suppression of spontaneous cortical discharges, in the absence of acoustic stimulation, is suggestive of direct auditory cortical inhibition in addition to a reduction in afferent signals from the brainstem. Further evidence is provided by multiple neurons recorded simultaneously from single electrodes. An example in Fig 2A shows a unit with large action potentials that was completely suppressed during vocalization while a unit with smaller action potentials increased its firing instead. The response properties of simultaneously recorded pairs were heterogeneous (Fig 2B). Completely suppressed responses, for example, were often recorded in the same electrode along with responses ranging from suppression to excitation. The absence of correlation between simultaneously-recorded neuron pairs supports the hypothesis that vocalization-induced inhibition is cortical in origin, as each pair likely comes from the same cortical column and, as such, receives similar subcortical inputs.

Our preliminary data indicate that vocalization-induced inhibition is not evenly distributed across cortical layers. In the upper cortical layers, neurons favoring suppression account for 75-80% of recorded units (Fig 2C). In contrast, deeper layers show more equal fractions of suppressed and excited neurons. The dominance of inhibition in upper layers suggests the action of local interneurons, possibly under the influence of long-range cortical-cortical connections, with GABAergic outputs in layers II/III.

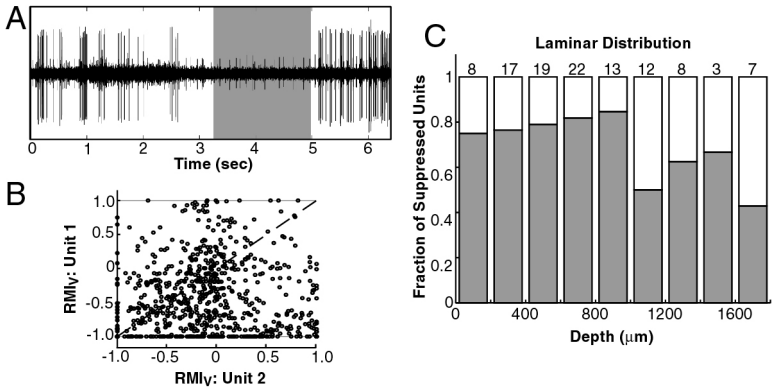


Fig. 2. A,B: Response properties of simultaneously-recorded neuron pairs. An example of two simultaneously-recorded units (A). A large sample of neuron pairs showed no correlation (r=0.02) of response modulations during vocalization (B). Vocalization responses were quantified by $RMI_V = (R_{voc}-R_{pre\text{-}voc})/(R_{voc}+R_{pre\text{-}voc})$, where $R_{pre\text{-}voc}$ and R_{voc} are the firing rates before and during vocalization (suppression: RMI<0, excitation: RMI>0). C: Laminar distribution of response properties. The fraction of units showing suppression is plotted versus the recording depth. Numbers on top of the histogram indicate the number of units recorded in each depth.

4 Comparison of single neuron responses and global activity

Although single-unit responses recorded in non-human primates have revealed the existence of two types of cortical responses, suppressed and excited, during vocalization, most human studies have thus far reported only dampened activation (MEG: Houde et al. 2002; PET: Paus et al. 1996; intra-operative electrocorticography: Crone, Hao, Hart, Boatman, Lesser, Irizarry, and Gordon 2001). Given the nature of the recording methods used in humans, the observations of those studies may reflect globally summed neural activities. In Fig 3, we compare population properties of well-isolated single neurons and multi-unit clusters. Single-unit recordings, based on large action potentials, show strongly suppressed responses in 75% of auditory cortical neurons (Fig 3A). Units of smaller action potential size, likely more distant from the recording electrode, show the same ratio of suppressed and excited neurons, but with reduced magnitude of inhibition. Poorly isolated multi-units, however, exhibit a further reduced suppression and an increase in the fraction and magnitude of excitation. This suggests that, while the percent of suppressed cortical neurons is relatively invariant, reduction in the quality of unit isolation allows the interference of field potentials resulting from the summed activity of local neurons. Summing together responses of all three categories, an overall excitatory response pattern is observed (Fig 3B), demonstrating that the results reported in human studies likely represent the aggregate activity of populations of both suppressed and excitatory neurons during vocalization.

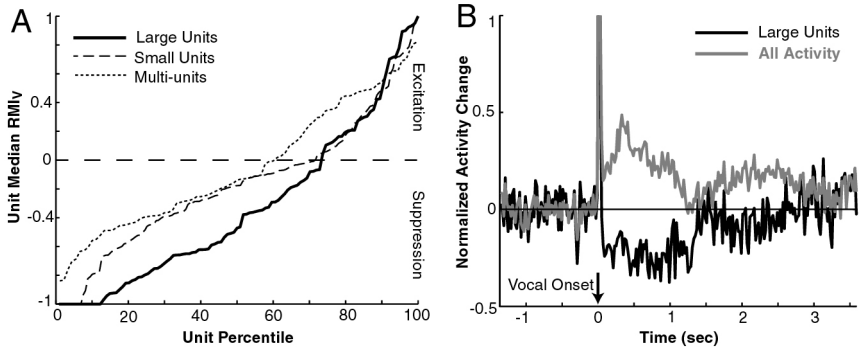


Fig. 3. A: The distribution of median response modulations index (RMI_V). "Large units": isolated single-units with larger action potential size. "Small-units": sorted single-units with smaller action potential size. "Multi-units": multi-unit clusters. B: The summed activity of all large units shows an onset response followed by sustained inhibition, while the summed activity of all three categories shows an overall excitation.

5 Functional models for auditory-vocal interaction

Sensory-motor interaction has been described in a number of neural systems. In each, a neural signal, termed an efference copy or corollary discharge, relays information from motor control areas to influence sensory activity. The exact form of this efferent signal is unclear, though it has been suggested to contain a representation of the expected sensory outcome of a motor action (Bell 1989; Poulet and Hedwig, 2003). These discharges act, almost universally, to inhibit sensory neurons. In the weakly electric fish, the best studied model of sensory-motor interaction, electric organ discharges have been shown to influence central sensory neurons through GABA-mediated inhibition (Bell 1989). Such findings parallel auditory cortical suppression and the suggested mechanisms of modulation during vocalization.

The function of efference copy mediated inhibition is twofold. First, it may play a role in separating self-generated from external sensory stimuli. Central electrosensory neurons in the fish perform a subtractive comparison between efferent and afferent signals, the output of which reflects environmental stimuli, but not the fish's own electric discharges. The cricket cercal system is suppressed during stridulation in order to prevent saturation by self-generated sounds and the resulting loss of acoustic sensitivity (Poulet et al. 2003). Efferent signal mediated inhibition is also seen in mammalian somatosensory and visual cortices and has been implicated in the processing of sensory information (Blakemore, Wolpert, and Frith 1998; Judge, Wurtz, and Richmond 1980). The auditory cortex is likely suppressed by a similar efferent mechanism, however the processing of external auditory stimuli during vocalization may not be related to such suppression. Limited evidence suggests that suppressed auditory neurons respond poorly to external sounds, while excited neurons, that presumably do not receive inhibitory inputs, respond to acoustic stimuli similarly during vocalization and quiet (Eliades and Wang 2003).

The second possible function for efferent-mediated sensory-motor interaction is self-monitoring for the control of motor behavior. Brainstem neurons in nuclei surrounding the bat lateral lemensicus are suppressed during the production of echolocation sounds (Metzner 1993), similar to what we have seen. These nuclei represent a specialized adaptation for echolocation and are involved in the control of produced spectra when presented with frequency-shifted feedback, a phenomenon known as Doppler-shift compensation (Smotherman, Zhang, and Metzner 2003). Sensory input also plays a role in controlling many other motor phenomena, including occulomotor control (Sperry 1950). While sensory feedback has an important function in regulating human and primate vocalization, the involvement of efferent signals is unclear. Subtraction of an expected sensory consequence of vocalization could result in a signal representing deviations in production or feedback, an error signal that could then be used to regulate future vocal production. Whether the suppression observed in the auditory cortex serves such a function remains to be seen. Efferent comparisons could just as easily be made in motor or vocal control centers, providing more direct behavioral access to the error information. If this were the case, suppression in the auditory cortex

might serve to properly format feedback information for later comparison. The generality of inhibition observed, including the suppression of external stimulus responses, is more consistent with this alternate hypothesis, but remains inconclusive.

6 Summary

We have observed sensory-motor interaction at the level of single neurons in the auditory cortex of primates during self-initiated vocalization. The predominant response observed was vocalization-induced suppression beginning before the onset of vocal production. Simultaneous recording of multiple neurons showed little correlation in vocalization-related modulations and this, coupled with the pre-vocal onset of suppression, suggests that inhibition is a product of local cortical circuits mediated by efferent signals from vocal control centers. A smaller fraction of neurons demonstrated excitation during vocalization, likely a result of sensory feedback. These two neural populations, suppressed and excited, may play functional roles in two important tasks during vocalization, self and environmental monitoring. Finally, by summing together the activity of neurons of all categories, we show that the resulting global cortical response during vocalization is similar to the dampened excitation during speech observed in human studies.

Acknowledgements

Our work in the marmoset model has been supported by a grant from the NIDCD, and by a Presidential Early Career Award for Scientists and Engineers (X. Wang). We thank Ashley Pistorio for assistance with animal training. Publications from our lab are available at: www.bme.jhu.edu/~xwang/papers.html.

References

Bell, C.C. (1989) Sensory coding and corollary discharge effects in mormyrid electric fish. J. Exp. Biol. 146, 229-253.

Blakemore, S.J., Wolpert, D.M. and Frith, C.D. (1998) Central cancellation of self-produced tickle sensation. Nat. Neurosci. 1, 635-640.

Burnett, T.A., Freedland, M.B., Larson, C.R. and Hain, T.C. (1998) Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 103, 3153-3161.

Crone, N.E., Hao, L., Hart, J., Boatman, D., Lesser, R.P., Irizarry, R. and Gordon, B. (2001) Electrocorticographic gamma activity during word production in spoken and sign language. Neurology 57, 2045-2053.

Eliades, S.J. and Wang, X. (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. J. Neurophysiol. 83, 2194-2207

Houde, J.F. and Jordan, M.I. (1998) Sensorimotor adaptation in speech production. Science. 279, 1213-1216.

Houde, J.F., Nagarajan, S.S., Sekihara, K. and Merzenich, M.M. (2002) Modulation of the auditory cortex during speech: an MEG study. J. Cog. Neurosci. 14, 1125-1138.

Judge, S.J., Wurtz, R.H. and Richmond, B.J. (1980) Vision during saccadic eye movements. I. Visual interactions in striate cortex. J. Neurophysiol. 43, 1133-1155.

Leonardo, A. and Konishi, M. (1999) Decrystallization of adult birdsong by perturbation of auditory feedback. Nature. 399, 466-470.

Metzner, W. (1993) An audio-vocal interface in echolocating horseshoe bats. J, Neurosci. 13, 1899-1915.

Müller-Preuss, P. and Ploog, D. (1981) Inhibition of auditory cortical neurons during phonation. Brain. Res. 215, 61-76

Osmanski, M.S., Dooling, R.J. and Venkatachalam, V. (2003) Effects of pitch-altered auditory feedback on budgerigar vocal production. Association of Research in Otolaryngology Abs. 26, 74

Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J. and Evans, A.C. (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. Eur. J. Neurosci. 8, 2236-2246.

Poulet, J.F. and Hedwig, B. (2003) A corollary discharge mechanism modulates central auditory processing in singing crickets. J. Neurophysiol. 89, 1528-1540.

Sinnott, J.M., Stebbins, W.C. and Moody, D.B. (1975) Regulation of voice amplitude by the monkey. J. Acoust. Soc. Am. 58, 412-414.

Smotherman, M., Zhang, S. and Metzner, W. (2003) A neural basis for auditory feedback control of vocal pitch. J. Neurosci. 23, 1464-1477.

Sperry, R.W. (1950) Neural basis of the spontaneous optokinetic responses produced by visual inversion. J. Comp. Physiol. Psych. 43, 482-489.

Suga, N. and Shimozawa, T. (1974) Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. Science 183, 1211-1213.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

From sound to meaning: Hierarchical processing in speech comprehension

Ingrid Johnsrude, Matt Davis, and Alexis Hervais-Adelman

Medical Research Council Cognition and Brain Sciences Unit,
{ingrid.johnsrude,matt.davis,alexis.hervais-adelman}@mrc-cbu.cam.ac.uk

1 Introduction

Cognitive models of spoken language comprehension postulate several processing stages as sound is mapped onto meaning (e.g. Gaskell and Marslen-Wilson 1997; McClelland and Elman 1986). Some of these stages, operating on sound information, may map onto the sequential, hierarchical organization observed for general auditory processing in macaques (see Kaas and Hackett 2000; Rauschecker 1998 for reviews), whereas higher-level processing stages operate upon more abstract representations of linguistic, rather than acoustic, information. However, the degree to which higher-level linguistic processes can be distinguished from less-specialized auditory and sound-form-based processes remains unclear (Remez, Rubin, Berns, Pardo, and Lang 1994; Scott, Blank, Rosen, and Wise 2000; Whalen and Liberman 1987).

In this study, we alter (distort) the specific surface properties of speech in three different ways, and use a correlational design to relate brain activity to intelligibility, using functional magnetic resonance imaging (fMRI). We operationalize "intelligibility" as the amount of a sentence that is understood: an aggregate measure of the multiple, hierarchically organized, processes involved in comprehension. Within areas that correlate with intelligibility, we can differentiate regions that are sensitive to the type of distortion used (form-dependent), and thus probably involved in acoustic analysis; and those that are insensitive to distortion type (form-independent); these areas may be involved in higher-level, linguistic processes.

2 Methods

Methods are described in detail in Davis and Johnsrude (in press), which presents data from 12 of the 27 listeners discussed here.

2.1 Stimuli

Stimuli were 190 declarative English sentences 5 to 17 words (1.7 to 4.3 seconds) long, digitized at a sampling rate of 22.1Khz. Three forms of distortion were applied to these sentences using Praat software [www.praat.org]. All three forms of distortion preserved the duration, amplitude and average spectral composition of the original sentences but markedly altered the acoustic form.

Segmented speech was created by dividing the speech waveform into short chunks at fixed intervals and replacing even-numbered chunks of speech with a signal-correlated noise version of the original speech (Bashford, Warren, and Brown 1996). Signal-correlated noise is a waveform with the same spectral profile and amplitude envelope as the original speech but consists entirely of noise, and is totally unintelligible (Schroeder, 1968). The duration of clear speech was fixed at 200ms and 500, 200, or 100 ms sections of speech were replaced by signal-correlated noise.

Noise-vocoded speech (Shannon, Zeng, Kamath, Wygonski, and Ekelid 1995) was created by dividing the speech signal between 50 and 8000 Hz into 4, 7 or 15 band-pass filtered frequency bands. Sentences were re-synthesised by replacing information in each frequency band with amplitude-modulated, bandpass noise.

Speech in noise was generated by adding a continuous speech-spectrum-noise background to sentences at three signal-to-noise ratios (-1, -4, or -6 dB). The overall amplitude of each speech-in-noise stimulus was reduced to match the amplitude of the original sentence.

Signal-correlated noise (SCN) was generated as a totally unintelligible baseline stimulus using the same algorithm as for segmented speech, but without periods of clear speech. (Schroeder, 1968).

2.2 Pilot study

In order to ensure that a continuum of intelligibility was obtained for each form of distortion, 18 native English speakers heard single stimulus sentences over closed-ear headphones (BeyerDynamic DT770) played from the soundcard of a Dell laptop PC. Participants were required to either type as many words as they could understand or to rate intelligibility (on a nine-point scale) immediately after each item. Sentences were pseudorandomly assigned to a type and level of distortion. Word-report performance (calculated as the proportion of words per sentence that were reported correctly) and rated intelligibility were averaged over 5 items per condition per subject: these were reliably correlated ($r=.99$, $p<.001$). A total of six levels of intelligibility were tested for each form of distortion. We selected three levels of each form of distortion: a low-intelligibility condition (approximately 20% of words reported correctly); a medium-intelligibility condition (65% words correct); and a high-intelligibility condition (90% correct).

2.3 Subjects

Twenty-seven right-handed volunteers aged between 18 and 42 were scanned in two experiments. All subjects were native speakers of English, without any history of neurological illness, head injury, or hearing impairment. The study was approved by the Addenbrooke's Local Research Ethics Committee and written informed consent was obtained from all subjects.

2.4 Scanning procedure

Stimuli were presented diotically using a high-fidelity auditory stimulus-delivery system incorporating flat-response electrostatic headphones inserted into sound-attenuating ear defenders (Palmer, Bullock, and Chambers 1998). To further attenuate scanner noise, participants wore insert earplugs (www.aearo.com), rated to attenuate by approximately 30 dB. Twelve subjects were asked to rate the intelligibility of each item using a four-alternative button press with their right hand, after presentation of each sentence (Davis and Johnsrude in press). The remaining 15 listened to the stimuli without performing a task.

We acquired imaging data using a Bruker Medspec (Ettlingen, Germany) 3-Tesla MR system. Echo-planar whole-brain image volumes (228 in total; resolution 2 x 2 x 4 mm) were acquired using a sparse imaging technique, in which stimuli are presented in the silent period between successive scans, minimizing acoustic interference (Edmister, Talavage, Ledden, and Weisskoff 1999; Hall, Haggard, Akeroyd, Palmer, Summerfield, Elliott, Gurney, and Bowtell 1999).

Each trial comprised a stimulus item followed by a tone pip and a single EPI volume. Stimuli were pseudorandomly drawn from the 11 experimental conditions (low- medium- and high- intelligibility conditions for each of three forms of distortion, plus signal-correlated noise and clear speech). There were 19 trials of each stimulus type and an additional 19 silent trials.

2.5 Data analysis

Data processing and analysis was accomplished using Statistical Parametric Mapping (SPM99, www.fil.ion.ucl.ac.uk/spm) Pre-processing steps included within-subject realignment, spatial normalization and spatial smoothing using a Gaussian kernel of 12 mm, suitable for random-effects analysis (Xiong, Rao, Jerabek, Zamarripa, Woldorff, Lancaster, and Fox 2000).

We first wished to identify areas within subjects in which activation correlated with intelligibility (as indexed by word-report scores from the pilot study; see Davis and Johnsrude in press). Within these intelligibility-sensitive areas, we then wished to differentiate between areas of form dependence (activation that was sensitive to the acoustic form of the stimulus) and areas of form independence (areas that responded equivalently to the different forms of distortion). In addition, we identify areas involved in a preliminary cortical stage of auditory processing as those exhibiting elevated response to signal-correlated noise over silence, without a correlation with intelligibility. Some spatial segregation among the three response types might indicate a hierarchy of processing within auditory cortices as stimulus

characteristics become more complex, such as has been observed in the macaque (Rauschecker, 1998).

Single-subject analyses in the two sets of subjects (12 subjects with task, 15 without) were followed by a random-effects analysis on all 27 in which Task was included as a factor. The significance threshold was set at $p < .05$, corrected for comparisons across the whole brain.

3 Results

The effect of Task was not significant in any of the analyses presented here; and so data from all 27 subjects are combined.

Comparison of SCN and silence across subjects yielded activation bilaterally in Heschl's gyrus and surrounding areas, consistent with recruitment of core and belt auditory cortex (even with areas sensitive to intelligibility excluded; Fig. 1a).

BOLD signal was positively correlated with word-report score in voxels along the length of the superior and middle temporal gyri in the left hemisphere, extending outwards from auditory cortex towards the temporal pole and the temporoparietal junction. Similar, less extensive, activation was observed in the right superior and middle temporal gyri. A portion of left inferior frontal gyrus also showed a positive correlation with intelligibility, as did the body of the left hippocampal complex.

These activation foci can be divided into those showing sensitivity to acoustic form (form dependence) and those that are insensitive to the acoustic properties of sound. The intelligibility-responsive region was masked by all six possible contrasts between pairs of the three distortion types. A form-dependent response (in which at least one of the 6 contrasts was significant) was observed in the superior temporal gyrus, bilaterally (Fig 1b). Intelligibility-responsive areas in which none of these contrasts reach significance at $p<.00851$ were considered to be form-independent: this response pattern was observed in the anterior middle temporal gyrus bilaterally, and in the left posterior superior temporal sulcus, left inferior frontal gyrus, left hippocampus and left precuneus (Fig. 1c).

4 Discussion

Our observation of intelligibility-sensitive regions in the lateral temporal lobe replicates and extends the findings of previous functional imaging studies (Binder, Frost, Hammeke, Bellgowan, Springer, Kaufman, and Possing 2000; Scott *et al*., 2000; Vouloumanos, Kiehl, Werker, and Liddle 2001). By using multiple, acoustically different, distortions and a correlational design, we were able to overcome an important methodological limitation of earlier studies; namely that differences in intelligibility were confounded with specific acoustic differences between intelligible and unintelligible stimuli.

The results point clearly to anatomical segregation consistent with hierarchical processing of speech. Sound (compared to silence) produced activation in the probable location of primary auditory cortex (e.g., Rademacher, Morosan, Schormann, Schleicher, Werner, Freund and Zilles 2001). Importantly, activation

here did not correlate reliably with intelligibility; instead, the bilateral temporal-lobe region in which activation correlated with intelligibility is adjacent to this initial processing area. The form-dependent portion of this intelligibility-sensitive region may include some core auditory cortex and probably includes both auditory belt and parabelt areas (and beyond), and so probably subserves more than one processing stage (e.g. Kaas and Hackett, 2000; Rauschecker, 1998), although our data cannot speak to further functional segregation.

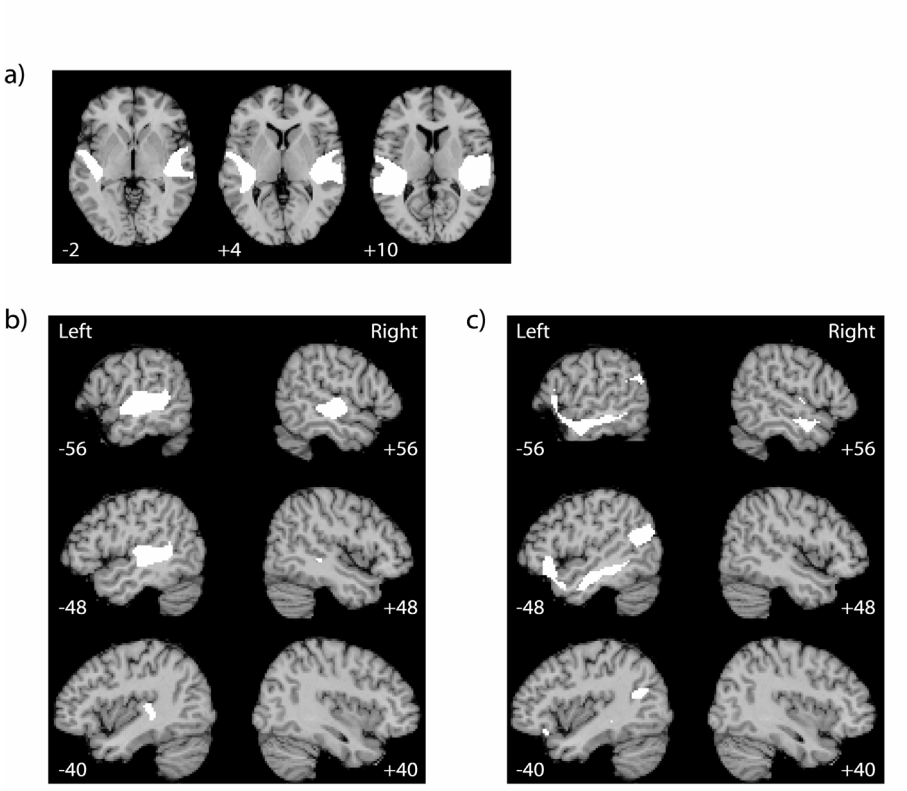


Fig. 1. Activations are shown superimposed on a canonical structural MR image from a single individual, and thresholded at $p < 0.001$, uncorrected for multiple comparisons. **(a)** Axial sections depicting areas in which activation was observed for signal-correlated noise relative to rest (excluding areas exhibiting a correlation with intelligibility) **(b)** Areas that correlate significantly with intelligibility, and show a significant difference in activation level across distortion types, indicating sensitivity to acoustic form. **(c)** Areas that correlate significantly with intelligibility, and do not differ among distortion types, indicating a lack of sensitivity to acoustic form.

Surrounding this periauditory form-dependent region anteriorly, posteriorly, and inferolaterally, we observed areas in which activation correlated significantly with intelligibility but was insensitive to acoustic differences among types of distortion. These areas may include some parabelt but are largely in what is probably polymodal cortex. We conclude that these form-independent areas are involved in processing speech at more abstract, non-acoustic levels of representation. The hierarchical structure that we infer from these results is consistent with cognitive accounts of spoken language comprehension (Gaskell and Marslen-Wilson 1997; McClelland and Elman 1986) in which lexical and semantic processes are driven by the output of lower-level acoustic and phonetic processes.

We also observed a form-independent, intelligibility-related response in left posterior superior temporal gyrus and left angular gyrus. These activations may be indicative of other, parallel streams of processing, extending posteriorly from auditory and form-dependent regions (Hickok and Poeppel 2000; Scott and Johnsrude 2003). Anatomical support for connections between auditory and inferior frontal cortex comes from studies of macaques (Hackett, Stepniewska and Kaas 1999; Romanski, Bates, and Goldman-Rakic 1999). Although the functional significance of these streams has yet to be firmly established, they may play a role in linking the perception and production of speech (Scott and Johnsrude 2003). In support of this account, a number of recent cognitive models have proposed separate processing pathways involved in phonological versus lexical processing of speech (e.g. Gaskell and Marslen-Wilson 1997).

Acknowledgements

We thank the staff of the Wolfson Brain Imaging Centre, University of Cambridge for their help with data acquisition and Matthew Brett and Ian Nimmo-Smith for advice on image processing and statistical analysis. This work was supported by the Medical Research Council of the UK

References

Bashford, J.A.J., Warren, R.M., Brown, C.A. (1996) Use of speech-modulated noise adds strong "bottom-up" cues for phonemic restoration. Percept. Psychophys. 58, 342-350.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., and Possing, E.T. (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512-528.

Davis, M.H., and Johnsrude, I.S. (in press) Hierarchical processing in spoken language comprehension. J. Neurosci.

Edmister, W.B., Talavage, T.M., Ledden, P.J., and Weisskoff, R.M. (1999) Improved auditory cortex imaging using clustered volume acquisitions. Hum. Brain Map. 7, 89-97.

Gaskell, M.G., and Marslen-Wilson, W.D. (1997) Integrating form and meaning: a distributed model of speech perception. Lang. Cog. Processes 12, 613-656.

Hackett, T., Stepniewska, I., and Kaas, J. (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. Brain Res. 817, 45-58.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., and Bowtell, R.W. (1999) "Sparse" temporal sampling in auditory fMRI. Hum. Brain Map. 7, 213-223.

Hickok, G., and Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. Trends Cog. Sci. 4, 131-138.

Kaas, J., and Hackett, T. (2000) Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. U. S. A. 97, 11793-11799.

McClelland, J.L., and Elman, J.L. (1986) The TRACE model of speech perception. Cog. Psychol. 18, 1-86.

Palmer, A.R., Bullock, D.C., and Chambers, J.D. (1998) A high-output, high-quality sound system for use in auditory fMRI. NeuroImage 7, S359.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J., and Zilles, K. (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. NeuroImage 13,669-683.

Rauschecker, J.P. (1998) Cortical processing of complex sounds. Curr. Opin. Neurobiol. 8, 516-521.

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S., Lang, J.M. (1994) On the perceptual organization of speech. Psych. Rev. 101,129-156.

Romanski, L., Bates, J., and Goldman-Rakic, P. (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J. Comp. Neurol. 403, 141-157.

Schroeder, M. R., (1968) Reference signal for signal quality studies. J. Acoust. Soc. Am. 44, 1735-1736.

Scott, S.K., and Johnsrude, I.S. (2003) The neuroanatomical and functional organziation of speech perception. Trends Neurosci. 26, 100-107.

Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J.S. (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400-2406.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., and Ekelid, M. (1995) Speech recognition with primarily temporal cues. Science 270, 303-304.

Vouloumanos, A., Kiehl, K.A., Werker, J.F., and Liddle, P.F. (2001) Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. J. Cog. Neurosci. 13, 994-1005.

Whalen, D.H., and Liberman, A.M. (1987) Speech perception takes precedence over nonspeech perception. Science 237, 169-171.

Xiong, J., Rao, S., Jerabek, P., Zamarripa, F., Woldorff, M., Lancaster, J., and Fox, P.T. (2000) Intersubject variability in cortical activations during a complex language task. NeuroImage 12, 326-339.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Effects of differences in the accent and gender of interfering voices on speech segregation

John F. Culling and Julia S. Porter

School of Psychology, Cardiff University, P.0. Box 901, Cardiff CF11 9BQ,
CullingJ@cf.ac.uk

1 Introduction

Bregman (1990) draws a distinction between "primitive" and "schema-based" grouping cues, the former being general regularities in auditory objects such as similarity in frequency, timbre or location, and the latter being learned consistencies such as the combinations of sounds that form words within a language, the structural rules of music and so forth. Research on primitive grouping has been extensive, but the potential value of schema-based grouping has been explored less (Bregman, 1990, p395).

The main evidence offered for schema-based grouping is concerned with enhancement of "stream coherence," rather than segregation from competing streams. For instance, it has been argued that sine-wave speech (Remez *et al*., 1981) is evidence for schema-based grouping. Sine-wave speech is constructed by modulating several concurrent sine waves so that they follow the formant tracks from a source utterance. Remez *et al.* (1994) argue that this effect is evidence that linguistic knowledge serves to fuse the sine waves into a unified percept. However, sine-wave speech is much more intelligible when the sine waves are grouped by a primitive grouping cue, such as coherent amplitude modulation (Carrell and Opie, 1992). The grouping cue offers no additional information to the listener, so this result suggests that the grouping afforded by their speech-like pattern of movement alone is weak. Furthermore, if listeners are presented with a segregation task in which they are expected to exploit linguistic knowledge in order to segregate one utterance of sine-wave speech from a simultaneous interfering one, then they perform very poorly (Barker and Cooke, 1999).

The present investigation examined the roles of speaker accent and vocal-tract length as grouping cues. Both of these cues have high ecological validity, since they relate to the identity of an individual speaker and individual voice characteristics may be useful to listeners in focusing their attention on that voice.

Accent is a clear case of a schema-based cue; listeners familiar with a particular accent may make use of their knowledge of the rhythmic patterning, characteristic vowel transformations and allophonic variants of a given accent. These cues may

help to group phonemes and words that have been spoken with the same accent and segregate them from phonemes and words which follow different rules of pronunciation.

Vocal tract length differs markedly across the sexes due to the lower position of the larynx in post-pubertal males, but the largest variations are due to differences in head size when comparing children and adults (Peterson and Barney, 1952). Acoustically, a shorter vocal tract results in an upward movement of all the formants in a voice. Vocal-tract length could be regarded as a primitive grouping cue, in that it affects the timbre of the voice. However, it could also be construed as a schema-based cue, since listeners may have to learn in childhood to differentiate voice timbre from phonetic content; one varied set of sounds can be the same voice, but saying different vowels, while another set can be different voices saying the same vowel. In order to decide whether such grouping is primitive or schema-based, therefore, one could determine whether newborn infants hear the same vowel produced by different vocal tracts as more similar phonetically than different vowels from the same vocal tract. This question is a methodologically difficult one to address, but Marean *et al.* (1992) have produced persuasive evidence that two-month-old infants will spontaneously generalize a learned phonetic contrast from a male to a female voice. Two-month-old infants performed similarly in this task to four- and six-month-old infants, so there was little sign of a perceptual learning process over this range of ages. This evidence supports an interpretation of vocal tract length as a potential primitive grouping cue.

We are not aware of any previous work on the effects of differences in accent on the perceptual separation of competing speech, but there is a limited amount of work on differences in vocal-tract length. Darwin and Hukin (2000) digitally manipulated vocal-tract length using a similar method to that employed here. They demonstrated that consistency of vocal-tract length across time could influence listeners' judgements of which of two simultaneous words belonged with a target carrier sentence. However, since their experiment involved the deliberate temporal alignment of alternative words, it is not clear how strong an influence such an effect may have upon the understanding of natural target speech against interfering speech, where such alignment might be a rare event. Assmann (1999) conducted a study that was better designed to answer this question, because it was based on recognition score. Although Assmann also observed an effect, he noted that the results he observed could have been the product of a spectral tilt in the stimuli. The software he used to perform the vocal-tract manipulation (STRAIGHT) used envelope scaling and this implementation produces a change in spectral tilt.

2 Experiment 1

The first experiment conducted a recognition study similar to Assmann's, but using the same software, Praat (www.praat.org), as Darwin. This software does not produce a spectral tilt, but it does move the spectrum up in frequency within a fixed spectral envelope (see Fig. 2).

2.1 Stimuli

In order to keep uncontrolled aspects of vocal quality as constant as possible, a single speaker, of Welsh origin, digitally recorded two versions each of 88 sentences from the Harvard sentence corpus (Rothauser *et al.*,1969). In one version he affected an English accent and in the other a Welsh accent somewhat stronger than his usual speech.

Recordings were made with a Sennheiser microphone (K6/ME62) in a single-walled IAC chamber. The signal was conditioned and D/A converted at 20-kHz sampling rate with Tucker-Davis Technologies equipment (MA2, FT6, DD1). The sentences were automatically trimmed by a computer program to eliminate leading and following silences, using threshold power within a sliding rectangular analysis window. The accuracy of trimming was assessed by concatenating the sentences into a continuous sequence and checking by ear for gaps or missing sections. Problem sentences were re-trimmed using altered threshold parameters. After trimming it was evident from the file durations that the Welsh-accent speech was about 20% slower. The Welsh speech was manipulated using Praat to equalize its speech rate with the English speech. The speech was analysed, dynamically time warped to equalize it with the English speech rate and resynthesised using the PSOLA technique. All sentences were then equalized in rms level.

Eighty sentences were nominated as targets and divided into eight lists of ten. The remaining eight sentences, selected to be longer than the targets, were designated as interferer sentences. Copies of the interferers were digitally manipulated using Praat to decrease the apparent vocal tract length by 13%, conveying the impression of a low-pitched female voice. This manipulation is achieved by decreasing the fundamental and the speech rate by the same factor and then resampling the waveform for a reduced sampling rate. When resynthesised and played back at the original sampling rate, the effect is to change vocal-tract length while keeping speech rate and fundamental frequency unaltered. In addition to these materials, practice sentences were drawn from standard recordings made at M.I.T. of the same sentence corpus.

2.2. Procedure

SRTs were measured for a male target voice with either a Welsh or an English accent against an interfering voice with either accent and either gender. Sounds were attenuated and mixed using a TDT AP2 array processor and then presented to listeners via a TDT psychoacoustics rig (DD1, FT6, PA4, HB6) and Sennheiser HD414 headphones in a single-walled IAC sound-attenuating chamber located in a sound-treated room. The listener made responses via a computer terminal, whose keyboard was placed within the booth and whose screen was visible through the booth window.

Sixteen listeners (Cardiff Psychology undergraduates) each attended a 75-minute experimental session. Ten speech reception thresholds (SRTs), two practice and eight experimental, were measured using the method originally described by Culling and Colburn (2000) and based upon that of Plomp (1986). The practice was intended only to familiarize the listener with the procedure. For each SRT, the same

interfering sentence was presented throughout the measurement. Initially, the first target sentence was presented against this interferer at a very adverse SNR and the listener pressed the 'return' key on the keyboard. The stimulus was repeated, each time the 'return' key was pressed at a 4-dB more favourable SNR until the listener judged that half the words of the targets sentence were audible. The listener then entered a transcript. When the listener's transcript was complete, the actual transcript was also displayed on the screen with five keywords in capital. The listener self-marked the keywords in his or her transcript and progressed to the next target sentence from a list of ten. Successive sentences were presented at different SNRs according to a 1-up/1-down adaptive threshold algorithm, which increased SNR by 2 dB if fewer then 3 keywords were correctly transcribed and otherwise decreased SNR by 2 dB. The last 8 SNRs derived in this way were averaged to yield each threshold value. The experimenter could watch the entire transaction on the computer monitor in order to ensure that the listener was performing as instructed.

2.3 Results

Figure 1 shows that SRTs were 2.5 dB lower for the female interfering voice ($F(1,15)=23.7$, $p<0.001$), suggesting that a difference in vocal tract length acts as a segregation cue. No other effects were significant. However, there was a noticeable non-significant trend towards lower thresholds for English-accent targets, regardless of interferer accent.

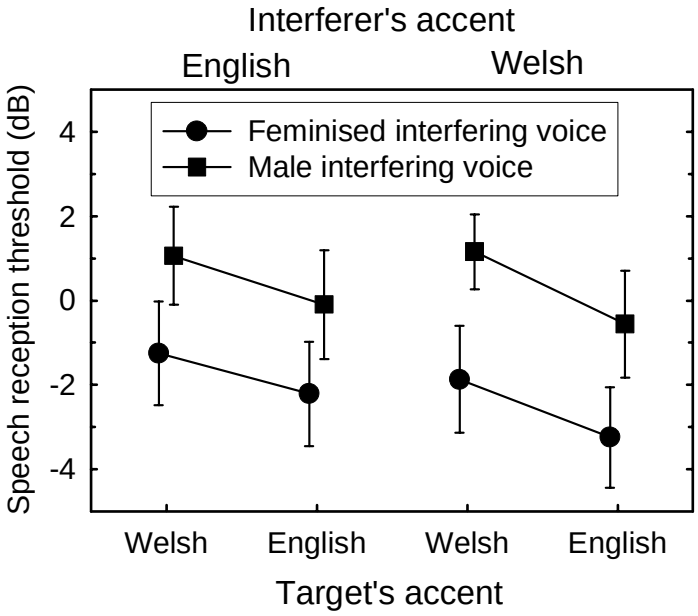


Fig. 1. Mean speech reception thresholds for English- and Welsh-accent target sentences, presented against a single English- or Welsh-accent interferer and for either the original (male) or a feminised interfering voice. Error bars are one standard error of the mean.

2.4 Discussion

The results of Expt. 1 suggest that vocal-tract length acts as a perceptual segregation cue. However, although the Praat software does not produce spectral tilts of the sort described by Assmann (1999) using STRAIGHT, it still has some effect upon the long-term spectrum. Figure 2 shows the difference in long-term excitation pattern (Moore and Glasberg, 1983) produced by the original and the feminised speech. The excitation pattern of the feminised speech is shifted upwards within a constant overall envelope. The result is that its spectral power at frequencies above about 1000 Hz is unchanged, but the spectral energy at lower frequencies is reduced. This attenuation of the masking stimulus at lower frequencies raises the possibility that the observed effect of vocal tract length difference is, in fact, attributable to a reduction in low-frequency masking.

The trend towards lower SRTs for English target sentences suggests that these were slightly more intelligible for our listeners. This trend is unsurprising given the predominance of English accents in the UK media and the high proportion of Cardiff undergraduates who come from England.

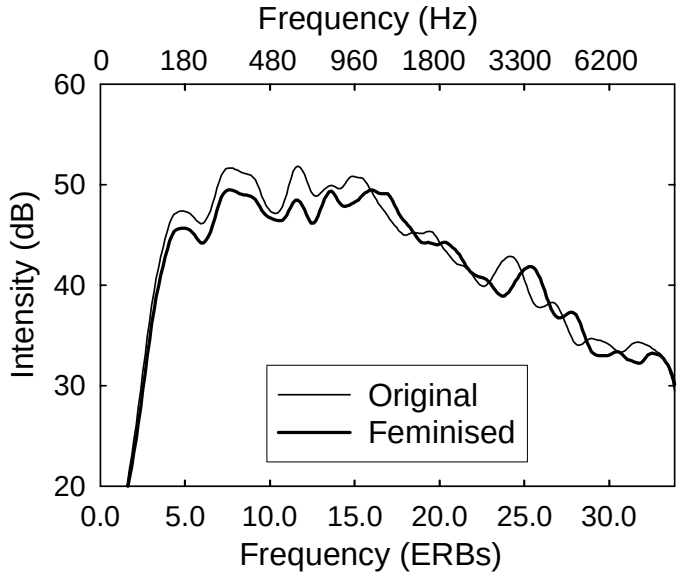


Fig. 2. Long term excitation patterns of the original (thin line) and feminised (thick line) interfering speech. All eight interferers were concatenated and analysed according to the formulae of Moore and Glasberg (1983).

3 Experiment 2

The results of Expt. 1 were somewhat ambiguous, in that, although an effect of vocal tract was found, there was a confounding difference in the spectra of the

original and feminised speech. Experiment 2 set out to quantify the effect of this spectral difference on the resulting SRTs.

3.1 Method

The long-term spectra of the original and feminized Welsh-accent interferers were used to design digital filters. These filters were applied to four different samples of Gaussian noise in order to produce speech-shaped-noise interferers of equivalent energetic masking to the original (original-noise) and feminized (feminised-noise) interferers used in Expt. 1.

SRTs were measured using the same method as Expt. 1. Twelve Cardiff undergraduates each attended an hour-long experimental session. During this time they did two practice SRTs similar to Expt. 1, but using one of the speech-shaped noise interferers, followed by eight experimental SRTs. Interferers based on the original and feminised voices were used for alternate SRTs. The pairing of original/feminised interferers with each set of target sentences also alternated.

3.2 Results and Discussion

The mean SRT in the feminised-noise condition, was significantly lower ($t(11)=3.63$, $p<0.005$) than in the original-noise condition. This 1.5-dB difference was equal to the difference in overall rms power, but is smaller than the 2.5-dB difference between original and feminised speech observed in Expt. 1. Differences in energetic masking power between the two types of interferer at low frequencies thus accounted for a substantial proportion of observed effect in Expt. 1. However, there remained a 1-dB residual effect that was not accounted for by this mechanism.

Although the effect of vocal tract length appears small from these experiments, the differences in length employed here would normally be caused by listening to competing male and female speech. Here the vocal-tract difference would be accompanied by a large difference in fundamental frequency (typically about an octave), which would provide an additional grouping cue. The present experiments did not address how vocal tract length and fundamental frequency might interact in these circumstances.

4. Conclusions

Experiments 1 and 2 indicate that differences in vocal tract length have a small, but measurable influence on the segregation of two simultaneous speech streams. On the other hand, there was no evidence from these experiments of an effect of differences in accent. Since the status of vocal-tract length as a schema-based cue is relatively uncertain, the present study can offer little support to the idea that schema-based grouping plays a substantial role in the segregation of simultaneous voices.

References

Assmann, P.F. (1999) Vocal tract size and the intelligibility of competing voices. J. Acoust. Soc. Am. 106, 2272.

Barker, J. and Cooke, M. (1999) Is the sine-wave speech cocktail party worth attending? Speech Comm. 27, 159-174.

Bregman, A.S. (1990) *Auditory Scene Analysis*. MIT Press.

Carrell, T.D. and Opie, J.M. (1992) The effect of amplitude comodulation on auditory object formation in sentence perception. Perc. and Psychophys., 52, 437-445.

Culling, J.F. and Colburn H.S. (2000) Binaural sluggishness in the perception of tone sequences and speech in noise. J. Acoust. Soc. Am. 107, 517-527.

Darwin, C.J. and Hukin, R.W. (2000) Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. J. Acoust. Soc. Am. 108, 335-342.

Marean, G.C. Werner, L.A. and Kuhl, P.K. (1992) Vowel categorization by very young infants. Developmental Psych. 28, 397-405.

Moore, B.C.J. and Glasberg, B.R. (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. 74, 750-753.

Peterson, G.E. and Barney, H.L. (1952) Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175-184.

Plomp, R. (1986) A signal-to-noise ratio method for the speech-reception SRT of the hearing impaired. J. Sp. Hear. Res. 29, 146-154.

Remez, R.E., Rubin, P.E., Berns, S.M., Pardo, J.S. and Lang, J.M. On the perceptual organization of speech. Psych. Rev. 101, 129-156.

Remez, R.E., Rubin, P.E., Pisoni, D.P., and Carrell, T.D. (1981) Speech perception without traditional speech cues. Science 212, 947-950.

Rothauser, E.H., Chapman, W.D., Guttman, N., Nordby, K.S., Silbiger, H.R., Urbanek, G.E. and Weinstock, M. (1969) I.E.E.E. recommended practice for speech quality measurements. IEEE Trans. Aud. Electroacoust. 17, 227-246.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

The Articulation Index is a Shannon channel capacity

Jont B. Allen

ECE Dept. and The Beckman Inst.
University of IL, Urbana IL.

1 Introduction

Articulation Index theory, created at Western Electric Research Labs by Harvey Fletcher in 1921, is a widely recognized method of characterizing the information-bearing frequency regions of speech (Allen, 1996). We shall show that the AI [denoted mathematically as $\mathcal{A}(snr)$] is similar to a *channel capacity*, an important concept from Shannon's information theory, defining the maximum amount of information that may be transmitted on a channel without error (Shannon, 1948).

The term *articulation*, in this context, is defined as the recognition of nonsense words. *Intelligibility* is defined as the recognition of meaningful words. Bell Labs articulation testing consisted of playing nonsense syllables, composed of 60% CVC, and 20% each of CV and VC sounds. These three types of speech sounds have been shown to compose 76% of all telephone speech (Fletcher, 1995). The use of balanced *nonsense* sounds maximizes the *entropy* of the corpus. This was an important methodology, first used around 1910 to control for context effects (Campbell, 1910), which were recognized as having a powerful influence on the recognition score. The speech corpus was held constant during these tests, to guarantee that the source entropy was constant. Even though information theory had not yet been formally proposed, these very basic concepts were clear.

The team consisted of 10 members, with 1 member acting as a *caller*. Three types of linear distortions were used, lowpass filtering, highpass filtering, and a variable *snr*. The sounds were typically varied in level to change the signal-to-noise ratio *snr*, to simulate the level variations of the telephone network.

The test consisted of the caller repeating context neutral *zero predictability* (ZP) sentences, such as "The first group is *na'v*." and "Can you hear *pōch*." All the initial consonants, vowels, and final consonants were scored, and several statistical measures were computed. For CVCs, the average of the initial $c_i(snr)$ and final $c_f(snr)$ consonant score (each score is the probability correct of identification of the nonsense phone) was computed as $c(snr) = (c_i + c_f)/2$, while the vowel recognition score was $v(snr)$. These numbers characterize the raw data. Next the data is modeled, and a *mean-CVC-syllable* score is computed from the triple product

$$\hat{S}(snr) = cvc. \tag{1}$$

Based on thousands of trials, they found that the *average nonsense phone recognition score*, defined as

$$s \equiv (2c + v)/3, \tag{2}$$

did a good job of representing nonsense CVC syllable recognition, defined as

$$S_3 \equiv s^3 \approx \hat{S}. \tag{3}$$

Similarly, nonsense CV and VC phone recognitions were well represented by

$$S_2 \equiv s^2 \approx (cv + vc)/2. \tag{4}$$

From a great number of measurements it was found that these models did a good job of characterizing the raw data (Fletcher, 1995, Figs. 175, 178, 196-218). These few simple models worked well over a large range of scores, for both filtering and noise degradations (Rankovic, 2002).

Note that these formulae only apply to nonsense speech sounds, *not* meaningful words. The exact specifications for the tests to be modeled with these probability equations are discussed in detail in Fletcher (1929, Page 259-262). The above models are necessary but not sufficient to prove that the phones may be modeled as being independent. Namely the above models follow given independence, but demonstrating their validity experimentally does not guarantee independence. To prove independence, all permutations of element *recognition* and *not-recognition* would need to be demonstrated (Bronkhorst, Bosman, and Smoorenburg, 1993).

2 Extensions to the frequency domain

Given the success of the average phone score Eq. 2, Fletcher immediately extended the analysis to account for the effects of filtering the speech into bands (Fletcher, 1921, 1929). This method later became known as *articulation index* theory, which many years later developed into the well known ANSI 3.2 AI standard. To describe this theory in full, we need more definitions.

The basic idea was to vary the signal-to-noise ratio *and* the bandwidth of the speech signal, in an attempt to idealize and simulate a telephone channel. Speech would be passed over this simulated channel, and the phone articulation $s \equiv P_c(\alpha, f_c)$ measured. The parameter α is the gain applied to the speech, used to vary the *snr*. The signal-to-noise ratio depends on the noise spectral level (the power in a 1 Hz bandwidth, as a function of frequency), and α. The consonant and vowel articulation [$c(\alpha)$ and $v(\alpha)$] and $s(\alpha)$ are functions of the speech level. The *mean phone articulation error* is $e(\alpha) = 1 - s(\alpha)$.

The speech was filtered by complementary lowpass and highpass filters, having a cutoff frequency of f_c Hz. The articulation for the low band is $s_{_L}(\alpha, f_c)$, while for the high band is $s_{_H}(\alpha, f_c)$. The nonsense syllable, word, and sentence intelligibility are $S(\alpha)$, $W(\alpha)$ and $I(\alpha)$, respectively.

Formulation of the AI. Once the functions $s(\alpha)$, $s_{_L}(\alpha, f_c)$ and $s_{_H}(\alpha, f_c)$ are known, it is possible to find relations between them. These relations, first derived by Fletcher in 1921, were first published by French and Steinberg (1947).

The key insight Fletcher had was to find a linearizing transformation of the results. Given the wideband articulation $s(\alpha)$, and the banded articulations $s_{_L}(\alpha, f_c)$ and $s_{_H}(\alpha, f_c)$, he sought a nonlinear transformation of probability $\mathcal{A}$, now called the *articulation index*, which would render the articulations additive, namely

$$\mathcal{A}(s) = \mathcal{A}(s_{_L}) + \mathcal{A}(s_{_H}). \tag{5}$$

This formulation payed off handsomely.

The function $\mathcal{A}(s)$ was determined empirically. It was found that the data for the nonsense sounds closely follows the relationship

$$\log(1-s) = \log(1-s_{_L}) + \log(1-s_{_H}), \tag{6}$$

or in terms of error probabilities

$$e = e_{_L} e_{_H}, \tag{7}$$

where $e = 1-s$, $e_{_L} = 1-e_{_L}$ and $e_{_H} = 1-s_{_H}$. These findings require $\mathcal{A}(s)$ of the form

$$\mathcal{A}(s) = \frac{\log(1-s)}{\log(e_{min})}. \tag{8}$$

This normalization parameter $e_{min} = 1 - s_{max}$ is the minimum error, while s_{max} is the maximum value of s, given ideal conditions (i.e., no noise and full speech bandwidth). For much of the the Bell Labs work $s_{max} = 0.986$ (i.e., 98.6% was the maximum articulation), corresponding to $e_{min} = 0.015$ (i.e., 1.5% was the minimum articulation error) [Rankovic and Allen (2000, MM-3373, Sept. 14, 1931, J.C. Steinberg), Fletcher (1995, Page 281) and Galt's notebooks, Rankovic and Allen (2000)].

Fletcher's simple two-band example illustrates Eq. 7: If we have 100 spoken sounds, and 10 errors are made while listening to the low band, and 20 errors are made while listening to the high band, then

$$e = 0.1 \times 0.2 = 0.02, \tag{9}$$

namely two errors will be made when listening to the full band. Thus the wideband articulation is 98% since $s = 1 - 0.02 = 0.98$, and the wideband nonsense CVC syllable error would be $S = s^3 = 0.941$.

In 1921 Fletcher, based on results of J.Q. Stewart, generalized the two-band case to $K = 20$ bands:

$$e = e_1 e_2 \cdots e_k \cdots e_K, \tag{10}$$

where $e = 1 - s$ is the wideband average error and $e_k \equiv 1 - s_k$ is the average error in one of K bands. Formula 10 is the basis of the *articulation index*. The K band case has never been formally tested, but was verified by working out many examples.

The number of bands $K = 20$ was an empirically choice that was determined after many years of experimental testing. The number 20 was a compromise that probably depended on the computation cost as much as anything. Since there were no computers, too many bands was prohibitive with respect to computation. Fewer bands were insufficiently accurate.

Each of the bands was chosen to have an equal contribution to the articulation (This represents a maximum entropy partition). Eventually they found that articulation bands, defined as having equal articulation, were proportional to cochlear critical bands. Each of the $K = 20$ articulation bands corresponds to approximately 1 mm along the basilar membrane (Fletcher, 1995). When the articulation is normalized by the critical ratio, as a function of the cochlear tonotopic axis, it was found that the articulation density per critical band, is constant (Allen, 1994, 1996). This property depends critically on the initial maximum entropy distribution of sounds used in the testing.

3 French and Steinberg (1947)

In 1947 French and Steinberg provided an important extension of the formula for the band errors by relating e_k (the k^{th} band probability of error) to the band signal-to-noise ratio SNR_k (in dB), by the relation

$$e_k = e_{min}^{SNR_k/K}, \tag{11}$$

which is the same as Eq. (10a) of the 1947 French and Steinberg paper, where SNR_k is the normalized signal-to-noise ratio, defined next.

In each articulation band the signal and noise power is measured, and the long term ratio is computed as

$$snr_k \equiv \frac{1}{\sigma_n(\omega_k)}\left[\frac{1}{T}\sum_{t=1}^{T}\sigma_s^2(\omega_k,t)\right]^{1/2}, \tag{12}$$

where $\sigma_s(\omega_k,t)$ is the short-term RMS of a speech frame and $\sigma_n(\omega_k)$ is the noise RMS, at frequency band k. The time duration of the frame impacts the definition of the *snr*, and this parameter must be chosen to be consistent with a cochlear analysis of the speech signal. It seems that the best way to established this critical duration is to use a cochlear filter bank, which is presently an uncertain quantity of human hearing (Allen, 1996; Shera, Guinan, and Oxenham, 2002). The standard method for calculating a perceptually relevant signal-to-noise ratio was specified in 1940 (Dunn and White, 1940).

Each band snr_k is converted to dB, and then limited and normalized to a range of 0 to 30, defined as

$$SNR_k \equiv \begin{cases} 0 & 20\log_{10}(snr_k) < 0 \\ 20\log_{10}(snr_k)/30 & 0 < 20\log_{10}(snr_k) < 30 \\ 1 & 30 < 20\log_{10}(snr_k). \end{cases} \tag{13}$$

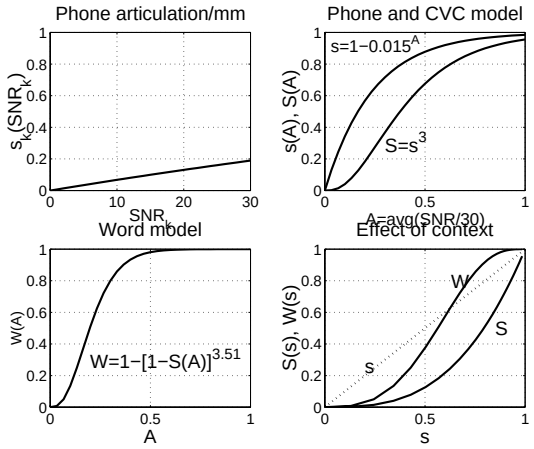


Fig. 1. Typical results for the French and Steinberg AI model, as defined in Allen (1994).

The factor 30 comes from the fact that speech has a 30 dB dynamic range in a given articulation band (French and Steinberg, 1947, Fig. 4, page 95).

The basic idea of this formula is that when the snr_k is less than 0 dB within each cochlear critical band, the speech is undetectable. When snr_k is greater than 30 dB, the noise has no effect. Between 0 and 30 dB SNR_k is proportional to $\log(snr_k)$.

Merging the formula for the total error Eq. 10 with that for the band errors SNR_k Eq. 13, the total error is related to the average SNR

$$\mathcal{A} \equiv \overline{SNR} = \frac{1}{K}\sum_k SNR_k \tag{14}$$

since

$$e = e_1 e_2 \cdots e_K = e_{min}^{\overline{SNR}} = e_{min}^{\mathcal{A}}. \tag{15}$$

The final articulation index formula, relating the articulation $s = 1 - e$ to the articulation index $\mathcal{A} \equiv \overline{SNR}$, is therefore

$$s = 1 - e_{min}^{\mathcal{A}}. \tag{16}$$

Note that as $snr_k \to 30$ dB in every band, $\mathcal{A} \to 1$ and $s \to s_{max}$. When $snr_k \to 0$ dB in all the bands, $\mathcal{A} \to 0$ and $s \to 0$. This formula for $s(\mathcal{A})$ has been verified many times, for a wide variety of conditions (Allen, 2004). However it is not perfect (Allen, 2004). Figure 1 shows typical results of articulations in a band $[s_k(SNR_k)]$, for phones $[s(\mathcal{A})]$, CVCs $[S(\mathcal{A})]$, words $[W(\mathcal{A})]$, and the effects of two types of context. For details, see (Allen, 1996, 2004).

3.1 The AI and the Channel Capacity

It is interesting that this band average is taken over the dB values $\sum_k SNR_k$ rather than the linear values $\sum_k snr_k$. This is a subtle and significant fact that has been

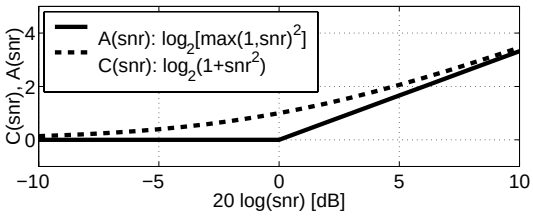


Fig. 2. Plot of $\log(1+snr^2)$ and $\log[\max(1,snr)^2]$ versus *SNR*$= 20*log(snr)$.

overlooked in discussions of the AI. The average over SNR_k, which are in log units, is proportional to the log of the *geometric mean* of snr_k, namely

$$\mathcal{A} \equiv \frac{1}{K}\sum_k SNR_k \propto \log\left(\prod_k snr_k\right)^{1/K}. \tag{17}$$

The geometric mean of the *snr* is used in information theory as a measure an abstract volume, representing the amount of information that can be transmitted by a channel. For example, if the integral is replaced by a finite sum, then the Shannon Gaussian channel capacity formula

$$C = \int_{-\infty}^{\infty} \log_2[1+snr^2(f)]df, \tag{18}$$

which is a measure of a Gaussian channel's maximum capacity for carrying information, is very similar to Eq. 14. From Fig. 2, we see that $\mathcal{A}(snr)$ is a straight–line approximation to to the Shannon channel capacity formula $C(snr)$. The figure shows the two functions $C(snr) \equiv \log_2[1+snr^2]$ and $A(snr) \equiv \log_2[\max(1,snr)^2]$, which is $\propto \mathcal{A}(snr)$.

The early idea of a channel capacity, as proposed by R. V. L. Hartley, was to count the number of intensity levels in units of noise variance (Hartley, 1928; Wozenkraft and Jacobs, 1965). This is a concept related to counting JNDs. It is interesting and relevant that Hartley, a Rhodes scholar well versed in psychophysical concepts, also proposed the decibel, which was also based on the intensity JND (Hartley, 1929, 1919). The expression

$$\log(1+snr^2) = \log\left(\frac{I+\Delta I}{I}\right) \approx \frac{\Delta I}{I}, \tag{19}$$

(the approximation holding when the ratio $\Delta I/I$ is small) where ΔI and I are the JND and intensity respectively, is closely related to counting JNDs. It has been shown, by George A. Miller (Miller, 1947), that noise is close to the first JND level if its presence changes the input stimulus by 1 dB, that is when $10\log_{10}(1+\Delta I/I) = 1$, or $\Delta I/I = 1/10$. Hence, the function $\log_2(1+snr^2)$ is related to the number of JNDs, in bits (French and Steinberg, 1947; Fletcher and Galt, 1950; Allen, 1997). The product of the number of articulation bands times the number of JNDs determines a volume, just as the channel capacity determines a volume.

References

Allen, J.B. (1994) How do humans process and recognize speech? IEEE Transactions on speech and audio, 2(4):567–577.

Allen, J.B. (1996) Harvey Fletcher's role in the creation of communication acoustics. J. Acoust. Soc. Am., 99(4):1825–1839.

Allen, J.B. and Neely, S.T. (1997) Modeling the relation between the intensity JND and loudness for pure tones and wide–band noise. J. Acoust. Soc. Am., 102(6):3628–3646.

Allen, J.B. (2004) Articulation and intelligibility. In David B. Pisoni and Robert E. Remez, editors, *Handbook of Speech perception*, chapter 6, page 80. Blackwell, Oxford, UK and Malden, MA. To appear.

Bronkhorst, A.W. , Bosman, A.J., and G.F. Smoorenburg (1993) A model for context effects in speech recognition. J. Acoust. Soc. Am., 93(1):499–509.

Campbell, G.A. (1910) Telephonic intelligibility. Phil. Mag., 19(6):152–9.

Dunn, H.K. and White, S.D. (1940) Statistical measurements on conversational speech. J. Acoust. Soc. Am., 11:278–288.

Fletcher, H. (1921) An empirical theory of telephone quality. AT&T Internal Memorandum, 101(6).

Fletcher, H. (1929) *Speech and Hearing*. D. Van Nostrand Company, Inc., New York.

Fletcher, H. (1995) Speech and hearing in communication. In Jont B. Allen, editor, *The ASA edition of Speech and Hearing in Communication*. Acoustical Society of America, New York.

Fletcher, H. and Galt, R.H. (1950) Perception of speech and its relation to telephony. J. Acoust. Soc. Am., 22:89–151.

French, N.R. and Steinberg, J.C. (1947) Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Am., 19:90–119.

Hartley, R.V.L (1919) The function of phase difference in the binaural location of pure tones. Phy. Rev., 13:373–385.

Hartley, R.V.L. (1928) Transmission of information. Bell System Tech. Jol., 3(7):535–563.

Hartley, R.V.L. (1929) "TU" becomes "DECIBEL". Telephone Engineering, 33:40.

Miller, G.A. (1947) Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. J. Acoust. Soc. Am., 19:609–619.

Rankovic, C.M. (2002) Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions (l). J. Acoust. Soc. Am., 111(6):2545–2548.

Rankovic, C.M. and Allen, J.B. (2000) *Study of Speech and hearing at Bell Telephone Laboratories: The Fletcher Years;* CDROM *containing Correspondence Files (1917–1933), Internal reports and several of the many Lab Notebooks of R. Galt*. Acoustical Society of America, Suite 1N01, 2 Huntington Quadrangle, Melville, New York.

Shannon, C.E. (1948) The mathematical theory of communication. Bell System Tech. Jol., 27:379–423 (parts I, II), 623–656 (part III).

Shera, C.A., Guinan, J.J. and Oxenham, A.J. (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc. Natl. Acad. Sci. USA, 99:3318–2232.

Wozencraft, J.M. and Jacobs, I.M. (1965) *Principles of Communication Engineering*. John Wiley, New York.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Comodulation masking release and the role of wideband inhibition in the cochlear nucleus

Ian M. Winter[1], Veronika Neuert[1], and Jesko L.Verhey[2]

[1] CNBH, Cambridge {imw1001, vn210}@cam.ac.uk
[2] University of Oldenburg, jesko@medi.physik.uni-oldenburg.de

1 Introduction

Human psychophysical studies have shown that the detection of a masked signal can be improved when the masker is coherently modulated over a wide frequency range. This phenomenon is referred to as comodulation masking release (CMR; Hall, Haggard, and Fernandes 1984). Experiments on CMR can be divided on the basis of masker type. In the first case the masker is a single bandpass noise spectrally centred at the signal frequency. In the second case the masker consists of several narrowband maskers (flanking bands); one at the signal frequency and one or more spectrally separate from the signal frequency.

Using a paradigm based on the second type of masker, Pressnitzer, Meddis, Delahaye, and Winter (2001) showed that some single units at the level of the ventral cochlear nucleus (VCN) could demonstrate a CMR. They further showed that their data were consistent with the idea that wideband inhibition was the underlying neural mechanism. However, the strongest wideband inhibition in the cochlear nucleus is found in the dorsal region (DCN) and in this study we wished to investigate whether the wideband inhibition hypothesis could be extended to the responses of single units in this region.

Here we show that for many single units in the DCN the addition of comodulated flanking bands reduced the response to the masker, and as a result the salience of the response to the signal was enhanced. These results are consistent with the hypothesis that wideband inhibition plays a role in the enhancement of signal detection in comodulated noisy environments and that this across-frequency processing takes place at a level as early as the cochlear nucleus.

2 Methods

Detailed methods may be found elsewhere (e.g. Winter and Palmer 1990; Pressnitzer *et al*. 2001) and will only be described in brief below.

2.1 Physiology

The data reported in this paper were recorded from pigmented guinea pigs anaesthetised with urethane (1.5g/kg ip). Supplementary analgesia was provided by Hypnorm (1 mg/kg im). Additional doses of urethane and the analgesic were given when required. The surgical preparation and stimulus presentation took place in a sound-attenuated chamber (IAC). All animals were tracheotomised and core temperature was maintained at 38oC with a heating blanket.

2.2 Complex stimuli

The stimuli (Fig. 1) were similar to those used in some psychophysical studies (Grose and Hall 1989; Moore Glasberg, and Schooneveldt 1990). The on-frequency masker (OFM) was a pure tone, 100% sinusoidally amplitude modulated (SAM) at a rate of 10Hz. The carrier frequency was chosen to be equal to each unit's best frequency (BF). Five modulation cycles were presented, giving a 500ms total duration. The level of the OFM was set at least 23 dB above the unit's pure tone threshold. The signal consisted of three, successive 50 ms tone pips presented in the last three dips of the OFM modulation. The first OFM dip was left without a signal to facilitate the visual interpretation of the physiological data.

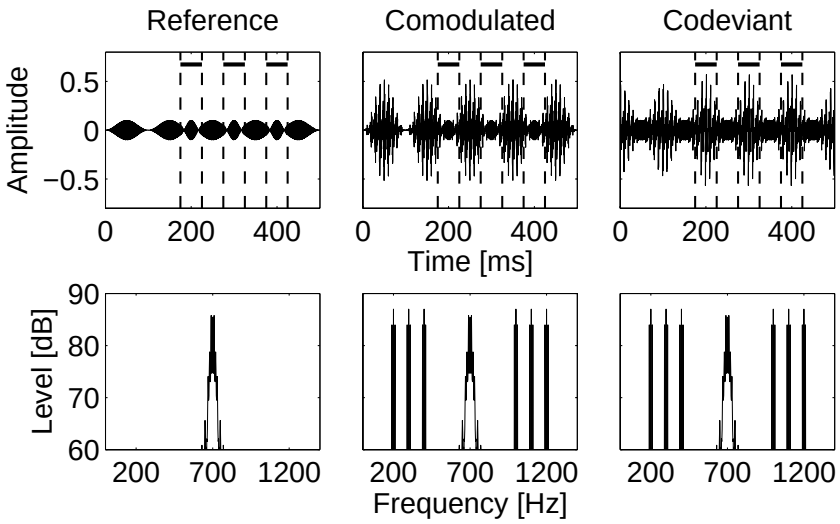


Fig. 1. Examples of the waveforms *(top ro*w) and spectra *(bottom ro*w) of the stimuli in the three different conditions: reference (*left column*), comodulated (*middle column*), and codeviant (*right column*). The signal-to-on-frequency-masker (OFM) ratio is 0 dB. The signal-to-OFM ratio is defined as the mean amplitude of the signal pip divided by the mean OFM amplitude. Signal and masker frequencies are 700 Hz. The frequency spacing of the flanking components is 100 Hz with one gap around the signal frequency.

The condition involving only the signal and the OFM is referred to as the 'reference' condition. In the 'comodulated' condition, flanking bands (FBs) were added to the OFM plus signal compound. The FBs were SAM pure tones modulated in phase with the OFM, with the same level as the OFM. In the third, 'codeviant' condition, the number and position of FBs was identical to the comodulated condition but they were amplitude-modulated 180 degrees out of phase with respect to the OFM. The codeviant condition yields higher psychophysical thresholds in humans in comparison with the reference condition (+10 dB), presumably because of across-channel masking (Moore *et al.*, 1990). Wherever possible the FBs were positioned in inhibitory regions of the unit's response area.

2.3 Analyses

Recordings were made using tungsten-in-glass microelectrodes (Merrill and Ainsworth, 1972). A wideband noise stimulus was used to locate the surface of the cochlear nucleus and to search for single units. Upon isolation of a single unit, estimates of best frequency (BF) and threshold were obtained using audio-visual criteria. The spontaneous discharge was measured over a ten-second period. Single units were classified by: (i) peri-stimulus time histogram shape in response to suprathreshold BF tone–bursts, (ii) inter-spike interval and discharge regularity, and (iii) receptive field.

The following observations suggest that all units reported in this study were located in the DCN: (i) the frequency of occurrence of a pause-build PSTH temporal adaptation patterns and type II or type IV response maps, (ii) all units were located in the appropriate position along the tonotopic axis of the cochlear nucleus (i.e. before the abrupt change from low to high best frequencies that indicates the DCN/PVCN border), and (iii) location of electrolytic lesions showed that all tracks coursed through the DCN.

3 Results

The results of this study come from the responses of 39 units in the DCN to the three stimulus conditions described above. In each case the signal and OFM were positioned at or near the units BF and the FBs were positioned in the units inhibitory sidebands.

Examples from two units are shown below. In Fig. 2 three FBs were located on either side of the best frequency with a spacing of 200 Hz and one gap below and 5 gaps above the signal frequency (i.e. 0.5, 0.7, 0.9, 2.5, 2.7, and 2.9 kHz). Both the OFM and FBs were set to a level of 48 dB above pure-tone threshold. This unit was classified as type II according to receptive field classification schemes.

In the comodulated condition a large response is observed at the temporal position of the signal whereas the response to the masker alone is considerably smaller than in the reference condition. For the two levels shown, the addition of the signal does not change the unit's response in the codeviant or reference condition.

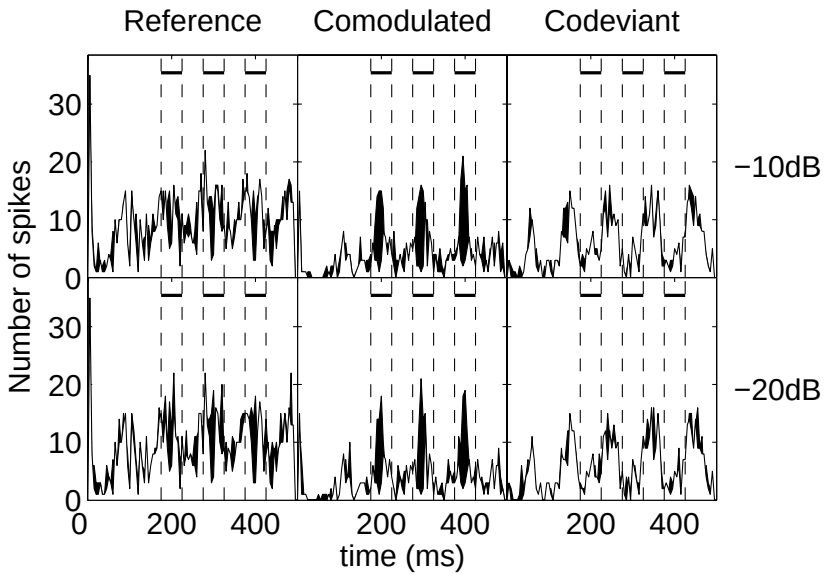


Fig 2. Poststimulus time histograms of the response to the three stimulus conditions for a Type II unit (1064009010). The binwidth is 5 ms. The unit's best frequency was 1.3 kHz. The signal and OFM frequencies were set to the best frequency of the unit. Responses to the reference, comodulated, and codeviant stimulus condition are shown in the *left*, *middle*, and *right column*s, respectively. The white area bordered with a thin black line shows the units response to the masker alone. The black areas show the change in the units response when a signal is added to the masker at the level of –10 dB (*top row*) and –20 dB signal-to-OFM level (*bottom row*). The temporal positions of the signal pips are indicated by the *dashed vertical lines* connected by *solid horizontal lines*.

The responses of a second unit are shown in Fig. 3. Again a clear response to the signal was seen in the comodulated condition. For this Type III unit there was a small, but noticeable increase in the response to the signal in the reference condition. This is in contrast to the codeviant condition where little or no response to the signal was evident in the PSTHs. However, perhaps the most noticeable change, for both units, was the reduced response to the OFM in the comodulated condition in comparison with the reference condition. A similar pattern of results was observed for all units in the DCN that showed an enhanced representation of the signal in the comodulated condition (24/39). The majority of these 24 units were classified as type II or type III according to the receptive field classification scheme.

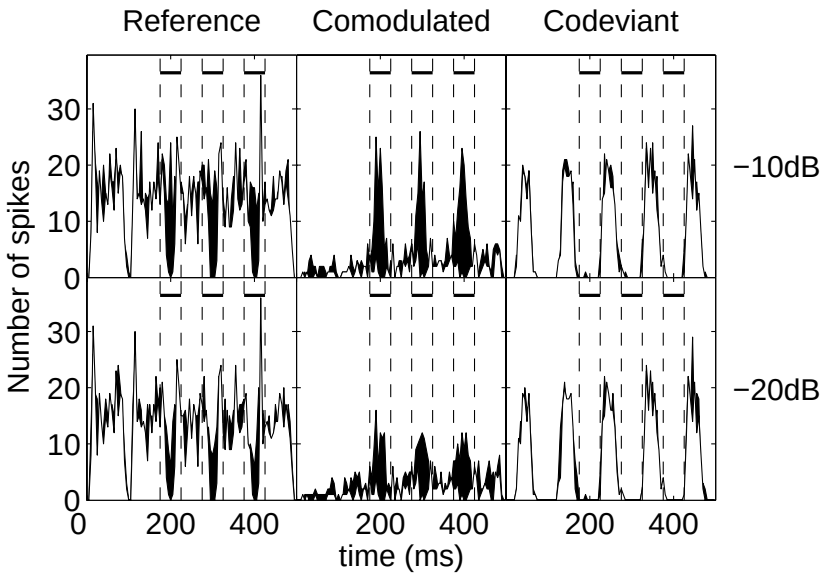


Fig. 3. Same format as Fig. 2 but for a type III unit (1042007025). The unit's best frequency was 1.8 kHz. The signal and OFM frequencies were set to the best frequency of the unit. Both the OFM and FBs were set to a level of 44 dB above pure-tone threshold. Three FBs were located on either side of the best frequency with a spacing of 200 Hz and 4 gaps (0.4, 0.6, 0.8, 2.8, 3.0, and 3.2 kHz).

4 Discussion

We have recorded from single units in the DCN of the anaesthetized guinea pig to investigate the effect of inhibition on signal detection in noise. We have shown (e.g. Figs 2 and 3) that the response to a modulated masker was strongly suppressed by the addition of comodulated FBs positioned in a unit's inhibitory sidebands. The suppression of the response to the masker was primarily responsible for making the neural response to the signal more salient in the comodulated condition.

In an earlier study in the VCN (Pressnitzer *et al.* 2001), we showed that a population of transient chopper units could also show a release from masking. This release, however, was relatively small and only occurred at relatively high signal/OFM ratios. Pressnitzer *et al.* (2001) also observed a group of onset units that showed the opposite type of response to the comodulated stimulus, i.e. a poor response to the signal but a strong response to the modulation. It was hypothesized that onset units might be performing a subtraction of the modulation from the transient chopper units. A computational model based on this idea was able to replicate the essential features of the physiology.

The hypothesis that wideband inhibition of chopper units contributes to CMR is supported by physiological (Ferragamo and Oertel 1998) and anatomical evidence

(Arnott, Wallace and Palmer 2001). That WBI units inhibit the principal cells in the DCN has long been hypothesized to help explain the exquisite notch sensitivity of these units (for a review, see Davis and Young 2002) and in this sense is not controversial (however, direct evidence for this interaction is still lacking).

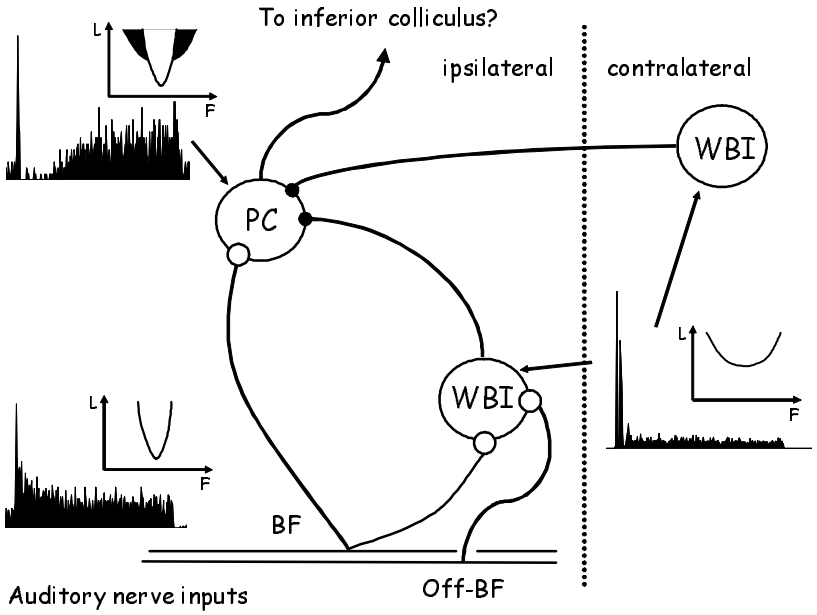


Fig. 4. Hypothesised circuit underlying across-frequency processing in the DCN of the guinea pig. The excitatory input to the circuit is provided by auditory-nerve fibres with a primary-like PSTH temporal adaptation shape (bottom left). The wideband inhibition (WBI) is believed to come from OC units while the PC (principal cell) in the DCN can be a variety of temporal adaptation shapes, including the Pauser PSTH shown in the upper Left. Schematic receptive fields (Frequency, F, versus Level, L) are shown inset on the PSTHs.

The simple circuit proposed by Pressnitzer et al (2001) is modified in Fig. 4 by replacing the transient chopper unit from the VCN with a pauser unit from the DCN. We do not imply that all pauser units have type III receptive fields. There is increasing anatomical evidence that the WBI units also project to the contralateral cochlear nucleus and this projection is added in Fig. 4. We have previously speculated that this contralateral projection may contribute to the CMR observed when the signal and OFM are in one ear and the flanking bands are placed in the opposite ear (Verhey, Pressnitzer, and Winter 2003). It is clear from our results that the cochlear nucleus is able to provide an enhanced representation of a signal when presented in comodulated noisy backgrounds. Whether this observation underlies or contributes to psychophysical CMR remains to be determined.

Acknowledgments

We thank the Wellcome Trust for supporting this work.

References

Arnott RH, Wallace, MN., Palmer, AR (2001) Innervation of the ventral and dorsal cochlear nuclei by an onset cell in the anteroventral cochlear nucleus. Brit. J. Audiol. (in press).

Ferragamo, MJ, Golding NL,, Oertel D (1998) Synaptic inputs to stellate cells in the ventral cochlear nucleus. J. Neurophysiol. 79:51-63.

Grose JH, Hall JW (1989) Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position. J. Acoust. Soc. Am. 85: 1276-1284.

Hall, JW, Haggard, MP, Fernandes MA (1984) Detection in noise by spectro-temporal pattern analysis J. Acoust. Soc. Am. 76:50-56.

Meddis, R., Delahaye, R, O'Mard, L, Sumner, C, Fantini D, Winter, IM, Pressnitzer D. (2002) Acta Acustica united with acustica 88, 387-398.

Merrill, EG, Ainsworth, A (1972) Glass coated platinum tipped tungsten microelectrodes. Med. Biol. Eng. 10: 662-672.

Moore, BCJ, Glasberg, BR, Schooneveldt, GP (1990) Across-channel masking and comodulation masking release. J. Acoust. Soc. Am. 87:1683-1694.

Pressnitzer, D., Meddis, R. Delahaye, R. Winter, IM (2001) Physiological correlates of comodulation masking release in the ventral cochlear nucleus. J. Neurosci. 21, 6377-6386

Verhey, JL Pressnitzer., D, Winter, IM (2003) The psychophysics and physiology of comodulation masking release. Exp. Brain Res. Submitted.

Winter, IM., Palmer, AR. (1990a) Responses of single units in the anteroventral cochlear of the guinea pig. Hear. Res. 44:161-178.

Young ED, Davis, K (2002) Circuitry and function of the dorsal cochlear nucleus. In *Integrative functions in the mammalian auditory pathway.*Chapter 5. Springer handbook of Auditory Research Vol. 15 Ed. Oertel, D, Fay RR and Popper, AN.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

The relevance of rate and time cues for CMR in starling auditory forebrain neurons

Georg M. Klump[1] and Sonja B. Hofer[1,2]

[1] Zoophysiology & Behaviour Group, Oldenburg University, 26111 Oldenburg, Germany, georg.klump@uni-oldenburg.de
[2] Max-Planck-Institute for Neurobiology, Am Klopferspitz 18a, 82152 Muenchen-Martinsried, Germany, sonja.hofer@neuro.mpg.de

1 Introduction

Psychoacoustic experiments have demonstrated that the auditory system of humans, of other mammals and of birds can exploit temporal coherence of amplitude fluctuations at different frequencies of a masking sound to improve the detection of a tone signal (e.g., Hall et al. 1984, Schooneveldt and Moore 1987, 1989, Wagner and Klump 2001, Klump and Langemann 1995). This effect has been termed "Comodulation Masking Release" (CMR, Hall et al. 1984). Since in the natural acoustic environment masking background noise often has marked envelope fluctuations (e.g., Nelken et al. 1999), CMR may be relevant for communication by increasing the distance over which signals can be perceived.

Psychoacoustic studies have suggested that two types of cues contribute to CMR (for a review see Moore 1992). On the one side, the auditory system can integrate the information across different auditory filters, i.e., use across-channel cues to achieve a release from masking. The underlying mechanism would require a comparison of the masker envelope in separate frequency channels. For example, correlated fluctuations of the masker envelope at various frequencies that are represented in different frequency channels of the auditory system would make signal detection more easy for the auditory system than in maskers with uncorrelated envelope fluctuations since the correlated masker envelope fluctuations could be cancelled out (e.g. Buus 1985). Psychoacoustic studies of CMR indicate, however, that only in some experimental conditions across-channel cues contribute substantially to CMR. Physiological investigations suggested that neural circuits in the cochlear nucleus could be involved in the across-channel comparisons (e.g., Pressnitzer et al. 2001). On the other side, it has been suggested that cues restricted to one frequency channel of the auditory system (i.e., within-channel cues) may be sufficient to largely explain the CMR effect in many experimental conditions (e.g. Schooneveldt and Moore, 1987, 1989, Verhey et al. 1999). Adding the signal to the masker presented in the same channel changes the

temporal pattern of envelope fluctuations. This change, if represented in the response of neurons in the auditory system, could provide the cue for signal detection (Schooneveldt and Moore 1987, 1989). Furthermore, a model proposed by Verhey et al. (1999) suggests that a combination of peripheral frequency filtering of the carrier and central filtering of the envelope of a sound by modulation filters may explain CMR by contributions of within-channel cues only. Results obtained in the primary auditory cortex of the cat demonstrating that neurons respond markedly differently to sounds with a more or less deeply modulated envelope have been interpreted as a neurophysiological correlate of CMR (Nelken et al. 1999).

Here we compare the relevance of different cues for generating a CMR effect in the responses of neurons in an area of the auditory forebrain of a bird, i.e., in field L2 of the European starling that corresponds to the mammalian primary auditory cortex. We apply a masking paradigm with two narrow-band noise maskers that have been used to study CMR psychoacoustically in humans (McFadden 1986, Schooneveldt and Moore 1987) and starlings (Klump et al. 2001). Here we concentrate on the analysis of single-unit responses. A more detailed comparison of multi-unit and single-unit data is provided by Hofer and Klump (2003).

2 Methods

A microdrive with four metal electrodes was chronically implanted in the auditory forebrain of wild-caught adult European starlings, *Sturnus vulgaris*, under halothane anesthesia (all procedures were in agreement with the local regulations for the treatment of the experimental animals). Both epoxide-insulated tungsten electrodes (FHC; 6 – 10 MΩ) and custom-made electrodes from teflon-coated platinum-iridium wires (A-M Systems; 4-10 MΩ) with an electro-chemically sharpened tip were used. By adjusting the microdrive, the electrodes could be positioned at multiple recording sites in the auditory forebrain area L2. The neurons in area L2 that receive projections from the auditory thalamus typically show primary-like phasic-tonic responses and the excitatory frequency-tuning curves are bordered by frequency areas in which the neural response is suppressed by tones (Nieder and Klump 1999). Multi-unit activity was recorded via radiotelemetry (FHC FM-transmitter and commercial FM-tuner) from the awake unrestrained birds in a cage inside a sound-proof echo-reduced booth (IAC 403A). Spikes from multi-unit recordings with a spike amplitude that was at least 3 times larger than the background activity were sorted into single-unit data applying a Baysian approach (Lewicki 1994).

The acoustic signals presented to the forebrain neurons time-locked with the recording of their multi-unit activity were 800-ms pure tones (including 50-ms Hanning ramps, 2s inter-stimulus interval, varying in amplitude from 0 to 70 dB SPL) of a frequency that was close to the neurons' characteristic frequency. 25-Hz-wide bands of noise were continuously presented as the masker. The masker level was between 10 and 20 dB above the neurons' response threshold. One band of masking noise was centered on the tone signal (on-frequency band, OFB). The OFB

was presented alone or together with a flanking band (FB) that was either close to the OFB (i.e., its center frequency differed by 0.5 critical bands from the tone frequency and was within the limits of the neuron's excitatory tuning curve) or that was remote from the signal frequency (i.e., its center frequency differed by 4 critical bands from the tone frequency positioning the FB in the suppression areas of the tuning curve). In the correlated condition, OFB and FB had identical envelopes that were produced by multiplying the same low-passed noise with sine waves of two different frequencies. In the uncorrelated condition, the envelopes of OFB and FB varied independently (i.e., different low-passed bands of noise were used to define the envelopes).

The data analysis was based on 10 artifact-free recordings of the neuronal response at a certain combination of tone amplitude and masker condition (for an example of the neuronal response pattern see Fig. 1). Rate thresholds of the neurons were determined by summing up the spikes over a time window of 100, 400 or 800 ms that started with tone onset and was corrected for the neurons' minimum latency. The neurons' activity driven by the masker alone was used as a baseline, and the tone threshold was defined as the sound-pressure level of the signal resulting in an activity that was 1.8 standard deviations above the baseline activity.

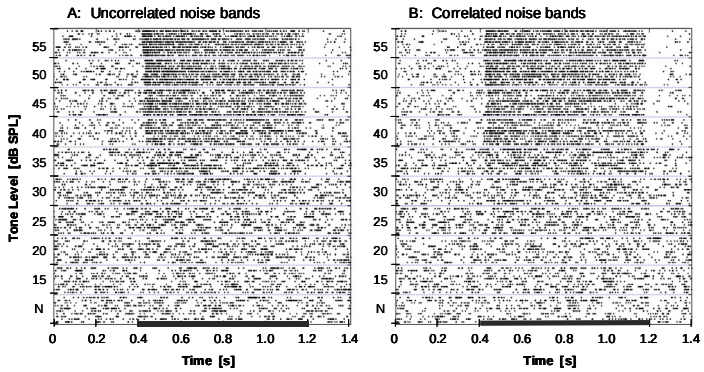


Fig. 1. Rasterplot showing the response of a starling forebrain neuron when presented with 800-ms tones (presentation marked by black bar) in a continuous noise masker with OFB and FB components exciting the neuron. The sub-panel marked N shows the response to the noise alone. Tone driven responses are obvious at a lower level in the correlated masker.

CMR[U-C] was calculated as the threshold difference between the correlated and uncorrelated OFB/FB condition. CMR[R-C] was calculated as the threshold difference between the correlated OFB/FB and the reference condition (OFB alone). To evaluate the neurons' potential to reflect the temporal properties of the stimulus envelope in their temporal response, periodograms (Matlab) were computed from the spike times in an 800-ms time window that was adjusted to the start of the signal as described above. A Friedman test comparing the amplitude of the spike periodogram for frequencies found in the masker envelope (5 to 15 Hz) and frequencies not found in the masker envelope (200-215 Hz) was used as one means to estimate the neurons ability to encode the slow envelope fluctuations. If a

neuron would be able to encode the envelope of the masker, a higher amplitude of the periodogram would be expected at the envelope frequencies than at other frequencies (e.g. as is evident in Fig. 2 showing a periodogram of a neuron's response driven by the masker alone). To test whether a change in the shape of the periodogram would occur if a signal was presented (indicating that the neurons could use this temporal cue to detect a signal) periodograms normalized to their average amplitude for masker alone and for masker plus signal were compared using a Kolmogoroff-Smirnov test.

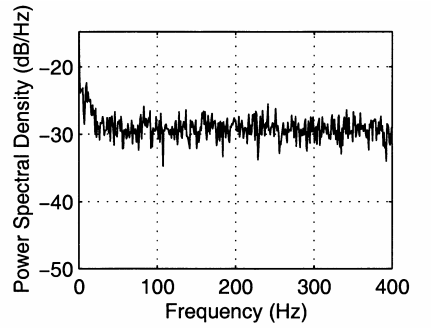


Fig. 2. Averaged periodogram of the spike trains recorded during ten 800-ms intervals when a masker with correlated OFB and FB positioned in the excitatory area of the neuron's tuning curve was presented without an embedded signal. The power spectral density is increased at frequencies below 25 Hz suggesting a representation of the masker envelope in the spike train periodicity.

3 Results and Discussion

CMR is evident in the neurons' rate response integrated over different time windows (Fig. 3). If the FB was positioned in the excitatory frequency range of the neuron's tuning curve, the neuronal signal-detection threshold in general was lower for pairs of correlated maskers than for pairs of uncorrelated maskers (i.e., a significant CMR[U-C] could be observed in the rate response). Adding a correlated FB to the OFB masker, however, did not result in a release from masking in this situation (i.e., no significant CMR[R-C] could be observed in the rate response). If the FB was presented in the area of the frequency-tuning curve indicating a suppression of the response to tones (probably due to inhibition acting across frequencies, see Nieder and Klump 1999) and the OFB and signal were presented close to the frequency at which the neuron responded most sensitively to an excitation, no significant CMR[U-C] was observed. A significant CMR[R-C] was only observed when integrating the neurons' response over a 400 ms time window and the FB was positioned in the suppression area found at frequencies below the excitatory area. In summary, CMR is evident in the neurons' rate response integrated over longer time windows. CMR[U-C] in the starling forebrain neurons does not appear to rely on suppressive effects across frequencies, whereas suppression at least for one condition appears to affect CMR[R-C]. This is in contrast to psychoacoustical data in the starling showing both CMR[U-C] and CMR[R-C]. How CMR relates to the neurons' activation by the masker is discussed in detail by Hofer and Klump (2003) analyzing predominantly multi-unit responses.

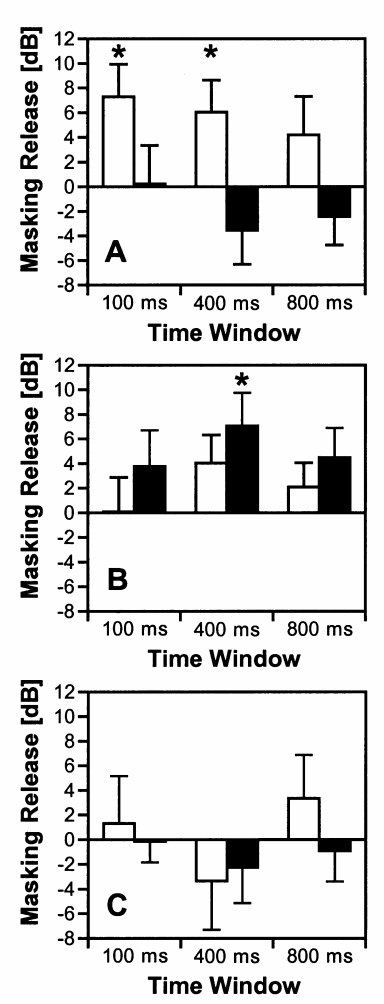


Fig. 3. Masking release of single units in the area L2 of the starling's auditory forebrain calculated as the difference between the neuronal detection threshold for a tone embedded in a masker with uncorrelated OFB/FB and in a masker with correlated OFB/FB (white bars) or as the threshold difference for neuronal detection of a tone in an OFB-only masker and in a masker with correlated OFB/FB (black bars). Error bars indicate standard error of the mean; stars indicate significant differences. Data are shown separately in each graph for three different time windows over which the neurons' spikes were integrated in the analysis. Sample sizes range from 15 and 23 cells at the various conditions. The OFB was always centered on the signal and presented close to the frequency at which the neurons responded most sensitively to an excitatory response. In the graph labeled (A) both the OFB and FB were presented at excitatory frequencies. In the graphs labeled (B) and (C) the FBs were presented in the lower and upper suppression areas, respectively, that bordered the excitatory tuning curve of the neurons.

Although neurons in the field L complex of the starling auditory forebrain have been shown to be able to synchronize well to the envelope of a sinusoidally amplitude-modulated signal up to frequencies of 100 Hz (above this frequency the synchronization of the response to the envelope deteriorates, see Knipschild et al. 1992), the shape of the periodograms provided no evidence of a systematic change that was related to adding the signal to the masker (neither the analysis with the Friedman test nor with the KS-test showed increasing deviations in the periodigram shape with increasing signal level). Thus, temporal within-channel cues appear not to be represented in the response of the forebrain neurons.

Acknowledgments

The study was funded by grants from the DFG (FG 306) and the University of Oldenburg.

References

Buus, S. (1985) Release from masking caused by envelope fluctuations. J. Acoust. Soc. Am. 76, 1958-1965.

Hall, J.W., Haggard, M.P. and Fernandes, M.A. (1984) Detection in noise by spectro-temporal pattern analysis. J. Acoust. Soc. Am. 76, 50-56.

Hofer, S.B and Klump, G.M. (2003) Within- and across-channel processing in auditory masking: A physiological study in the songbird forebrain. J. Neurosci., in press

Lewicki, M.S. (1994) Bayesian modeling and classification of neural signals. Neural Comp. 6, 1005-1030.

McFadden, D. (1986) Comodulation masking release: effects of varying the level, duration and time delay of the cue band. J. Acoust. Soc. Am. 80, 1658-1667.

Moore, B.C.J. (1992) Across-channel processes in auditory masking. J. Acoust. Soc. Jpn. 13: 25-37.

Nelken, I., Rotman, Y. and Yosef, O.B. (1999) Responses of auditory-cortex neurons to structural features of natural sound. Nature 397, 154-157.

Nieder, A., Klump, G.M. (1999) Adjustable frequency selectivity of auditory forebrain neurons recorded in freely moving songbird via radiotelemetry. Hear. Res. 127, 41-54.

Klump, G.M., Langemann, U., Friebe, A. and Hamann, I. (2001) An animal model for studying across-channel processes: CMR and MDI in the European starling. In: Breebart, D.J., Houtsma, A.J.M., Kohlrausch, A., Prijs, V.F., Schoonhoven, R. (eds.) *Physiological and psychophysical bases of auditory function.* Shaker Publishing BV, Maastricht, pp. 266-272.

Knipschild, M., Dörrscheidt, G.J. and Rübsamen, R. (1992) Setting complex tasks to single units in the avian auditory forebrain. I: Processing of complex artificial stimuli. Hear. Res. 57, 216-230.

Pressnitzer, D., Meddis, R., Delahaye, R. and Winter, I.M. (2001) Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J. Neurosci. 21, 6377-6386.

Schooneveldt, G.P. and Moore, B.C.J. (1987) Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band. J. Acoust. Soc. Am. 82, 1944-1956.

Schooneveldt, G.P. and Moore, B.C.J. (1989) Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration. J. Acoust. Soc. Am. 85, 273-281.

Wagner, E. and Klump, G.M. (2001) Comodulation masking release in Mongolian gerbils (*Meriones unguiculatus*) studied with narrow-band maskers. In: Elsner, N. and Kreutzberg, G.W. (eds) *Göttingen Neurobiology Report 2001.* Thieme-Verlag, Stuttgart New York, 415.

Verhey, J.L., Dau, T., and Kollmeier, B. (1999) Within-channel cues in comodulation masking release (CMR): experiments and model predictions using a modulation-filterbank model. J. Acoust. Soc. Am. 106, 2733-2745.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Effects of concurrent and sequential streaming in comodulation masking release

Torsten Dau[1], Stephan D. Ewert[1], and Andrew J. Oxenham[2]

[1] Carl von Ossietzky Universität Oldenburg, Medizinische Physik, Oldenburg, Germany
[2] Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

1 Introduction

Across-frequency comparisons of temporal envelopes are likely to be a general feature of auditory pattern analysis and may play an important role in extracting signals from noisy backgrounds, or in separating competing sources of sound. Comodulation of different frequency bands in background noise facilitates the detection of tones in noise, a phenomenon known as comodulation masking release (CMR). There has been recent interest in elucidating the possible physiological basis for CMR (Nelken, Rotman, and Bar Yosef 1999; Pressnitzer, Meddis, Delahaye, and Winter 2001) and in confirming its presence in non-human species (e.g., Klump and Langemann 1995).

CMR has been measured in two ways. The first is to use a single band of noise, centered around the signal frequency, as a masker and to compare thresholds for modulated and unmodulated maskers as a function of the masker bandwidth (e.g., Hall, Haggard, and Fernandes 1984; Haggard, Hall, and Grose 1990; Schooenveldt and Moore 1989; Carlyon, Buus, and Florentine 1989). The second method is to use a masker consisting of several narrow masker bands, one at the signal frequency (on-frequency band) and one or more other bands (flanking bands) spectrally separated from the on-frequency band (e.g., Hall *et al.* 1984; Hall, Grose, and Haggard 1990; Schooneveldt and Moore 1987).

CMR is usually assumed to depend on comparisons of the outputs of different auditory filters. However, especially in conditions using a single band of noise, within-channel cues can also facilitate the detection of a signal in modulated noise, leading to a measured CMR without any necessary involvement of across-channel mechanisms (Schooneveldt and Moore 1989; Verhey, Dau, and Kollmeier 1999). In order to establish a physiological basis for CMR, it is important to separate within-channel from across-channel effects. Similarly, it is important to determine the extent to which CMR is affected by auditory grouping processes, which are unlikely to be peripheral in nature. For instance, Pressnitzer *et al.* (2001) recently presented physiological data at the level of the ventral cochlear nucleus (VCN) in

support of a possible physiological implementation for a model of CMR at a relatively peripheral processing stage. In contrast, at least one psychophysical study has suggested that CMR may interact with auditory object formation, commonly associated with higher-level processes (Grose and Hall 1993).

The present study attempts to clarify the extent to which CMR reflects within- and/or across-channel processes by assessing the influence of concurrent and sequential streaming cues on CMR. Three spectral configurations were considered: (1) A "narrowband" configuration where all flanking bands were located close to the on-frequency band; (2) a "broadband" condition where the same number of bands were widely separated in frequency; and (3) an "intermediate" configuration which was very similar to the one investigated by Grose and Hall (1993). The hypothesis was that perceptual segregation or grouping may affect across-channel CMR while it should not affect within-channel CMR. If so, the differential effect of grouping cues may provide a functional definition of within- and across-channel CMR.

2 Method

2.1 Subjects

Five normal-hearing listeners ranging in age from 25-39 years participated. All received several hours of listening experience in CMR tasks prior to the final data collection. Two of them were the first and the second author.

2.2 Stimuli and procedure

The signal was a 1000-Hz pure tone, 187.5 ms in duration including 20-ms raised-cosine ramps. The composite noise masker consisted of five bands of noise 20 Hz wide. Figure 1 shows the different stimulus configurations. In the narrowband configuration, the noise bands were centered at 794, 891, 1000, 1123 and 1260 Hz representing a sixth-octave spacing around the signal frequency. In the broadband condition, the noise bands were centered at 250, 500, 1000, 2000 and 4000 Hz, covering a frequency range of 4 octaves with a one-octave spacing between the bands. In the intermediate configuration, the masker consisted of seven noise bands centered at the 3rd through 15th odd harmonics of 125 Hz, i.e., at 375, 625, 875, 1125, 1375, 1625, and 1875 whereby the signal frequency was at 1125 Hz. In all configurations, the noise bands were generated in the time domain and restricted to the appropriate bandwidth in the Fourier domain. Comodulated noises were frequency-shifted versions of the on-frequency band. The spectrum level of the noise was 60 dB SPL/Hz.

Thresholds for the 1000-Hz tone masked by the noise band centered at 1000 Hz were measured in nine conditions, many of which are illustrated in Fig. 2: (1) Single band condition (SB), where the on-frequency band was presented alone; (2) Random flanking bands (R); (3) Comodulated flanking bands (C); (4) Random flanking bands with fringe (FR), where the flankers were gated on earlier and gated off later (by 100 ms) than the on-frequency band; (5) Comodulated flanking bands

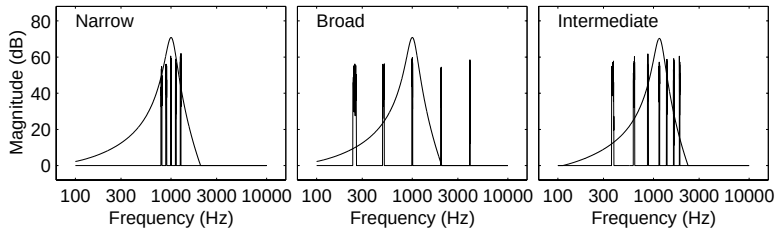


Fig. 1. Left: Narrowband configuration where the masker bands were closely spaced in frequency around the signal frequency. Middle: Broadband configuration where a larger spacing between the components was used. Right: Intermediate configuration with a masker as that used in Grose and Hall (1993). In addition, the magnitude transfer function of the Gammatone filter at the signal frequency is shown in all panels. See text for details.

with fringe (FC); (6) Random flanking bands, each preceded by four precursor bands (PR); (7) Comodulated flanking bands preceded by the precursors (PC); (8) Random flankers followed by four "postcursors" (PoR), or following bands and (9) Comodulated flankers followed by four postcursors (PoC). The four pre- or postcursors were all the same duration as the on-frequency band and were separated by gaps of 62.5 ms, giving an overall repetition rate of 250 ms.

An adaptive three-interval 3AFC procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point of the psychometric function. The initial step size was 8 dB, which was reduced to 4 and 2 dB after the second and fourth reversals, respectively. Threshold was defined as the mean of the levels at the last six reversals of a threshold run. Four threshold estimates were obtained from each listener in each condition.

3 Results

The left panel of Fig. 3 shows the mean data for the nine conditions. The filled circles indicate thresholds obtained for the narrowband configuration and the open squares represent thresholds for the broadband configuration. The triangle in the leftmost condition of the left panel represents the threshold for the single on-frequency band (SB). The right panel of Fig. 3 shows the threshold differences obtained with random versus comodulated flanking bands. These differences represent the amount of CMR.

In the narrowband configuration (filled circles), thresholds are always lower in the comodulated condition than in the corresponding random condition (left panel), leading to consistently positive CMRs (right panel). The effect of condition was investigated using a one-way within-subjects analysis of variance (ANOVA). Using all conditions except SB, a highly significant main effect was found; $F(7,28) = 48.97$, $p < 0.001$. Post hoc comparisons based on Tukey's honestly significant difference (HSD) criterion showed a highly significant difference between each of the comodulated conditions and the corresponding random condition ($p < 0.001$). In contrast, there were no significant differences between thresholds within either the comodulated conditions or the random conditions.

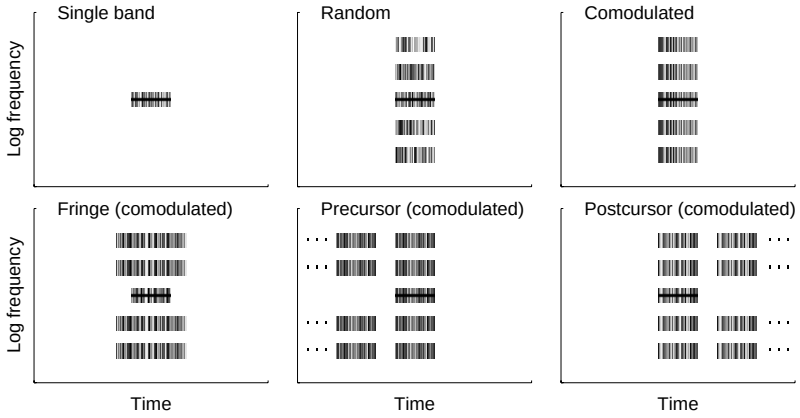


Fig. 2. Experimental conditions: The upper row shows the single band (SB), the standard random (R) and comodulated (C) conditions. The bottom row shows some of the conditions that promote the perceptual segregation of the on-frequency band: Comodulated flankers with fringes (FC), comodulated flankers with pre- (PC) and postcursors (PoC).

Overall, a highly significant CMR effect was observed in the narrowband configuration, independent of condition.

The results are different for the broadband configuration (open squares). While there is a clear CMR effect of about 6 dB in the standard condition, no CMR is observed in any of the other conditions. A one-way within-subjects ANOVA was used to investigate the effect of condition. Using all conditions except SB, the main effect was highly significant; $F(7,28) = 6.19$, $p < 0.001$. Post hoc comparisons based on the HSD criterion showed that the threshold in condition C was significantly ($p < 0.05$) different from all other conditions except for PC. No difference between the other conditions was found. The ANOVA was repeated for conditions restricted to [R FR FC PR PC PoR PoC] where no significant main effect was found. Thus, as can been seen by the squares in the right panel of Fig. 3, CMR was essentially eliminated in conditions promoting perceptual segregation of the on-frequency and flanker bands.

Figure 4 shows the mean data for the intermediate spectral configuration, corresponding to that used in the Grose and Hall (1993) study (open triangles), together with the data from the previous experiments. Only two of the five subjects from the previous experiments participated in this experiment. For the intermediate configuration, CMR is slightly larger than for the narrowband configuration. In the fringe condition, the CMR effect is only slightly reduced relative to the narrowband configuration. In the streaming conditions (PR versus PC and PoR versus PoC) results are essentially identical to those obtained in the narrowband configuration.

4 Discussion

The results from the broadband configuration show a very strong influence of factors designed to affect the perceptual grouping of the masker and flankers. All

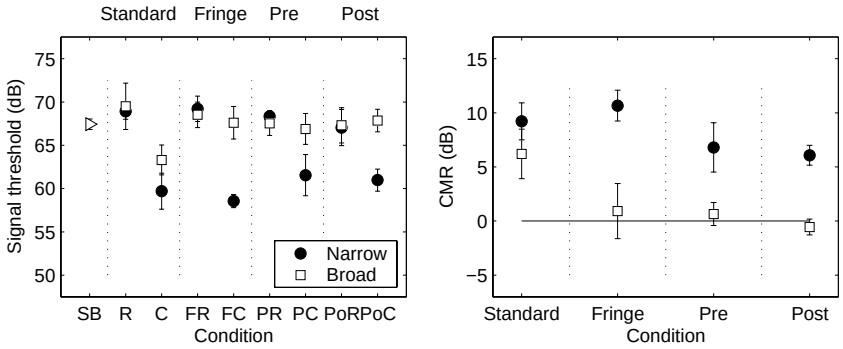


Fig. 3. Left: Mean masked thresholds and standard deviations for the signal as a function of the stimulus condition. Thresholds for the narrowband configuration are indicated as filled symbols. Open symbols represent thresholds for the broadband condition. Right: Threshold differences (CMR) for the standard, fringe, precursor and postcursor conditions.

attempts to perceptually segregate the on-frequency masker from the flankers resulted in a complete elimination of CMR, suggesting that "true" CMR does not act in isolation from the processes that give rise to auditory object formation. This is probably not compatible with peripherally based explanations of CMR in terms of, for instance, processing in the cochlear nucleus (Pressnitzer *et al*. 2001). Especially difficult to explain in peripheral terms is the strong influence of the following bands, or "postcursors," which seem to influence the perception of the masker and flankers *after* their presentation. This is unlikely to be related to backward masking, as the time scale is much greater in this case; backward masking is generally negligible within about 20 ms (e.g., Oxenham and Moore, 1994). In contrast, the results from the narrowband configuration show no effect of grouping manipulation.

Taken together, these results support the initial hypothesis that two different mechanisms contribute to CMR: results from the broadband conditions reflect across-channel mechanisms, which are susceptible to grouping manipulations, while the results from the narrowband conditions reflect within-channel mechanisms, which do not depend on variations of the acoustical context. Thus, the introduction of stimulus components and sequences as described here allow one to separate within- and across-channel contributions to CMR. The results obtained in the intermediate configuration indicate that the within-channel contributions to CMR can be dominant even when only a few components are processed within the peripheral channel at the signal frequency.

The intermediate configuration was designed to be as similar as possible to the spectral configuration used by Grose and Hall (1993), who also found influences of grouping on CMR. Their data are compared with the present findings in the left panel of Fig. 4. Their finding of only a small effect of precursor tones on CMR is replicated here, suggesting that their masker spacing may have been too narrow to observe the full effect. However, unlike the present finding of only a small effect of gating asynchrony in the fringe condition, they found that CMR was virtually eliminated. It is unclear what accounts for this apparent discrepancy.

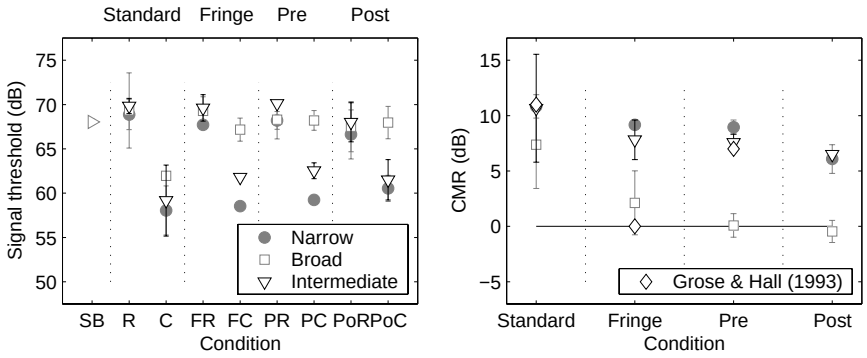


Fig. 4. Left: Open triangles show thresholds for the intermediate configuration (mean data of only two subjects). The gray symbols represent the mean data for the original two configurations for the same two subjects. Right: Difference values (CMR) as in Fig. 3. The open diamonds represent thresholds obtained in Grose and Hall (1993). The errorbars indicate standard deviations.

Finally, it is interesting to note that another putative across-channel effect, modulation detection interference (MDI), has also been found to be highly dependent on manipulations of auditory grouping (Oxenham and Dau 2001). It may be a general auditory organizational principle that access to higher-order percepts, such as modulation strength, is generally afforded only after the outputs of peripheral frequency channels are combined to form auditory objects or streams. In other words, access to sensations is via objects rather than outputs of peripheral auditory channels.

5 Summary

The effects of introducing a gating asynchrony or a stream of preceding or following flanker bands were studied for conditions of CMR. The results and conclusions of this study can be summarized as follows:

1) Using widely spaced flanking bands (broadband configuration), CMR effects were eliminated by introducing a gating asynchrony between the on-frequency masker and the flanking bands, by introducing precursor flanking bands, and by introducing following flanking bands.

2) Using narrowly spaced flanking bands (narrowband configuration), CMR was not affected by any of the stimulus manipulations tested here.

3) The results suggest two forms of CMR. One is based on within-channel mechanisms (Schoonevelt and Moore 1989; Verhey *et al.* 1999), determined by the stimulus envelope statistics, which is peripheral in nature and thus not susceptible to manipulation by auditory grouping constraints. The other is based on across-channel comparisons and is highly dependent on auditory grouping constraints. Specifically, any manipulation that perceptually segregates the flankers from the on-frequency masking band leads to an elimination of CMR.

4) The dependence of across-channel CMR on auditory grouping places strong constraints on potential neural substrates for CMR. In particular, it seems unlikely that the effects can be accounted for by processing in auditory brainstem or below.

References

Buus, S. (1985) Release from masking caused by envelope fluctuations. J. Acoust. Soc. Am. 78, 1958-1965.

Carlyon, R.P., Buus, S., and Florentine, M. (1989) Comodulation masking release for three types of modulator as a function of modulation rate. Hear. Res. 42, 37-46.

Cohen, M.F. and Schubert, E.D. (1987) Influence of phase synchrony on the detection of a sinusoid. J. Acoust. Soc. Am. 81, 452-458.

Grose, J. H. and Hall, J. W. (1989) Comodulation masking release using SAM tonal complex maskers: effects of modulation depth and signal position. J. Acoust. Soc. Am. 85, 1276-1286.

Grose, J. H. and Hall, J. W. (1993) Comodulation masking release: Is comodulation sufficient?. J. Acoust. Soc. Am. 93, 2896-2802.

Haggard, M.P., Hall, J.W., and Grose, J.H. (1990) Comodulation masking release as a function of bandwidth and test frequency. J. Acoust. Soc. Am. 88, 113-118.

Hall, J.W., Haggard, M.P. and Fernandes, M.A. (1984) Detection in noise by spectro-temporal pattern analysis. J. Acoust. Soc. Am. 76, 50-56.

Hall, J.W., Grose, J.H. and Haggard, M.P. (1990) Effects of flanking band proximity, number, and modulation pattern on comodulation masking release. J. Acoust. Soc. Am. 87, 269-283.

Klump, G. M. and Langemann, U. (1995). Comodulation masking release in a songbird. Hear. Res. 87, 157-164.

Nelken, I., Rotman, Y. and Bar Yosef, O. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. Nature 397, 154-157.

Oxenham, A. J., and Dau, T. (2001). Modulation detection interference: Effects of concurrent and sequential streaming. J. Acoust. Soc. Am. 110, 402-408.

Oxenham, A. J., and Moore, B. C. J. (1994). Modeling the additivity of nonsimultaneous masking. Hear. Res. 80, 105-118.

Pressnitzer, D., Meddis, R., Delahaye, R. and Winter, I. (2001) Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J. Neurosci. 21, 6377-6386.

Schooneveldt, G.P. and Moore, B.C.J. (1987) Comodulation masking release (CMR): effect of signal frequency flanking band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band. J. Acoust. Soc. Am. 82, 1944-1956.

Schooneveldt, G.P. and Moore, B.C.J. (1989) Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth and signal duration. J. Acoust. Soc. Am. 85, 273-281.

Verhey, J.L., Dau, T. and Kollmeier, B. (1999) Within-channel cues in comodulation masking release (CMR): experiments and model predictions using a modulation filterbank model. J. Acoust. Soc. Am. 106. 2733-2745.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Effects of contralateral sound stimulation on forward masking in the guinea pig

Ray Meddis[1], Christian Sumner[2], and Susan Shore[2]

[1] Department of Psychology, University of Essex, Colchester, UK, rmeddis@essex.ac.uk
[2] Kresge Hearing Research Institute, University of Michigan Medical School, Ann Arbor, Michigan 48109

1 Introduction

Psychophysical forward masking (psyFM) refers to a raised threshold for a stimulus following a previous stimulus. It has been possible to develop explanations of psyFM in terms of a reduced signal to noise ratio by positing an integration temporal window that includes both the stimulus and the masker, (Moore et al. 1988). On the other hand, physiological measurements in the auditory nerve and brainstem also reveal reduced responsiveness to a tone preceded by another tone. Physiological forward masking (physFM) at the level of the auditory nerve (AN) is normally attributed to adaptation whose origin is in the IHC/AN synapse (Harris and Dallos, 1979). However, at the level of the cochlear nucleus (CN), suppression of responses has also been attributed to descending inhibitory effects that can be disrupted surgically (Shore, 1998).

It is possible that physFM and psyFM are linked. However, the idea has its challengers (e.g., Relkin and Turner, 1988) and it is certainly true that current explanations for the two phenomena are very different in type (adaptation and inhibition on the one hand with temporal integration and reductions in signal to noise ratios on the other hand). In this study we seek to establish a useful parallel through the mechanism of 'release from forward masking' using binaural stimuli.

Less psyFM is observed when remote frequencies are added to maskers (Moore and Vickers, 1997). This might be explained at the physiological level either in terms of mechanical two-tone suppression within the broadband masker or release from neural inhibition. Release from masking is still observed to a lesser degree if a broadband stimulus is presented contralaterally and simultaneously with the masker (Moore and Glasberg, 1982). In the latter paradigm, two-tone suppression should be ruled out as an explanation. Release from inhibition has also been invoked to explain some aspects of comodulation masking release (CMR) where a broadband modulated stimulus is a less effective masker than a narrowband masker for targets presented in the modulation dips (Meddis et al., 2002). CMR is also a binaural

phenomenon and, recently, contralateral release from psyFM has been observed using modulated broadband releasers (Delahaye, 2002).

A tentative model of the physiological basis of masking release has been developed to allow us to explore the process in more detail. Figure 1 shows a hypothetical neural circuit in the cochlear nucleus (CN) that simulates release from forward masking. The focus of attention is a VCN projection unit (shaded gray) that receives excitatory input from the auditory nerve and inhibitory input from two different sources. One inhibitor has a wide receptive field and produces inhibition with fast onset and offset. The other has a narrow receptive field and delivers inhibition that is slow to initiate and slow to dissipate (such as might be expected at GABAergic synapses). This latter inhibition outlasts the original stimulus and is hypothesized to be responsible for the neural component of physFM. The model proposes that the delayed inhibitor is itself subject to suppression by the fast acting broadband inhibition. A release from forward masking is most likely to occur when a broadband masker is used because this provides the most effective input to the fast inhibitor. The model further proposes that the fast broadband inhibitor acts both ipsilaterally and contralaterally. In the current study, we aim to provide support for the model by demonstrating that at least some units in the VCN show release from physFM when broadband contralateral releasers are used. We report the results of an exploratory study designed to investigate this model.

2 Physiological measurements

2.1 Materials and methods

Experiments were performed on 2 healthy female, adult pigmented guinea pigs (NIH outbred strain) with normal Preyer's reflexes, weighing 393g. All procedures were performed in accordance with the NIH guidelines for the care and use of laboratory animals (NIH publication No. 80-23), and guidelines provided by the University of Michigan (UCUCA).

2.2 Surgical preparation

The guinea pigs were anesthetized with ketamine (120 mg/kg) and Xylazine (16 mg/kg) and held in a stereotaxic device (Kopf) with hollow ear bars for the delivery of sounds. Rectal temperature was monitored and maintained at 38° $\pm$ 0.5°C with a thermostatically controlled heating pad. The bone overlying the cerebellum and posterior occipital cortex was removed to allow placement of a recording electrode into the VCN, after aspirating a small amount of cerebellum to visualize the surface of the DCN. All unit recordings were made in a sound-attenuating double-walled booth. Sixteen-channel electrodes fabricated by the University of Michigan Electrical Engineering and Computer Science Department were used for recording unit responses, enabling us to record from many units simultaneously. The multi-channel neuronal acquisition processor is designed to facilitate both waveform recording and spike sorting. Initially spikes were detected by a threshold crossing algorithm set automatically just above the noise floor. The waveform for each spike

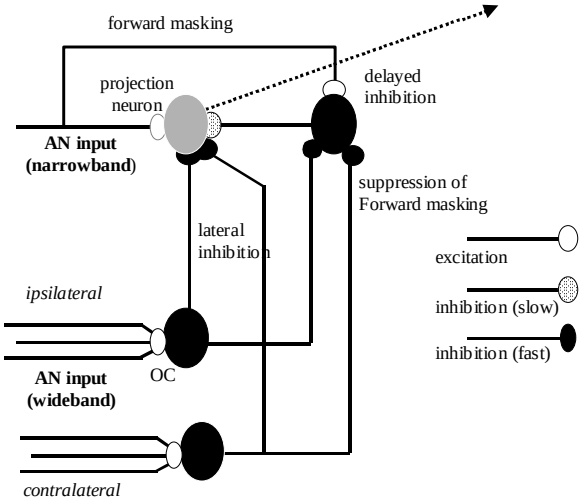


Fig. 1. Hypothetical neural circuit that demonstrates forward masking (see text).

was stored for later sorting. In many cases it was not possible to sort spikes belonging to a single unit. In these cases, multi-unit data are presented. However, all recordings to be reported below are based on channels showing narrow receptive fields and a transient chopping PSTH. Best frequencies were assessed using a 55 dB sweep tone or a receptive field routine that collected responses to different levels and frequencies in randomized order.

2.3 Stimuli

The stimulus paradigm is presented in Fig 2. In all conditions, the probe tone was presented before (unmasked) and after (masked) a masker tone. All three tones were presented at the same frequency. In the experimental conditions, a contralateral broadband noise was presented simultaneously with the masker at a range of levels. All signal components had a 1.5 msec $\cos^2$ onset and offset ramp. The level of the broadband noise was varied across a limited number of values. Stimuli were randomized across all dimensions. Typically 100 repetitions were given for each condition.

Binaural stimuli were used to avoid complications associated with mechanical suppression. The release from forward masking is expected to be less using these stimuli than monaural stimuli but simpler to interpret in terms of the model.

3 Results

Only channels showing a substantial response to the unmasked probe and unambiguous forward masking were chosen for further study. Fig.2 shows an illustrative result for a single unit. Masking is measured as a reduction in the response of a unit to the masked probe compared to the unmasked probe. In this example, the average response to the unmasked probe over six conditions was 388 spikes (or 194 sp/s). In the control condition the response to the masked probe was 175 spikes. In the first experimental condition (contralateral noise level 40 dB), the response to the masked probe increased to 190 spikes, equivalent to a rate change of 7.5 sp/s. An increase (15, 14, 29, 9 spikes) is seen across all levels of contralateral noise. The effect is, therefore, consistent across conditions. Units typically showed either a consistent increase (9/23) or a consistent decrease (6/23) when the contralateral masker was added.

Fig. 2B shows a second kind of analysis for the same unit where the control PSTH is *subtracted* from the experimental PSTHs. Firing rate is reduced during the masker and this effect is greater for stronger contralateral stimuli. At the time of the

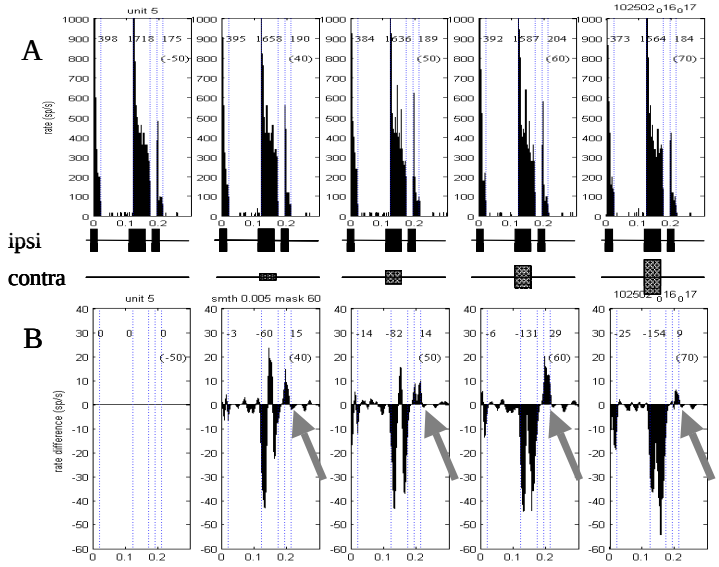


Fig. 2. Stimulus paradigm and response measurement. **A.** penetration 1, trial 1, unit 5, BF=9.1 kHz. The ipsilateral stimulus is an initial probe followed by a 100-ms gap, a masker and a second probe 20 ms after the masker (maskers 60 dB / probes 25 dB, all 12 kHz). Contralateral wideband noise was presented at 40, 50, 60, and 70 dB. The PSTHs show unit response rate for 100 repetitions. The numbers at the top of the PSTH specify the total spike counts for the probe, masker and masked probe periods respectively. **B.** Difference functions obtained by subtracting control PSTH from experimental PSTH and smoothing with a 5-ms Gaussian window. Arrows indicate the position of the masked probe. Difference counts are computed before smoothing.

masked probe, however, an *increase* in responding can be seen in all four experimental conditions. Despite considerable variability, this overall pattern could be seen in many units. Even when there is no overall release from masking, there is often a brief increase in rate during the probe period. Fig. 3 illustrates variations across units. Fig. 3A shows a release from masking that is consistent across four levels of contralateral noise. Figs. 3B and 3C show the same pattern. Fig. 3D is a *counter* example; a channel showing no *net* masking release However, a brief release from masking is still observable as a small positive spike even though a second suppressive effect at the beginning and end of the probe creates a net reduction in response (i.e. no masking release).

Figure 4 shows a larger data set for a unit tested at three masker levels and three masker-probe gaps. In this case the, contralateral noise increases the response to the masker and it is difficult to see if masking release has occurred at the 10 ms gap. However, possible masking release can be seen at the longer gaps.

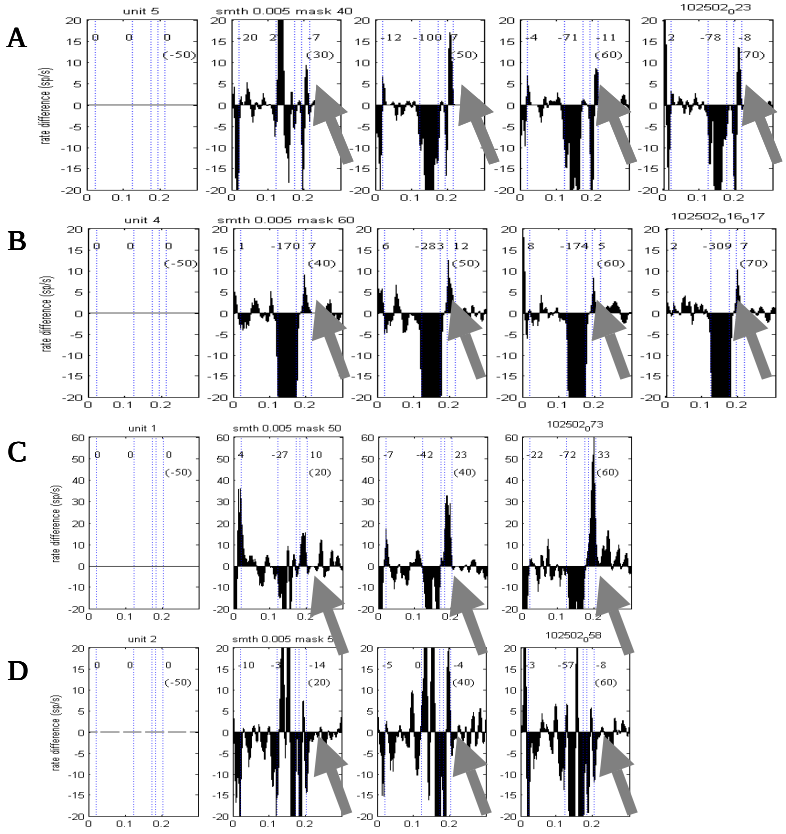


Fig. 3. Difference PSTHs for four units **A.** Penetration 1, trial 1, unit 4, BF=12. 8 kHz. masker 60 dB /probe 25 dB, 9.1 kHz. **B.** Penetration 2, trial 1 unit 16, BF=5.5 kHz; masker 40 dB /probe 20 dB, 12 kHz. **C.** Penetration 6, trial 2 unit 1, BF=5 kHz; masker 50 dB /probe 35 dB, 5.5 kHz. **D.** Penetration 4, trial 3 unit 2, masker 50 dB /probe 30 dB, 7.3 kHz. Contralateral noise levels are shown in parentheses in individual figures.

4 Discussion

The masking release we have observed is small. This was expected for two reasons. Firstly, binaural stimuli were used. Secondly, the reduction of response is only concerned with neural suppression. That portion of the response that is attributable to auditory nerve adaptation will not be affected. However, when present, the effect can be seen across all levels of contralateral stimulation. This encourages us to believe that further observations are warranted and are likely to continue to conform to the model proposed in the introduction.

Contralateral suppression of the response to the masker was also very noticeable in this study. Contralateral neural inhibition of the activity of ventral CN neurons has been previously demonstrated (Shore et al, 2003). Approximately 30% of neurons recorded in that study showed a suppression of spontaneous rate by

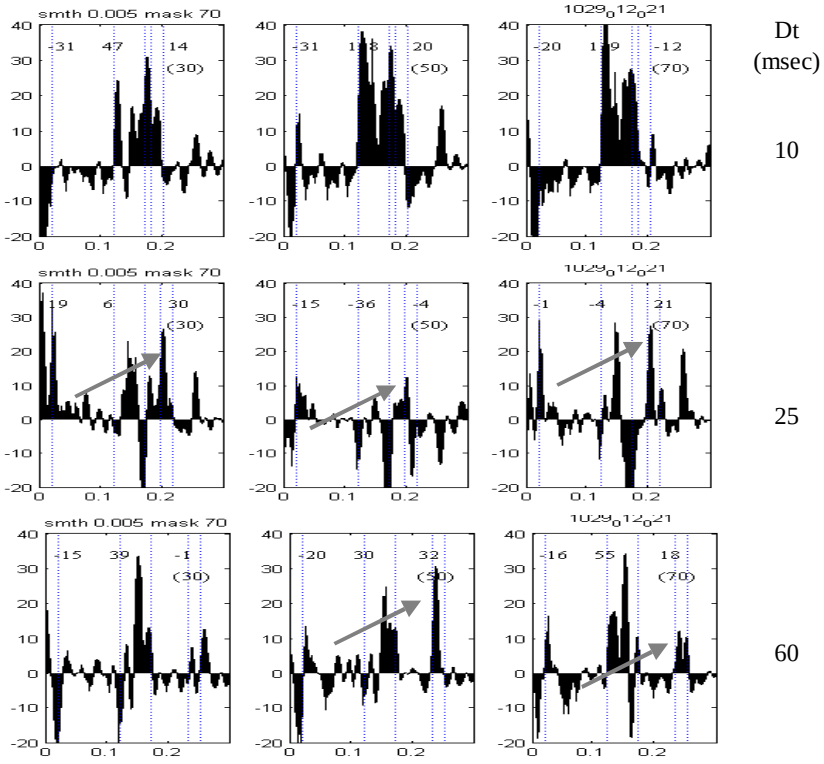


Fig. 4. Difference functions for animal 2, penetration 1, channel 2 (BF=528). Masker frequency = 300 Hz, level= 70 dB throughout. Masker-probe gap is increased from top to bottom (10, 25 and 60 ms). The level of the contralateral noise is increased from left to right (30, 50 and 70 dB). The functions are the difference between the PSTH for this stimulus and a control condition without contralateral noise. Arrows indicate possible masking release.

contralateral noise presentation. The latency of this suppression ranged from 2- to 6 ms greater than latencies of responses to ipsilateral stimulation (Shore et al, 2003). While it is not the focus of this report, it is an important reminder that more than one inhibitory process is at play. The increases in probe rate with contralateral stimulation were not well correlated with suppression of the masker suggesting that the masking process here can be attributed to inhibition and not adaptation. Forward masking of the probe in VIIIth nerve is correlated with masker firing rate, and attributed to adaptation (Harris and Dallos, 1979). The model assumes that contralateral stimulation produces a rapid and direct inhibition that is additional to the delayed inhibition that may be causing physFM.

The data from these preliminary experiments are encouraging. The experimental results are consistent with the model shown in Fig. 1. The small size of the results urges caution even if it was expected. Contralateral noise produced a rapid and often substantial reduction in the response to the ispilateral masker. This is evidence in favour of a contralateral inhibitor driven by wideband stimulation. The small but consistent release from masking observed in some units provides tentative support for a release from masking circuit at an early stage in brainstem auditory processing. The size of the effect is less important than the fact of its existence. Masking release is more important for ipsilateral rather than contralateral stimuli. However, a physiological demonstration of the ipsilateral effect would be open to an interpretation in terms of mechanical suppression. A neural masking release circuit showing physFM could offer a new way of thinking about psyFM.

References

Delahaye, R. (2002) Contralateral inhibition in a release from forward masking. Hear. Res. 166, 44-53.

Harris, D.M. and Dallos, P. (1979) "Forward masking of auditory nerve fibre responses,", J. Neurophysiol., 42, 1983-1107

Meddis, R., Delahaye, R., O'Mard, L., Sumner, C., Fantini, D.A., Winter, I and Pressnitzer, D. (2002) "A model of signal Processing in the Cochlear Nucleus: Comodulation masking Release," Acustica, 88, 387-398.

Moore, B.C.J. and Glasberg, B.R. (1982) Contralateral and ispilateral cueing in forward masking. J. Acoust. Soc. Am., 71, 1982

Moore, B.C.J., Glasberg, B.R., Plack, C.J. and Biswas, A.K. (1988) The shape of the ear's temporal window, J. Acoust. Soc. Am., 83, 1102-1116.

Moore, B.C.J. and Vickers, D.A. (1997) The role of spread of excitation and suppression in simultaneous masking, J. Acoust. Soc. Am., 102, 2284-2290.

Relkin, E.M., and Turner, C.W. (1988) "A reexamination of forward masking in the auditory nerve", J. Acoust. Soc. Am., 84, 2, 584-591

Shore, S.E. (1998) " Influence of centrifugal pathways on forward masking of ventral cochlear nucleus neurons", J. Acoust. Soc. Am., 104, 378-389

Shore SE., Sumner, C., Bledsoe SC, and Lu, J. (2003) Binaural integration in the guinea pig ventral cochlear nucleus. In: Central Auditory Processing – Integration of Auditory and Non-Auditory Information. Accepted for Experimental Brain Research.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Inhibition in models of coincidence detection

H. Steven Colburn, Yi Zhou, and Vasant Dasika

Department of Biomedical Engineering - Boston University,
{colburn,yizhou,vdasika}@bu.edu

1 Introduction

Discussion of the role of inhibition in models of coincidence detection started early, involving the original model of Jeffress (1948) and the first anatomically confirmed recordings from the MSO (Goldberg and Brown 1969). In Jeffress's mechanism for interaural time sensitivity, a network of coincidence detector neurons represents interaural time delay as spatial patterns of activity. Jeffress did not specifically exclude inhibition and coincidence detection, in its simplest form, does not require it. The MSO recordings showed that the antiphasic case (when the interaural delay was equal to half the period of the stimulus) gave lower rates than a monaural stimulus, suggesting inhibitory activity; however, this observation does not require inhibition when there is spontaneous activity (Colburn, Han, and Culotta 1990). The role of inhibition in coincidence detection is addressed in Sec. 4. First, the relation of ITD sensitivity to the neural patterns generated by coincidence detectors is discussed, and then some factors in purely excitatory models that influence the critical rate-ITD curve for single neurons are addressed. Finally, the role of inhibition in coincidence detection is discussed in light of empirical and modeling results.

2 Interaural time delay coding and discrimination

2.1 Information in auditory-nerve patterns

Behavioral thresholds on the order of microseconds can be achieved at low frequencies in an auditory system with broad individual pulses and relatively large time jitter — both factors are on the order of a millisecond at low frequencies. In an analysis of the limits on interaural time discrimination imposed by the firing patterns of auditory-nerve fibers, Colburn (1973) found that just-noticeable differences (JNDs) of a fraction of a microsecond would be possible with optimum processing. Even if processing is severely restricted in time (to coincidences with narrow windows) and in place (to matching input fibers), and even if only the number of coincidences

for each detector is used for decisions, there is still enough information to allow observed performance. With these restrictions, the resulting processing becomes a realization of the Jeffress coincidence mechanism.

2.2 Performance limits imposed by coincidence detectors

Three issues are central in considering sensitivity to interaural time with a network of coincidence detectors: first, the information provided by a single detector with a fixed characteristic delay (i.e., by a single cell); second, the distribution of characteristic delays over the population; and third, the way that cells' responses are combined. The single-detector information depends primarily on the shape of its rate-ITD curve, and the rate-ITD curve is the focus of the modeling described subsequently. A useful statistic for the amount of information is R (Colburn 1969), corresponding roughly to the square of d' from detection theory and to the inverse of the squared JND using a single fiber. Because R is additive over statistically independent coincidence detectors, it can be used to calculate optimum performance possible with a population of cells. In Fig. 1, this factor is shown for a single, 500-Hz coincidence detector with a characteristic delay of zero. Specifically, the periodic function $R(\tau)$ reflects (inversely) the dependence of the estimated JND on the interaural delay τ of a sinusoidal stimulus. The shape is roughly independent of frequency for a narrow coincidence windows at low frequencies. The calculations for R in Fig. 1 assumed a single pair of excitatory inputs, Poisson statistics, and an exponentiated-sinusoid input rate function.

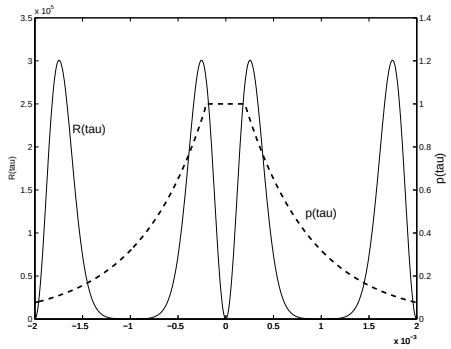


Fig. 1. Performance metric R (solid) versus interaural time delay for a single coincidence detector along with $p(\tau)$ (dashed), the distribution of internal delays from Colburn (1977).

The second critical function, the distribution of characteristic delays over the population, is represented by $p(\tau)$. The example $p(\tau)$ shown in Fig. 1 was chosen to give a good fit to the binaural detection data [specifically, the "flattening" effect (Durlach 1972)] assuming that $p(\tau)$ is independent of frequency. This assumption, later modified by Stern and Shear (1996), minimized the number of parameters in

the modeling. The need for large internal delays at low frequencies was demonstrated empirically by van der Heijden and Trahiotis (1999) in masking experiments. In light of recent data (e.g., McAlpine, Jiang, and Palmer 2001) indicating that the distribution of delay varies with characteristic frequency, $p(\tau)$ and the consequences for modeling detection data should be revisited.

Optimal performance results from combining single detector information over a population of neurons with internal delay distribution $p(\tau)$. The total information is equal to the convolution $R(\tau) \star p(\tau)$. This calculation leads to JND functions like those shown in Fig. 2. The left panel shows data for the $p(\tau)$ of Fig. 1, and the right panel for another $p(\tau)$ corresponding to all internal delays at approximately $\pm 45^{\circ}$. The results relating $p(\tau)$ to ITD JND predictions are related to the recent work of Harper and McAlpine (2003) who showed that the optimum $p(\tau)$ distribution for achieving the best performance for single source, free-field conditions is a pair of narrow bands at low frequencies. Altogether these results suggest that performance at larger reference delays may give insight into the internal delays available.

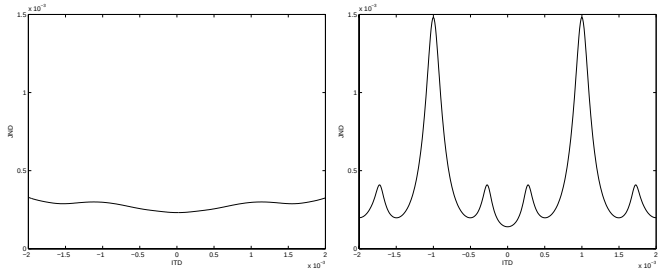


Fig. 2. JND for 500-Hz tone calculated by optimally combining single fiber information. Left panel uses the characteristic delay distribution in Fig. 1 and the right panel uses a two-impulse distribution.

The third issue that determines performance with a population of coincidence detectors is the way that the outputs are combined. It has been suggested (Colburn and Latimer 1978) that a non-optimal combination of information is necessary in order to describe the dependence of psychophysical performance on combinations of interaural intensity difference (IID) and interaural time differences ITD). In general, psychophysical data indicate that the variables are combined such that poorer resolution is found when the variables are "naturally" combined (loud ear leads) than when the same magnitudes are put in opposition. Since a coincidence mechanism is fundamentally symmetric and indifferent to which input has higher rates, there must be an additional mechanism to create the measured asymmetry. Colburn and Latimer (1978) postulated a decision variable that used the IID to weight summated coincidence cells on the two sides differentially. This combined and intensity-weighted statistic is an example of a non-optimum combination and also serves as an example of the IID-dependent data that are a challenge for some binaural models.

3 Pure excitation cell models for physiological responses

Although it is clear that inhibition is present and has a role in the functioning of the MSO, much of the available *in vivo* MSO data can be described without inhibitory processes (Colburn 1996) and the study of purely excitatory models can help to understand how assumptions about the cell models relate to predicted response patterns and can demonstrate where the model fails. In particular, two aspects of the *in vivo* data have been difficult to match with simple models: Rate-ITD curves obtained from a single MSO cell over a wide range of frequencies (Yin and Chan 1990) are difficult to reproduce with a single set of cell parameters. Also, the effects of intensity, both overall intensity and IID, on rate and on rate-ITD functions, have not been successfully described with excitation-only models. In the next section, we consider intensity effects in more complex models that include inhibition.

The *in vitro* data can also be addressed by coincidence models. The simplest model is a leaky-integrate-and-fire artificial neuron with a small number of parameters, including the decay time-constant, the number of inputs from each side, and the threshold. Figure 3 shows, parametrically, the effect of increasing the number of input fibers for perfectly synchronized inputs at a single frequency. The left panel shows physiological results *in vitro* and the right panel shows model results. It is clear that larger numbers of inputs increase the modulation in the phase dependence of the output patterns.

Despite trying with various numbers of inputs and various values of time constant, the rate-ITD curve shapes across frequency were not achievable (data not shown). Thus inhibition and/or adaptation may be influencing the degree of modulation, at least for some frequencies.

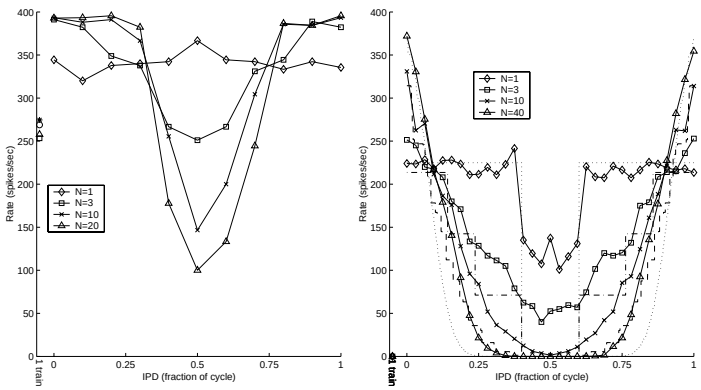


Fig. 3. Effects of increasing the number of input pairs N on modulation in rate-ITD curves. The left panel shows data from Reyes, Rubel, and Spain (1996) and the right panel shows model results, both for a fixed frequency of 400 Hz.

4 Cell models with excitation and inhibition

This section considers results from a biophysical model that receives both excitatory and inhibitory inputs. The basic cell structure, which is realized in the NEURON simulation environment, includes passive symmetric dendrites and an active cell body. Active ionic channels are located in the soma and the axon. Excitatory synapses are distributed along the dendrites, and the inhibitory synapses are located near the soma. The dynamics of inhibitory synapses are specified to have a decay time constant of approximately two milliseconds. This is relatively fast in comparison to most inhibitory synapses in the auditory system, and relatively slow in comparison to excitatory synapses in the model. The input spike patterns are chosen to be consistent with those of input fibers to the MSO, and output spikes patterns generated by the model are compared with MSO recordings.

Within the context of this model, inhibition is required to describe the dependence of the rate of firing on the levels, and the variations of the shape of the period histograms with interaural level differences. In Fig. 4, firing rates are shown for a purely excitatory model, for a model with both excitation and inhibition, and as measured by Goldberg and Brown (1969). The non-monotonic dependence seen in the data for the out-of-phase and the monaural case is not seen in the purely excitatory (EE) model, but is achieved by the model with inhibition (EE/II). Similar difficulties with the EE model arise when interaural level differences are considered.

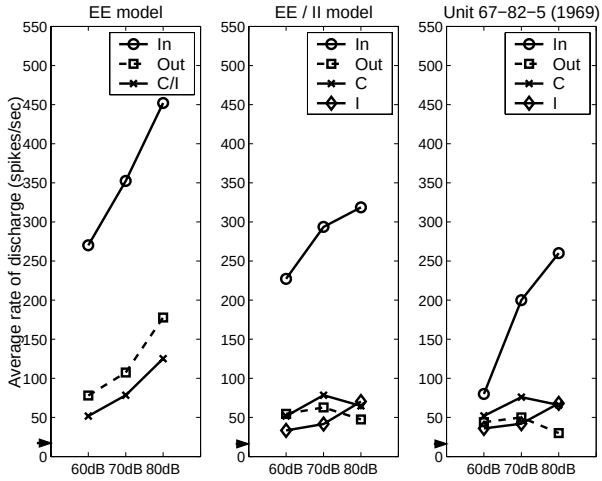


Fig. 4. Comparison of in-phase, out-of-phase and monaural rates at three sound levels (dB SPL). Left column: results from the EE model, in which the increase of average input rates of each fiber raises firing rates for all in-phase, out-of-phase and M cases. Middle column: results for the EE/II model. Right column: unit 67-82-5 from Goldberg and Brown (1969). The arrows indicate spontaneous output rates for each case.

The effects of level combinations on the shapes of period histograms are shown in Fig. 5 for the out-of-phase stimulus case in the Goldberg and Brown (1969) data and in the model. The combination of the changes in the shapes of the normalized histograms and the rates for different intensity combination make it clear that inhibition is necessary to achieve the observed reduction in the responses at the ear with its level fixed as the level at the opposite ear increases.

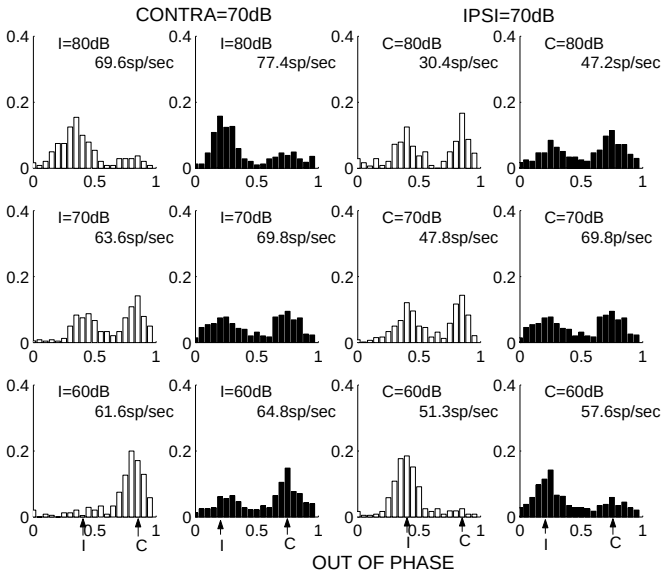


Fig. 5. Period Histograms for out-of-phase stimuli with various IID combinations for the EE/II model (filled bars) and for the data (open bars) of Goldberg and Brown (1969). Arrows on the ordinate indicate the locations of the maxima for monaural stimulation.

5 Summary

Coincidence detection is a key ingredient in the extraction of information about timing information in the stimulus. The processing of interaural time delay is almost certainly dependent on the quality of the coincidence detection mechanisms in the auditory brainstem; however, the role of inhibition in the operation of the coincidence mechanism is not yet clear. There are many factors that improve ITD sensitivity in coincidence mechanisms, some seen in purely excitatory models and some seen only in with inhibitory inputs; nevertheless, it is not yet clear which effects are most important.

Acknowledgments

This work was supported by the U.S. Public Health Service (NIDCD grant number R01 DC00100).

References

Colburn, H.S. (1973) Theory of Binaural Interaction Based on Auditory-Nerve Data. I. General Strategy and preliminary results on interaural discrimination. J. Acoust. Soc. Am. 54, 1458-1470

Colburn, H.S. (1977) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. J. Acoust. Soc. Am. 61, 525-533.

Colburn, H.S. (1969) *Some physiological limitations on binaural performance.* Dissertation, Electrical Engineering Department, M.I.T., Cambridge MA.

Colburn, H.S. (1996) Computational models of binaural processing. In: H. Hawkins and T. McMullen (Eds.), *Auditory Computation.* [Volume In: A. Popper and R. Fay (Eds.) Springer Handbook on Auditory Research.] Springer, New York, pp. 332-400.

Colburn, H.S., Han, Y., and Culotta, C.P. (1990) Coincidence Model of MSO Responses. Hear. Res. 49, 335-346.

Colburn, H.S. and Latimer, J.S. (1978) Theory of binaural interaction based on auditory-nerve data. III. Joint dependence on interaural time and amplitude differences in discrimination and detection. J. Acoust. Soc. Am. 64, 95-106.

Durlach, N.I. (1972) Binaural signal detection. In: J.V. Tobias (Ed.) *Foundations of Modern Auditory Theory*, Vol. 2. New York: Academic Press, 405-466.

Goldberg, J.M. and Brown, P. B. (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization, J. Neurophys. 32, 613-636.

Harper, N. and McAlpine, D. (2003) Ideal population coding of interaural phase. Assoc. Res. Otolaryn. 26, 237.

Jeffress, L. A. (1948) A place theory of sound localization. J. Comp. Physiol. Psychol. 41, 35-39.

Stern, R.M. and Shear, G.D. (1996) Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. J. Acoust. Soc. Am. 100, 2278-2288.

McAlpine D, Jiang D, Palmer AR (2001) A neural code for low-frequency sound localization in mammals. Nat. Neurosci. 4, 396-401.

Reyes, A.D. Rubel, E.W. Spain, W.J. (1996) In vitro analysis of optimal stimuli for phase-locking and time-delayed modulation of firing in avian nucleus laminaris neurons. J. Neurosci. 16, 993-1007.

van der Heijden, M. and Trahiotis, C. (1999) Masking with interaurally delayed stimuli: the use of “internal” delays in binaural detection. J. Acoust. Soc. Am. 105, 388-399.

Yin, T.C.T. and Chan, J.C.K. (1990) Interaural time sensitivity in medial superior olive of cat. J. Neurophysiol. 64, 465-488.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

What can auditory evoked potentials tell us about binaural processing in humans?

Birger Kollmeier and Helmut Riedel

Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany
birger.kollmeier@uni-oldenburg.de

1 Introduction

Binaural signal processing provides us with the remarkable ability to precisely localize sounds based on tiny interaural time differences (ITD) and interaural level differences (ILD). Even though much research has been done on the psychophysical and physiological bases of these mechanisms, we are still lacking a thorough understanding on how these physical differences are extracted and utilized in the early stages of the human auditory system. While some models stress the importance of ITD processing using a Jeffress-type delay-line model without a special mechanism for ILD processing, other models like the equalization and cancellation model utilize interaural subtraction mechanisms that are sensitive to interaural intensity (and phase) differences as well. Since combined models for ITD and ILD processing have been introduced and discussed recently, the current contribution tries to provide more experimental data on the interplay between ITD and ILD processing in humans.

Evoked potentials have been widely used to investigate directional hearing in humans. The majority of the studies dealing with the dependence of auditory brain stem responses (ABRs) on ITD and ILD focused on the analysis of the waveforms of single EEG channels (e.g., McPherson and Starr 1995; Brantberg *et al.* 1999, Riedel and Kollmeier 2002a). A more sophisticated approach to localize active neural tissue in the brain is dipole source analysis from multi-channel EEG measurements (Scherg 1990; Mosher *et al.* 1992). Scherg and von Cramon (1985) proposed a model of six fixed current dipoles to describe the five waves of the monaurally evoked ABR. This model not only used a fixed location and orientation for each dipole, but also heavily constrained the time course of the dipole moment magnitude. The active structures at the latency of the largest deflection, wave V, were identified as the superior olive and the lateral lemniscus.

The present study focuses on the source analysis of monaurally and binaurally evoked ABRs. In order to analyze the influence of lateralization, different combinations of ITDs and ILDs are used. Since the same lateralization can be generated by different combinations of the interaural differences, the variation of

both cues allows us to draw conclusions about the representation of the stimulus laterality in the brain stem.

2 Methods

Twelve normal-hearing subjects (3f, 9m) aged between 25 and 36 years volunteered in this study. Rarefaction click stimuli were produced by applying rectangular voltage pulses of 100 µs duration to Etymotic Research ER-2 insert earphones. The time interval between two successive stimuli varied randomly and was equally distributed between 62 and 72 ms, yielding an average stimulation rate of approximately 15 Hz. 15 stimulus conditions were tested: 9 binaural and 6 monaural. The monaural clicks were presented at the levels 53, 59 and 65 dB nHL. The binaural stimuli were the nine possible combinations of 3 ITDs (-0.4, 0 and 0.4 ms) and 3 ILDs (-12, 0 and 12 dB), see Fig. 2 for the naming of the stimuli. ABRs were recorded from 32 sites according to the extended 10-20-system using a DC-coupled differential amplifier (Synamps 5803) at a samplingrate of 10 kHz with 16-bit resolution. 10000 single sweeps for all of the 15 stimuli were recorded in an interleaved manner. Details of the recording procedure can be found elsewhere (Riedel *et al.* 2002a, Riedel and Kollmeier 2002c). Before averaging, the single sweeps were filtered with a linear-phase FIR bandpass with 200 taps and corner frequencies of 100 and 1500 Hz (Granzow *et al.* 2001). An iterated weighted average of the filtered sweeps was computed for all subjects and stimulus conditions (Riedel *et al.* 2001).

Two alternative models were analyzed to explain the data: a single rotating dipole versus a pair of fixed dipoles with hemispheric symmetry. The distance of the fixed dipoles was set to 2 cm corresponding roughly to the distances of the superior olives and lateral lemniscii. To fit current dipoles to the data, a software package was written in MATLAB. The inverse problem was solved by means of generalized maximum likelihood estimation taking the noise covariance matrix into account (Lütkenhöner 1998a, 1998b). Confidence regions of the dipole parameters and the goodness-of-fit by means of a χ^2–test were determined. A detailed description of the source analysis strategy can be found elsewhere (Riedel *et al.* 2002c, Riedel 2002d).

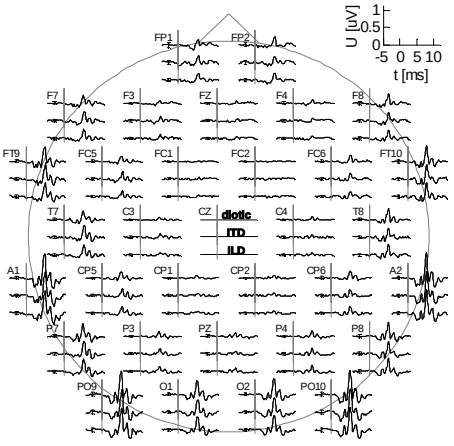


Fig. 1. ABRs recorded from 32 channels for subject DJ. Data are plotted in recording reference, the reference electrode was Cz. Top trace: binaural response to the diotic click (C00). Middle trace: response to the stimulus with an ITD of 0.4 ms (R0+). Bottom trace: response to the stimulus with an ILD of 12 dB (R+0). The error bars indicate ±3 standard errors of the mean.

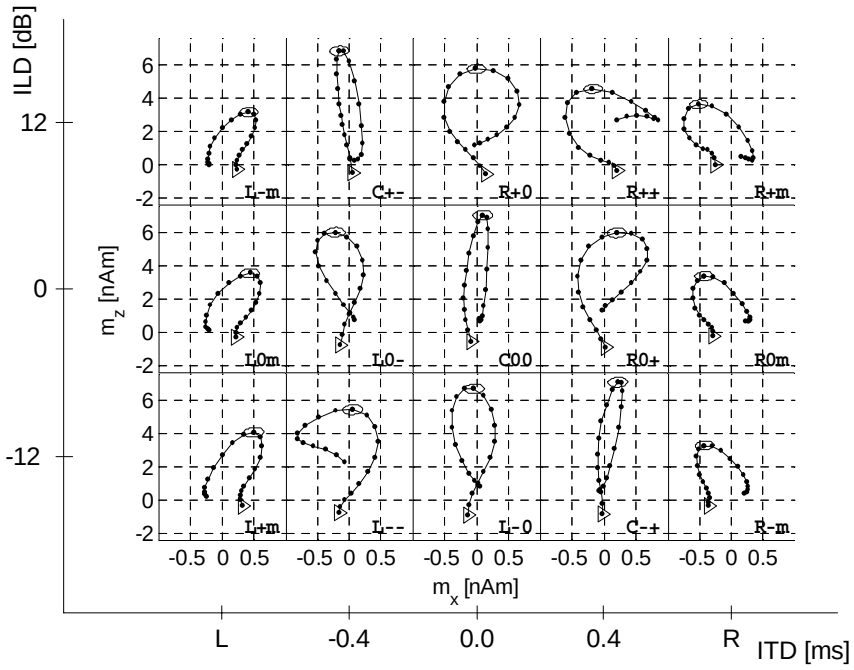


Fig. 2. Dipole moment trajectories of a rotating dipole for the 15 stimulus conditions in the frontal plane, mean over subjects.The x-coordinate points to the right, z to the top. The fit interval lasted 2 ms, from 1 ms before (triangles) to 1 ms after peak V. At the latency of wave V, ellipses denoting the 95% confidence regions for the dipole moment are drawn.

3 Results

Figure 1 shows ABRs of one subject (DJ) for all 32 measured channels in recording reference. For each channel three responses are depicted: the response to the diotic stimulus (C00) in the top trace, the response to the stimulus with an ITD of 0.4 ms (R0+) in the middle trace and with an ILD of 12 dB (R+0) in the bottom trace.The residual noise for each channel and condition is shown by the error bars (±3 S.E.M.). Channels near the reference electrode (Cz) generally exhibit smaller responses. The three traces in this example look very similar, however, systematic differences between stimulus conditions can be unveiled by dipole source analysis.

Figure 2 shows the dipole moment trajectories of the rotating dipole fitted in the interval from 1 ms before to 1 ms after peak V in the frontal plane. Data are averaged over subjects. The triangle denotes the start of the trajectory at t_V-1ms. The ellipses drawn at t_V are the 95% confidence regions for the dipole moment. The trajectories for the central stimuli, i.e., the diotic stimulus C00 and the antagonistic stimuli (C+- and C-+), exhibit the largest dipole moments in the vertical direction (m_z). With growing lateralization, m_z decreases and is smallest for the monaural

conditions. The moment trajectories allow us to distinguish between left and right conditions. For stimuli which are lateralized to the right, clockwise trajectories are observed, while left-lateralized stimuli exhibit counter-clockwise trajectories. Furthermore, for the conditions with ITD only and ILD only, the trajectories look very similar. The trajectories for R+0 and R0+ on the one hand, and for L-0 and L0- on the other hand, bear a strong resemblance to each other. This is striking since they are produced by different physical stimuli but elicit the same subjective lateralization.

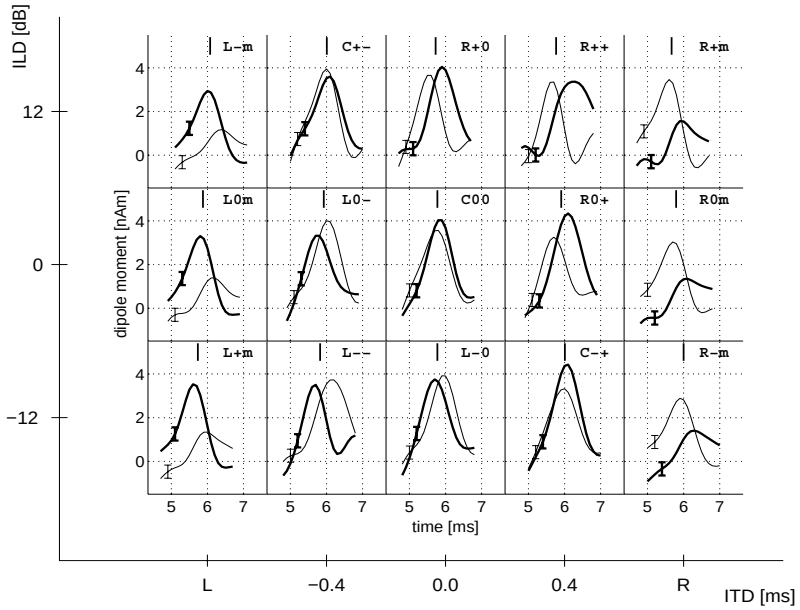


Fig. 3. Dipole moment magnitudes of two constrained fixed dipoles for all stimulus conditions in the frontal plane, mean over subjects. **Thin curves:** dipole in the left hemisphere. **Thick curves:** dipole in the right hemisphere. The error bars denote the 95% confidence regions of the moment magnitudes, the vertical bars mark the mean latency of wave V.

An alternative, physiologically motivated model, assumes two fixed dipoles with hemispheric constraints, i.e., mirrored x-component of the location and mirrored azimuth. The y- and z-component of the location and the elevation were forced to be identical for both dipoles. Unfortunately, fitting two fixed dipoles using these constraints leads to solutions with nearly identical locations of the dipoles in the vast majority of the cases, i.e., the fitting algorithm converges to a single dipole solution. To ensure the separation of the two dipoles, a stronger constraint for the location parameters was introduced. For every subject, the locations of the rotating dipole to all 15 stimulus conditions were averaged. The locations of the two fixed dipoles were defined by shifting the x-component of the rotating dipole 1 cm to the left and 1 cm to the right, respectively. The superior olive (SO) and the nucleii of the lateral lemniscii (NLL) are considered as the likely generator sites of wave V (Scherg *et al.* 1985). According to an anatomical atlas of the brain (Nieuwenhuys *et*

al. 1988), the distance between left and right SO is about 1.6 cm. The nuclei of left and right NLL reside at a distance of roughly 2.2 cm. Additionally, it was assumed that the current direction in the neural tissue is the same for all stimulus conditions. Therefore, the orientation constraint was extended by fitting a common dipole orientation for all responses. Only the dipole moment magnitude varied between conditions. The fitted orientation of the dipoles is mainly vertical, they are inclined about 9° in the frontal plane (left dipole to the left, right dipole to the right) and roughly 5° to the front in the sagittal plane. The time courses of the dipole magnitudes of the two fixed dipoles are presented in Fig. 3 for the mean over subjects. The error bars mark the 95% confidence regions of the moment magnitudes corresponding to ±2 standard errors of the mean. While the maximal moments for the binaural conditions are similar, for the monaural conditions the dipole contralateral to the side of stimulation clearly exhibits a larger moment. For the monaural stimuli, the maximum of the contralateral dipole is reached 0.4 ms earlier than the maximum of the ipsilateral dipole. This time difference is similar for conditions R0+, R+0, L0- and L-0 and slightly larger for the synergistic stimuli. For the central stimuli the latencies of the maximal dipole moment coincide.

4 Discussion

The aim of this work is to analyze the correspondence between psychophyical lateralization and the neural generators of potentials evoked by lateralized stimuli. Since the generators of early auditory evoked potentials are deep, i.e., reside in the brain stem, data exhibit a relatively low SNR (signal-to-noise ratio). Therefore a large number of sweeps per stimulus condition (10000) was collected to ameliorate the signal quality.

To evaluate the data in a consistent way, equivalent current dipole locations and dipole moments were fitted based on different assumptions about the underlying dipole sources (Riedel, 2002d). Despite the small extension of the brain stem, significantly different dipole locations were detected for monaural and binaural stimuli in the case of the single rotating dipole. For binaural stimuli, a centered source with a 95% confidence region radius as small as 2 mm is found for subjects with high SNR. For monaural stimuli, the fitted dipole position is found in the contralateral hemisphere. This is physiologically meaningful because the majority of the auditory fibers projects to contralateral nuclei in the brain stem (Nieuwenhuys *et al.* 1988). However, the distance between left and right fitted sources maximally amounts to 1 cm. Given the anatomical distance of the likely generators of wave V, namely 1.6 cm for the superior olives and 2.2 cm for the nuclei of the lateral lemniscii, the fitted distances appear too small. Two reasons are conceivable to explain this discrepancy. Firstly, the homogeneous sphere which served as head model may be too simple because it does not model the attenuation effect of the skull. Compared to the brain tissue and the skin, the conductivity of the skull is about 80 times smaller (Cuffin and Cohen 1979). Ary *et al.* (1981) compared the homogeneous sphere with a three-shell head model. They showed that a dipole in the 3-shell head model must have a larger excentricity to generate approximately the same EEG as an identically oriented dipole in a homogeneous

sphere. Secondly, for monaural stimulation the ipsilateral generators will also be activated, albeit more weakly. The fitting algorithm had to optimize a single source that must explain two sources of different strengths. It consequently found the best matching position between both sources which is located nearer to the stronger source. Additionally, the rotating dipole fit unveils characteristics of the generators. Centrally perceived stimuli cause trajectories of the dipole moment in the frontal plane that mainly extend in the vertical direction. Lateral stimuli generate trajectories with smaller vertical but larger horizontal extension. This corresponds well to the results from Riedel *et al.* (2002b), taking into account that the single channels A1, A2, PO9 and P010 are orientated predominantly vertically and therefore map the vertical component m_z of the source dipole. The laterality of the stimulus is coded in the direction of rotation of the trajectory. The moment trajectories of the rotating dipole do not code the ITD or ILD alone, but show a striking correlation with the lateralization of the stimuli (see Fig. 2), i.e., stimuli with similar lateralization cause similar dipole moment trajectories. This means that ITD and ILD are not processed independently in the brain stem.

As alternative source model, a pair of hemispherically symmetric fixed dipoles was chosen. The attribute 'fixed' means that their orientations were fitted but required to be constant during the fitting interval. This constraint is physiologically motivated by the idea that in an activated brain area the direction of the current should remain constant since the orientation of the nerve fibers does not change. However, from a mathematical point of view, each rotating dipole can be considered as a superposition of three perpendicular fixed dipoles at the same location. For most subjects and stimulus conditions, the two fixed dipoles converged to nearly the same location representing a single rotating dipole that could rotate in only two dimensions. A separation of the two dipoles could only be ensured by constraining the dipoles to have a position known *a priori* which was determined from physiological constraints. A distance of 2 cm was chosen to reflect the distance of the nuclei involved in the generation of wave V. For the monaural conditions, the resulting moments show a stronger activation of the contralateral dipole which is physiologically plausible. For the lateralized binaural stimuli, the latency of the maximal dipole moment is larger for the ipsilateral dipole pointing to a faster signal transduction in the contralateral pathway.

With the current method no difference can be detected between the processing of interaural time and interaural level differences at the level of the human brain stem. This might partly be due to an encoding of intensity cues into timing/latency cues using the well-known level-latency characteristic of auditory evoked potentials. However, the level-latency relation for monaural ABR recordings (about 0.7 ms latency decrease for an intensity increase from 50 to 60 dB nHL) does not coincide with the time-intensity tradeoff found here (0.4 ms for 12 dB). Also, a lateralization due to an interaural time shift can psychophysically be distinguished from a lateralization due to an interaural intensity shift. Hence it has to be assumed that ITD and ILD are both used by an early binaural signal processing stage to derive a fast, but still not detailed estimate of the direction of an incident sound source. The two-dipole fit would also support the notion that the differential activation of both early binaural processing nuclei in the brain stem plays a key role for sound localization in humans.

Acknowledgements

The present work was supported by the *Deutsche Forschungsgemeinschaft* through the *Sonderforschungsbereich Neurokognition* (SFB 517).

References

Ary, J.P., Klein, S.A. and Fender, D.H., (1981) Location of sources of evoked scalp potentials: corrections for skull and scalp thicknesses. IEEE Trans. Biomed. Eng. 28(6), 447-452.

Brantberg, K., Hansson, H., Fransson, P.A. and Rosenhall, U. (1999) The binaural interaction component in human ABR is stable within the 0- to 1-ms range of interaural time differences. Audiol. Neurootol. 4, 88-94.

Cuffin, B.N. and Cohen, D. (1979) Comparison of the magnetoencephalogram and electroencephalogram. Electroenceph. clin. Neurophysiol. 47(2), 132-146.

Granzow, M., Riedel, H. and Kollmeier, B. (2001) Single-sweep-based methods to improve the quality of auditory brain stem responses. Part I: Optimized linear filtering. Z. Audiol. 40(1), 32-44.

Lütkenhöner, B. (1998a) Dipole source localization by means of max. likelihood estimation I. Theory and simulations. Electroenceph. clin. Neurophysiol. 106(4), 314-321.

Lütkenhöner, B. (1998b) Dipole source localization by means of max. likelihood estimation. II. Experimental evaluation. Electroenceph. clin. Neurophysiol. 106(4), 322-329.

McPherson, D.L. and Starr, A. (1995) Auditory time-intensity cues in the binaural interaction component of the auditory evoked potentials. Hear. Res. 89, 162-171.

Mosher, J.C., Lewis, P.S. and Leahy, R.M. (1992) Multiple dipole modeling and localization from spatio-temporal MEG data. IEEE Trans. Biomed. Eng. 39(6), 541-557.

Nieuwenhuys, R., Vogel, J. and van Huijzen, C. (1988) *The Human Central Nervous System - a Synopsis and Atlas*. Springer, Berlin.

Riedel, H., Granzow, M. and Kollmeier, B. (2001) Single-sweep-based methods to improve the quality of auditory brain stem responses. Part II: Averaging methods. Z. Audiol. 40(2), 62-85.

Riedel, H. and Kollmeier, B. (2002a) Auditory brain stem responses evoked by lateralized clicks: Is lateralization extracted in the human brain stem ? Hear. Res. 163(1-2), 12-26.

Riedel, H. and Kollmeier, B. (2002b) Comparison of binaural auditory brain stem responses and the binaural difference potential evoked by chirps and clicks. Hear. Res. 169(1-2), 85-96.

Riedel, H. (2002c) Dipole source analysis of auditory brain stem responses evoked by lateralized clicks. Z. Med. Phys., in press.

Riedel, H. (2002d) Analysis of early auditory evoked potentials elicited by stimuli with directional information. Ph.D. thesis, University of Oldenburg, Germany, Oldenburg. http://docserver.bis.uni-oldenburg.de/publikationen/dissertation/2002/rieana02/rieana02.html

Scherg, M. and von Cramon, D. (1985) A new interpretation of the generators of BAEP waves I-V: results of a spatio-temporal dipole model. Electroenceph. clin. Neurophysiol. 62, 290-299.

Scherg, M. (1990) Fundamentals of dipole source potential analysis. In: Grandori, F., Hoke, M. and Romani, G.L. (Eds.), *Auditory Evoked Magnetic Fields and Electric Potentials*. Karger, Basel.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Sensitivity to changes in interaural time difference and interaural correlation in the inferior colliculus

Trevor M Shackleton and Alan R Palmer

MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, U.K.,
{trevor, alan}@ihr.mrc.ac.uk

1 Introduction

For humans, the azimuthal localization of sounds below 1500 Hz is mediated primarily by sensitivity to the small difference in travel time to the two ears (interaural time difference: ITD). The smallest change in the ITD of 500-Hz pure tones detectable by humans is 20 μs (Klumpp and Eady, 1956), for cat it is 32 μs (Wakeford and Robinson, 1974), and for macaque it is 98 μs (Houben and Gourevitch, 1979). Converting the behavioural minimum audible angles for wide-band stimuli (Heffner and Heffner, 1992) yields ITD discrimination thresholds between 10-40 μs for common experimental animals. Since the variability in synaptic delay is much larger than the psychophysically determined discrimination threshold, it has been widely assumed that thresholds as low as these could only be achieved if the responses from many neurons were combined.

There are, however, instances where the resolution measurable from single neurons reflects the psychophysical thresholds. Individual neurons in the visual and somatosensory systems carry sufficient information to achieve performance comparable with psychophysical thresholds (see Skottun et al., 2001 for citations). Individual auditory nerve fibers also carry sufficient information to exceed human capability in tone detection and forward masking (Relkin and Pelli, 1987; Relkin and Turner, 1988) and to equal human capability in detecting tones in noise (Young and Barta, 1986). In light of these findings, which show that some individual neurons can match psychophysical thresholds, it is important to reexamine the assumption that behavioral ITD thresholds require pooling across many neurones.

We previously found that the best single neurones in the guinea pig inferior colliculus (IC) are capable of discrimination performance equivalent to humans and other mammals by using multiple repeats and Receiver Operator Characteristic (ROC) analysis on the measured distributions in a simulated 2-IFC task (Fig. 1: Shackleton et al., 2003a; Skottun et al., 2001). The data parallel the 1-s duration data as a function of frequency (thin line), and the best neurones show very similar thresholds to the human results also obtained with a duration of 50 ms (thick line). It is apparent that single neurones are, indeed, capable of discrimination performance equivalent to that obtained behaviourally.

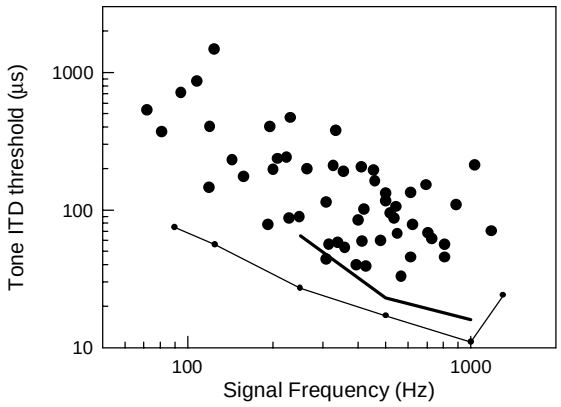


Fig. 1. Best Neural ITD thresholds for 50-ms tone bursts (filled circles). Thick line: human psychophysical thresholds with a duration of 50 ms (Hafter et al., 1979). Thin line: human psychophysical thresholds using the much longer duration of 1 s (Klumpp and Eady, 1956).

In theories of sensitivity to ITD, interaural correlation plays a central role, for example, it is the change in correlation which is detected when a dichotic tone is detected in diotic noise. This sensitivity is normally modelled using an array of coincidence detectors, tuned to a particular ITD (Jeffress, 1948), which perform the basic operation required to measure interaural correlation. Given this tight theoretical link between ITD and interaural correlation sensitivity, we might expect neural and behavioural interaural correlation discrimination thresholds to match each other as well as tone ITD discrimination thresholds match. This was tested using 50-ms long noise bursts that were equal to the guinea pig critical band centred on the best frequency of the neurone (Evans, 2001). In order to control for the token-to-token variability in interaural correlation inherent in such narrow, short, stimuli we measured the interaural correlation of the waveforms and recorded the spike counts against the measured correlation. We then performed ROC analysis in order to determine the threshold (Shackleton et al., 2003b). The thresholds are shown in Fig. 2. It is obvious that the neural thresholds do not match the behavioural thresholds, being an order of magnitude worse at references of ±1 and about double at a reference of 0.

The best neural tone ITD thresholds are comparable with human psychophysical thresholds. However, the neural interaural correlation thresholds are much worse than human thresholds. This could be because the neural substrate is being used in different ways. Assuming that the set of neurones to use for ITD discrimination can be identified, then the behavioural performance can easily be predicted from single neurone performance. If the neurones were physically ordered according to characteristic delay (i.e. within the classical “Jeffress delay line” (Jeffress, 1948)) then the optimum neurones to use would be located on the edge of the activity pattern. It does not require such a map in order for this to be true, though, any “labelling” of the neurones with their appropriate best ITDs would work. However, changes in interaural correlation are typically reported as changing the image width of the perception. Such a cue would require comparison across many neurones, so we would not expect single neurone thresholds to match behavioural performance.

Before accepting this interpretation, however, it is necessary to consider whether factors inherent to the signals used could create the observed discrepancy. The ITD thresholds were obtained using tones, whereas the interaural correlation

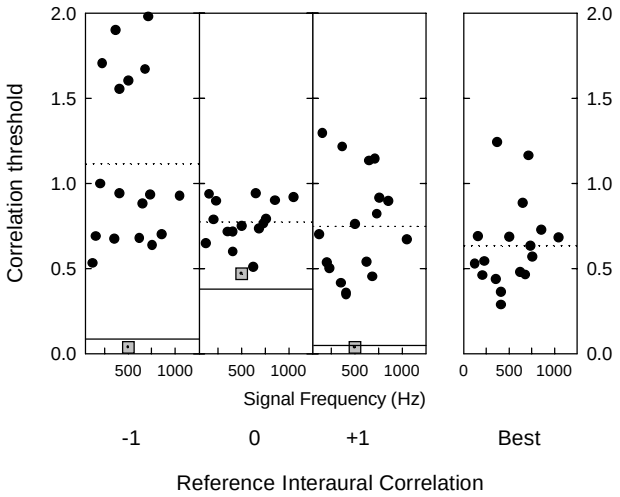


Fig. 2. Neural Interaural Correlation Thresholds. Individual panels show thresholds as a function of frequency. From left to right the panels show the thresholds away from a reference correlation of -1, 0, and 1 respectively; on the right are shown the best thresholds where the reference was chosen to minimise the threshold. The dotted lines show the mean for each reference. The squares show the human interaural correlation threshold interpolated to a duration of 50 ms with a bandwidth of 100 Hz, centred on 500 Hz (Bernstein and Trahiotis, 1997); it is assumed that the threshold for a reference correlation of –1 was the same as for +1 since this was not measured. The long, thin, solid lines show human thresholds for 400-ms duration, broadband noises where the threshold for a reference of –1 was approximately twice that for a correlation of +1 (Boehnke et al., 2002)

thresholds were obtained using noise. Although these stimuli were equated in level (20 dB above their respective detection thresholds), the neural response to noise was highly variable, probably because the neural firing patterns were often dependent upon envelope shape which is highly variable in noise stimuli. We used freshly computed noise for each trial, so it is possible that this variability introduced extra variance into the firing rate distributions and hence increased thresholds.

In order to determine whether this was a possible explanation, we remeasured the tone ITD and interaural correlation discrimination thresholds, together with *noise* ITD thresholds. A limited number of tokens of frozen noise were used and spike rate distributions were obtained for each token. This allowed both within and across token comparisons to be made.

2 Methods

Single-unit recordings using tungsten-in-glass microelectrodes were made from units with best frequencies below 1.5 kHz in the central nucleus of the right IC of Guinea pigs which were anaesthetised with urethane (1.3 g/kg in 20%

solution, i.p.), supplemented by Hypnorm (0.2 ml, i.m.) on indication. Sounds were presented in closed-field using methods previously published (e.g. Shackleton et al., 2003a). All stimuli were 50 ms long and presented every 200 ms at 20 dB above the rate threshold for that stimulus. A cosine-squared gating with rise-fall times of 2 ms was applied simultaneously to each ear after stimulus synthesis.

A coarse noise ITD function was first obtained in order to determine the best delay for the unit (either at the peak for peak units or at the trough for trough units), and the range over which to obtain the best-frequency tone and noise fine ITD functions (from peak to trough in 0.01 or 0.02 cycle steps). The noise stimuli had a bandwidth equal to the guinea pig critical band centred on the best frequency of the neurone (Evans, 2001). For each neurone, 10 noise tokens were synthesised. For noise ITD functions the same token was presented to both ears, but with that to the right ear delayed by the desired ITD. Interaural correlation functions were obtained in 0.1 steps between ±1. For each token presented to the left ear, a second, independent, noise token was synthesised and adjusted to be completely uncorrelated with the left-ear token using the Gram–Schmidt procedure (Culling et al., 2001). A fraction of this signal was then added to a fraction of the left-ear token delayed by the best ITD of the unit and presented to the right ear. The fractions of the signals to add together were obtained from the "two-independent noise generator" equation (Jeffress and Robinson, 1962). The advantage of the Gram–Schmidt procedure is that it removes any correlation between the generator noises thus ensuring that the sample interaural correlation is exactly as desired.

Responses to between 20 and 50 repeats of each token were obtained for tone ITD, noise ITD and interaural correlation conditions in blocks of 10 repeats. A block of 10 repeats was collected for each condition before the next block was collected. Since recording times of up to 120 mins were typical this interleaving was essential to ensure that changes in neural responsiveness did not affect comparisons between conditions. Results were monitored online and checked offline to ensure that the neural characteristics did not change during recording. It was felt preferable to reduce the number of repeats rather than collect nonhomogeneous data.

To determine the threshold for both ITD and interaural correlation, we first chose a reference point and calculated the standard separation (which is a form of d' in the unequal variance case, calculated as the difference in mean firing rate between target and reference, divided by the geometric mean of the standard deviations). Results based on ROC and standard separation analysis have previously been shown to give nearly identical results (Shackleton et al., 2003a) so we chose to use the simpler technique. For each possible reference point a neurometric function could thus be constructed, showing the standard separation for all possible targets. For each reference point, the threshold for a standard separation of 1 (approximately 75% correct) was determined from the function by interpolation. The process was repeated for all possible reference points. For ITD we determined the lowest threshold possible (best threshold) by shifting the reference point, and for interaural correlation the threshold obtained with a reference of +1.

For each unit, the threshold was computed for each token individually and for all tokens pooled together (i.e. ignoring which token produced which spike count). The thresholds for individual tokens shows performance which is only limited by

the intrinsic neural variability. The thresholds for all tokens pooled together shows performance which is limited by both intrinsic variability and the variability in the stimulus.

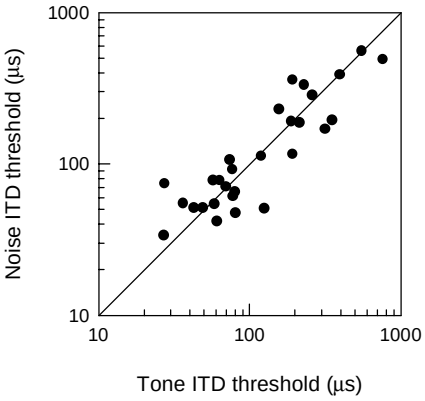


Fig. 3. Comparison of tone ITD and noise ITD thresholds. The noise thresholds were computed for all tokens pooled together. The diagonal line shows equality.

3 Results

3.1 Comparison of tone and noise ITD thresholds

If stimulus variability was elevating interaural correlation thresholds, then we would also expect it to elevate noise ITD thresholds relative to tone ITD thresholds. The comparison between tone thresholds and noise thresholds computed for all tokens pooled together is shown in Fig. 3. The thresholds are approximately equal, so there is little evidence for stimulus variability elevating thresholds.

3.2 Contribution of stimulus variability to overall variance

The total variance of each spike rate distribution in response to noise stimulation comprises two sources, the variability due to different noise tokens and intrinsic neural variability. This variance can be decomposed by comparing the variance of the mean firing rate across different tokens with the total variance pooled across all tokens. For the noise ITD functions the stimulus variance was, on average, 18% of the total variance. For the interaural correlation functions the stimulus variance was, on average, 27% of the total variance. These figures show that, although not trivial, stimulus variability is the lesser contributor to the overall variability limiting discrimination performance.

3.3 Noise ITD thresholds

Noise ITD thresholds can be computed for each individual token considered independently of the others, and for all tokens pooled together. The mean of the individual token thresholds indicates the average performance of the system if it was able to reduce variability by recognising different tokens and analysing them separately. These thresholds are plotted for each unit along the abscissa of Fig. 4, along with the range of individual token thresholds. On the ordinate are plotted the thresholds for all tokens pooled together. If stimulus variability was greatly increasing thresholds then we would expect the pooled thresholds to lie far above the line of equality. In fact, most of the points lie very close to the line, showing that stimulus variability has very little effect on the noise ITD thresholds.

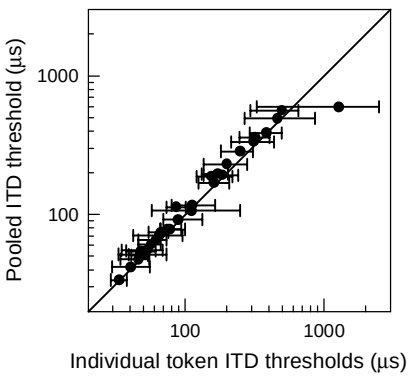


Fig. 4. Noise ITD thresholds for all noise tokens pooled together compared against thresholds for each token considered individually. The average threshold is plotted as a filled circle, and the range is indicated by the lines. The diagonal line shows equality.

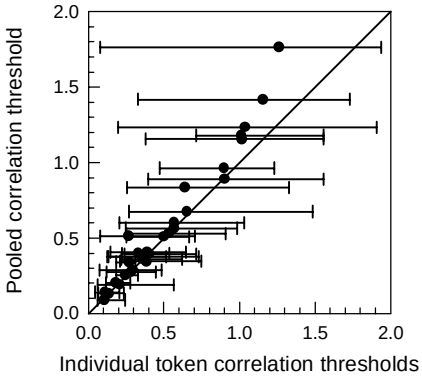


Fig. 5. Interaural correlation thresholds for all noise tokens pooled together compared against thresholds for each token considered individually. Details as Fig. 4.

3.4 Interaural correlation thresholds

Figure 5 shows the relationship between the pooled correlation thresholds and the average threshold for individual tokens analysed separately. The range of individual token thresholds is wider than for the corresponding ITD thresholds, indicating that there is substantial between token variability. However, this does not greatly elevate the pooled thresholds. The majority of thresholds are still close to the line of equality, except where the thresholds are very large.

4 Conclusions

Tone ITD discrimination thresholds measured from neurones in the IC are comparable with those measured psychophysically in humans and other mammals. However, interaural correlation discrimination thresholds are much worse. We considered the hypothesis that stimulus variability due to different noise tokens having very different waveforms elevates the correlation thresholds. Response variability is increased because of stimulus variability, but controlling for stimulus variability is not sufficient to greatly decrease thresholds. On the assumption that only the firing rates of neurones are used, then single neurones are sufficient to mediate behavioural ITD discrimination (if they are identified correctly: see introduction). However, single neurones cannot mediate interaural correlation discrimination, where presumably a comparison across the population needs to be made. The use of other coding strategies, such as the use of spike timing, might also resolve this discrepancy.

References

Bernstein, L.R. and Trahiotis, C. (1997) The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation. J. Acoust. Soc. Am. 102, 1113-1120.

Boehnke, S.E., Hall, S.E. and Marquardt, T. (2002) Detection of static and dynamic changes in interaural correlation. J. Acoust. Soc. Am. 112, 1617-1626.

Culling, J.F., Colburn, H.S. and Spurchise, M. (2001) Interaural correlation sensitivity. J. Acoust. Soc. Am. 110, 1020-1029.

Evans, E.F. (2001) Latest comparisons between physiological and behavioural frequency selectivity. In: A.J.M. Houtsma, A. Kohlraush, V.F. Prijs, R. Schoonhoven, (Eds.), *Physiological and psychophysical bases of auditory function*. Shaker Publishing BV, Maastricht. pp. 382-387.

Hafter, E.R., Dye, R.H. and Gilkey, R.H. (1979) Lateralization of Tonal Signals Which Have Neither Onsets Nor Offsets. J. Acoust. Soc. Am. 65, 471-477.

Heffner, R.S. and Heffner, H.E. (1992) Visual factors in sound localization in mammals. J. Comp. Neurol. 317, 219-232.

Houben, D. and Gourevitch, G. (1979) Auditory lateralization in monkeys: An examination of two cues serving directional hearing. J. Acoust. Soc. Am. 66, 1057-1063.

Jeffress, L.A. (1948) A place theory of sound localization. J. Comp. Psychol. 44, 35-39.

Jeffress, L.A. and Robinson, D.E. (1962) Formulas for the coefficient of correlation for noise. J. Acoust. Soc. Am. 34, 1658–1659.

Klumpp, R.G. and Eady, H.R. (1956) Some measurements of interaural time difference thresholds. J. Acoust. Soc. Am. 28, 859-860.

Relkin, E.M. and Pelli, D.G. (1987) Probe Tone Thresholds in the Auditory Nerve Measured by Two-interval Forced-choice Procedures. J. Acoust. Soc. Am. 82, 1679 - 1691.

Relkin, E.M. and Turner, C.W. (1988) A Reexamination of Forward Masking in the Auditory Nerve. J. Acoust. Soc. Am. 84, 584 - 591.

Shackleton, T.M., Skottun, B.C., Arnott, R.H. and Palmer, A.R. (2003a) Interaural Time Difference Discrimination Thresholds for Single Neurons in the Inferior Colliculus of Guinea Pigs. J. Neurosci. 23, 716-724.

Shackleton, T.M., Arnott, R.H. and Palmer, A.R. (2003b) Sensitivity to changes in Interaural Correlation in the Inferior Colliculus of the Guinea Pig. J. Neurosci., in preparation.

Skottun, B.C., Shackleton, T.M., Arnott, R.H. and Palmer, A.R. (2001) The ability of inferior colliculus neurons to signal differences in interaural delay. Proc. Natl. Acad. Sci. U. S. A. 98, 14050-14054.

Wakeford, O.S. and Robinson, D.E. (1974) Lateralization of tonal stimuli by the cat. J. Acoust. Soc. Am. 55, 649-652.

Young, E.D. and Barta, P.E. (1986) Rate Responses of Auditory Nerve Fibers to Tones in Noise Near Masked Threshold. J. Acoust. Soc. Am. 79, 426 - 442.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Processing of interaural temporal disparities with both "transposed" and conventional stimuli

Leslie R. Bernstein and Constantine Trahiotis

Dept. of Neuroscience and Dept. of Surgery (Otolaryngology)
University of Connecticut Health Center, Farmington, CT 06032 USA

1 Introduction

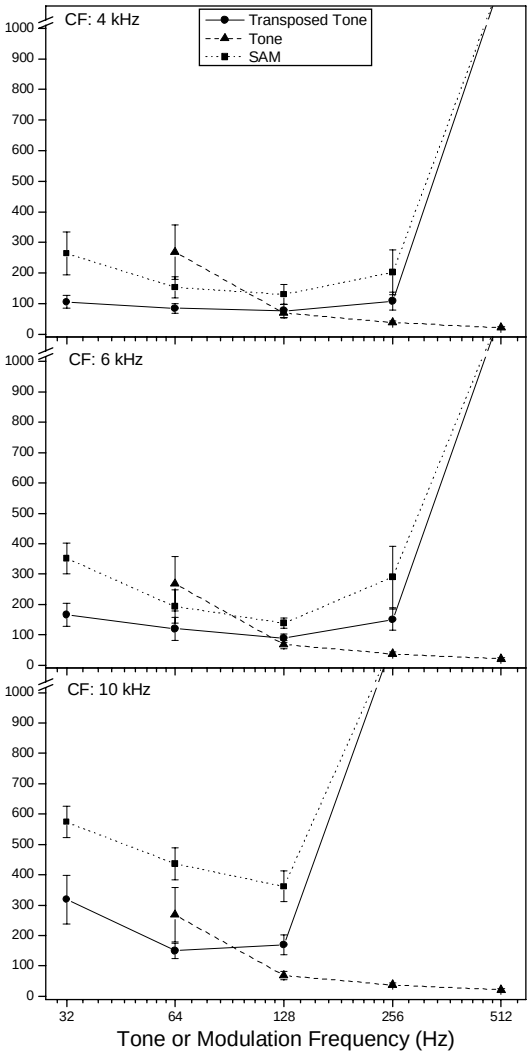
At the 12th International Symposium on Hearing, we presented preliminary data demonstrating that the processing of interaural temporal disparities (ITDs) at high frequencies could be enhanced by using "transposed stimuli" similar to those first employed by van de Par and Kohlrausch (1997). The stimuli were designed to provide the high-frequency channels of the binaural processor with envelope-based information that mimics waveform-based information normally available in low-frequency channels. Since then, we have collected comprehensive sets of data concerning threshold-ITDs and extents of laterality produced by high-frequency transposed stimuli, high-frequency "conventional" stimuli, and low-frequency, conventional stimuli. The purpose of this presentation is to discuss how quantitative analyses of those data have provided new insights regarding the processing of envelope-based binaural information at high frequencies. For example, in order to account for threshold ITDs as a function of rate of modulation, it appears that one must include a stage of low-pass filtering of the envelopes within the high-frequency channels of each ear. In addition, although stimulus-based consistency of ITD ("straightness") produces similar lateralization effects at high and low frequencies, it appears that those effects are produced in different manners. Specifically, at low frequencies, it appears necessary to postulate *across*-auditory-filter interactions in order to account for the data. At high frequencies however, it appears that only *within*-auditory-filter interactions suffice.

2 Threshold ITDs

Threshold ITDs were measured with a two-cue, two-alternative, forced choice adaptive task. Three types of stimuli were employed: 1) low-frequency pure tones 2) 100% sinusoidally amplitude-modulated (SAM) high-frequency tones and 3) "transposed" high-frequency stimuli the envelopes of which were designed to provide the high-frequency channels with envelope-based interaural timing information

similar to that normally conveyed via low-frequency tonal stimuli. Transposed stimuli were generated by multiplying rectified, low-pass filtered (2 kHz cutoff) low-frequency tones by high-frequency carriers (Bernstein and Trahiotis, 2002). Four normal-hearing adults served as listeners and estimates of thresholds were calculated by averaging the individual thresholds obtained from six separate adaptive runs.

Figure 1 displays the mean threshold ITDs computed across the four listeners when the center frequency of the SAM and transposed stimuli was either 4 kHz, 6 kHz, or 10 kHz. The thresholds are plotted as a function of either the frequency of the low-frequency pure tone or the frequency of modulation of the high-frequency SAM and transposed stimuli. By "frequency of modulation" of a transposed stimulus, we mean the frequency of the pure tone that was used to generate it. Error bars represent ± 1 standard error of the mean. The breaks in the lines connecting the data and in the vertical axes indicate rates of modulation for which two or more listeners could not perform the task, even with ITDs of up to 1 ms.



The patterning of the data indicates that thresholds obtained with high-frequency transposed stimuli: 1) are consistently smaller than those obtained with high-frequency SAM tones and 2) at frequencies of modulation of 128 Hz and 64 Hz, are as small or smaller than thresholds obtained with low-frequency pure tones. More important for our purposes, the data obtained at all three center frequencies indicate that, in general, thresh-

Fig. 1. Threshold ITDs averaged across the four listeners as a function of the modulation or pure-tone frequency. The center frequency of the high-frequency SAM and transposed stimuli was 4, 6, or 10 kHz. The error bars represent ± standard error of the mean.

olds increased as the rate of modulation was increased beyond 128 Hz. Furthermore, thresholds increased more rapidly with rate of modulation, and were more often unmeasurable, as center frequency of the SAM and transposed stimuli was increased to 10 kHz.

These latter effects *cannot* be explained solely by assuming that peripheral bandpass filtering reduces depth of modulation as rate of modulation is increased, thereby degrading the binaural processing of ITDs. If that were true, then increasing center frequency would improve performance at the higher rates of modulation where the auditory filters are broader.

The data are consistent with a mechanism that limits the ability to "follow" or to encode high rates of fluctuation of the envelope. Data and arguments supporting the existence of such a "rate limitation" can be found in several binaural investigations (e.g. McFadden and Pasanen, 1976; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1992a, 1992b, 1994). Additional empirical evidence that an envelope rate-limitation is manifest at high spectral frequencies can be found in recent studies using monaural and/or diotic stimuli (Kohlrausch et al., 2000, Ewert and Dau, 2000). In those studies, temporal modulation transfer functions (TMTFs) were measured at various center frequencies. The results led those authors to include in their quantitative models a low-pass filter that attenuates fluctuations of the envelope that are more rapid than 150 Hz, independent of the center frequency of the stimulus. It appears from those studies and from a more recent study by Moore and Glasberg (2001) that the low-pass filtering of the envelope information is a monaural process that occurs prior to binaural interaction.

2.1 Quantitative analyses

The model used to make the predictions for the data in Fig. 1 included "envelope compression" (exponent = 0.23), square-law rectification, and low-pass filtering at 425 Hz to capture the loss of neural synchrony to the fine-structure of the stimuli that occurs as the center frequency is increased (see Bernstein and Trahiotis, 1996b; Bernstein et al., 1999). The model included a peripheral stage of bandpass filtering via Gammatone filters (see Patterson et al., 1995) which, like the stimuli, were centered at either 4, 6, or 10 kHz. It was assumed that the listener's threshold ITD for stimuli at a given center frequency, reflects a constant change of normalized interaural correlation.

The three panels of Figure 2 display the predictions of the model for SAM and transposed stimuli centered at 4, 6, and 10 kHz, respectively. The dotted lines indicate predicted thresholds of the model. Thresholds for SAM and transposed stimuli having rates of modulation less than 256 Hz are accounted for successfully. The model fails at the higher rates of modulation where threshold were either substantially elevated or could not be obtained.

Note, however, that augmenting the model with a 150-Hz low-pass filter (solid lines) allows it to capture virtually all aspects of the data. The augmented model accounts both for the elevated thresholds found at the higher rates of modulation and, to us quite amazingly, for instances where the average listener was essentially unable to perform the task. The amount of variance in the data that was accounted for by the augmented model was 86% at 4 kHz, 96% at 6 kHz, and 77% at 10 kHz.

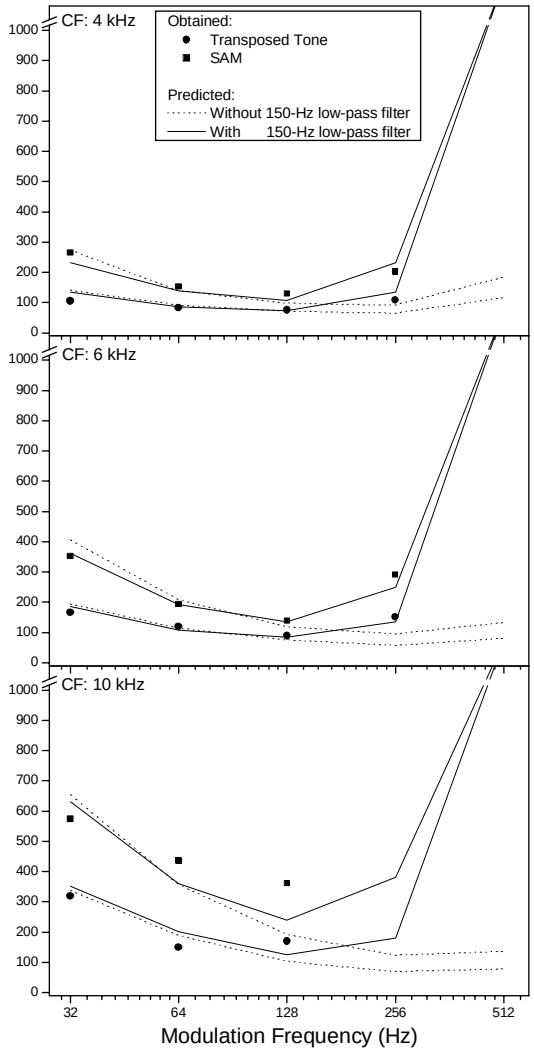


Fig. 2. Threshold ITDs (symbols) and predictions (lines) for the SAM and transposed stimuli.

3 Lateralization and consistency of ITD

Measures of ITD-based extents of laterality obtained with *low-frequency stimuli* have revealed that across-frequency consistency of interaural timing information can influence greatly the intracranial position of acoustic images (e.g., Jeffress, 1972; Stern et al., 1988; Trahiotis and Stern, 1989). Stern and his colleagues accounted for these effects using interaural cross-correlation-based models of binaural processing. Their explanation rests upon two aspects of a putative internal representation of the binaural stimuli. One is termed "straightness" and refers to the extent to which maxima of the internal cross-correlation of the stimuli occur at the same internal delay over a range of frequencies. The second aspect, termed "centrality," refers to the extent to which those maxima are located at small internal delays. Straightness and centrality are cast as two, sometimes conflicting, weighting functions that determine the relative salience of individual peaks of the cross-correlation function.

The data shown in Figure 3 were obtained in an acoustic pointing task. Listeners adjusted the interaural intensitive disparity (IID) of a 200-Hz-wide noise centered at 500 Hz (the pointer) so that its intracranial position matched that of the experimenter-controlled stimuli that served as "targets" (see Bernstein and Trahiotis, 1985). The targets were either low-frequency bands of Gaussian noise centered at 250 Hz (open symbols) or their transposed counterparts centered 4 kHz (closed symbols).

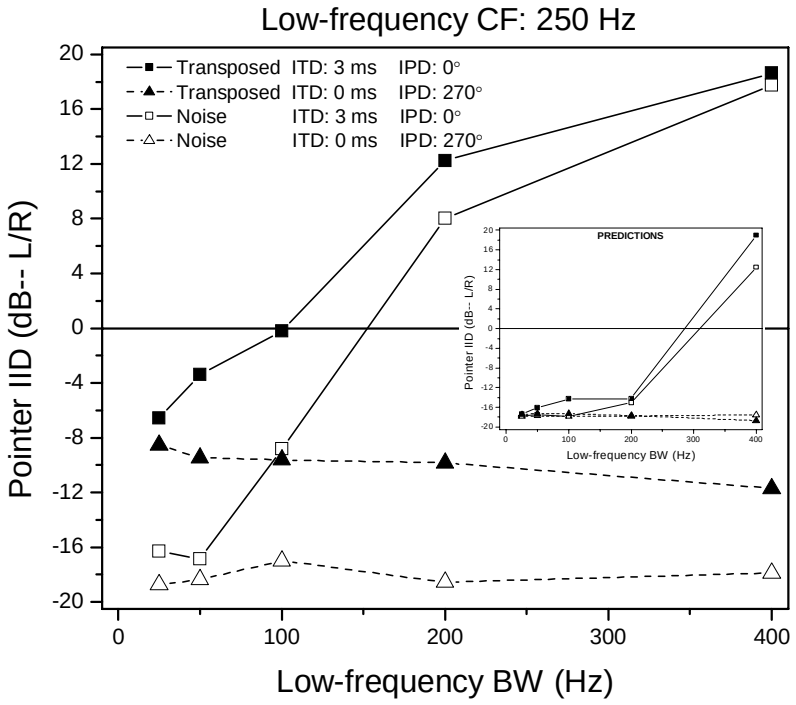


Fig. 3. IID of the pointer as a function of the bandwidth of the target. The data points represent the mean values computed across the four listeners. Inset: Model Predictions

Consider data obtained with low-frequency bands of noise having an ITD of 3.0 ms and an IPD of 0° (open squares). They indicate that, increases in bandwidth moved the intracranial image away from the lagging ear, across the midline and, when bandwidth reached 400 Hz very far toward the leading (left) ear. In contrast, when the same bandwidths of noise had an ITD of 0 ms and an IPD of 270° (open triangles), the intracranial images remained very far toward the right (lagging) ear. These outcomes replicate the "straightness/centrality" effects reported by Trahiotis and Stern (1989) for bands of noise centered at 500 Hz.

Remarkably, the data obtained with the transposed stimuli reveal parallel effects. Note, however, that transposed noises with an ITD of 0 ms and an IPD of 270° are not lateralized as far as are low-frequency bands of Gaussian noise.

The principal finding is that the effects on laterality produced by combinations of bandwidth, ITD, and IPD are very similar for low-frequency noises and their transposed counterparts. As will be seen, this does not mean that straightness and centrality along the internal cross-correlation surface influence lateral position in the same manner for both types of stimuli.

3.1 Quantitative analyses

The model described in section 2.1 was supplemented by an extended bank of Gammatone filters. The central binaural processor was represented as a cross-correlogram with one dimension representing frequency and a second dimension representing values of delay (τ). The third dimension was the value of the cross-products integrated over the 100-ms duration of the stimuli. Laterality was defined as the value of τ associated with the most central peak of activity of the across-frequency averaged cross-correlogram. This definition was based on the knowledge that, all other things being equal, peaks of the cross-correlation function that are closest to midline dominate perceived lateral position (e.g., Stern et al., 1988).

The predictions of the model, in units of τ (*internal* delay), were transformed to units of IID of the acoustic pointer using functions relating IID of the pointer to ITDs obtained with low-frequency noises (see Bernstein and Trahiotis, 2003).

The predictions are in the inset of Fig. 3. The model correctly predicts that for the 3.0 ms/0° case, the narrowest bands of low-frequency noise and their 4-kHz-centered transposed counterparts are each lateralized toward the lagging ear and that the 400-Hz-wide band of noise and its transposed counterpart are lateralized toward the leading ear. For the ITD/IPD combination of 0 ms/270°, the model also correctly predicts that low-frequency Gaussian noises and their transposed counterparts are lateralized toward the lagging ear. The model does not capture the gradual changes in the locus of the intracranial image seen in the 3.0 ms/0° data as bandwidth increases from 50 to 200 Hz. This seems to result from the absence of functions designed to "weight," differentially activity within the cross-correlogram. Several attempts to define suitable high-frequency functions proved unsuccessful.

The similarities in the behavioral data found with low-frequency noises and their transposed counterparts almost surely do not result from similar processing of their respective cross-correlograms. At low center frequencies, the combining of binaural timing information for the 3.0 ms/0° stimulus condition *must occur across auditory filters*. This is so because the bandwidths of the external stimuli that affect laterality exceed the bandwidths of those filters. In contrast, the vast majority of the energy of the transposed stimuli centered at 4 kHz is contained within the auditory filter centered at that frequency. Therefore, no across-auditory-filter integration is required. This notion was verified in that predictions of a model employing *only one* pair of left/right auditory filters centered at 4 kHz were virtually identical to those obtained using the entire bank of Gammatone filters.

Acknowledgments

This work was supported by research grants DC-04147, DC-04073, and DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References

Bernstein, L. R. and Trahiotis, C. (1992a). Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies. J. Acoust. Soc. Am. 91, 306-316.

Bernstein, L. R. and Trahiotis, C. (1992b). Detection of antiphasic sinusoids added to the envelopes of high-frequency bands of noise. Hearing Research, 62, 157-165.

Bernstein, L. R. and Trahiotis, C. (1994). Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise. J. Acoust. Soc Am. 95, 3561-3567.

Bernstein, L. R. and Trahiotis, C. (1996). The normalized correlation: Accounting for binaural detection across center frequency. J. Acoust. Soc Am. 100, 3774-3784.

Bernstein, L. R., and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli. J. Acoust. Soc Am. 112, 1026-1036.

Bernstein, L. R., and Trahiotis, C. (2003). Enhancing interaural-delay-based extents of laterality at high frequencies by using "transposed stimuli". J. Acoust. Soc Am, in press.

Bernstein, L. R., and Trahiotis, C. (1985). Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities. J. Acoust. Soc. Am. 77, 1868-1880.Beveridge, H.A. and Carlyon, R.P. (1996) Effects of aspirin on human psychophysical tuning curves in forward and simultaneous masking. Hear. Res. 99, 110-118.

Bernstein, L. R., Par, Steven van de, and Trahiotis, C. (1999). The normalized correlation: Accounting for NoSπ thresholds obtained with Gaussian and "low-noise" masking noise. J. Acoust. Soc Am. 106, 870-876.

Ewert, S. D., and Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. J. Acoust. Soc Am. 108, 1181-1196.

Jeffress, L. A. (1972). Binaural signal detection: Vector theory. In: J. V. Tobias (Ed.), *Foundations of Modern Auditory Theory (Vol 2.).* Academic Press, New York, pp. 349-368.

Kohlrausch, A., Fassel, R., and Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J. Acoust. Soc Am. 108, 723-734.

McFadden, D. and Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. J. Acoust. Soc. Am. 59, 634-639.

Moore, B. C. J., and Glasberg, B. R. (2001). Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. J. Acoust. Soc. Am. 110, 1067-1073.

Nuetzel, J. M. and Hafter, E. R. (1981). Discrimination of interaural delays in complex waveforms: Spectral effects. J. Acoust. Soc. Am. 69, 1112-1118.

Par, S van de, and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. J. Acoust. Soc Am. 101, 1671-1680.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. J. Acoust. Soc Am. 98, 1890-1894.

Stern, R.M., Zeiberg, A.S., and Trahiotis, C. (1988). Lateralization of complex binaural stimuli: A weighted image model. J. Acoust. Soc. Am. 84, 156-165.

Trahiotis, C., and Stern, R. M. (1989). Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. J. Acoust. Soc Am. 86, 1285-1293.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Sound localization in the frontal horizontal plane by post-lingually deafened adults fitted with bilateral cochlear implants

D. Wesley Grantham, Daniel H. Ashmead, and Todd A. Ricketts

Vanderbilt Bill Wilkerson Center, Vanderbilt University School of Medicine, USA,
d.wesley.grantham@vanderbilt.edu

1 Introduction

Multi-channel cochlear implants have enabled many patients with severe-to-profound hearing loss to achieve near-normal levels of communication in some quiet listening situations (Helms, Müller, and Schön 1997). Bilateral implantation, which has been increasingly performed in recent years, affords the additional potential advantage of an increased awareness of auditory space. Initial studies of persons with bilateral implants have reported that localization of sources in the horizontal plane is superior when both implants are active than when either implant is turned off (Tyler, Gantz and Rubinstein 2002; van Hoesel, Ramsden, and O'Driscoll 2002; Nopp, Schleich, and D'Haese 2003). The superior performance under bilateral-implant conditions suggests that subjects are able to take advantage of interaural differences produced by their cochlear implant devices.

The present study extends previous work by testing horizontal-plane localization performance in a large number of bilaterally-implanted persons (eventually 24), employing a large number of sound sources spanning the frontal horizontal plane. In addition, in order to determine if a speech stimulus may be more readily localizable than a noise burst, the current investigation measured localization performance with both types of signals.

2 Method

2.1 Participants

Participants were nine severely-to-profoundly post-lingually deafened adults who were bilaterally implanted with the MED-EL C40+ cochlear device. All participants were fit using standard clinical procedures. There were four females and five males, ranging in age from 25 to 77 years. Eight of the nine had had both devices

implanted in a single surgical procedure, 5-7 months prior to testing. The ninth participant (G) was implanted in two separate surgeries, 14 months apart (the second ear was implanted 29 months prior to testing). All participants were USA residents, and were flown to the Vanderbilt Bill Wilkerson Center in Nashville, TN, for the two-day series of tests [testing included distance perception and speech recognition tasks, not reported here].

Participants each took part in two sessions (on succeeding days), each of which lasted about two hours. Frequent breaks were allowed within a session to avoid fatigue. Subjects were instructed to select the program and volume settings on their devices that they were most accustomed to using. Sensitivity was adjusted to maximum prior to testing.

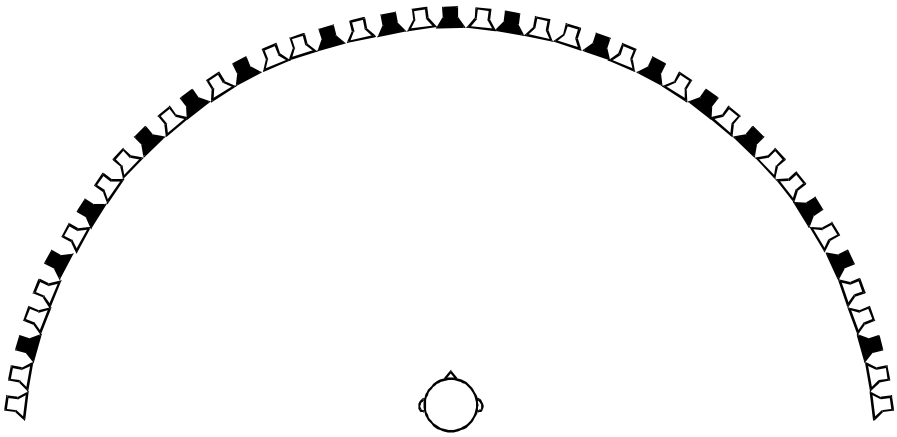


Fig. 1. Loudspeaker configuration in the anechoic chamber.

2.2 Testing environment and stimuli

Participants were tested individually in a lighted anechoic chamber. During a session, the participant was seated in the center of the room and instructed to maintain an upright, forward orientation. Forty-three stationary loudspeakers (JBL 8110) were positioned in a horizontal arc at ear level, 1.8 m in front of the listener, extending from –90° to +90° azimuth (Fig. 1). Although 43 loudspeakers were available, only 17 loudspeakers were used for this experiment (shown as the filled speakers in Fig. 1).

Two stimuli were employed: (1) NOISE, a 200-ms Gaussian noise burst, bandpass filtered from 100-4000 Hz; (2) SPEECH, a 200-ms sampling of a male voice uttering the word "hey!" also bandpass filtered from 100-4000 Hz. Signals were output at a rate of 25 ksps via a Tucker-Davis Power-Dac (PD1). Nominal stimulus level was 70 dB SPL in all conditions, as measured at the position of the participant's head. The intensity level for each stimulus presentation was randomly varied over a 10-dB range around the nominal stimulus level to thwart the use of intensity cues.

2.3 Procedure

A modified source identification task was employed (e.g., Rakerd and Hartmann, 1985; "modified" since only 17 of the 43 sources were employed). The participant was seated in the center of the room facing the center of the loudspeaker array, as shown in Fig. 1. A run consisted of the presentation of the stimulus 68 times (4 times from each of the 17 loudspeakers employed) in a random order. For each run of 68 trials, lasting approximately 7 minutes, the stimulus (NOISE or SPEECH) was held constant.

Prior to each stimulus presentation the participant directed his or her gaze to center of the array (Speaker #22) and pushed a button on a lap-held response box when ready. After the stimulus presentation the subject responded by calling out the loudspeaker number s/he believed produced the sound, swiveling in the chair, if necessary, to read the loudspeaker label. Since the stimulus duration was 200 ms, the participant was not able to move his or her head prior to the termination of the signal. The investigator (monitoring via intercom from the adjacent control room) keyed the response into the computer. Feedback was not provided. The participant was unaware that only 17 of the loudspeakers were employed.

On each day of testing, following some practice trials, nine runs were completed, consisting of three successive runs with the left implant only (LEFT), three successive runs with the right implant only (RIGHT), and three successive runs with both implants (BOTH). The order of conditions (LEFT, RIGHT, BOTH) was fully counterbalanced across participants, and for a given participant, the same conditions were visited in reverse order on the second day of testing. The stimulus (NOISE or SPEECH) was alternated from run to run through the complete set of 18 runs conducted across the two days. Thus, for each condition (LEFT, RIGHT, BOTH) and for each stimulus (SPEECH, NOISE), there were 12 responses per loudspeaker.

3 Results and discussion

Data from one participant (H) for the SPEECH stimulus are shown in Fig. 2. Overall rms error $\underline{D}$ for this participant in the BOTH condition was 18.7° ($\underline{D}$ ranged from 18.1° to 39.0° across the 9 participants). For comparison, data from a normal-hearing listener are shown in the inset at the lower right ($\underline{D}$ = 8.7°).

The three large panels in Fig. 2 display performance for the LEFT, RIGHT, and BOTH conditions. In each panel, mean response azimuth is plotted as a function of stimulus azimuth, with error bars indicating ±1 standard deviation computed across the 12 responses given for each stimulus. The diagonal line represents perfect performance. As can be seen, participant H's mean responses in the BOTH condition lie fairly close to the diagonal for stimulus azimuths between ±40°, but for more laterally presented stimuli, responses diverge from the diagonal. Variability (shown by the error bars) is greater in all conditions than that shown by the normal listener.

For the unilateral (LEFT and RIGHT) conditions, it can be seen that this participant could not localize the stimulus: the sound was always perceived to come

from the side of the active implant. This complete failure to localize signals under unilateral conditions occurred for all 9 listeners (see below).

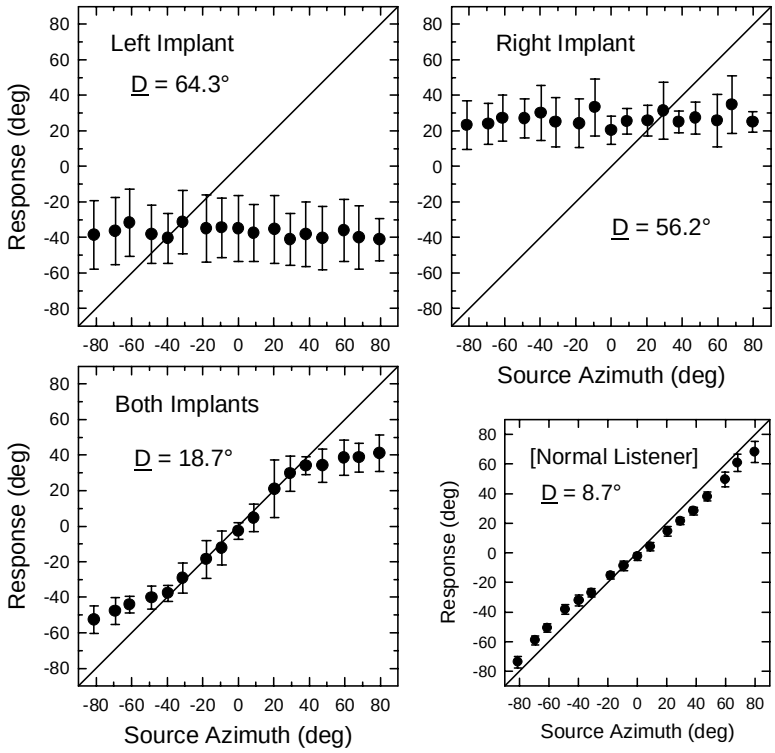


Fig. 2. Localization responses for participant H. Inset at lower right shows data from a normal-hearing listener.

3.1 Analysis of random and constant errors

The overall rms error *D*, shown in the panels of Fig. 2, represents the combined effects of random error (i.e., the standard deviation of responses, indicated by the error bars) and the unsigned constant error (the rms deviation of the mean responses from the diagonal) (Rakerd and Hartmann, 1985). To fully understand the results, it is necessary to consider each of these measures separately. [Here we employ an *adjusted* unsigned constant error, which is independent of the laterality of the overall mean response averaged over all stimulus azimuths.]

3.2 Unilateral performance

Adjusted constant error is shown in Fig. 3 for all participants, for LEFT and RIGHT conditions. Since performance was equivalent for the NOISE and SPEECH stimuli,

error rate is shown for both stimuli combined. The upper dashed line represents chance performance, and the lower dashed line shows the measure computed for the normal listener whose data are shown in Fig. 2.

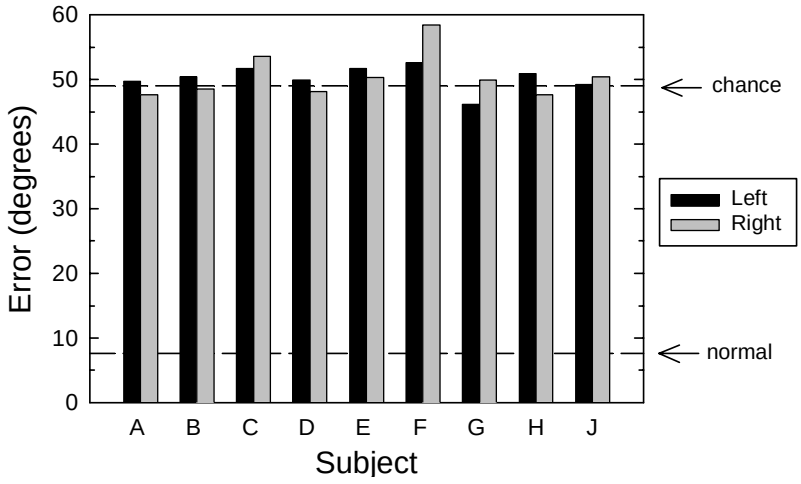


Fig. 3. Unsigned constant error for the nine participants in the unilateral conditions.

It is clear that all participants responded at chance level when listening unilaterally. This result is consistent with previous reports that most bilaterally implanted persons (but not all) respond at chance level when listening unilaterally (van Hoesel *et al* 2002; Tyler *et al.* 2002; Nopp *et al.* 2003). Under unilateral conditions, sound sources all appear to emanate from the side of the active device.

On the other hand, there was considerable variability among our participants, both in how extreme their lateral responses were under unilateral conditions, and in the variability of their responses. Due to space limitations, these data are not displayed or discussed here.

3.3 Bilateral performance

Unsigned constant error is plotted for the BOTH condition for all nine subjects in Fig. 4, for both the SPEECH and NOISE stimuli. As in Fig. 3, horizontal lines in the figure indicate chance level (upper line) and performance from one normal listener (lower line). Participants' performance spans the range between these two limits: for example, G shows near-normal performance, while C shows near-chance performance).

All participants showed a smaller constant error with the SPEECH stimulus than with the NOISE stimulus, although for most participants the difference was slight. Interestingly, G, the best localizer (reaching normal levels for the SPEECH stimulus) was the one who had received his implants in two separate surgeries (separated by 14 months). He is also unique in that testing occurred over 2 years

after the most recent surgery (the delay was 5-7 months for all others). Future study will be required to determine whether duration of use of the cochlear implants will lead to improved performance in other users.

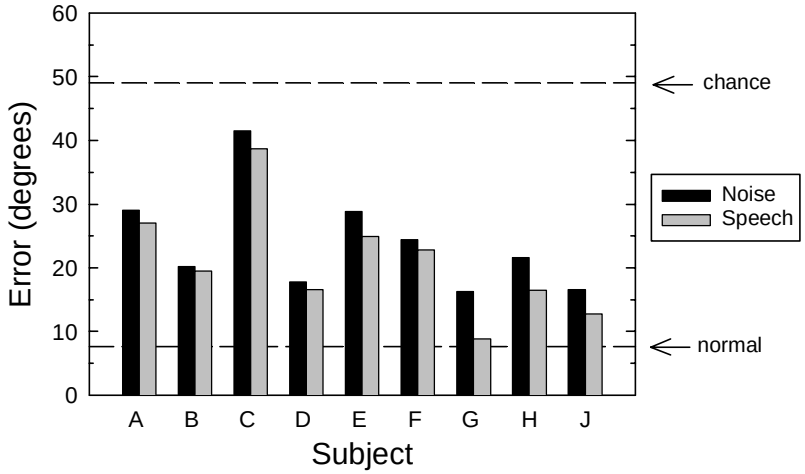


Fig. 4. Unsigned constant error for the nine participants in the bilateral condition.

4 Conclusions

These results replicate and extend those of van Hoesel *et al* (2002), Tyler *et al.* (2002), and Nopp *et al.* (2003). Given that unilateral performance was at or near chance level for all our participants, we can conclude that there is little or no monaural contribution to horizontal localization performance in our task with these nine bilaterally implanted individuals. Thus, the performance they exhibited in the BOTH condition must have been based on the availability and use of interaural difference cues – specifically, on interaural time differences (ITDs) and/or interaural level differences (ILDs). The cochlear implant devices are clearly able to provide one or both types of these cues, in order that all participants, to some degree, were able to achieve localization performance above chance. Additionally, there was some indication that the SPEECH stimulus was more localizable than the NOISE stimulus (all 9 participants had lower constant error with the SPEECH), although it is not yet known if this is a reliable difference.

There have been reports that bilaterally implanted individuals are quite sensitive to ILD cues, but relatively insensitive to ITD cues (van Hoesel, Tong, Hollow, and Clark 1993; van Hoesel *et al.* 2002). For example, best ITD thresholds obtained from bilaterally implanted persons have been on the order of 300-500 μs, more than an order of magnitude larger than ITD thresholds obtained in normal listeners (Grantham, 1995). Based on these reports, it has been surmised that the better-than-

chance localization performance shown by cochlear implantees must be based on the processing of ILD cues.

It is difficult without collecting further data to determine which interaural cues underlie horizontal-plane localization in bilaterally implanted persons. It should be noted that previous investigations that have measured ITD sensitivity in bilateral patients have employed direct electrical stimulation of one or two electrode pairs on the two devices. One may argue that this type of direct stimulation would provide the best opportunity to observe binaural interaction; however, it would be of interest to measure ITD thresholds employing broadband acoustic signals as employed in this and previous localization experiments. Possibly the use of wideband signals that simultaneously stimulate several channels would result in lower ITD thresholds than have been reported to date, thus revealing ITD cues to be potential contributors to localization performance.

Acknowledgments

This research was supported in part by grants from NIDCD and Med-El Corporation. The authors are indebted to Lindsay Russell for collecting the data and for other assistance, and to the participants for donating their time to this project.

References

Grantham, D. W. (1995) Spatial hearing and related phenomena. In: B.C.J. Moore (Ed.), *Handbook of Perception and Cognition: Hearing.* Academic Press, San Diego, pp. 297-345.

Helms, J., Müller, J., Schön, F., Moser, L., Arnold, W., Janssen, T. *et al*. (1997) Evalutation of performance with the COMBI40 cochlear implant in adults: A multicentric clinical study. ORL J. Otorhinolaryngol. Relat. Spec. 59, 23-35.

Nopp, P., Schleich, P. and D'Haese, P. (2003) Sound localization in bilateral users of MED-EL COMBI 40/40+ implants," Ear and Hear. In press.

Rakerd, B. and Hartmann, W. M. (1985) Localization of sound in rooms. II: The effects of a single reflecting surface. J. Acoust. Soc. Am. 78, 524-533.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., Lowder, M. W. (2002) Three-month results with bilateral cochlear implants. Ear and Hear. 23, 80S-89S.

van Hoesel, R. J. M., Tong, Y. C., Hollow, R. D. and Clark, G. M. (1993) Psychophysical and speech perception studies: A case report on a binaural cochlear implant subject. J. Acoust. Soc. Am. 94, 3178-3189.

van Hoesel, R., Ramsden, R. and O'Driscoll, M. (2002) Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. Ear and Hear. 23, 137-149.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise

Steven van de Par[1], Armin Kohlrausch[1,2], Jeroen Breebaart[1], and Martin McKinney[1]

[1] Philips Research Laboratory, Eindhoven, The Netherlands{Steven.van.de.Par, Armin.Kohlrausch, Jeroen.Breebaart, Martin.McKinney}@philips.com
[2] Eindhoven University of Technology

1 Introduction

Binaural listening leads to considerable improvements in detection thresholds in masked listening conditions compared to monaural listening. When an interaurally out-of-phase tonal signal is masked by an in-phase noise masker, detection thresholds can be as much as 25 dB lower than when detecting an in-phase tone within the same noise masker provided that the masker is narrow-band (e.g. Zurek and Durlach, 1987).

When high-level auditory processing of speech for example is considered, detectability of target signals may not be the most relevant perceptual parameter. The extent to which specific properties (e.g. temporal envelope) of the target signal are audible may be more relevant for high-level auditory processing. This notion motivated our investigation of the discriminability of pairs of target signals with different temporal envelope structures that are partially masked by noise.

Studies on discriminability on the basis of frequency differences (Henning, 1973) and interaural time differences (Cohen, 1981; Stern, Slocum, and Phillips, 1983) have been done. Henning showed that for a low signal-to-noise ratio, frequency discrimination between sinusoidal signals in the presence of white diotic noise improved when the sinusoids were presented interaurally out-of-phase rather than in-phase. In these experiments, detection thresholds and frequency discrimination thresholds were both lower in the dichotic condition. It was concluded that apparently the information needed to discriminate frequency is preserved beyond the first stages of binaural interaction.

Cohen (1981) found that midline interaural time delay jnds for a tone were smallest when presented in diotic noise, increased with interaurally uncorrelated noise, but were largest when presented in interaurally out-of-phase noise. In contrast to Hen-

ning's findings, in these experiments ITD discrimination thresholds were highest in conditions where detection thresholds were lowest.

In the present study we explore to what extent and under which conditions listeners have a binaural advantage in the processing of temporal information of target signals. For this purpose we measured the discriminability between bandlimited noise and harmonic-complex-tone targets presented in diotic noise. The targets were presented either interaurally in-phase or out-of-phase. In addition, we measured detection thresholds for the noise target and the harmonic-complex-tone target.

2 Experiment I: the effect of target-signal bandwidth

2.1 Method and stimuli

Within each interval of a discrimination task, a 300 ms target was presented, temporally centred in a 400 ms masker. The masker was a 2-kHz low-pass noise that was interaurally in-phase with an SPL of 65 dB. Two types of targets were used: bandpass noise (BPN) with a flat spectral envelope; and a harmonic tone complex (HTC) with the same band-pass characteristic as the BPN, component spacings of 20 Hz, and a sinusoidal phase spectrum. In narrow-band conditions, the targets were 100 Hz wide and centred at 1 kHz, the first of five components of the HTC started at 960 Hz. In broad-band conditions, the targets were 1500 Hz wide and also centred at 1 kHz. The first of 75 components of the HTC started at 260 Hz. Both targets were either presented interaurally in-phase (N_0S_0) or out-of-phase (N_0S_π) depending on the measurement condition. Both masker and target had raised-cosine on and off-set ramps with 30 ms duration.

In the discrimination task, three intervals were presented on each trial. Two intervals contained the BPN target plus masker, while one interval contained the HTC target plus masker. In all intervals the target levels were the same. The subjects' task was to indicate which interval contained the HTC target. Target levels were controlled with a 2-down 1-up adaptive tracking procedure. At the start of each run, levels were adjusted with 8 dB steps and stepsizes were halved after each second reversal until a minimum step-size of 1 dB was reached after which another 8 reversals were measured. The median of the last 8 reversals was used as the measured threshold of that run. For each condition at least four thresholds were measured.

In addition to the discrimination task, subjects also performed a *detection* task, where only one of the three intervals contained a target. These detection conditions served as reference conditions to the corresponding discrimination conditions.

2.2 Results and discussion

The results showed good agreement between the five subjects, therefore we present here only the mean results in Fig. 1.

For the narrowband (100 Hz) N_0S_0 condition, detection thresholds for BPN (left open circle) and HTC (left open square) are nearly equal, while discrimination thresholds (left open diamond) are about 10 dB higher. This indicates that in the

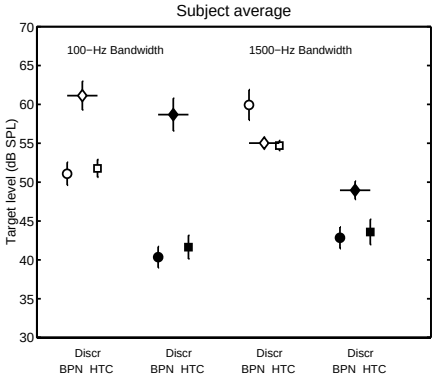


Fig. 1. Average discrimination thresholds (diamonds) and detection thresholds for BPN (circles) and HTC (squares) are shown for N_0S_0 conditions (open symbols) and N_0S_π conditions (filled symbols). The left half of the panel shows thresholds for 100-Hz wide targets, the right half for 1500-Hz wide targets. Vertical lines indicate averaged standard deviations based on individual subject standard deviations.

presence of the masking noise, targets need to be clearly audible in order to discriminate between the spectro-temporal structures of the BPN and the HTC.

The filled symbols on the left side of the panel show the same measurement conditions for an N_0S_π condition. Detection thresholds decrease by about 10 dB revealing the expected binaural advantage for a target at 1 kHz. The discrimination thresholds, however, are only about 3 dB lower than in the N_0S_0 condition. Apparently, the binaural cues present in the N_0S_π condition do not contribute much to the discriminability between BPN and HTC.

The same conditions presented above were repeated with wide-band (1500 Hz) targets and results are shown at the right side of the panel. For the N_0S_0 condition, detection thresholds are clearly lower for the HTC than for the BPN. It is possible that the temporal structure of the HTC present over a wide range of auditory filters contributes to the improved detectability, although it is not clear what across-channel process is responsible for the improvement.

In this same condition, the discrimination threshold is close to the HTC detection threshold. This can be understood by considering that at low target levels, in the discrimination task, the subject can only hear the HTC and not the BPN and thus is effectively performing a detection task.

For the N_0S_π condition we see again a clear binaural advantage for the detection thresholds. The masking level differences are 17 dB and 10 dB for the BPN and HTC, respectively. In this condition there is no indication for a difference in detection thresholds for the two types of targets. For the discrimination task there is an improvement of about 6 dB compared to the N_0S_0 condition. In the binaural condition, discrimination is now possible at a level where, with monaural listening, neither of the targets would be audible.

Comparing the narrow-band and the wide-band thresholds (for both N_0S_0 and N_0S_π) there is a clear reduction in the discrimination thresholds for the wide band conditions, while detection thresholds increase. This result seems to suggest that across spectrum integration contributes more to the discrimination task than to the detection task, at least for the dichotic condition.

3 Experiment II: the effect of centre frequency

In the previous experiment it was found that discrimination thresholds were lower for larger target bandwidths, while detection thresholds showed the opposite trend. These differences may be due to the target bandwidth but could also be related to frequency specific sensitivity. Here we performed separate measurements with narrowband targets at low and high frequencies in order to differentiate between the contributions of low-, mid-, and high-frequency auditory channels.

3.1 Method and stimuli

This experiment is similar to Experiment I except that the centre frequencies of the targets are 280-Hz and 1650-Hz with bandwidths of 60 Hz and 200 Hz, respectively. For the 280-Hz center frequency, the first of 3 components started at 260 Hz, for the 1650-Hz center frequency, the first of 10 components started at 1560 Hz. Three of the subjects from the first experiment participated in this experiment.

3.2 Results and discussion

The mean results of Experiment II are shown in Fig. 2 together with the mean narrowband results for the same subjects from Experiment I.

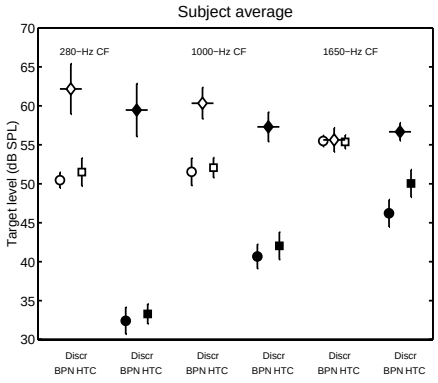


Fig. 2. Average discrimination thresholds (diamonds) and detection thresholds for BPN (circles) and HTC (squares) are shown for N_0S_0 conditions (open symbols) and N_0S_π conditions (filled symbols). The left, middle and right parts of the panel show thresholds for 280-Hz, 1-kHz and 1.65-kHz centre-frequency targets, respectively. Vertical lines indicate averaged standard deviations based on individual subject standard deviations.

For the N_0S_0 conditions (open circles and squares), detection thresholds increase somewhat with center frequency. This result is consistent with the idea that critical bandwidth increases towards high frequencies, which results in more masker energy per critical band. For the same conditions, discrimination thresholds (open diamonds) show a decrease at high frequencies, where they equal detection thresholds.

For the N_0S_π conditions (filled circles and squares), thresholds increase much more towards high-frequencies, in agreement with previous data that show a reduced binaural advantage for high frequencies (e.g. van de Par and Kohlrausch, 1999). The

high sensitivity at the lowest frequency suggests that the wideband N_0S_π detection thresholds of experiment I may be based on listening only to the lowest frequency bands containing binaural cues.

Discrimination thresholds (filled diamonds) for N_0S_π show some decrease towards high center frequencies, although this trend differed slightly among subjects. Compared to the N_0S_0 discrimination thresholds, the binaural advantage is largest at low-frequencies. For none of the centre-frequencies, do the N_0S_π discrimination thresholds reach the same low threshold values that are reached in the wide band case of experiment I. Therefore, the low thresholds in the wideband condition cannot be attributed to a high sensitivity in binaural discrimination for one particular frequency range. Apparently, across-spectrum integration of binaural cues leads to a considerable advantage in the discrimination task.

4 Experiment III: the effect of harmonic-tone-complex phase spectrum and F0

From the previous experiments it was concluded that binaural discrimination performance for wideband conditions is better than can be expected based on the individual narrowband conditions. In this experiment we investigate specific properties of the temporal structure of HTC targets that could contribute to the discriminability in wideband conditions.

4.1 Method and stimuli

Two experiments were conducted, similar to the previous experiments, to investigate how changes in the temporal structure of the wide band (1500-Hz bandwidth) HTC target affect discrimination performance under N_0S_π conditions.

In the first set of experiments, the phase relations between the sinusoidal components were varied. In addition to the standard sinusoidal phase structure, also random phase, Schroeder-positive, and Schroeder-negative phase were used. For the latter two conditions the frequency sweep rate resulted in one sweep per period of the HTC across the spectrum of the complex. The random-phase condition was chosen to examine whether or not a target with reduced peakedness would increase discrimination thresholds. The targets with Schroeder phase were chosen to examine the effect of asynchrony. With the rather small separation between partials of 20-Hz, Schroeder-phase targets are peaked in an asynchronous way across auditory filters. If across-channel synchrony is important for discrimination performance, this stimulus is expected to show higher discrimination thresholds than the sinusoidal phase conditions.

In the second set of experiments the influence of frequency separation between partials (F0) was investigated. In addition to the 20-Hz separation, also 40- and 80-Hz separations were used.

In both experiments three subjects of the first experiment participated.

4.2 Results and discussion

The results of Experiment III are shown in Fig. 3. The left panel shows data for different phase conditions of the HTC target. Detection thresholds do not differ significantly for the different phase conditions. The discrimination thresholds, however, show a marked elevation for the random-phase condition, while the Schroeder-phase conditions show only slightly higher discrimination thresholds than the sine-phase condition. Because the random-phase stimuli are the only targets with a low peak-factor, these data suggest that the peakedness of the HTC target is the critical property that enables the low discrimination thresholds observed for wide-band targets. The Schroeder-phase stimuli contain periodic linear frequency sweeps. The data show that synchronicity of the waveform peaks across auditory filters is not needed for a good discrimination performance.

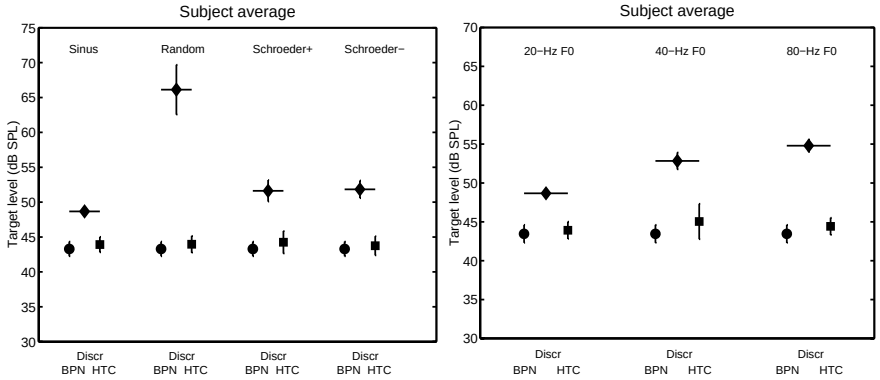


Fig. 3. Average discrimination thresholds (diamonds) and detection thresholds for BPN (circles) and HTC (squares) are shown for N_0S_π conditions. The clusters of data in the left panel show sinusoidal-phase, random-phase, Schroeder-positive, and Schroeder-negative phase conditions. The clusters of data in the right panel show thresholds for targets with F0s of 20-Hz, 40-Hz and 80-Hz, respectively. Vertical lines indicate averaged standard deviations based on individual subject standard deviations.

In the right panel of Fig. 3, N_0S_π data are shown as a function of frequency separation between partials (F0) of the sine-phase HTC. It is clear that detection thresholds are almost constant as a function of F0. Discrimination thresholds on the other hand show an increase at higher F0s where there is an increase in the rate of variation in binaural cues. A rate limitation in the binaural processing of dynamic interaural disparities, which would result in some kind of temporal averaging of these disparities, might explain the increase in discrimination thresholds because short-term changes in the interaural disparities may be the cues used for discriminating the two target types. With such an assumption, *detection* thresholds are not expected to change because the average amount of interaural disparities will not change depending on F0. Such a rate limitation is observed, e.g., in the study by Grantham

and Wightman (1979) where the interaural correlation of a masking noise was varied sinusoidally over time and a short interaurally out-of-phase sinusoidal probe had to be detected. For correlation modulations faster than 4 Hz, probe position relative to the masker correlation phase did not influence the measured thresholds. This led the authors to define a minimum binaural integration time of about 44-243 ms. Such an integration time is just long enough to resolve the target with the lowest periodicity (20 Hz) in our experiments.

A second reason for the increase in discrimination thresholds with increasing F0 may be the reduction in peakedness of HTC targets because fewer sinusoidal components will fall within one auditory filter. Again, detection thresholds are not expected to be influenced by such an effect because peakedness is not a factor that seems to influence detection thresholds.

5 Conclusions

It appears that the across-spectrum integration of binaural and monaural information is much better for discrimination conditions than for detection conditions. For detection conditions, thresholds increase towards larger bandwidths while the opposite tends to be the case for the discrimination experiments. An important factor for the rather good discriminability between the two types of wide-band binaural targets seems to be the temporal peakedness of the targets. For random-phase HTC targets with rather weak peakedness, discrimination from noise targets was only possible at very high target levels.

References

Cohen, M.F. (1981) Interaural time discrimination in noise. J. Acoust. Soc. Am. 70, 1289-1293.

Henning, G.B. (1973) Effect of interaural phase on frequency and amplitude discrimination. J. Acoust. Soc. Am. 54, 1160-1178

van de Par, S., and Kohlrausch, A. (1999) Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. J. Acoust. Soc. Am. 106, 1940-1947.

Stern Jr., R.M., Slocum, J.E., and Phillips, M.S. (1983) Interaural time and amplitude discrimination in noise. J. Acoust. Soc. Am. 73, 1714-1722.

Grantham, D.W., and Wightman, F.L. (1979) Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. J. Acoust. Soc. Am. 65, 1509-1517.

Zurek, P.M. and Durlach, N.I. (1987) Masker-bandwidth dependence in homo-phasic and antiphasic tone detection. J. Acoust. Soc. Am. 81, 459-464.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

A cat's cocktail party: Psychophysical, neurophysiological, and computational studies of spatial release from masking

Courtney C. Lane[1], Norbert Kopco[2], Bertrand Delgutte[1], Barbara G. Shinn-Cunningham[2], and H. Steven Colburn[2]

1 Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA, USA {court, bard}@epl.meei.harvard.edu

2 Hearing Research Center, Boston University, Boston, MA, USA {kopco, shinn, colburn}@bu.edu

1 Introduction

Masked thresholds can improve substantially when a signal is spatially separated from a noise masker (Saberi et al. 1991). This phenomenon, termed "spatial release from masking" (SRM), may contribute to the cocktail party effect, in which a listener can hear a talker in a noisy environment. The purpose of this study is to explore the underlying neural mechanisms of SRM.

Previous psychophysical studies (Good, Gilkey, and Ball 1997) have shown that for high-frequency stimuli, SRM was due primarily to energetic effects related to the head shadow, but for low-frequency stimuli, both binaural processing (presumably ITD processing) and energetic effects contributed to SRM. The relative contributions of these two factors were not studied for broadband stimuli.

Previous physiology studies have identified possible neural substrates for both the energetic and ITD-processing components of SRM. For the energetic component, our group has shown that some inferior colliculus units, "SNR units," have masked thresholds that are predicted by the signal-to-noise ratio (SNR) in a narrowband filter centered at the unit's CF (Litovsky et al. 2001). For the ITD component, a series of studies (e.g. Jiang, McAlpine, and Palmer 1997) shows that ITD-sensitive units can exploit the differences between the interaural phase difference (IPD) of a tone and masker to improve the neural population masked thresholds. These studies did not describe how the units' masked thresholds change when a broadband signal and masker are placed at different azimuths.

Here, we examine the contributions of energetic effects and binaural processing for broadband and low-frequency SRM using psychophysical experiments and an idealized population of SNR units. We also show that a population of ITD-sensitive units in the auditory midbrain exhibits a correlate of SRM. Finally, a model of ITD-sensitive units reveals that the signal's temporal envelope influences the single-unit masked thresholds.

2 Psychophysics and modeling of SRM in humans

2.1 Methods

SRM was measured for three female and two male normal-hearing human subjects using lowpass and broadband stimuli. Azimuth was simulated using non-individualized head-related transfer functions (Brown 2000). Stimuli consisted of a 200-ms 40-Hz chirp train (broadband: 300-12,000 Hz; lowpass: 200-1500 Hz) masked by noise (broadband: 200-14,000 Hz, lowpass: 200-2000 Hz). The spectrum-level for the signal was fixed at 14 dB re 20 μPa/√Hz (56 dB SPL for the broadband signal). The masker level was adaptively varied using a 3-down, 1-up procedure to estimate the signal-to-noise ratio (SNR) yielding 79.4% correct detection performance. Stimuli were delivered via insert earphones to subjects in a sound-treated booth.

Inspired by the SNR units described above, predictions from a simple, "single-best-filter" model were used to evaluate if the SNR in the best narrow-frequency band can explain how masked threshold varies with signal and noise locations. The model analyzes SNR as a function of frequency, but does not allow for any across-frequency integration of information or any binaural processing. The model consists of a bank of 60 log-spaced gammatone filters (Johannesma 1972) for each ear. For each spatial configuration, the root-mean-squared energy at the output of every filter is separately computed for the signal and noise. The model assumes that the filter with the largest SNR (over the set of 120) determines threshold. The only free parameter in the model, the SNR yielding 79.4% correct performance, was fit to match the measured threshold when signal and noise were at the same location.

2.2 Results

Figure 1 shows measured (solid lines) and predicted (broken lines) thresholds as a function of noise azimuth for three signal azimuths (arrows). Two sets of model predictions are shown. Dash-dot lines show both lowpass and broadband predictions generated jointly for the model parameter fit to the broadband threshold measured with signal and masker co-located. Dotted lines show lowpass predictions generated with the model parameter fit to the measured lowpass threshold separately. Overall, performance is better for broadband (BB) stimuli than for lowpass (LP) stimuli (BB thresholds are always lower than LP). Further, the amount of SRM, the improvement in threshold SNR compared to the thresholds when signal and noise are co-located, is larger for broadband than lowpass stimuli (30 dB and 12 dB, respectively).

When the model parameter is fit separately for broadband and lowpass stimuli, predictions are relatively close to observed thresholds although lowpass predictions consistently underestimate SRM. These results suggest that for the chirp-train signals used, 1) the main factor influencing SRM for both lowpass and broadband stimuli is the change in SNR in narrow frequency bands, and 2) binaural processing increases SRM for lowpass, but not broadband stimuli.

When the same threshold SNR parameter is used to predict broadband and lowpass results (dash-dot lines), predicted thresholds are equal when signal and

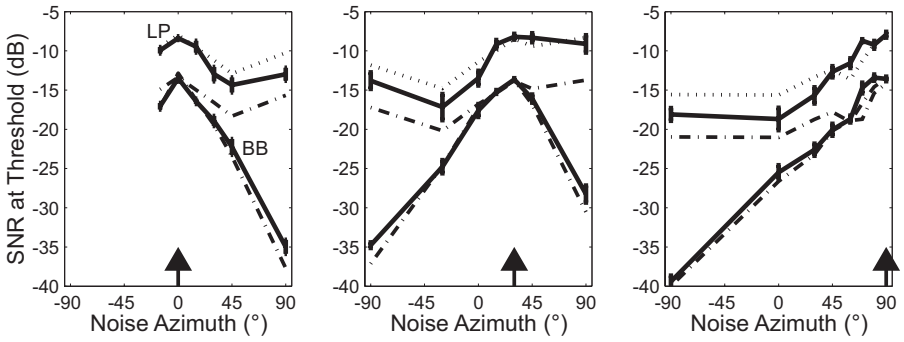


Fig. 1. SRM for human subjects for broadband (BB) and lowpass (LP) stimuli. Measured (subject mean and standard error) and predicted thresholds as a function of noise azimuth for three signal azimuths (arrows). Dash-dot line: lowpass and broadband model fit with same parameter; dotted line: lowpass data fit separately.

noise are co-located, regardless of stimulus bandwidth (because the SNR is constant across frequency when signal and noise are co-located). However, measured performance is always worse for the lowpass stimuli compared to the broadband stimuli. This result suggests that the listener integrates information across frequency, leading to better performance for broadband stimuli.

3 Neural correlates of SRM in the cat auditory midbrain

As shown above, the single-best-filter model underestimates the SRM for low frequencies. Here, thresholds for a population of ITD-sensitive neurons are measured to determine if these units can account for the difference between the single-best-filter model and behavioral thresholds.

3.1 Methods

Responses of single units in the anesthetized cat inferior colliculus were recorded using methods similar to those described in Litovsky and Delgutte (2002). The signal was a 40-Hz, 200-msec chirp train presented in continuous noise; both signal and noise contained energy from 300 Hz to 30 kHz. The chirp train had roughly the same envelope as the one used in the broadband psychophysical experiments. The signal level was fixed near 40 dB SPL, and the noise level was raised to mask the signal response. Results are reported for 22 ITD-sensitive units with characteristic frequencies (CFs) between 200 and 1200 Hz.

3.2 Results

Figure 2A shows the temporal response pattern for a typical ITD-sensitive unit as a function of noise level for the signal in noise (first 200 msec) and the noise alone (second 200 msec). The signal and noise were both placed at +90° (contralateral to the recording site). At low noise levels, the unit produces a synchronized response to the 40-Hz chirp train. As the noise level increases, the response to the signal is

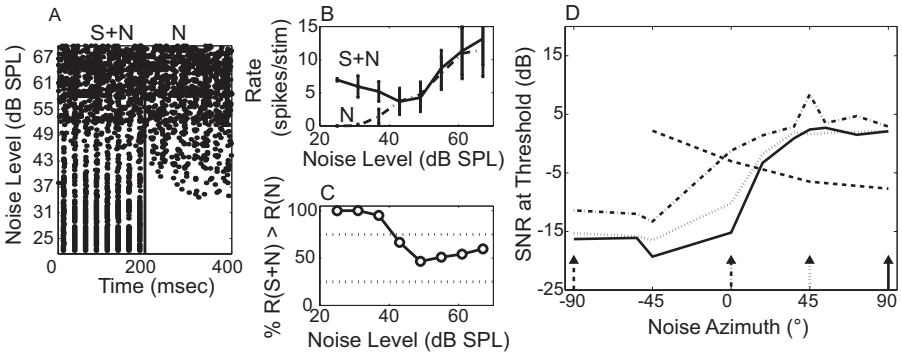


Fig. 2. A: Single-unit response pattern for signal in noise (S+N, 0-200 msec) and noise alone (N, 200-400 msec) for signal and noise at 90°. Signal level is 43 dB SPL. B: Rate-level functions for S+N and N from A. C: Percent of stimulus presentations that have more spikes for S+N compared to N. Threshold is the SNR at 75% or 25% (dotted lines). D: Same unit's masked thresholds as a function of noise azimuth for four signal azimuths (arrows indicate signal azimuth, arrow tail indicates corresponding threshold curve).

overwhelmed by the response to the noise (A, B). For this unit, +90° is a favorable azimuth so both the signal and the noise excite the unit. When placed at an unfavorable azimuth, the signal can suppress the noise response or vice versa.

Threshold is defined for single units as the SNR at which the signal can be detected through a rate increase or decrease for 75% of the stimulus repetitions (75% and 25% lines in Fig. 2C). Thresholds for this unit are shown in D as a function of noise azimuth for four signal azimuths. For three of the signal azimuths (-90°, 45°, and 90°), moving the noise away from the signal can improve thresholds by more than 15 dB. However, when the signal is at 0°, thresholds become slightly worse as the noise moves from the midline to the contralateral (positive azimuth) side. In other words, although some SRM is seen for some signal azimuths, no direct correlate of SRM can be seen in this, or any other, individual unit's responses for all signal and noise configurations.

A simple population threshold is constructed based on the same principle as the single-best-filter model (Section 2). For each signal and noise configuration, the population threshold is the best single-unit threshold in our sample of ITD-sensitive

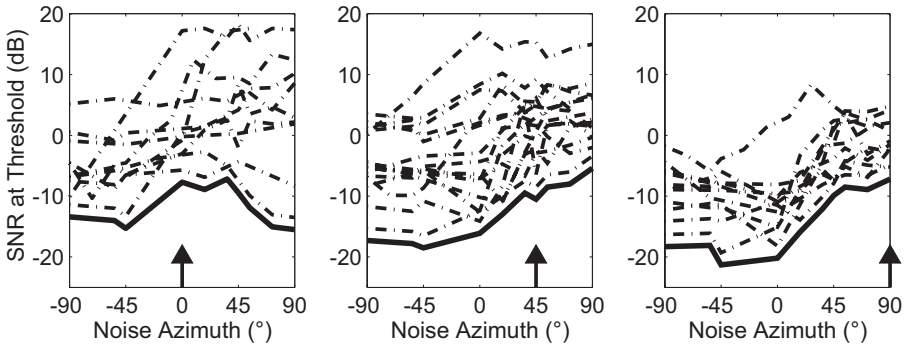


Fig. 3. Neural population thresholds for three signal azimuths (arrow). Dash-dot lines: single unit thresholds; solid lines: population thresholds (offset by 2 dB).

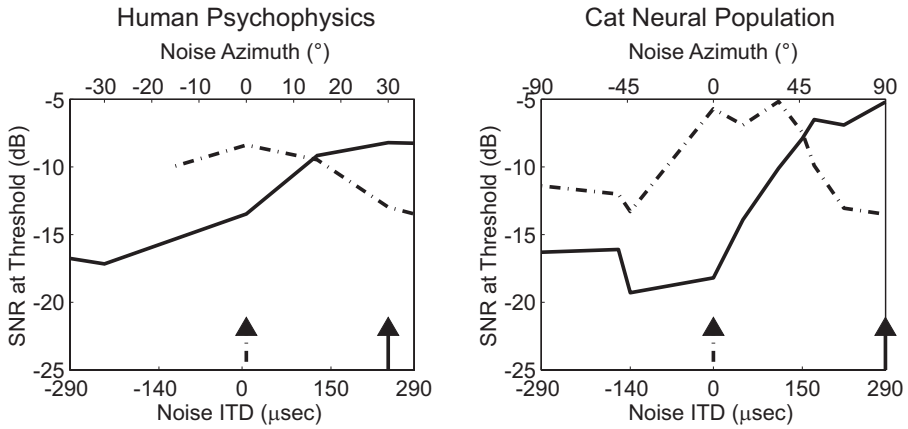


Fig. 4. Human psychophysical thresholds (left) and cat neural population thresholds (right) for two signal azimuths (arrows indicate signal azimuth, arrow tail indicates corresponding threshold curve) as a function of noise ITD (lower axis) and azimuth (upper axis).

units. Figure 3 shows the population thresholds (solid lines) as a function of noise azimuth for three signal azimuths (arrows). Unlike single unit thresholds (dot-dash), the population thresholds show SRM in that thresholds improve when the signal and noise are separated.

Figure 4 compares the low-pass human psychophysical thresholds (left) to the cat neural population thresholds (right). In order to compare the two thresholds despite the difference in species headsize, the axes are matched for noise ITD (lower axis) rather than noise azimuth (upper axis). The neural population thresholds are similar to the human behavioral thresholds, indicating that these ITD-sensitive units could provide a neural substrate for the binaural component of SRM.

3.3 Neural modeling of single-unit thresholds

Because our population consists of ITD-sensitive units, we attempted to model the unit responses using an interaural cross-correlator model similar to Colburn (1977). Figure 5A shows the thresholds for five units for which we measured thresholds for the signal at their best azimuths (+90°, squares) and their worst azimuths (-90°, circles). The noise was placed at the ear opposite the signal. For the data, the best-azimuth thresholds are better or equal to the worst-azimuth thresholds. In contrast, the cross-correlator model predicts that the worst-azimuth thresholds are better (Fig. 5B) because the largest change in interaural correlation occurs when the signal decreases the overall correlation. The cross-correlator, although able to predict the noise-alone response, failed to predict the response to the signal (not shown). The primary difference between the chirp-train signal and the noise is that the signal has a strong 40-Hz amplitude modulation while the noise envelope is relatively flat. Because many units in the IC have enhanced responses to modulated stimuli (Krishna and Semple 2000), we added an envelope processor that changes the rate response in proportion to the energy in the 40-Hz Fourier component of the cross-correlator's output. With envelope processing (Fig. 5C), best-azimuth thresholds are

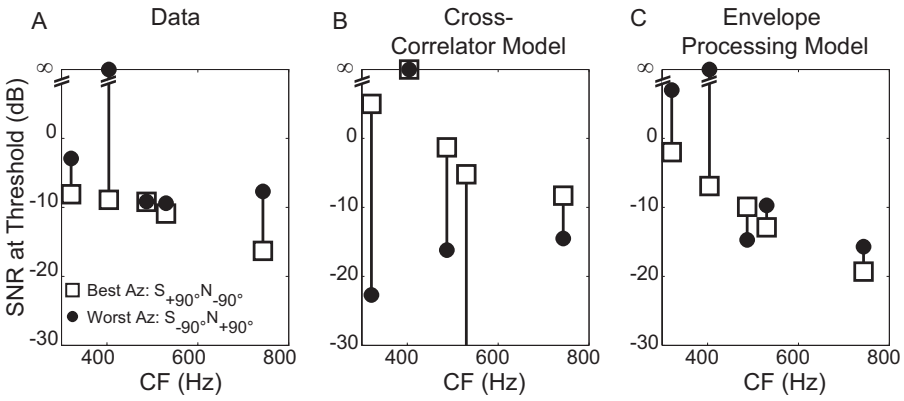


Fig. 5. A: Masked thresholds for 5 units. Best-azimuth thresholds (squares): signal at +90°, noise at -90°; worst-azimuth thresholds (circles): signal at -90°, noise at +90°. B, C: As in A for cross-correlator model (B) and cross-correlator model with envelope processor (C).

about the same or better than worst-azimuth thresholds, consistent with the data, because the envelope processor only changes the responses for favorable azimuths. These results suggest that 1) a traditional cross-correlator model cannot account for neural responses in the IC, 2) the temporal envelope can affect the detectability of signals in inferior colliculus neural responses, and 3) envelope processing is necessary to predict which units are best for signal detection (discussed below).

4 Discussion

Human listeners exhibit a large amount of SRM for both broadband and lowpass 40-Hz chirp-train signals. For broadband stimuli, the SNR in a single high-frequency filter predicts the amount of SRM, indicating high-frequency narrow-band energetic changes determine the SRM. SNR units, which have thresholds that are predicted by the SNR in a narrowband filter, could detect these changes.

For the lowpass condition, the single-best-filter model predicts some SRM, but underestimates the total amount by several dB. A correlate of the lowpass SRM is evident in the population response of ITD-sensitive units in the IC. It is possible, then, that there are two populations of neurons that can give SRM at low frequencies: an ITD-sensitive population and an SNR-unit population. When a listener is able to use the ITD-sensitive population, thresholds should improve by a few dB. When this population cannot be used (such as when the signal and masker are co-located or when listening monaurally), the SNR-unit population would determine performance, resulting in worse masked thresholds for some spatial configurations. These two hypothesized neural populations may respond differently to different stresses. For example, because the SNR population response depends on a neural population with narrow tuning and a wide range of CFs, relying on this population might be especially difficult for listeners with hearing impairment.

The envelope-processing model predicts that different ITD-sensitive populations, in either the left IC or the right IC, will dominate signal detection performance for different stimuli. The best single-unit thresholds for both the data

and the envelope-processing model occur when the chirp-train signal is positioned at a unit's best azimuth. Thus, for modulated signals, the IC contralateral to the signal yields better thresholds than the ipsilateral IC. However, for unmodulated signals, the model predicts that the best thresholds occur for the signal placed at the unit's worst azimuth. This prediction is consistent with previous studies (e.g. Jiang, McAlpine, and Palmer 1997) showing that the best single-unit thresholds for tones in noise occurred when the tone had an unfavorable IPD. Therefore, different ICs seem to be used for signal detection depending on the signal envelope.

Finally, human broadband thresholds are better than lowpass thresholds for all spatial configurations. Because this improvement is evident for co-located signals and maskers, the auditory system seems to integrate information across frequency. Because units in the IC are relatively narrowly tuned, auditory centers above the IC are also likely to be involved in the detection of broadband signals.

In summary, SRM seems to depend on binaural and energetic cues, which may be processed by separate neural populations. Neural processing related to SRM can be observed in the auditory midbrain, but centers higher than the midbrain also seem necessary for the integration of information across frequency.

References

Brown, T. J. (2000). "Characterization of acoustic head-related transfer functions for nearby sources," unpublished M.Eng. thesis. Electrical Engineering and Computer Science, MIT, Cambridge, MA.

Colburn, H. S. (1977) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. J. Acoust. Soc. Am. 61, 525-533.

Good, M.D., Gilkey, R.H., and Ball, J.M. (1997) The relation between detection in noise and localization in noise in the free field. In R.H. Gilkey and T.R. Anderson (Eds), Binaural and Spatial Hearing in Real and Virtual Environments. Lawrence Erlbaum Associates, Mahwah, N.J, pp 349–376.

Jiang, D., McAlpine, D., and Palmer, A.R. (1997) Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. J. Neurosci. 17, 9331-9339.

Johannesma, P.I.M. (1972) The pre-response stimulus ensemble of neurons in the cochlear nucleus. In: B.L. Cardozo, E. de Boer, and R. Plomp (Eds.), IPO Symposium on Hearing Theory. IPO, Eindhoven, The Netherlands, pp. 58-69.

Krishna, B.S. and Semple, M.N. (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. J. Neurophysiol. 84, 255-73.

Litovsky, R.Y. and Delgutte, B. (2002) Neural correlates of the precedence effect in the inferior colliculus: Effect of localization cues. J. Neurophysiol. 87, 976-994.

Litovsky, R.Y., Lane, C.C., Atencio, C., and Delgutte, B. (2001) Physiological measures of the precedence effect and spatial release from masking in the cat inferior colliculus. In: D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven (Eds). Physiological and Psychophysical Bases of Auditory Function. Shaker, Maastricht, pp. 221-228.

Saberi, K., Dostal, L., Sadralodabai,T., Bull, V., and Perrott, D.R. (1991) Free-field release from masking. J. Acoust. Soc. Am. 90, 1355-1370.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Localization of noise in a reverberant environment

Brad Rakerd[1] and William M. Hartmann[2]

[1] Department of Audiology and Speech Sciences, Michigan State University, East Lansing, MI, 48824, USA
[2] Department of Physics and Astronomy, Michigan State University East Lansing, MI, 48824, USA, hartmann@pa.msu.edu

1 Introduction

The effect of room acoustics on the human ability to localize a broad-band noise was studied in a variable-acoustics concert hall, the Espace de Projection (ESPRO) at the Institut de Recherche et Coordination Acoustique/Musique (Hartmann, 1983). In that study the noise was turned on slowly to eliminate attack transients. Listeners proved able to localize such slow-onset noise accurately, substantially more accurately than they could localize a complex tone with a slow onset. The explanation given for this advantage with noise was that its fine structure is transient and can trigger the precedence effect, a perceptual process that enhances the accuracy of sound localization in rooms (Wallach, Newman, and Rosenzweig, 1949; Litovsky, Colburn, Yost and Guzman, 1999). A question left unanswered at that time was whether the availability of an attack transient might further enhance the localization of noise in a room. That question is addressed here.

The ESPRO study also revealed an interesting decision-making strategy with slow onsets. The listeners tended to make their decisions early, while the noise envelope was still rising. During this interval, the reverberant field of the room had not fully formed and the direct sound stood out by comparison. It was conjectured that listeners chose to exploit this early advantage, even though the direct sound was still relatively faint. As a result, it is possible that the ESPRO experiments did not fairly represent the ability of the listeners to localize noise based on steady-state conditions only. The present study introduces a new onset method whereby only the steady-state sound field is available to the listeners.

Still another finding of the ESPRO study was that listeners localized noise less accurately in a reverberant configuration of the hall than in an absorbing configuration. To account for the increased error with increased reverberation, it was proposed that reverberated sound acts as a masker that obscures the direct sound, an effect that could be quantified by a direct-reverberant sound power ratio. That ratio is a key parameter in the present study. In sum, the present study was undertaken to learn more

about the localization of broadband noise in rooms – about the role of onsets and about the strategies that listeners use when onset transients are not available.

2 Experiment 1: Noise onsets

Experiment 1 addressed two questions about noise onsets. The first question was whether an attack transient enhances the localization of noise in a room. To answer this question, we asked listeners to localize two broadband noises, one turned on slowly and the other turned on abruptly. The second question was whether, with slow-onset noise, listeners benefit from making their localization decisions early while the noise onset is increasing and before the reverberant field in a room is fully developed. To answer this question we compared the localization of slow-onset noise with localization of an identical noise that was masked for its first few seconds.

2.1 Methods

The experiment was run in a reverberation room (IAC 107840) with dimensions 7.7 m (wide) $\times$ 6.4 m (long) $\times$ 3.6 m (high) and with a reverberation time of 4 s at speech frequencies. Localization ability was measured using the source identification method (Hartmann, Rakerd and Gaalaas, 1998) with a source array of 24 loudspeakers. The speakers were matched in frequency response, extended to 17 kHz, and flattened in one-third octave bands by an equalizer. The speakers were secured with velcro mounting to the top of two 2.4-m rails, with an inter-speaker separation of 2 degrees. As sources, the speakers were numbered from left to right with speakers 12 and 13 on either side of the listener's forward direction. Thus, the array spanned a total angle of 46 deg, half to the listener's left and half to the right.

An important experimental parameter was the ratio of direct to reverberant sound, controlled by varying the source-listener distance, either 3 m or 6 m. Whenever we changed this distance, we also moved the sources along the rails to maintain the 2-degree separation of adjacent sources. For the 3-m distance the measured C-weighted direct-reverberant ratio was -7 dB. For the 6-m distance it was -13 dB. These differ by 6 dB, as expected if the reverberant level in the room is independent of listener location.

Noise onsets

The stimulus for this experiment was white noise with a steady-state level of 55 dBA at the listener's chair. There were three different noise onset conditions – abrupt, slow, and masked. For the abrupt-onset stimulus, the noise was turned on with a step-function envelope. It remained on until the subject gave a localization response. Following the response, the stimulus was turned off with a 500-ms ramp. The slow-onset stimulus was the same except that the noise was turned on gradually with a linear amplitude envelope, 2 s in duration. For the masked-onset stimulus, a trial began with a masking noise from a loudspeaker directly behind the listener's neck. The masking noise was uncorrelated with the stimulus source noise, and it

was sufficiently intense (85 dBA) to render the source undetectable. The masker was turned on with a 100-ms ramp. After 250 ms, one of the source speakers was turned on gradually. After the source onset was completed, the masking noise was removed (500-ms offset ramp), leaving only the source noise sounding.

Listeners and procedure

There were seven listeners in the experiment. Five of them (A,B,C,D and E) were students, 17-30 years old, with audiometrically normal hearing in both ears. The other two listeners (F and G) were middle-aged men with audiograms that showed modest bilateral high-frequency hearing loss, approximately 25 dB at 8 kHz.

Listeners were tested one at a time. The chair height was adjusted to put a listener's ears 1.17 m from the floor, which matched the height of the speakers in the source array. To assure that all subjects received the same stimuli, we obliged them to sit still and to make their localization judgments while facing straight ahead. An L-shaped bar, connected to the back of the chair, pressed against the crown of the head as a guide to keep the head stationary. All of the sources of the array could be viewed by moving the eyes without moving the head.

On each trial, a noise stimulus was presented from one of the 24 sources, selected at random. The listener then reported the number of the source (1 through 24) that was heard to have sounded. Test trials were blocked into runs of 48 trials, two presentations from each of the 24 sources. Altogether, a listener did three runs for each onset condition at the 3-m listening distance and three runs at the 6-m distance. The order of these runs varied randomly across listeners.

2.2 Results and discussion

Response plots – Typical listener

A detailed picture of a subject's performance is provided by response plots, as shown in Fig. 1 for listener D, the subject whose response pattern was most similar to the average. The panels of Fig. 1 give listener D's results for the six different conditions of the experiment: tests at 3 and 6 meters with abrupt, slow, and masked onsets. Comparisons among the panels show evidence of three notable effects.

(1) Listener D was sensitive to the direct-reverberant sound power ratio in the room. Function $R(k)$ was closer to the 45-degree line at the 3-m distance than at 6 m for every onset condition. Also, the error bars were smaller at 3 m. We attribute this difference to the fact that the direct-reverberant ratio was more favorable by 6 dB at 3m.

(2) Listener D benefited from an attack transient, particularly when listening at 6 m. The responses in the abrupt-onset condition were visibly closer to the 45-degree line than the responses in the slow-onset condition, and the error bars were generally smaller.

(3) Finally, listener D was able to make successful localization decisions before the reverberant field had reached its steady state. Error bars were much smaller for the

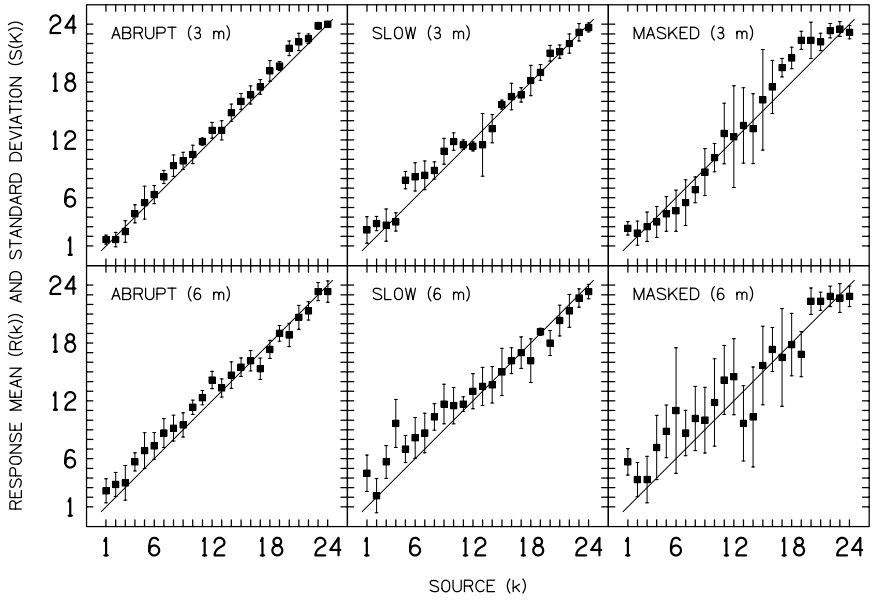


Fig. 1. Response plots for listener D: Statistic $R(k)$ is the mean response number for each source as a function of the source number, k. Perfect performance corresponds to a 45-degree line. The error bars represent one standard deviation about the mean response to a source, statistic $s(k)$. [See Rakerd and Hartmann (1986) for a detailed description of these statistics.] A separate plot is given for each source distance and onset condition of Experiment 1.

slow-onset conditions where the listener was able to hear onsets building in the room, than for the masked-onset conditions where the build-up was inaudible.

RMS error – All listeners

An overall measure of localization ability is statistic $\overline{D}$, the root-mean-square error averaged over all 24 sources. Figure 2 shows how $\overline{D}$ varied with distance to the sources and with the various onset types. Functions are plotted separately for each listener. Analysis of variance on the results showed significant effects of both distance to the sources [$F(1,6) = 51.26$, $p < 0.001$], with the listeners overall more accurate at 3 m than at 6 m, and onset type [$F(2,12) = 34.62$, $p < 0.001$], with listeners most accurate for noise with an abrupt onset, intermediate for noise with a slow onset, and least accurate for noise with a masked onset. (Post-hoc tests showed all of the pairwise difference among the onset means to be significant; $p < 0.01$). The results of Experiment 1 support two conclusions about the role of noise onsets. The first conclusion is that the presence of an attack transient does increase the accuracy of localization of noise in a room. The second conclusion is that when listening to noise with a slow onset listeners benefit by listening in advance of the buildup of the reverberant field of a room. Figure 2 shows that this benefit is reduced when the source-listener distance is increased from 3 meters to 6 meters. We conjecture

that this effect occurs because increasing the distance reduces the delay between the direct sound and early reflections.

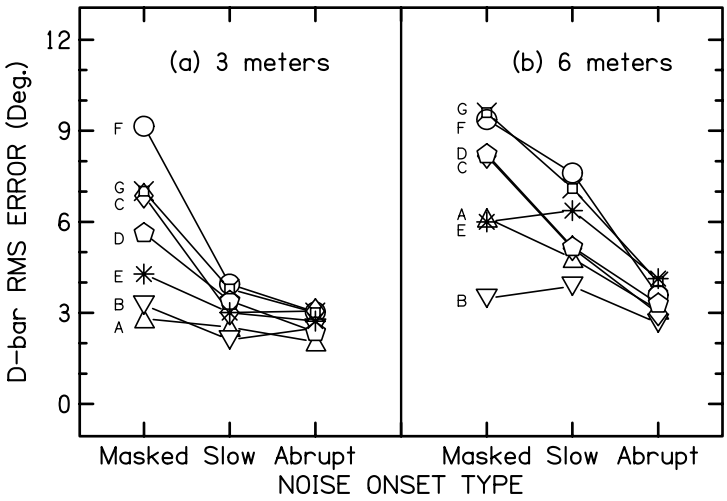


Fig. 2. RMS error $\overline{D}$ for seven listeners depending on listening distance (direct-reverberant power ratio) and onset type.

3 Experiment 2: Head/body turn

In Experiment 1 subjects were required to hold their heads and bodies motionless so that stimuli from individual sources were reproducible. In everyday life, listeners are free to move, and it seemed possible that movement could improve localization ability here by allowing integration of information obtained from different perspectives. Experiment 2 was run to see whether rotating the head and body would enhance the ability to localize noise in a fully formed reverberant sound field.

3.1 Method

The stimulus for this experiment was the masked-onset noise. It was presented at the 6-m distance where the direct-reverberant ratio was most disadvantageous. Two good localizers (A and C) and two poor localizers listeners (F and G) did three interleaved runs under each of two conditions: (1) while sitting still and facing straight ahead, (2) while free to rotate the head and body. In condition (2), a seated listener was minimally required to rotate the trunk around once and to move the head back and forth once before making a localization response. The listener was then free to make any additional movements as desired, so long as he remained seated. All of the listener elected to move extensively, devoting 5 to 10 minutes per (48-trial) run to the exercise.

3.2 Results and discussion

Figure 3 shows the results of the experiment. In the absence of a head/body turn, both of the older listeners (F and G) made large localization errors, as in the prior experiments with masked-onset noise. When they were allowed to move, F and G both improved their localization accuracy substantially (mean decrease in $\overline{D}$ = 3.3 deg, or 35 percent). The two younger listeners (A and C) were much more accurate than the older subjects when sitting still. One of them (A) was helped substantially by moving, even relative to this good baseline, improving accuracy by 1.6 deg, which exceeded the error bars. The other young listener (C) was unaffected. A comparison of 12 stationary-moving run pairs (4 listeners $\times$ 3 run pairs per condition) showed an advantage for moving (accuracy improved by 1 deg or more) on ten of the twelve (sign test, $p < 0.005$). We conclude that head/body motion can aid a listener in localizing noise in a room, particularly when the listener is substantially challenged by reverberation or by some limitations on hearing.

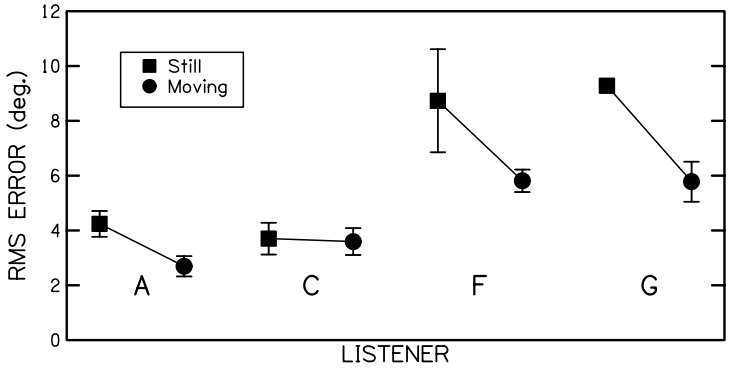


Fig. 3. RMS error $\overline{D}$ for four listeners when seated in a fixed position (squares) and when allowed to move (circles). Error bars show $\pm$ one standard deviation over runs.

4 Summary and conclusions

This study examined the localization of broadband noise in a reverberant environment. Experiments were primarily devoted to the role of noise onsets. In Experiment 1, we varied the onset characteristics of a broadband (white) noise in two ways: (a) An abrupt onset and a slow onset contrasted in that the former invited precedence effect with an attack transient and the latter did not. (b) The slow onset and a masked onset contrasted in that the former was audible during the development of the reverberant sound field and the latter was not. We also varied the ratio of direct to reverberant sound by positioning a listener at different distances from the sources. In Experiment 2 we studied the possibility that localization performance might improve

if listeners were allowed some motion. The results of the experiments, as described in Figs. 2 and 3, lead to the following conclusions:

(1) Localization of noise is enhanced by an attack transient. An attack transient appears to be particularly helpful when the direct-reverberant ratio is low. Attack transients give an advantage over slow onsets when the reflections are not much delayed *re* the direct sound. By contrast, attack transients are of only marginal value when noise is presented by headphones or tones are presented in an anechoic room (Tobias and Schubert, 1959; Rakerd and Hartmann, 1986).

(2) Onsets are a great leveler among individuals. Whereas the ability to localize steady steady-state sounds varies greatly among listeners, the ability to localize sounds with an onset transient shows best to worst differences less than 1.5 degrees among our seven listeners.

(3) Given a slowly increasing direct sound, listeners make localization decisions early, before the reverberant response of a room has fully formed. During this interval the power of the direct sound is low, but the ratio of direct sound power to reverberant sound power is favorable. Thus, the ESPRO experiments (Hartmann, 1983) do not fairly represent the ability of listeners to localize in a steady-state reverberant field.

(4) Listeners can use head and trunk motions to improve localization of sounds in the steady state. It is interesting to try to imagine what kind of computations the auditory system actually performs on the binaural sound field to obtain improved performance.

Acknowledgements

This project was supported by the National Institutes of Health through NIDCD grant DC00181.

References

Hartmann, W.M. (1983) “Localization of sound in rooms,” J.Acoust. Soc. Am. 74, 1380-1391.

Hartmann, W.M., Rakerd, B. and Gaalaas, J.B. (1998) “On the source identification method,” J. Acoust. Soc. Am. 104, 3546-3557.

Litovsky, R.Y., Colburn, H.S., Yost, W.A. and Guzman, S.J. (1999) “The precedence effect,” J. Acoust. Soc. Am. 106, 1633-1654.

Rakerd, B. and Hartmann, W.M. (1986) “Localization of sound in rooms, III: Onset and duration effects,” J. Acoust. Soc. Am. 80, 1695-1706.

Tobias, J.V. and Schubert, E.R. (1959) “Effective onset duration of auditory stimuli,” J. Acoust. Soc. Am. 31, 1591-1594.

Wallach, H., Newman, E.B., and Rosenzweig, M.R. (1949) “The precedence effect in sound localization,” Am. J. Psychol. 52, 315-336.

Listening in real-room reverberation: Effects of extrinsic context

Anthony J. Watkins

School of Psychology, The University of Reading. syswatkn@reading.ac.uk

1 Introduction

When speech is heard in a real room, it is distorted by reverberation. As reflections mix with the direct sound, gaps in the signal are filled and offsets are extended, so that there is distortion of the waveform's amplitude envelope. The Fourier transform of the amplitude envelope, the modulation spectrum, captures this distortion as an attenuation that is more pronounced towards the higher modulation frequencies. Measures of a communication channel's attenuation of modulation frequencies have led to reliable predictors of speech intelligibility in diverse surroundings, particularly of the articulation loss for consonants (Houtgast and Steeneken 1973).

Phoneme distinctions can be brought about by changing only the sound's amplitude envelope, e.g., 'sir' can be amplitude modulated so that 'stir' is heard. Perception of such sounds is sensitive to the effects of reverberation, as this opposes the effects of amplitude modulation to some extent and causes a test word such as 'stir' to be heard as more like 'sir'. There appears to be some perceptual compensation for this distortion when neighboring speech is distorted in a similar way. Thus, when test words like these are embedded in a 'context' utterance, such as 'next you'll get _ to click on', the effects of reverberation on the test-word distinction are less evident when both the context and the test word are distorted by the same reverberation. Such results have been taken as evidence of a perceptual compensation mechanism, which is informed about the distortion from its presence in neighboring speech that is 'extrinsic' to the test word itself (Watkins 1992).

This evidence of an extrinsic perceptual compensation comes from experiments that used rather severe synthetic reverberation. Such results have not been verified in real-room listening conditions where reverberation may generally be milder. In realistic listening conditions reverberation is reduced because peoples' mouths and ears have directionally dependent characteristics. This can serve to enhance the sound that arrives directly from the source, as the contribution from reflected sound that travels in directions that differ from the direct sound's direction is reduced. One aim of the perceptual experiments here is to ask whether compensation for reverberation can still be observed in the mild reverberation that is obtained in

typical real rooms, and with transducers that have the directional characteristics of human heads.

In compensating for *spatial* distortions by reflections, Clifton (1987) has proposed that listeners use binaural information to create a model of the room that they are listening in, and use it to select the aspects of subsequent sounds that are most informative about the source's point of origin in that space. So it may be that compensation for effects of reverberation on speech is similarly mediated through a binaural representation of the listening space that derives from preceding sounds. To test these ideas here, compensation is measured when the reverberation in the context and in the test word come from different rooms. Compensation is also measured in monaural as well as in dichotic listening conditions, and with contexts that either precede or succeed the test word.

Speech recognition automata have traditionally been designed to try and cope with the influences of 'low level' distortions, such as from reverberation, by postponing decisions to higher levels, where typically the machine holds representations of words (Huckvale 1996). There seem also to be perceptual phenomena that arise from analogous word representations. For example, when a noise such as a cough replaces one of the phonemes in a word, there is the illusory perception of the missing phoneme, accompanied by the sound of the cough (Warren 1970). Extrinsic perceptual compensation for reverberation may similarly be mediated through representations of the words of an utterance, perhaps providing a basis for segregating the context's speech from the distortion. To test this idea here, contexts are played backwards so that a sequence of words is not heard. Such reversed contexts should give reduced compensation if it is mediated by way of word-level representations.

Finally, compensation is measured when the context's spectro-temporal variations are removed. This is done by reversing the polarity of a randomly selected half of the signal's samples, thereby turning the speech into signal-correlated noise. The noise is then 'speech shaped' to give it the same long term average spectrum as the original. This preserves the broad-band amplitude-envelope but reduces the modulation present in narrow frequency bands, such as those received by individual auditory-nerve fibers. An effect of this manipulation is to reduce extrinsic compensation for distortion of the *spectral* envelope, 'coloration' (Watkins, 1991) and it is used here to test whether spectro-temporal variation is also important in compensation for effects of reverberation on the waveform's amplitude envelope.

2 Method

Sounds were delivered to listeners through Sennheiser HD420 headphones while they listened in the otherwise quiet conditions of an IAC 1201 booth. The sounds were processed to give the effect of real-room listening. Binaural room impulse responses, BRIRs (Zahorik 2002) were obtained at distances of 1.25 m, or 10 m in a corridor and in an L-shaped room of a disused office building. These impulse responses were then convolved with 'dry', non-reverberant sounds to provide stimuli for different experimental conditions.

BRIRs were obtained by 'deconvolution', whereby the FFT of a measurement sound *after* it has been played in the room is divided it by the FFT of the same sound *before* it is played in the room. The measurement sound was spectrally dense, being a maximum length sequence derived from a centre-tapped 24-bit register. Linear (as opposed to circular) deconvolution was effected with the sequence-doubling method (Gardner and Martin 1994). The measurement sounds were played from a dummy head with an acoustic transducer at its mouth (B&K HATS), while recordings were made using a dummy-head receiver with microphones in its ears (KEMAR). This results in BRIRs that can match the headphone-delivered sounds used in experiments to the sound at the eardrums of a listener in the room that is measured.

Noise in the impulse response measurements was checked by way of departures from linearity in the energy decay curves, plotted as dB vs. time, which are computed by reverse integration of the impulse response (Schroeder 1965). These were found to be suitably linear down to a -42-dB noise floor, as shown by the examples in Fig.1.

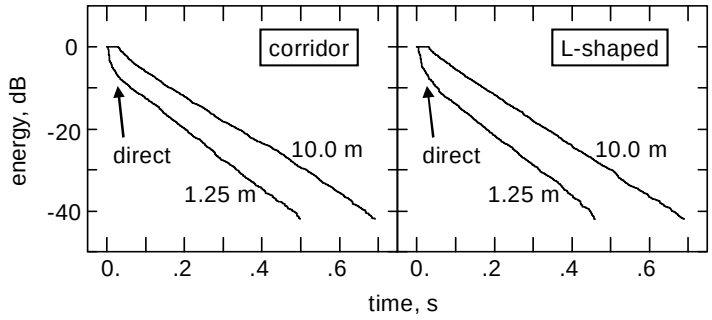


Fig. 1.The ordinate shows the BRIR's energy, found by integrating it from the time shown on the abscissa. These left-ear BRIRs were obtained at 1.25 m or 10 m in the L-shaped room or the corridor. The change in slope near the beginning is the direct sound's energy giving way to the linear decay of the reflected sound's energy, and this effect is more prominent for nearer sources. The flat part right at the start is the travel time from source to receiver.

A male talker (AJW) was recorded saying, "next you'll get sir to click on" and "next you'll get stir to click on", and then an amplitude modulation technique was used to form an 11-step continuum of test-words between the "sir" and "stir" tokens. These words were first isolated from their context and time aligned at the onset of the periodic, voiced part of the utterances, where the noise-like frication of the consonant segment gives way to the following vowel. The amplitude envelopes of both words were obtained and each point in the envelope of "stir" was divided by the temporally corresponding point in the envelope of "sir" to obtain the envelope ratio. A value of interpolation, k was chosen, from the range between 0.0 and 1.0. Each point in the envelope ratio was multiplied by k to obtain a modulation function. The original recording of sir was then multiplied by the modulation function and then added to a version of the original "sir" recording that was attenuated by multiplying it by 1-k. This gave test-words at steps of a continuum

numbered with the integers, n, from zero to 10. The corresponding values of k were chosen such that $2\text{asin}((1+k)/2)^{0.5}$ was $\pi/2 + n(\pi/20)$.

Test words from the resulting continuum are heard as "sir" at the lower steps and as "stir" at higher steps. A continuum's steps were presented just once each in a randomized sequence to each of 6 listeners. The step corresponding to the category boundary was found by subtracting 0.5 from the total number of sir responses, giving a boundary step-number between -0.5 and 10.5

Category boundaries were measured with test words played in their original context, i.e. after being re-embedded into "next you'll get _ to click on", or into transformed versions of this context. This re-embedding was actually performed by adding two waveforms, one of these being the context with silence replacing the original "sir", the other being the test word with silence where there was originally the context. In this way it was possible to apply by different BRIRs to the context and the test word, giving same-distance as well as different-distance conditions.

The characteristics of the dummy head talker and of the headphones were removed from the experimental sounds with filters that had the inverse of the corresponding frequency-response characteristics. The resulting sounds were delivered to listeners through these headphones at a peak level of 42 dB SPL so that they sounded as if they were being played where the BRIR was measured.

3 Results

Identification functions and mean category boundaries for the entire original context in same-distance conditions are shown in the left side panels of Fig. 2. Here

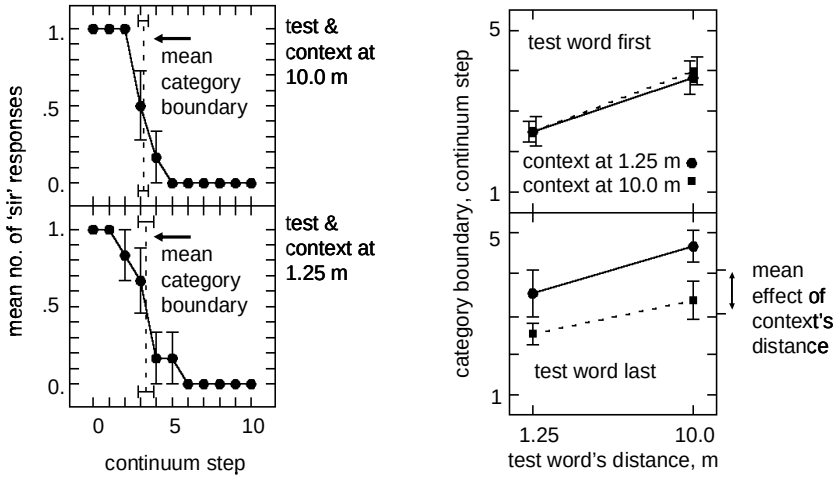


Fig. 2. Left side: Mean identification functions and mean category boundaries for the entire original context in same-distance conditions and for both distances. Right side: Mean category boundaries in contexts that were the same or different to that of the test word, and where the test word was first or last. The mean effect of the context's distance indicates the extent of perceptual compensation for reverberation. Bars are standard errors in all cases.

and elsewhere the room is the corridor, unless otherwise stated. The test word's identity is little affected by the different amounts of reverberation in these same-distance conditions.

If the context's distance affects the test word's identity, then the category boundaries should change between same- and different-distance conditions. Moreover, if this influence normally acts to compensate for effects of reverberation, then the category boundary should be at a higher step number when the context is nearer (test words sound more like sir), and at a lower step number when it is further. This is shown to be the case for the last word conditions shown in the right side panels of Fig. 2, but not for the first word conditions. The mean of these two effects of changing the context's distance indicates the extent of perceptual compensation for reverberation. This is shown for all the experimental conditions in Table 1.

condition	mean effect of context's distance	standard error	t(5)	p<
test-word first	-0.08	0.33	0.24	ns.
test-word last	1.17	0.28	4.18	.005
context and test word monaural	1.17	0.44	2.66	.025
context reversed	1.08	0.30	3.60	.01
context signal-correlated noise	0.50	0.37	1.37	ns.
context L-shaped, test-word corridor	1.67	0.21	7.95	.005

Table 1. For different experimental conditions the table shows the mean effect of the context's distance along with its standard error. The quotient of these quantities, t, is also shown, along with the associated maximum probability of this value under H_0 when t is significant.

4 Discussion and conclusions

The context's reverberation influences the position of the category boundary when only the first part of the context is present. This acts to compensate for the effects of reverberation on the test word. This compensation effect is not observed when only the second part of the context is present as here perception of the test-word is affected by the distortion whatever the reverberation in the context. This indicates an extrinsic perceptual compensation that uses information from preceding sounds, but not from subsequent sounds.

Realistically mild reverberation is therefore capable of producing perceptually salient distortions of speech sounds, while the extrinsic compensation observed here serves to ameliorate these effects.

This compensation does not seem to be brought about by way of a binaurally informed representation of the listening space. This is because its effects are still evident in monaural conditions, as well as in conditions where the context's reverberation is from a different listening space.

Compensation does not seem to be a result of the postponement of decisions to a 'word level', as it is still evident when contexts are reversed so that no words are heard.

These results indicate a fairly general compensation mechanism. However, the absence of compensation with contexts that are signal correlated noise precludes an explanation solely in terms of a response to effects that reverberation from different distances has on the broad-band modulation of the context's envelope.

Envelope modulation in narrow frequency bands *is* reduced when speech signals are changed into signal correlated noise. It is possible that this in some way reduces the effectiveness of these noise contexts. This might happen if the changes in modulation that reverberation from different distances brings about are for some reason less salient when this narrow-band modulation is lower than it is in the original speech signal.

The compensation mechanism might pick up features from the preceding context in order to assess the reverberation present in the listening space. This process may be hindered when the context is noise as it may depend upon a perceptual segregation of reverberant 'tails' from other simultaneous parts of the speech signal. Such segregation can happen in speech when there are transitions between noise-like and periodic segments, but would be less easily effected when the entire signal is noise.

Acknowledgements

Supported by BBSRC (UK). Elizabeth Hallum and Nigel Holt helped record rooms.

References

Clifton, R. K. (1987) Breakdown of echo suppression in the precedence effect. J. Acoust. Soc. Am. 82, 1834-1835.

Gardner, B. & Martin, K. (1994) HRTF measurements of a KEMAR dummy-head microphone. MIT Media Lab. Perceptual Computing - Technical Report #280.

Houtgast, T. and Steeneken, H. J. M. (1973) The modulation transfer function in acoustics as a predictor of speech intelligibility. Acustica 28, 66-73.

Huckvale, M. (1996) Learning from the experience of building automatic speech recognition systems. University College London, Dept. of Phonetics and Linguistics, Speech Hearing and Language: Work in Progress 9, 133-147.

Schroeder, M. (1965) New method of measuring reverberation time. J. Acoust. Soc. Am. 37, 409-412.

Warren, R. M. (1970) Perceptual restoration of missing speech sounds. Science 167, 392-393.

Watkins, A.J. (1991) Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. J. Acoust. Soc. Am. 90, 2942-2955.

Watkins, A.J. (1992) Perceptual compensation for the effects of reverberation on amplitude-envelope cues to the 'slay'- 'splay' distinction Proc. IoA. 14, 125-132.

Zahorik, P. (2002) Assessing auditory distance perception using virtual acoustics. J. Acoust. Soc. Am. 111, 1832-1846.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Psychophysical and physiological studies of the precedence effect and echo threshold in the behaving cat

Daniel J. Tollin[1,2], Micheal L. Dent[1], and Tom C.T. Yin[1]

[1] Department of Physiology, University of Wisconsin-Madison
[2] tollin@physiology.wisc.edu

1 Introduction

The *precedence effect* (PE) describes several perceptual phenomena that occur when similar sounds are presented from different spatial locations and separated by a delay (Wallach, Newman, and Rosenzweig 1949). During s*umming localization*, for delays between ±1 ms, a single 'fused' sound is perceived between the sources but biased towards the leading source ('lead'). During *localization dominance* (Litovsky, Colburn, Yost, and Guzman 1999), for delays of 1-10 ms, the fused sound is localized near the lead only, with little effect of the lagging source ('lag'). The *echo threshold* (Blauert 1997) is the delay at which the two separate source locations are first perceived.

We have been using stimuli that elicit these phenomena to study the neural and psychophysical bases of localization (Yin 1994; Litovsky and Yin 1998; Populin and Yin 1998; Tollin and Yin 2003). We have focused our physiological studies on the inferior colliculus (IC), a site of major convergence of inputs from virtually all brainstem nuclei (Irvine 1986). Many of the inputs themselves are sensitive to the acoustical cues to location (Yin 2002). Reflecting the importance of the IC for localization, lesions of its output pathways or the IC itself results in deficits in localization performance in animals (Jenkins and Masterton 1982) and humans (Litovsky, Fligor, and Tramo 2002). Finally, many IC units are sensitive to changes in location and also encode the individual cues to location themselves (Irvine 1986).

We (Yin 1994; Litovsky and Yin 1998), along with others (Fitzpatrick, Kuwada, Kim, Parham, and Batra 1999), have identified correlates of summing localization and localization dominance in the IC. For delays of a few ms, responses to the lag were found to be substantially reduced compared to responses to the same stimulus presented in isolation from the lagging location, yet the responses to the lead were generally unchanged. Since the responses to the paired and single-source stimuli were similar for these delays, to the extent to which those cells code location, localization performance would be expected to also be similar. With increasing delays, the responses to the lag recovered to those found for single sources.

All of our early studies were performed in barbiturate-anesthetized animals. More recent studies in unanesthetized but non-behaving rabbits have suggested that anesthesia might affect the responses of IC cells to these stimuli (Fitzpatrick et al. 1995). And the responses of IC cells can also depend on the behavioral state of the animal (Groh, Trause, Underhill, Clark, and Inati 2001). To circumvent these effects, we recorded from IC cells of cats that were actively participating in a sound localization task using stimulus configurations that were expected to elicit the PE.

2 Methods

Methods for the psychophysical experiments can be found in Tollin and Yin (2003). Five adult female cats were used. The cats were placed in the center of a dimly-lit sound-attenuating chamber with their heads held fixed and facing a bank of 15 loudspeakers. The cats were trained to indicate via saccadic eye movements the apparent two-dimensional location of auditory targets placed within their ocular-motor range (~$\pm 25^o$). Eye position was recorded using the scleral search coil technique. All procedures were approved by the University of Wisconsin Animal Care and Use Committee and also complied with the NIH guidelines for animal use.

The stimuli consisted of five (sometimes 10) identical broadband noisebursts, each 10-ms in duration and gated by a rectangular window, presented at 5 Hz for a total of 1 or 2 sec. On occasion, 100-μs clicks were used. This stimulus was presented either from single speakers (single source) or with equal level from two different speakers connected in phase (paired source) but with a delay, the inter-stimulus delay (ISD), between the onsets. Here the locations of the paired sources were held constant at $\pm 18^o$ on the horizontal plane.

Psychophysical data was taken from the *saccade* task, where cats were required to make saccades from an initial LED located at (0^o,0^o) to acoustic targets presented from single sources and paired sources with varying ISDs. Details on reward contingencies are found in Tollin and Yin (2003). For the physiology, the cats also performed *sensory probe* tasks. Here, the initial fixation LED remained illuminated during the stimulus presentation and the cats had to maintain fixation throughout the stimulus. Physiological data was taken during *sensory probe* tasks to reduce the possible effects of eye position on the responses (Groh et al. 2001).

Standard extracellular recording techniques were used to record the sound-evoked discharges of well-isolated single units or, on occasion, multi-unit clusters using metal microelectrodes. For each unit, the level of the stimulus was fixed at 10-30 dB above threshold (measured from a contralateral source) for all trials.

3 Results

The goals of these experiments were to 1) determine the ISDs over which cats experience the PE and 2) to characterize the responses of a population of single cells in the IC for the same stimulus and under similar behavioral conditions.

3.1 Psychophysics

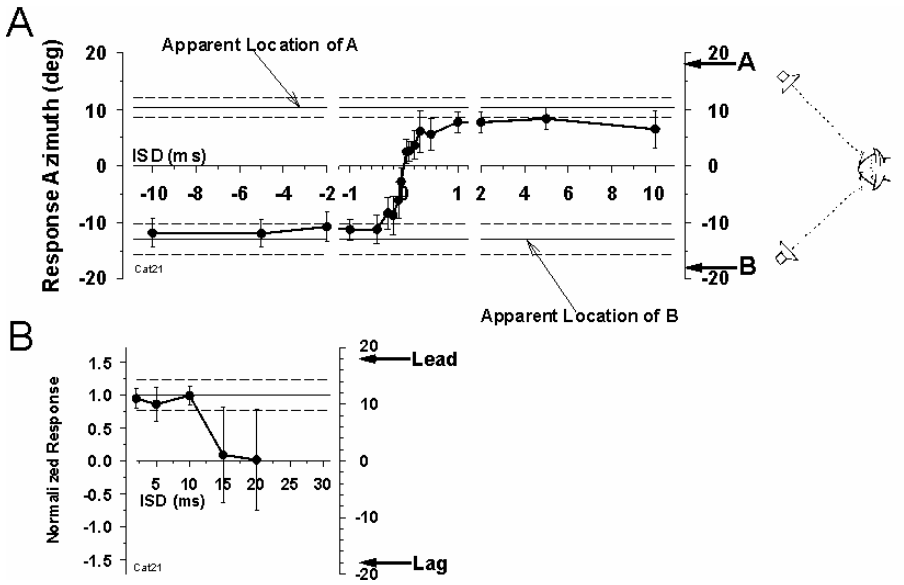


Fig. 1. (A) Response azimuth as a function of ISD between the stimuli presented from locations A and B (solid circles) for Cat21. Response azimuth for single sources presented from locations A and B are also shown (solid lines), +/- 1 SD (dashed lines). (B) Normalized response as a function of ISD between lead and lag stimuli for longer ISDs.

Figure 1A shows the performance for one cat with single and paired sources. Similar patterns of behavior were observed in the four other cats. Consistent with summing localization, as the ISD was varied between ±400 μs, the mean response azimuth was between the two sources, but became systematically biased towards the lead as the delay was increased. Consistent with localization dominance, for ISDs beyond 400 μs, response azimuth and elevation (not shown) was near the response to the lead in isolation and largely independent of ISD up to 10 ms. For ISDs > 10 ms, the cat made saccades towards the lag on some trials, suggesting that the echo threshold was exceeded and that the cat perceived the lagging stimulus. To quantify this, Fig. 1B shows the mean normalized response towards the lead (irrespective of which side was leading) as a function of ISD; the normalizing factor was the unsigned mean response azimuth to the two single sources (i.e., solid horizontal lines in Fig. 1A). A value of 1.0 indicates the orienting responses to the paired sources were identical in magnitude to that of the 'leading' single source location. For ISDs < 10 ms, responses were always at the lead. For ISDs of 15 and 20 ms, the mean responses were near 0^o but with large standard deviations that reflected the bimodal distribution of responses, some trials towards the lead (1.0) and some towards the lag (-1.0). Thus, our data suggest that echo threshold for these cats and these stimuli is between 10-15 ms.

3.2 Physiology: Responses to single sources

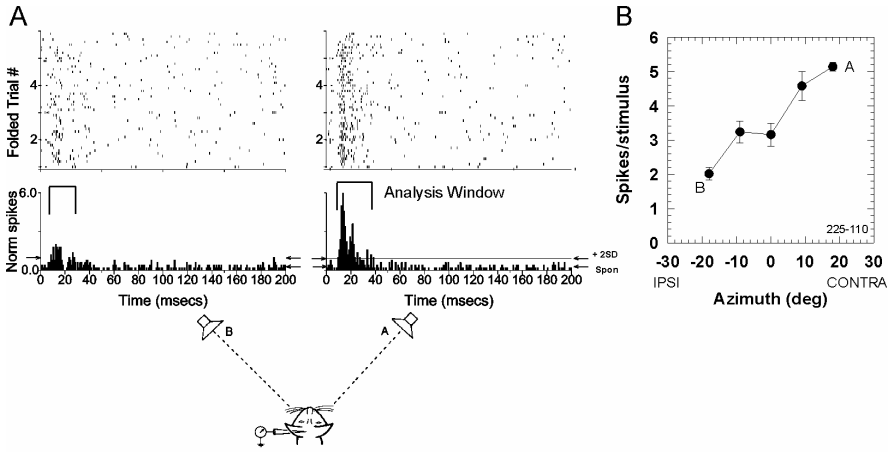


Fig. 2. (A) Responses of one IC unit to a single source located at –18° (left panels) and +18° (right panels). (B) Responses of the same unit as a function of stimulus location.

Over 150 units were sampled from five cats, and detailed recordings were taken from 52. Histology has not yet been done in all animals, but based on the orderly increase in best frequency (BF) with electrode depth, we believe that all units were from the central nucleus (ICC). All cells here had BFs > 1.5 kHz. Fig. 2A shows the response of one unit to the single sources at ±18° on the horizontal plane. The response rasters (top panels) have been 'folded' on the 5-Hz period of the stimulus, both within a trial and across trials, and the histogram (bottom panels) shows the result of this folding. For each location, the number of spikes was counted in an 'analysis' window whose onset and duration was defined by the post-stimulus time at which the instantaneous discharge rate (1-ms bins) exceeded 2 SD (upper arrow, right side) of the mean spontaneous rate (lower arrow) computed 500 ms prior to the onset of each trial. Fig. 2B shows the responses/stimulus as a function of source azimuth. Like many of the neurons we recorded from, the response for this unit was low for ipsilateral sources and was higher for contralateral sources.

3.3 Physiology: Responses to paired-sources

Figure 3 shows responses of the same cell to paired-source stimuli at three different ISDs. The left column shows conditions where the contralateral led the ipsilateral source, and vice versa for the right column. As was the case for most units, for large ISDs, clear responses were usually seen to both the lead and the lag, but with decreasing ISD, responses to the lag were typically diminished. For each unit, the responses to the lag at varying ISDs were compared to the responses to single-source stimuli (e.g., Fig. 2) from the same location as the lag in the following way.

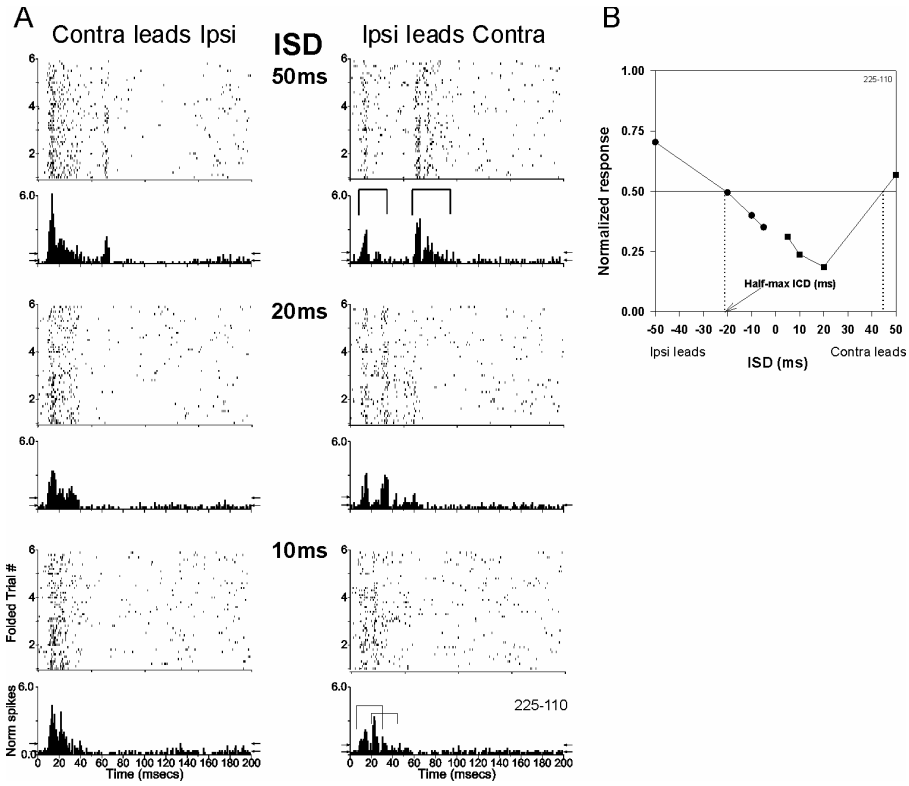


Fig. 3. (A) Response rasters and folded histograms from the unit in Fig. 2 to paired stimuli located at +/–18^{o} with Contra leading (left panels) or Ipsi leading (right panels) as a function of ISD. (B) Normalized response rate as a function of ISD for the same unit.

The lag response was normalized by dividing it by the number of spikes elicited by the 'lag' in the single source condition. For larger ISDs, the leading and lagging analysis windows did not overlap (e.g., 50-ms ISD) and we were able to compute the response to the lag independently. When the ISDs were small, the windows for the lead and lag overlapped, so we subtracted the single-source 'lead' response from the paired-source response as computed through a composite analysis window whose onset and offset were determined from the lead and lag analysis windows, respectively. Any increase or decrease in the overall response was attributed to the presence of the lag.

Figure 3B shows the normalized recovery of the lag response as a function of ISD for the same unit. The half-maximal ISD (Yin 1994), the ISD at which the recovery reached 50%, is a convenient indicator of the location of the recovery function along the ISD axis. As was the case for most units, at small ISDs the response to the lag was 'suppressed' to rates well below normal (as indicated by normalized responses < 1.0), but recovered with increasing ISD.

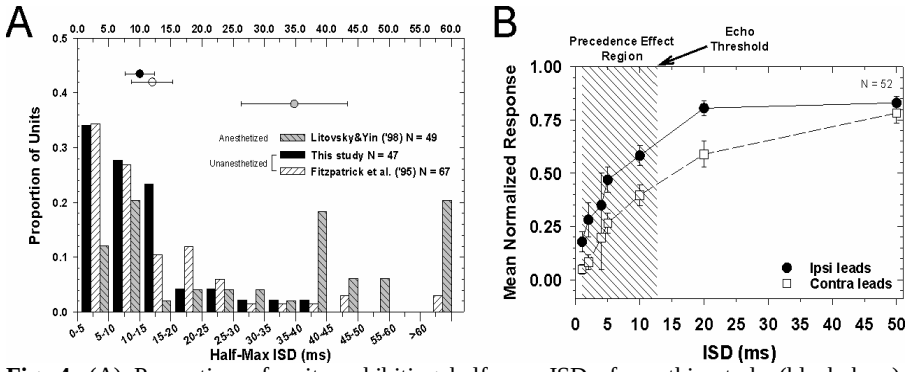


Fig. 4. (A) Proportion of units exhibiting half-max ISDs from this study (black bars), Litovsky and Yin (1998; grey bars), and Fitzpatrick et al. (1999; white bars). The circles represent average half-max ISDs from each study. (B) Mean normalized response as a function of ISD with the Ipsi stimulus leading (circles) or Contra stimulus leading (squares).

Figure 4 shows a summary histogram of these half-max ISDs for the ipsilateral-leading condition from this study, and for a similar study from our lab, but in a barbiturate-anesthetized preparation (Litovsky and Yin 1998). Also shown are results from the IC of the un-anesthetized, but non-behaving rabbit, using similar stimuli (Fitzpatrick et al. 1999). The half-max ISDs in the behaving cat IC were significantly different from those in the anesthetized cat (Mann-Whitney U-268, $p<0.0001$), but were not different from that in the rabbit ($U=726$, $p=0.684$).

Finally, Fig. 4B shows the mean recovery function of the population of IC units constructed by averaging the individual unit recovery functions (e.g., Fig. 3B). In essence, these two functions, one each for the ipsi- and contra-leading conditions, may be taken to represent the aggregate normalized response of both the ipsilateral and the contralateral inferior colliculi to the lag as a function of ISD. A two-factor ANOVA showed a significant main effect on the normalized response of ISD [$F(1,6) = 40.14$, $p<0.0001$] and of whether the lead was presented from the ipsilateral or contralateral side [$F(1,6) = 20.16$, $p<0.0001$]. Note that the population response to the lag was significantly reduced relative to normal for ISDs < ~15 ms.

4 Discussion

These are the first comprehensive psychophysical measurements of the PE in cats. Cats experienced summing localization for ISDs < 400 μs, localization dominance from 400 μs to ~10 ms, and echo thresholds at ~10-15 ms. The ISD ranges for localization dominance and echo threshold are similar to those found in humans and other species with simular stimuli (Litovsky et al. 1999). But the range for summing localization was smaller in cats, occurring for ISDs of ~800 μs in humans, in accordance with the difference in head widths.

The responses of IC cells showed several correlates of the various PE phenomena. For ISDs encompassing localization dominance, the responses to the lag were suppressed, yet responses to the lead were generally unaffected. If these

units code for sound location, then the localization performance of the cats would be expected to be similar in both the single-source and the paired-source conditions. The psychophysical data support this hypothesis. For ISDs at or greater than the echo threshold, the responses to the lag had recovered substantially and were nearly the same as those to the lag presented by itself. At the level of the IC, it appears that a nearly-normal response to the lag is related to its perception as an independent sound source that can be localized. Finally, comparisons of our data to that collected using similar stimuli, but under barbiturate anesthesia, reveal that anesthesia results in larger half-maximal ISDs and prolonged recovery times.

Acknowledgments

We thank Dr. Luis Populin for assistance on earlier portions of this work. Supported by NIH NIDCD grants DC02840, DC00116 (TCTY), DC006124 (MLD), and DC00376 (DJT).

References

Blauert, J. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA.

Fitzpatrick, D.C., Kuwada, S., Kim, D.O., Parham, K., and Batra, R. (1999) Responses of neurons to click-pairs as simulated echoes: Auditory nerve to cortex. J. Acoust. Soc. Am. 106, 3460-3472.

Groh, J.M., Trause, A.S., Underhill, A.M., Clark, K.R., and Inati, S. (2001) Eye position influences auditory responses in primate inferior colliculus. Neuron 29, 509-518.

Irvine, D.R.F. (1986) *The Auditory Brainstem: Processing of Spectral and Spatial Information*. Berlin, Springer-Verlag, pp. 1-276.

Jenkins, W.M. and Masterton, R.B. (1982) Sound localization: Effects of unilateral lesions in central auditory system. J. Neurophysiol. 52, 819-847.

Litovsky, R.Y., Colburn, H.S., Yost, W.A., and Guzman, S.J. (1999) The precedence effect. J. Acoust. Soc. Am. 106, 1633-1654.

Litovsky, R.Y., Fligor, B.J., and Tramo, M.J. (2002) Functional role of the human inferior colliculus in binaural hearing. Hear. Res. 165, 177-188.

Litovsky, R.Y. and Yin, T.C.T. (1998) Physiological studies of the precedence effect in the inferior colliculus of the cat. I. Correlates of psychophysics. J. Neurophysiol. 80, 1285-301.

Populin, L.C. and Yin, T.C.T. (1998) Behavioral studies of sound localization in the cat. J. Neurosci. 18, 4233-4243.

Tollin, D.J. and Yin, T.C.T. (2003) Spectral cues explain illusory elevation effects with stereo sounds in cats. J. Neurophysiol.

Wallach, H., Newman, E.B., and Rosenzweig, M.R. (1949) The precedence effect in sound localization. Amer. J. Psychol. 57, 315-336.

Yin, T.C.T. (1994) Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. J. Neurosci. 14, 5170-5186.

Yin, T.C.T. (2002) Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In: D. Oertel, R.R. Fay, and A.N. Popper (Eds.), *Integrative Functions in the Mammalian Auditory Pathway*. Springer, New York, pp. 99-159.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Some similarities between the temporal resolution and the temporal integration of interaural time differences

Michael A. Akeroyd

MRC Institute of Hearing Research (Scottish Section), Glasgow, maa@ihr.gla.ac.uk

1 Introduction

In many real-world environments the interaural properties of a sound vary across time. New sounds may come or go, masking or revealing the interaural properties of other sounds. Any measurements made over a relatively short time will minimize errors from such variations or changes and reduce the confounding effect of adjacent sounds. But if measurements can be made over a longer time then they will be more accurate, especially if the target is only slowly changing. The former can be characterized as the "temporal resolution" of binaural processing: the shortest duration over which a measurement can be made and also the degree to which surrounding sounds can be rejected. The latter is "temporal integration": the longest time over which static information can be included to advantage.

The effective duration for temporal integration for interaural time differences (ITD) is many hundreds of milliseconds (e.g., Houtgast and Plomp 1968). For temporal resolution, since Grantham and Wightman's (1978, 1979) studies of the resolution of time-varying changes in interaural parameters, it has been common to characterize the binaural system as "sluggish". To take just one example, Culling and Summerfield (1998) measured an effective duration of about 70 ms for the temporal resolution of the masking-level difference at 500 Hz (this value is the "equivalent rectangular duration" of an exponential function characterizing the temporal resolution). Nevertheless, Bernstein, Trahiotis, Akeroyd and Hartung (2001), using a design based on Wagner's (1991) studies of owls, showed that the temporal resolution of ITD can be remarkably short. They measured an equivalent rectangular duration of a mere 2 ms. Thus, the binaural system may not always, nor in all situations, be sluggish, but there may also be task-specific effects and theoretical dependencies in the analyses (for instance, Bernstein et al. showed that their extremely-short temporal-resolution function could also account for some of the original sluggish data from Grantham and Wightman, 1978).

The primary aim of this report is to show that Bernstein et al.'s data has intriguing similarities with Houtgast and Plomp's (1968) data on temporal *integration* for signals in noise. Both are fitted by a power-law function, as though later segments of a sound convey progressively less information than do earlier segments (Hafter and Dye 1983). Moreover, a power-law function—albeit with a

smaller exponent—also describes data on temporal integration of ITD for signals in quiet (Tobias and Zerlin 1959; Houtgast and Plomp 1968; Hafter and Dye 1983). It is also shown that a weighted-*d'* model—essentially a quantitative implementation of Houtgast and Plomp's ideas—can account for many of the similarities.

2 Signal-detection theory and temporal integration

Predictions of the "rate of integration" of ITD—or the rate of change of threshold ITD with the duration of the signal—can be obtained from signal-detection theory (Houtgast and Plomp 1968). The signal is divided into a train of discrete, independent "segments", for instance from 0-10 ms, 10-20 ms, and so on. The auditory system measures the ITD of each segment, but, because of internal noise inherent to the processing, with a slight inaccuracy and uncertainty. Accordingly, the ITD of each segment can be represented as a random variable, with a mean equal to the physical ITD of the stimulus and a standard deviation determined by the internal noise. If the segments are all equivalent, in that they (1) have the same mean ITD, (2) have the same standard deviation σ, and (3) are all weighted equally, and as (4) *d'* is proportional to ITD (e.g., Bernstein and Trahiotis 1982; Saberi 1995), then the detectability of a change in ITD ΔT carried by n segments will be given by (see Appendix 9-A of Green and Swets 1974)

$$d' = n\Delta T/(n\sigma^2)^{0.5} = n^{0.5}\Delta T/\sigma \quad (1)$$

The result is that a two-fold increase in signal duration will lead to a two-fold increase in the number of segments n and therefore a square-root-of-two increase in *d'*. Accordingly, at threshold (*d'*=1), the ITD will be less by the square-root of two. This model predicts a power-law relationship between signal duration and threshold ITD, with an exponent of –0.5. When plotted on coordinates of log threshold-ITD *vs.* log signal-duration, the result is a straight line with a slope of –0.5.

3 Temporal integration for noise stimuli in quiet

Tobias and Zerlin (1959) measured the temporal integration of threshold ITD for wideband noise lowpass filtered at 5000 Hz, presented at about 65 dB SPL. Threshold ITDs were measured as a function of the duration of the noise. They found (their Fig. 2) that threshold ITDs generally improved as the duration increased, asymptoting at a duration of about 0.5-1 seconds. Their results are redrawn in the left panel of Fig. 1 as log threshold-ITD *vs.* log signal-duration. The data can be well fitted by the dotted line, whose slope is approximately –0.3.

Houtgast and Plomp (1968) measured the temporal integration of threshold ITD for 1-octave bands of noise centered on 500 Hz, presented at a sensation level of about 50 dB. Threshold ITDs were measured as a function of the duration of a signal band of noise that was presented in the absence or presence of another continuous, masking band of noise, at signal-to-masker ratios (S/M) between –10 to +20 dB. They observed that the rate of integration depended upon S/M (their Fig. 5). The results for the signal-in-quiet condition are redrawn as open circles in the left panel of Fig. 1, again as log threshold-ITD *vs.* log signal-duration (their results

for a S/M of +5 dB will be considered later). The data can be well fitted by the solid line, whose slope is approximately –0.2.

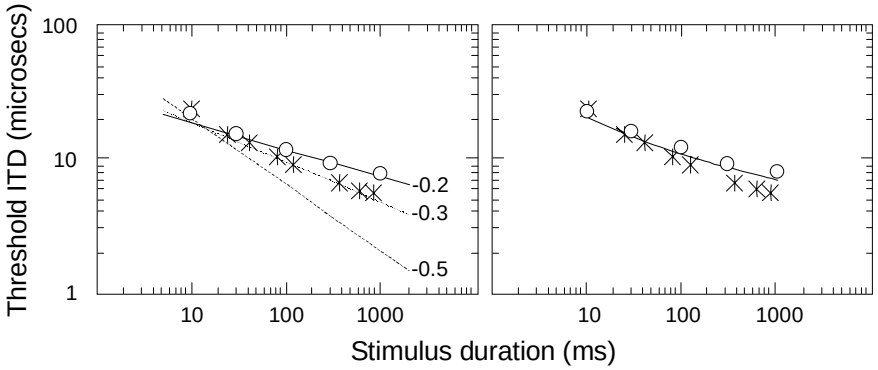


Fig. 1. The symbols plot threshold ITDs for noise stimuli in silence (asterisks, from Tobias and Zerlin 1959, and open circles, from Houtgast and Plomp 1968). In the left panel the lines plot various power-law relationships between threshold and duration (see eq. 2), with the slope *p* marked. In the right panel the line plots the predictions of the weighted-*d'* model.

In both datasets the relationship between threshold ITD and stimulus duration can be approximately characterized by a power function:

$$\textit{Threshold ITD} \propto \textit{signal-duration}^{p} \qquad (2)$$

where *p* is the slope of the straight line in the plots. The data suggests that *p* has the value of approximately –0.2 to –0.3. That is, the rate of integration is described by a power function with an exponent of –0.2 to –0.3. This rate is substantially smaller than the prediction of –0.5 (as shown by the dashed line in Fig. 1).

Houtgast and Plomp suggested that this effect was due to more emphasis being placed on the onset of the sound than elsewhere. This idea is supported by many of the experiments studying the precedence effect (summarized by Litovsky, Colburn, Yost and Guzman 1999). The idea can be quantified by incorporating a weighting of each stimulus segment into the earlier *d'* model. It was noted above that the prediction of a slope of –0.5 is based on all segments of the signal being weighted equally. But if early segments—and especially the first one, marking the onset of the sound—are weighted more then the *d'* model will predict a slope shallower than –0.5. One example is shown in the right panel of Fig. 1, for which the weight reduced in inverse proportion with time. Here, it was assumed that (1) the stimulus was divided into a train of concatenated segments, each 10 ms in duration; (2) the weight applied to each segment was given by $w = t^{-1}$, where *t* is the time of the end of the *n*th segment, but (3) the weight applied to the first segment was further emphasized by 2x, and (4) a value of *d'* was calculated by optimally combining the individually-weighted segments (e.g., Green and Swets 1974; Buell and Hafter 1991). The curved line in the plot shows the prediction. It has a rate-of-integration of –0.25 (at least for a duration up to about 100 ms), and is a fair characterization of the experimental data.

4 Temporal integration for click trains

Hafter and Dye (1983) studied the temporal integration of ITD carried by 1-octave click trains centered on 4000 Hz, presented at a level of each click of 40 dB SPL in quiet. Threshold ITDs were measured as a function of the click rate and the number of clicks in the train. The results (their Fig. 4) showed that the rate of integration depended upon the click rate, it increasing as the click rate decreased. The results are redrawn in Fig. 2 in terms of the overall duration of their stimuli. In order to do so, it was assumed that the effective duration of each click was 0.5 ms; for instance, for a click rate of 1000 Hz (that is, a click spacing of 1 ms), a 1-click stimulus was 0.5 ms long, a 2-click stimulus was 1.5 ms, and so on. The straight line plotted on the data gives a reasonable fit. Its slope is –0.2.

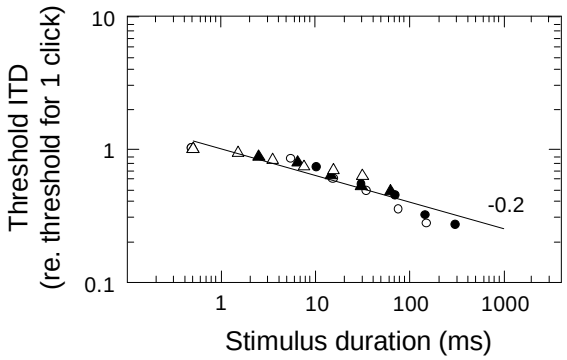


Fig. 2. The symbols plot threshold ITDs for click-train stimuli from Hafter and Dye (1983), with the parameter being the click rate (solid circles = 100 Hz; open circles = 200 Hz; solid triangles = 500 Hz; open triangles = 1000 Hz). The line plots a power-law relationship between threshold ITD and duration, with a slope of –0.2.

It would be reasonable to believe that a common process is responsible for temporal integration of ITD for both noise stimuli and click stimuli. The experimental data supports this notion: the temporal integration of threshold ITD (at least for signals in quiet) for both stimuli can be described by the power-law function of Eq. 2 with an exponent of about –0.2 to –0.3.

Hafter and Dye (1983) demonstrated that their "binaural saturation" model could elegantly account for the click-train data. It too effectively puts more weight on the onset of a sound than subsequent parts. More formally, it assumed that later clicks convey progressively less information than do earlier clicks, such that the total neural events N evoked by n clicks was given by $N \propto n^{k}$, where k was a parameter dependent upon the click rate, and was found to lie between 0.3 (1000-Hz click rate) and 0.8 (100-Hz click rate). This model predicts the dependence of threshold ITD on stimulus duration to be the same power function as Eq. 2, but with an exponent p equal to $-k/2$. The value of p would thus lie between –0.15 and –0.4.

The binaural-saturation and weighted-*d'* models are both variations on the same theme. At present, however, the first is limited to click stimuli and the second to noise stimuli. More work is needed to get one model to predict the data from the other's domain. One possible approach will be to equate the number of segments of

the weighted-*d'* model to the number of clicks in the binaural-saturation model, but it remains to be tested if this approach is successful by, for instance, predicting the dependency of *k* on click rate.

5 Temporal integration and resolution for signals in noise

It was noted earlier that Houtgast and Plomp (1968) demonstrated that, for a 1-octave-wide noiseband centered at 500 Hz presented against a continuous masking noise, the rate-of-integration depends on the signal-to-masker ratio (S/M). Their results for a S/M of +5 dB are redrawn in the left panel of Fig. 3 as open circles.

The asterisks in Fig. 3 plot data redrawn from Bernstein et al. (2001, Fig. 1). They studied threshold ITDs carried by broadband (0-8.5 kHz) bursts of noise. Threshold ITDs were measured as a function of the duration of a "probe" segment of noise embedded in the temporal center of a diotic noise. Importantly, the levels of the probe and flanking noise were the same, so there was no monaural information as to the timing of the probe: accordingly, there was no "onset" to the probe. For the data replotted here, the overall duration of the stimulus was fixed at 100 ms, with probe durations between 2 and 64 ms.

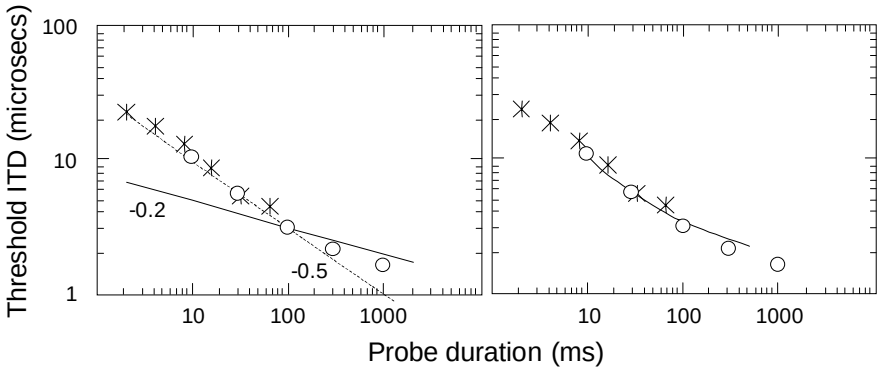


Fig. 3. Same as Fig. 1, except that the symbols plot threshold ITDs for noise stimuli in fringes or continuous maskers (asterisks, from Bernstein et al. 2001, and open circles, from Houtgast and Plomp 1968).

There is a clear convergence between the two sets of data. They overlap remarkably and they have the same slope. It is as though Bernstein et al.'s data is an extension of Houtgast and Plomp's to shorter durations. The dashed line accurately characterizes both datasets, at least up to a "breakpoint" at a duration of about 300 ms. Its slope is –0.5, which matches the prediction of the first *d'* model.

It would be reasonable to believe that a common process is responsible for both sets of results. The following paragraphs show that both Bernstein et al.'s window function and the weighted-*d'* model can be modified to account for the data.

Bernstein et al. interpreted their data in terms of the temporal resolution of the binaural system, and, like some of the studies demonstrating binaural sluggishness in temporal resolution (e.g., Grantham and Wightman 1979; Culling and

Summerfield 1998), they obtained a shape and duration for the averaging function or temporal window. A window constructed from the sum of two exponential-shaped functions (their Fig. 5) could account for 96% of the variance in the threshold ITDs. This window had the remarkably short equivalent-rectangular duration of only 2 ms—the time constant of the sharp, strong "peak" was only 0.1 ms while that of the shallow, and much weaker, "skirt" was 14 ms—leading Bernstein et al. to note "… the binaural system may not be necessarily or intrinsically sluggish" (p. 1614). The ability of this window to integrate ITD information is limited to stimuli durations less than about 100 ms (which was the longest duration used by Bernstein et al.). Further integration up to longer durations can only be achieved by including an additional, even-longer (100- or 200-ms time constant) and even-less-intense, skirt to the window function. This "triple" window can account for the commonality of the both sets of data. It gives temporal integration up to hundreds of milliseconds, but it has an equivalent-rectangular duration with the decidedly "non-sluggish" value of only about 5-10 ms.

Bernstein et al. placed uninformative, diotic fringes immediately before and after the noise probe, with no change in level across fringe or probe. Houtgast and Plomp surrounded their noise probes with an uninformative continuous masker, with the presence of the probe marked by a 6.2 dB gain in level. In the former design, there was no onset to the probe, and in the latter, the size of the onset—and the offset too—would have been substantially reduced compared to any signal presented in quiet. Maybe, then, for stimuli without clear, large, or abrupt onsets, the temporal integration of *ongoing* threshold ITD can be described by the power-law function of Eq. 2 but with an exponent of about –0.5 (Houtgast and Plomp 1968). The right panel of Fig. 3 shows that incorporating this reduction in the strength of the onset into the weighted-*d'* model leads to good predictions. The predictions were obtained by changing two aspects of the weighted model that was described in Section 2.: (1) the weight on the first 10-ms segment was not increased by 2x, and (2) all segments up to 500 ms were included, no matter what the actual duration of the probe. The first change reflects the reduction in the strength of the onset cue. The second change reflects the concomitant reduction in the offset cue (as though, when there is no clear offset, the system does not know when to *stop* integrating); 500 ms is a somewhat arbitrary value. This model successfully predicts a slope of about –0.5 for probe durations less than about 100 ms.

That both approaches give similar predictions suggest that they too are variations on a theme. Nevertheless, like in the analyses of signals in quiet, there are certain complications that need to be addressed. For instance, the weighted-*d'* model gives a breakpoint at a shorter duration than was observed. Also, is it fair to use a segment duration of 10 ms, when some of the probes used by Bernstein et al. were only 2, 4 or 8 ms? Third, in Bernstein et al.'s experiment, there was *no* change in level to mark either the onset or the offset of the probe. Can the auditory system know when to *start* integrating? (Bernstein et al. resolved this question by using the maximum output of a temporal window, which, for their stimuli, is equivalent to using a window placed at the center of the stimulus). Finally, Fig. 3 only shows Bernstein et al.'s data for a overall duration of 100 ms. Analysis of their data for shorter durations (20 and 40 ms) showed that the rate of integration

depended slightly on overall duration, reaching –0.6 for the 20-ms condition. It remains to be seen if the model can account for this small effect.

6 Summary

These analyses have outlined some links between measurements of the temporal integration of ITD and the temporal resolution of ITD. They indicate (1) for signals in noise, threshold ITD decreases with stimulus duration at a rate of approximately –0.5, yet (2) for signals in quiet, the rate is reduced, being approximately –0.2 to -0.3. These differences may be related to the clearness of the onset of the stimuli. It was also shown that a weighted-*d'* model, in which the weight was inversely proportional to time, can account for many of the effects, but it remains to be seen if this approach offers theoretical or practical advantages over others.

Acknowledgments

I would like to thank Quentin Summerfield, Les Bernstein, and Stuart Gatehouse for their beneficial criticisms of earlier drafts of this report.

References

Bernstein, L.R., Trahiotis, C., Akeroyd, M.A., and Hartung, K. (2001) Sensitivity to brief changes of interaural time and interaural intensity. J. Acoust. Soc. Am. 109, 1604-1615.

Bernstein, L.R. and Trahiotis, C. (1982) Detection of interaural delay in high-frequency noise. J. Acoust. Soc. Am. 71, 147-152.

Buell, T.N. and Hafter, E.R. (1991) Combination of binaural information across frequency bands, J. Acoust. Soc. Am. 90, 1894-1900.

Culling, J.F. and Summerfield, Q. (1998) Measurements of the binaural temporal window using a detection task. J. Acoust. Soc. Am. 103, 3540-3553.

Grantham, D.W. and Wightman, F.L. (1978) Detectability of varying interaural temporal differences. J. Acoust. Soc. Am. 63, 511-523.

Grantham, D.W. and Wightman, F.L. (1979) Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. J. Acoust. Soc. Am. 65, 1509-1517.

Green, D.M. and Swets, J.A. (1974). *Signal Detection Theory and Psychophysics*. Wiley, New York.

Hafter, E.R. and Dye, R.H. (1983) Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number. J. Acoust. Soc. Am. 73, 644-651.

Houtgast, T. and Plomp, R. (1968). Lateralization threshold of a signal in noise. J. Acoust. Soc. Am. 44, 807-812.

Litovsky, R.Y., Colburn, H.S., Yost, W.A., and Guzman, S.J. (1999). The precedence effect. J. Acoust. Soc. Am. 106, 1633-1654.

Saberi, K. (1995). Some considerations on the use of adaptive methods for estimating interaural-delay thresholds. J. Acoust. Soc. Am. 98, 1803-1806.

Tobias, J.V. and Zerlin, S. (1959). Lateralization threshold as a function of stimulus duration. J. Acoust. Soc. Am. 31, 1591-1594.

Wagner, H. (1991). A temporal window for lateralization of interaural time differences by barn owls. J. Comp. Physiol. A 169, 281-289.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Binaural "sluggishness" as a function of stimulus bandwidth

Caroline Witton[1], Gary G. R. Green[2], and G. Bruce Henning[3]

[1] Neurosciences Research Institute, Aston University, c.witton@aston.ac.uk.
[2] School of Neurology, Neurobiology, and Psychiatry, University of Newcastle upon Tyne, gary.green@ncl.ac.uk.
[3] Sensory Research Unit, Department of Experimental Psychology, Oxford University, bruce.henning@psy.ox.ac.uk.

1 Introduction

It has been suggested that the binaural auditory system is sluggish, meaning that it has lower temporal resolution than the monaural auditory system (Grantham and Wightman 1978, 1979; Grantham 1982, 1984). Grantham and Wightman's demonstration of sluggishness used band-limited frequency-modulated stimuli presented either diotically or 180 degrees out-of-phase so that the interaural time difference (ITD) was modulated. Their subjects discriminated the dichotic from the diotic FM, and thresholds increased sharply as modulation rates increased from 2.5 to 10 Hz; poorest performance occurred at modulation rates between 10 and 40 Hz. This sharp increase in threshold provided evidence that the auditory system was unable to follow the increasingly rapid changes in, presumably phase-locked, temporal information required to distinguish modulated ITDs from a zero-ITD stimulus.

Evidence from other experiments was soon provided to support the hypothesis that the binaural system is sluggish. Studies of sound motion detection in the horizontal plane, which relies in the main on the binaural processing of modulations in interaural time and level differences, showed that minimum audible movement angles were larger than minimum audible angles for static sounds (e.g., Harris & Sargeant 1971). Motion sensitivity is heavily dependent on the duration of movement, with minimum integration times as long as 200 ms (Grantham 1986; Chandler and Grantham 1992), indicating that the binaural temporal window is long (see also Kollmeier and Gilkey 1990).

Nevertheless, there have also been some reports that the binaural system may not be entirely sluggish. Human listeners can be sensitive to very brief changes in ITDs, and Bernstein, Trahiotis, Akeroyd and Hartung (2001) provide a double-exponential model of the binaural temporal window that includes a heavily weighted short time constant that can account for this apparent non-sluggishness.

Their model also includes a longer time constant that accounts for Grantham and Wightman's (1978) discrimination data. Witton, Simpson, Henning, Rees and Green (2003) also report apparently non-sluggish behaviour in the detection and discrimnation of linear ramp modulations in ITD and ILD, for tonal carriers.

A final piece of evidence for non-sluggishness in the binaural system is that sensitivity to FM in a tone is increased by up to an order of magnitude when it is accompanied by a tone of the same carrier in the contralateral ear (thus generating modulated ITDs; e.g., Henning & Zwicker 1984; Witton et al. 2000). This "dichotic advantage", similar to a binaural masking level difference, is greatest at low modulation rates but persists when the modulation rate is increased to 40 or 60 Hz. Interestingly, these findings appear to be at odds with the observations of Grantham and Wightman (1978), who used band-limited (250-3000 Hz) stimuli to show that discrimination of dichotic from diotic FM was increasingly impaired as modulation rates increased beyond 10 Hz. Paradoxically, for tonal stimuli, subjects' dichotic advantage in detection suggests that *discriminating* dichotic from diotic FM should be trivially easy and that thresholds should be similar to dichotic FM detection thresholds up to modulation rates of at least 40 Hz (i.e., approximately constant when expressed as frequency modulation index; Witton et al. 2000).

In the experiments described here, we measured thresholds for discriminating dichotic from diotic FM for a 500-Hz tonal carrier and then, using the same discrimination task, with carriers of wider bandwidth (41 Hz and 241 Hz).

2 Methods

Two experienced listeners with normal hearing participated in a series of psychophysical experiments to measure thresholds for discriminating dichotic from diotic FM across a range of modulation rates and carrier types.

Stimuli were generated with Tucker-Davis Technologies System II equipment. The required waveforms were created digitally and scaled to fill the dynamic range of two independent 16-bit digital-to-analog converters. The sampling rate was 100 kHz and the signals for each ear were low-pass filtered at 12 kHz, separately attenuated and used to drive calibrated headphones working in phase.

The frequency-modulated carrier could be tonal or band-limited, and all stimuli in each psychophysical trial were constructed from the same sinusoidally frequency-modulated carrier, s(t):

$$s(t) = A\,(t)\,\sin[2\pi f_c t + \theta(t) + \beta\,\sin\,(2\pi f_m t)], \qquad (1)$$

where A(t), the amplitude of the signal, is constant for a tonal carrier, f_c (Hz) is the carrier frequency, θ(t), the phase of the signal, is also constant for a tonal carrier, f_m (Hz) is the rate of sinusoidal modulation, and β is the modulation index for a tonal carrier but the amplitude of the added modulation for noise waveforms.

The noise waveforms consisted of the sum of a number of components, each separated by 1 Hz. The components were of random phase (uniform) and random

amplitude (Rayleigh); both from the same random distributions. They thus formed a Fourier-series band-limited white Gaussian noise. The waveform can be considered as an amplitude, $A(t)$, and phase ($\theta(t)$, (t)) modulated tone at the carrier frequency (Davenport and Root, 1958). We added to the phase term a sinusoidal phase modulation of frequency f_m and amplitude $\pm\beta$ (Eq. 1).

In the diotic FM condition, the same waveform was presented in phase to each ear. For the dichotic FM condition, the added modulation (governed by β) was presented with a 180-degree phase difference between the ears.

The duration of the stimuli was 1000 ms, with a 500-ms inter-stimulus interval. All sounds were presented at approximately 50 dB SL.

Discrimination thresholds were measured using a standard 2-alternative forced-choice method. One observation interval contained diotic FM, and the other contained dichotic FM, and subjects reported the interval in which the modulation was dichotic. Full psychometric functions were obtained, with at least 100 trials at each point. Thresholds (the FM depth at which subjects achieved 75% correct performance) were calculated using the fitting software provided by Wichmann and Hill (2001a, 2001b), which also provides bootstrap-based estimates of the ± one-standard deviation confidence intervals about the thresholds.

3 Results and Discussion

3.1 Experiment 1: Discrimination thresholds for a tonal carrier

In the first experiment, thresholds for discriminating diotic from dichotic FM of a 500-Hz tonal carrier were measured as a function of modulation rate. Data for each subject can be seen in Fig. 1 (clear circles). Thresholds are reported in frequency modulation index (β; right-hand ordinate) and the peak ITD in the dichotic condition (μs; left-hand ordinate). For both subjects, thresholds initially decrease as modulation rate is increased from 2 to 20 Hz. Thresholds then peak at 40 Hz for Subject 1 and 60 Hz for Subject 2. This pattern of results clearly differs from the shape of Grantham and Wightman's (1978) curve for a broader band-width of carrier.

3.1 Experiment 2: Broadening the bandwidth of the carrier

The main difference between Grantham and Wightman's (1978) study and our Expt. 1 lies in the bandwidth of the stimuli, so our Expt. 2 measured dichotic/diotic FM discrimination thresholds for carriers with a slightly broader bandwidth (41 Hz), estimated to lie within a monaural critical bandwidth centred on 500 Hz, and one to extend beyond it (241 Hz). Thresholds for two subjects, as a function of modulation rate, are also shown in Fig. 1 (41 Hz: triangles, 241 Hz: squares).

At the slowest modulation rate (2 Hz), thresholds are lower for the noise waveforms than for the tonal carrier. As modulation rate is then increased to 10 Hz, thresholds increase and become higher than thresholds for tonal carriers. Thresholds drop slightly at 20 Hz, and a second peak was then observed around the 40-Hz modulation rate. For Subject 1, at all rates tested above 2 Hz, thresholds for

the 241-Hz carrier are higher than for the 41-Hz carrier. For Subject 2, the pattern is less clear, but thresholds for the band-limited carriers are higher than for tonal carriers at all rates beyond 2 Hz.

Broadening the bandwidth of the carrier therefore does affect the relationship of thresholds with modulation rate. Although the increase in thresholds is not as steep as that reported by Grantham and Wightman, the curves relating threshold to modulation rate do begin to take on a characteristic that is broadly similar to their classic data, even though our carriers had a much narrower band-width than theirs. The evidence for sluggishness in processing dynamic ITDs therefore seems to emerge as the bandwidth of the stimulus is increased.

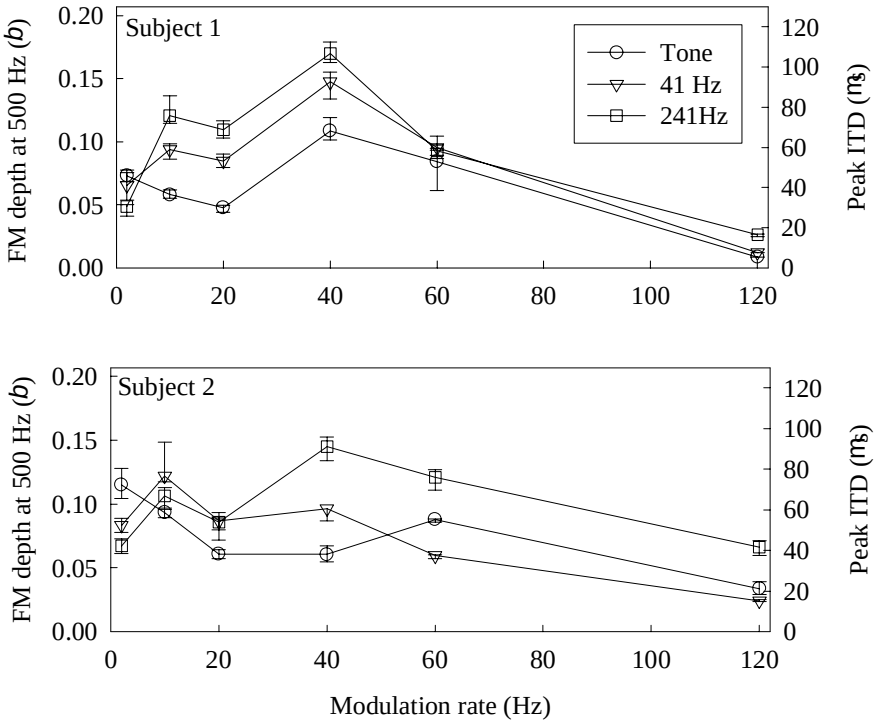


Fig. 1. Thresholds for discriminating dichotic from diotic FM for each of three carriers: a 500-Hz tone, a 41-Hz band-width and a 241-Hz band-width carrier, for two subjects. Thresholds are reported in terms of the FM depth at the 500 Hz centre frequency (β; left ordinate), and as peak ITD for the dichotic condition (μs; right ordinate). Error bars show the ± one-standard deviation confidence intervals of the threshold estimate.

For all carrier types, subjects reported a similar percept of dichotic modulation at each rate (other than the obvious changes in percept for both intervals associated with increasing the bandwidth). For 2-Hz modulation, subjects reported hearing smooth motion of the internalized sound source, back and forth between the ears.

For modulation rates between 10 and 60 Hz, subjects reported hearing a roughness that split in location and was lateralized to both ears. For the faster rates (60 - 120 Hz for Subject 1 and 40 - 120 Hz for Subject 2), subjects reported hearing a faint pitch in the dichotic intervals that was lateralized to one side of the head.

The results of this study indicate that the evidence for sluggishness in the binaural system depends on the type of stimuli used to measure it and what is asked of the subject. None of our subjects, for example, report hearing movement with modulation rates much above 2 Hz (Witton et al., 2000). Other evidence that the auditory system's use of temporal information (either monaural or binaural) is sluggish comes from studies of the mechanisms of FM detection. It has been argued that at modulation rates below about 10 Hz, listeners can use temporal information from phase-locking to detect FM, but above this rate they must rely on place cues, presumably because phase-locking in no longer sufficiently accurate (Moore and Sek 1996). We can clearly use temporal cues to discriminate dichotic from diotic FM at rates faster than 10 Hz, but this is not done on the basis of perceiving movement.

For modulations above about 10 Hz in the dichotic condition, subjects report hearing two images, one at each ear. Accurate representations of temporal information would be necessary for a distinct moving spatial image, but may not be necessary to discriminate dichotic from diotic stimuli; discriminating these modulations may not require the same level of phase-locking accuracy as does motion detection (or FM detection). Even with a degree of “aliasing” between temporal information and its neural representation, we may have enough information to know that the signal is dichotic, despite being unable to accurately estimate either motion or frequency change. Therefore it may be true to say that the auditory system is sluggish if it can only use temporal information to signal motion (or FM) below a 10-Hz modulation rate.

In this sense, auditory motion is like third-order motion sensitivity in vision (Lu and Sperling, 1995).

For dichotic/diotic discrimination, however, temporal information at much higher rates appears to be available. Above somewhere between 40 and 60 Hz (for our stimuli), dichotic modulation produces a clearly localised “distortion product”. Observers readily use this cue to make the discrimination. Between modulation rates of 10 and 40 Hz, subjects' performance depends on the bandwidth of the carrier.

For the tonal carrier, thresholds initially drop as modulation rates increase. This pattern of results is similar to those predicted by a model recently proposed by Breebaart, van de Par, and Kohlrausch (2001). They describe a computational model of binaural processing which can account for psychophysical performance on a wide range of binaural tasks but which does not predict the non-monotonic curve observed by Grantham and Wightman, instead predicting a gradual decrease in thresholds as modulation rate increases. However this prediction probably results from the model's high sensitivity to peak ITDs that occur near to the start of the stimulus, and cannot explain the increase in thresholds that was observed in the present study when bandwidth was increased.

One possible explanation for the effect of bandwidth on discrimination thresholds is the introduction of temporally modulated amplitude and phase

envelopes with the noise stimuli. Both modulations will result in intermittent cessation of the stimuli in random bands and moreover may well capture some of the phase-locking otherwise devoted to the characteristic we manipulate. In inferior colliculus, for example, even with very wide bandwidths, responses remain locked to characteristics of the envelope (Wake, Johnson, Green and Rees, 2003).

In conclusion, psychophysical evidence for binaural sluggishness seems to emerge as the bandwidth of the binaural stimulus is increased. Increasing the spectral content of the stimuli appears to reduce our ability to make use of binaural temporal information to discriminate dichotic from diotic FM between modulation rates of about 10 and 40 Hz.

References

Bernstein, L.R., Trahiotis, C., Akeroyd, M.A., and Hartung, K. (2001). Sensitivity to brief changes of interaural time and interaural intensity. J. Acoust. Soc. Am. 109, 1604-1615.

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001). Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. J. Acoust. Soc. Am. 110, 1105-1117

Chandler, D.W. and Grantham, D.W. (1992). Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth and velocity. J. Acoust. Soc. Am. 91, 1624-1636.

Davenport, W.B. and Root, W.L. (1958). Random signals and noise. McGraw- Hill, New York.

Grantham, D.W. (1982). Detectability of time-varying interaural correlation in narrow-band noise stimuli. J. Acoust. Soc. Am. 72, 1178-1184

Grantham, D.W. (1984). Discrimination of dynamic interaural intensity differences. J. Acoust. Soc. Am. 76, 71-76.

Grantham, D.W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J. Acoust. Soc. Am. 63, 511-523.

Grantham, D.W. and Wightman, F.L. (1978). Detectability of varying interaural temporal differences. J. Acoust. Soc. Am. 63, 511-523.

Grantham, D.W. and Wightman, F.L. (1979). Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. J. Acoust. Soc. Am. 65, 1509-1517

Harris, J.D. and Sargeant, R. (1971). Monaural/binaural minimum audible angle for a moving sound source. J. Speech Hear. Res. 14, 618-629.

Henning, G.B. and Zwicker, E. (1984). Binaural masking level differences with tonal maskers. Hear. Res. 16, 279-290.

Kollmeier, B. and Gilkey, R.H. (1990). Binaural forward and backward masking: Evidence for sluggishness in binaural detection. J. Acoust. Soc. Am. 87, 1709-1719.

Lu, Z.-L. and Sperling, G. (1995). Attention-generated apparent motion. Nature. 377, 237-239.

Moore, B. C. J. and Sek, A. (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. J. Acoust. Soc. Am. 100, 2320-2331.

Wake, G., Johnson, S., Green, G., and Rees, A. (2003). Encoding of second-order amplitude modulation in the inferior colliculus. Abstracts of the twenty-sixth annual mid winter research meeting of the Association for Research in Otolaryngology, February 2003.

Wichmann, F.A. and Hill, N.J.. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. Percept. Psychophys. 63, 1293-1313.

Wichmann, F.A. and Hill, N.J.. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. Percept. Psychophys. 63, 1314-1329.

Witton, C., Green, G.G.R., Rees, A., and Henning, G.B. (2000). Monaural and binaural detection of sinusoidal phase-modulation of a 500-Hz tone. J. Acoust. Soc. Am. 108, 1826-1833.

Witton, C., Simpson, M.I.G., Henning, G.B., Rees, A., and Green, G.G.R. (2003). Detection and direction-discrimination of diotic and dichotic ramp modulations in amplitude and phase. J. Acoust. Soc. Am. 113, 468-477.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Auditory thresholds re-visited

Peter Heil and Heinrich Neubauer

Leibniz Institute for Neurobiology Magdeburg, peter.heil@ifn-magdeburg.de

1 Introduction

A thorough understanding of any sensory system's operation requires knowledge of how it reaches threshold. Surprisingly, confusion exists as to how auditory thresholds are best defined and by which process they are reached. At the neuronal level, thresholds are routinely specified in terms of the sound pressure level (SPL) of the signal needed to evoke a response, exemplified by tuning curves, implying that thresholds are independent of stimulus duration. However, at the perceptual level, the threshold SPL decreases as stimulus duration increases in every species examined (Fay, 1993), a trade-off consistent with the idea that the auditory system integrates sound over time. The physical quantity commonly believed to be integrated by the system is sound intensity, *I(t)*, i.e., the sound power transmitted per unit area, which, for pure tones, is proportional to the square of the peak pressure or the pressure envelope, *P(t)*. Sound power integrated over time yields acoustic energy, so that the common interpretation of the perceptual data implies that the auditory system has a threshold that is best specified in terms of the sound´s acoustic energy density, and not its pressure. Threshold energy, in contrast to threshold SPL, generally increases with increasing stimulus duration, a finding which has often been attributed to leaky integration of *I(t)*, although the physiological foundations of this interpretation are unclear. Very long time constants, up to hundreds of milliseconds, are needed to describe the increase in threshold energy with increasing duration (for summaries see Gerken, Bhat and Hutchison-Clutter 1990; O´Connor, Barruel, Hajalilou and Sutter 1999). Such long time constants contrast with the short membrane time constants of neurons in the auditory periphery (e.g. Clock, Salvi, Saunders and Powers 1993) and are also difficult to reconcile with the high temporal resolution of the auditory system. Several ideas have been proposed to resolve this resolution-integration paradox. Among them are suggestions that the intensity integrator resides high up in the central auditory system (Zwislocki 1969; Watson and Gengel 1969; Gerken et al. 1990) or that the system only acts as if it were a long-term integrator, but instead takes "multiple looks" (Viemeister and Wakefield 1991).

We have previously shown that thresholds of cat auditory-nerve (AN) fibers are reached by temporal integration of *P(t)*, and not of *I(t)* as assumed for perceptual

thresholds (Heil and Neubauer 2001). Here we demonstrate, by re-analysis of published data, that perceptual thresholds are also much better explained by temporal integration of *P(t)* than of *I(t)* and that AN fiber and perceptual integration thresholds for *P(t)* vary with stimulus duration or integration time in similar ways. Thus, the first synapse in the auditory pathway constitutes the most likely location of the integrator and we discuss its mode of operation. A longer account of this topic is provided elsewhere (Heil and Neubauer 2003).

2 Perceptual thresholds are due to temporal integration of *P(t)*

To distinguish between integration of *P(t)* and of *I(t)* at the perceptual level, we first re-analyzed perceptual thresholds, measured by Gerken et al. (1990) and Solecki and Gerken (1990), for both humans and cats tested with single- and multiple-burst tones (of 3.125 kHz in humans and 6.25 kHz in cats) that differed in duration and envelope characteristics (Fig. 1a). We calculated grand means for all subjects from the individual mean threshold SPLs (11 humans, 4 tested twice; 5 cats) by minimizing the sum of the total variance across subjects without altering the mean SPL across all subjects and stimuli. The SPL corresponding to the peak amplitude of threshold stimuli decreased with increasing number of bursts (open circles, stimuli 1-5 in Fig. 1a), plateau duration (filled circles, stimuli 1,6-14), and onset–offset times for a single burst without plateau (plus signs; stimuli 1,15-18) (Fig. 1b). These trading relationships between threshold SPL and stimulus duration are all consistent with thresholds reached by temporal summation. Threshold SPLs did not differ for stimuli that were temporal mirror images (plus signs; stimuli 19-20) or for multiple-burst stimuli differing in interburst interval (crosses; stimuli 3, 21-24) (the five crosses representing the threshold SPLs for these stimuli in Fig. 1b fall on top of each other). The latter result is inconsistent with the leaky integration hypothesis, which would predict an increase in threshold SPL with increasing interburst intervals (Gerken et al. 1990). Threshold SPL for stimuli in the multiple-burst, single-burst and onset–offset series diverged as duration increased (Fig. 1b). Thus, threshold SPL is not an invariant function of stimulus duration.

We next calculated the energy densities of these threshold stimuli, i.e., the temporal integral of *I(t)* (in W s m^{-2}). Energy density increased as the number of bursts, plateau duration, and stimulus onset–offset times increased, again with substantial divergence of the thresholds between the series (Fig. 1c). This shows that threshold energy is also not an invariant function of stimulus duration, contrary to what would be expected from a temporal integrator of intensity.

In contrast, the thresholds for stimuli in the different series were very closely aligned when expressed as the temporal integral of *P(t)* (in Pa s) and plotted against stimulus duration (Fig. 1d), again, as for the calculation of energy densities, without any correction of onset–offset versus plateau duration. These results show clearly that integration of *P(t)* provides a much better fit to the data than integration of *I(t)*.

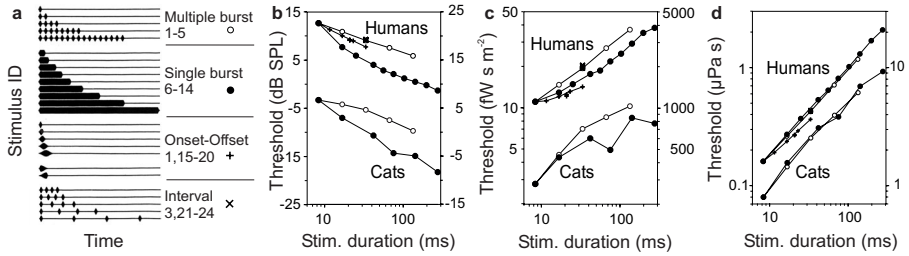


Fig. 1. Perceptual thresholds result from temporal integration of *P(t)* rather than of *I(t)*. **a**. Pressure envelopes of the stimuli for which detection thresholds were obtained by Gerken et al. (1990) and Solecki and Gerken (1990). Abscissa duration: 275 ms. **b-d**. Grand mean thresholds for humans (right ordinate) and cats (left ordinate) plotted against stimulus duration (excludes interburst intervals). Symbols as in **a**. Thresholds are expressed as SPLs (**b**), as temporal integrals of *I(t)* (i.e. as energy densities, calculated with a specific impedence of 414 Pa s m^{-1}) (**c**), and as temporal integrals of *P(t)* (**d**). Note the close alignment of thresholds at any given duration and the power law increase in **d**.

3 Pressure envelope integration thresholds increase with time according to a power law

Perceptual pressure envelope integration thresholds, $T(t_s)$, increase almost linearly with stimulus duration, t_s, on double-log scales for both cats and humans (Fig. 1d). Consequently, the power law:

$$T(t_s) = \int_0^{t_s} P(t)^1 \, dt = k \cdot {t_s}^m \tag{1}$$

provides an excellent descriptor of the data. It leaves only 0.84% unexplained variance for humans and 1.11% for cats, which is remarkably little for psychophysical data, particularly since the slope *m* and the scaling factor *k* are the only two free parameters. The unexplained variance was, on average, an order of magnitude smaller for each subject, as well as for the grand mean data, than that of the equivalent power law relating threshold energy and t_s. This also emphasizes that integration of *P(t)* provides a much better explanation for threshold than that of *I(t)*.

A re-analysis of available perceptual thresholds from other studies of temporal summation in various mammals, birds and fish (see legend Fig. 2) reveals that in each species the dependence of pressure envelope integration thresholds on stimulus duration can be well described by Eq 1 (Fig. 2a), indicating a common and conserved mechanism for temporal integration. Unexplained variances range from as low as 0.3% (chinchilla) to 3% (parakeet).

The scaling factor *k*, or a derived measure of sensitivity, varies with tone frequency and resembles an animal's audiogram, as shown by a re-analysis of data from the mouse (Ehret 1976) (Fig. 2c,d). In contrast, the power *m*, i.e. the slope of the increase in pressure envelope integration thresholds with duration in double-log

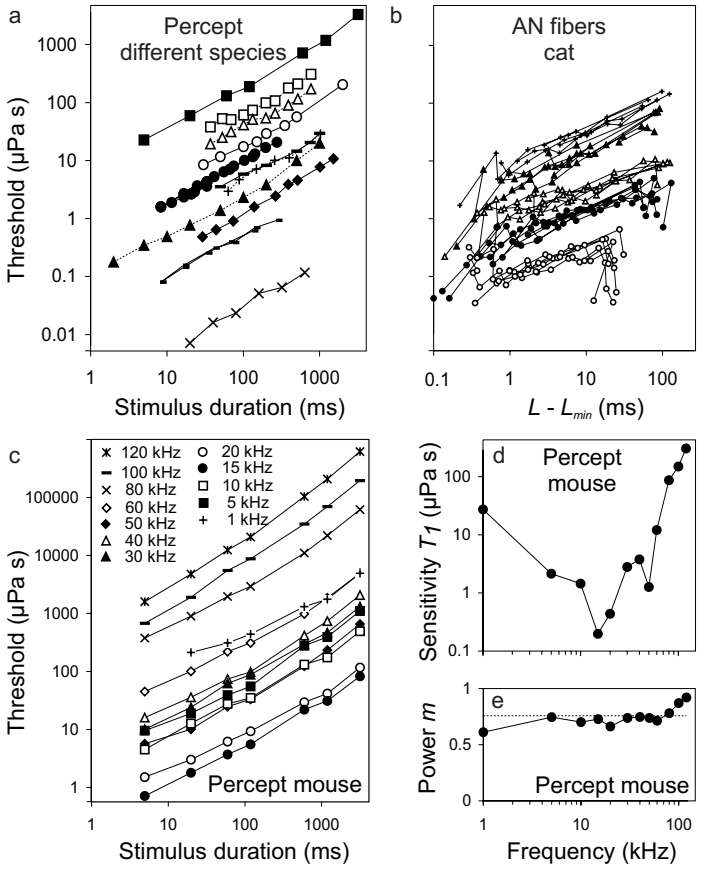


Fig. 2. Comparison of perceptual (**a**) and neuronal (**b**) pressure envelope integration thresholds. **a**. Pressure envelope integration thresholds (means across subjects and tone frequencies) plotted against stimulus duration for a range of vertebrate species, calculated from published data. For the porpoise (filled triangles; raw data from Johnson 1968), parakeet (open squares; Dooling 1979) and field sparrow (open triangles; Dooling 1979), absolute thresholds could not be determined from the original articles and thus the position on the ordinate is arbitrary. Other species shown are: mouse (filled squares; Ehret 1976), European starling (open circles; Klump and Maier 1990), human (filled circles; Gerken et al. 1990), mangabey (plus signs; Brown and Maloney 1986), cat (long dashes; Costalupes 1983; short dashes; Solecki and Gerken 1990), chinchilla (filled diamonds; Clark and Bohne 1986), goldfish (crosses; Fay and Coombs 1983). Thresholds in each species follow power laws with similar power, with no indication of a maximum integration time. **b**. Plot of thresholds against integration time, $L\text{-}L_{min}$, for 5 cat AN fibers of different CFs (different symbols). Lines connect thresholds derived from tones of different SPLs but with the same onset function and onset time. **c**. Perceptual pressure envelope integration thresholds of mice at different frequencies (raw data from Ehret 1976). **d-e**. Sensitivity ($T_1 = T(t_s)$ for $t_s = 1$ms) (**d**) and power m (**e**; horizontal line: mean m), derived from fits of Eq 1 to data in **c**, plotted against frequency.

scales, is remarkably independent of frequency (Fig. 2c,e). We obtained values of m from 0.67 to 0.79 in the different species and, after weighting them with the number of individuals, frequencies, and stimuli tested, a mean m of 0.75.

The adequacy of this power law may be obscured in the conventional plot of threshold SPL vs. log duration, particularly when only single burst stimuli with fixed onset and offset times are used. For such stimuli, threshold SPLs increase more rapidly the shorter the stimulus duration (see human data in Fig. 1a). This curved shape could suggest an exponential function with a long time constant, but actually results from the concomitant increase in the proportion of onset and offset times with respect to the entire duration of the tone.

The power law can also well describe the functions relating the pressure envelope integration thresholds to integration time of cat AI neurons (Heil and Neubauer 2003) and AN fibers (Fig. 2b), at least over the time range corresponding to the durations tested for perception in that species (cf. Fig. 2a). AN fiber thresholds were derived from the mean latency to the first spike, L, following the onsets of characteristic frequency tones of different SPLs, onset times and onset functions (Heil and Neubauer 2001). The analysis was based on the rationale that the first spike is triggered when the the stimulus reaches the fiber's threshold, and that the spike occurs with some transmission delay, L_{min}, thereafter. L_{min} is constant for a given fiber and tone frequency. Hence, the integration time is given by L-L_{min} which replaces t_s in Eq 1. In some AN fibers (e.g. open circles in Fig. 2b) there appear to be deviations from the power law at long integration times. They are due to spontaneous activity. This fiber-specific activity causes an upper limit for latency, producing roughly constant latencies to low-SPL stimuli, and results in a nearly vertical drop of threshold estimates for such stimuli.

4 The location and possible nature of the integrator

The functions relating the pressure envelope integration threshold to integration time for AN fibers and AI neurons or to stimulus duration for perception in the same species (cat) are very similar (cf. Fig. 2a-b). Thus, the most parsimonious conclusion is that the integrator is peripheral to the site of spike generation in the AN fibers. A more central origin, as proposed by others (Zwislocki 1969; Watson and Gengel 1969; Gerken et al. 1990; Viemeister and Wakefield 1991; Dau et al. 1996), seems unnecessary, although central processes could modify thresholds (e.g., Dai and Wright 1995). We can also conclude that the integrator must be central to the processes determining the inner hair cell's (IHC) membrane potential. This potential almost instantaneously follows the fine structure of the stimulus at low frequencies and the pressure envelope, $P(t)$, at higher frequencies, without changes of its DC component when $P(t)$ is constant (Kros 1996). These observations agree well with the short IHC membrane time constants. Thus, the integration of $P(t)$ over the observed long time scales cannot have been accomplished at, or peripheral to, the level of the receptor potential. Also, the compelling evidence that each IHC is innervated by 10-30 AN fibers of different spontaneous rates and thresholds (Liberman 1982), argues against an integrator identical for all afferent fibers of a given IHC (Heil and Neubauer 2001). This limits the possible location of the

integrator to the first synapse in the auditory pathway, between the IHC and the distal dendrite of a single AN fiber. This synaptic region is structurally diverse, which could readily account for the range of AN fiber sensitivities, as there are systematic relationships between morphological and physiological properties (Liberman 1982; Merchan-Perez and Liberman 1996).

Equation 1 can be reformulated to:

$$\overline{R} \cdot t_s = const. \tag{2}$$

Here $\overline{R} = c \cdot (\overline{P})^{\alpha}$, $\alpha = (1\text{-}m)^{-1}$, $c = const. \cdot (1/k)^{\alpha}$, and $\overline{P} = \left[\int_0^{t_s} P(t)\, dt \right] / t_s$ is the mean amplitude of $P(t)$ during t_s. $\overline{R}$ (in s^{-1}) can be interpreted as a mean rate of individual events, possibly point processes. Thus, the higher (lower) the mean rate of these events, the shorter (longer) the time needed to accumulate the number necessary for threshold. Such an accumulation process of individual events is not equivalent to the temporal integration of a continuously changing quantity. The probability of the occurrence of individual events, proportional to $\overline{P}^{\alpha}$, can be viewed as a conditional probability that results from the interaction of α sub-events, the probability of each of those occurring being proportional to $\overline{P}$. From the slopes m of the perceptual data values of α between 3 and 5 are derived. Thus, it seems meaningful to search for events, which are mediated by 3, 4 or 5 sub-events. One candidate for such events is exocytosis at the IHC-AN fiber synapse, for which, in mouse, 4-5 Ca^{2+}-binding steps are necessary (Beutner, Voets, Neher and Moser 2001). This supralinear dependence of exocytosis on the intracellular Ca^{2+}-concentration could constitute a molecular basis for the mechanism of temporal summation, likely conserved in evolution, because similar numbers (between 3 and 5) have been reported, or are inferred, for the giant synapse in the squid, the bipolar neuron in the goldfish retina, the neuromuscular junction in the frog, and the calyx of Held in the rat auditory brainstem (for review see Meinrenken, Borst and Sakmann 2003).

Acknowledgments

Many thanks to George M. Gerken and Dexter R.F. Irvine. Supported by grants of the Deutsche Forschungsgemeinschaft to P.H.

References

Beutner, D., Voets, T., Neher, E. & Moser, T. (2001) Calcium dependence of exocytosis and endocytosis at the cochlear inner ahir cell afferent synapse. Neuron 29, 681-690.

Brown, C.H. and Maloney, C.G. (1986) Temporal integration in two species of Old World monkeys: Blue monkeys (*Cercopithecus mitis*) and grey-cheeked mangabeys (*Cercocebus albigena).* J. Acoust. Soc. Am. 79**,** 1058-1064.

Clark, W.W. and Bohne, B.A. (1986) Cochlear damage. Audiometric correlates. In: M.J. Collins, T.J. Glattke and L.A. Harker (Eds.), *Sensorineural Hearing Loss*. University of Iowa, Iowa City, pp. 59-82.

Clock, A.E., Salvi, R.R., Saunders, S.S. and Powers, N.L. (1993) Neural correlates of temporal integration in the cochlear nucleus of the chinchilla. Hearing Res. 71, 37-50.

Costalupes, J.A. (1983) Temporal integration of pure tones in the cat. Hearing Res 9**,** 43-54.

Dai, H. and Wright, B.A. (1995) Detecting signals of unexpected and uncertain durations. J. Acoust. Soc. Am. 98, 708-896.

Dooling, R.J. (1976) Temporal summation of pure tones in birds. J. Acoust. Soc. Am. 65, 1058-1060.

Ehret, G. (1976) Temporal auditory summation for pure tones and white noise in the house mouse (*Mus musculus)*. J. Acoust. Soc. Am. 59, 1421-1427.

Fay, R.R. (1992) in *The Evolutionary Biology of Hearing*, eds. Webster, D.B., Fay, R.R. and Popper, A.N. (Springer, New York), pp. 229-263.

Fay, R.R. and Coombs, S. (1983) Neural mechanisms in sound detection and temporal summation. Hearing Res. 10**,** 69-92.

Gerken, G.M., Bhat, V.K.H. and Hutchison-Clutter, M. (1990) Auditory temporal integration and the power function model. J. Acoust. Soc. Am. 88, 767-778.

Heil, P. and Neubauer, H. (2001) Temporal integration of sound pressure determines thresholds of auditory-nerve fibers. J. Neurosci. 21**,** 7404-7415.

Heil, P. and Neubauer, H. (2003) A unifying basis of auditory thresholds based on temporal summation. Proc. Natl. Acad. Sci. USA (in press).

Johnson, C.S. (1968) Relation between absolute threshold and duration-of-tone pulses in the bottlenosed porpoise. J. Acoust. Soc. Am. 43**,** 757-763.

Klump, G.M. and Maier, E.H. (1990) Temporal summation in the European Starling (Sturnus vulgaris). J. Comp. Psychol. 104**,** 94-100.

Kros, C.J. (1996) Physiology of mammalian cochlear hair cells. In: P. Dallos, A.N. Popper and R.R. Fay (Eds.), *The Cochlea.* Springer, New York, pp. 318-385.

Liberman, M.C. (1982) Single-neuron labeling in the cat auditory nerve. Science 216**,** 1239-1241.

Meinrenken, C.J., Borst, J.G.G. and Sakmann, B. (2003) The Hodgkin-Huxley-Katz Prize Lecture: Local routes revisited: the space and time dependence of the Ca2+ signal for phasic transmitter release at the rat calyx of Held. J. Physiol. Jan 31; epub ahead of print.

Merchan-Perez, A. & Liberman, M.C. (1996) Ultrastructural differences among afferent synapses on cochlear hair cells: correlations with spontaneous discharge rate. J. Comp. Neurol. 371, 208-221.

O`Connor, K.N., Barruel, P., Hajalilou, R. and Sutter, M.L. (1999) Auditory temporal integration in the rhesus macaque (*Macaca mulatta*). J. Acoust. Soc. Am. 106, 954-965.

Solecki, J.M. and Gerken, G.M. (1990) Auditory temporal integration in the normal-hearing and hearing-impaired cat. J. Acoust. Soc. Am. 88, 779-785.

Viemeister, N.F. and Wakefield, G.H. (1991) Temporal integration and multiple looks. J. Acoust. Soc. Am. 90, 858-865.

Watson, C.S. and Gengel, R.W. (1969) Signal duration and signal frequency in relation to auditory sensitivity.J. Acoust. Soc. Am. 46, 989-997.

Zwislocki, J.J. (1960) Theory of temporal auditory summation. J. Acoust. Soc. Am. 32**,** 1046-1060.

Discrimination of temporal fine structure by birds and mammals

Marjorie Leek[1], Robert Dooling[2], Otto Gleich[3], and Micheal Dent[4]

[1] Walter Reed Army Medical Center, Marjorie.Leek@na.amedd.army.mil
[2] Department of Psychology, University of Maryland, dooling@psyc.umd.edu
[3] ENT Department, University of Regensburg, otto.gleich@klinik.uni-regensburg.de
[4] Department of Physiology, University of Wisconsin, dent@physiology.wisc.edu

1 Introduction

In a series of studies involving masking and discrimination of variants of Schroeder-phase waveforms, we have reported that birds not only demonstrate nearly identical masking by the positive- and negative-Schroeder-phase maskers, in contrast to humans, but that their ability to discriminate the fine structure in the positive- and negative-phase stimuli is maintained for very short fundamental periods. While humans require periods to be on the order of 3-4 ms, several bird species can make these discriminations over periods as short as 1-2 ms. We have suggested that the discrimination of fine structure over very short periods may reflect a generally enhanced capability in birds to process extremely precise temporal differences (Dooling, Leek, Gleich, and Dent 2002). In that earlier paper, we argued that the differences in compound action potential (CAP) responses in birds to positive and negative Schroeder complexes with different fundamental frequencies parallel the discriminability between positive and negative Schroeder complexes. Here, we extend those earlier studies by asking how the distribution of energy throughout a harmonic period may be discriminated. Such discriminations may be based either on changes in on-off temporal ratios within the waveform periods (duty cycle) or as rates of change of instantaneous frequency over time (rates of frequency sweeps).

2 Methods

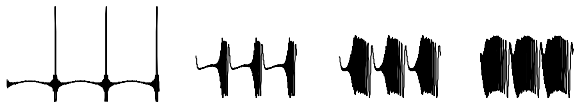
2.1 Stimuli

All stimuli were created digitally, at a sampling rate of 40 kHz, and stored as files for playback during the experiments. Harmonic complexes were generated with 49 equal-amplitude components with component frequencies from 200 to 5000 Hz. The fundamental frequency was 100 Hz. Stimulus duration was 260 ms, including

a 20-ms cosine-squared rise-fall time. Component starting phases were selected according to a modification of the algorithm given by Schroeder (1970):

$$\theta_n = C\pi n(n+1)/N \qquad (1)$$

where θ_n represents the phase of the n^{th} harmonic, N is the total number of harmonics, and C is a scalar. The original negative and positive Schroeder-phase waveforms used in previous studies of masking by harmonic complexes (e.g., Leek, Dent, and Dooling 2000) are generated by assigning C a value of –1 or +1, respectively, and assigning a value of 0 to the scalar produces a cosine-phase wave, with all component phases set to 0 degrees. Selections of scalars between 0 and ±1.0 produce periodic temporal waveforms with ever-decreasing silent intervals within each period. Fig. 1 shows samples of positive-Schroeder waveforms constructed with several scalar values. The negative-phase waveforms of the same scalar are identical, but reversed in time.



Scalar=0.0 Scalar=0.2 Scalar=0.8 Scalar=1.0

Fig. 1. Examples of waveforms created by varying C in Eq 1**.**

The instantaneous frequencies within these waveforms increase (negative scalars) or decrease (positive scalars) as harmonic frequency increases. As the low-energy portions of the stimuli increase in duration, this sweep in frequency occurs over shorter time sections of the period, producing increasing rates of frequency sweep. Smaller scalars produce faster frequency sweeps, and larger scalars produce slower frequency sweeps.

Temporal characteristics may be extracted from the scalar stimuli in order to translate the data into a more intuitive value of perception of the distribution of energy across periods. In order to assign a duty cycle to each stimulus, duration of the frequency sweep for each stimulus was assumed to be the "on-time" within the period. The ratio of the on-time to the full period (10 ms for these stimuli) was taken as the duty cycle, reflecting the distribution of energy in the waveforms across each period. Similar values could be extracted by calculating the envelope of each scaled waveform using a Hilbert transform.

2.2 Behavioral and physiological procedures

Three zebra finches (*Taeniopygia guttata*), two budgerigars (*Melopsittacus undulatus*), and three humans were tested on discrimination among selected waveforms. Birds were trained by operant conditioning and tested in a go/no-go task using a repeating background procedure and the method of constant stimuli (see Dooling and Okanoya 1995, for details of these procedures). Human listeners

were tested using the same procedures as the birds except that sounds were heard through earphones, and the "pecking" response was made by pushing buttons on a response box. On each block of 100 trials, waveforms with either 0-scalar or ±1.0-scalar were assigned as the background. The comparison stimuli on each block were intermediate scalar values, maintaining the sign of the scalars when the background scalar was ±1.0. All sounds were played through a loudspeaker in the free field at a sound pressure level that was randomly varied between 75 and 85 dB measured at the location of the bird's head. The level of each stimulus was roved over the 10 dB range in order to reduce possible loudness cues to the discriminations. On each block of 100 trials, only one of the three background sounds (0, +1.0 or -1.0 scalars) was presented and the other sounds were used as targets.

Cochlear microphonics (CM) and compound action potentials (CAP) in response to each of the scaled Schroeder-phase waves were recorded in three budgerigars, one zebra finch, and three gerbils using standard procedures described previously (Dooling *et al.* 2002). Responses to each scalar stimulus were averaged over 124 presentations, occurring at 2 per second. Following each set of presentations, responses to an inverted version of the stimulus were recorded. The CAP was extracted by adding the normal and inverted responses and scaling the sum by half, thereby cancelling the CM component. The CM was subsequently determined by subtracting the derived CAP from the recorded trace to the normal stimulus. The response was then averaged over the 10-ms fundamental period and response amplitudes were measured from these period histograms. Both the peak-to-peak amplitude for the CAP and the total root-mean-square amplitude (rms) for the CM were calculated for each stimulus.

3 Results

3.1 Behavioral measures of discrimination

Discrimination of the scaled stimuli from either a repeating background of the 0-scalar stimuli or ±1-scalars is shown in Fig. 2. (Note that although only one 0-scaled stimulus was used, for convenience the data are referred to as positive or negative 0, to indicate the signs of the comparison stimuli in each set). Data are averaged across birds (circles), showing the contrast between small birds and humans (triangles) on these discriminations. The duty cycles are shown on the abscissa (associated scalars are shown at the top) and the percent correct is plotted on the ordinate. Responses to positive and negative scaled waveforms are shown separately for each group (solid and open symbols, respectively). For both ends of the scalar continuum, birds can make finer discriminations than humans until the scalar values reach about 4-5 scalar units different from the background, when all subjects performed near 100% correct. The slopes of the psychometric functions are shallower for humans than for birds. Differences between the positive and negative scalar sets are small for the birds, but show a little more difference in the human data for a background scalar of |1|. For that end of the continuum, in

humans, the discrimination is better for negative scalars than for positive, for nearly all target scalars.

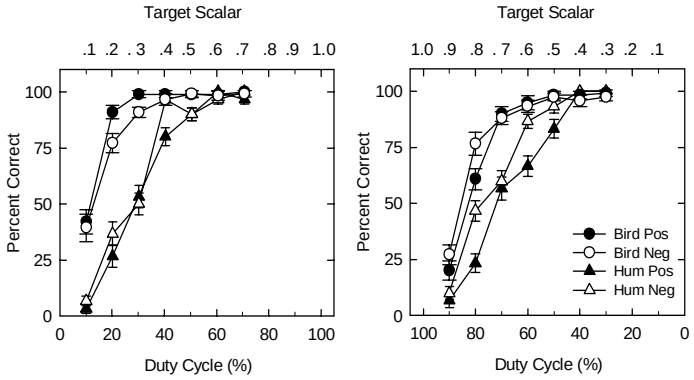


Fig. 2. Percent correct discriminations as a function of duty cycle. The scalar values are shown along the top. Left panel shows discrimination from a scalar of 0.0; right panel shows discrimination from a scalar of ±1.0

Thresholds (50% correct discriminations) are shown in Fig. 3 for each subject group and condition in terms of the duty cycle of the stimuli. Thresholds for comparisons to the ±1-scalars are subtracted from 100% for this display. Thresholds for all conditions for the birds fall in the range of duty cycles of 10-20%, while humans require a 30% - 40% duty cycle to support the discrimination. Separate ANOVAs for the two types of background indicate that there is a significant difference in discrimination performance between birds and humans ($F(1,7)=25.30$, $p<.005$ for 0-scalar background, $F(1,7)=7.09$, $p < .05$ for |1|-scalar background), but there were no significant differences between responses to positive and negative scalars, nor of the interaction between sign of the stimuli and species ($p > .05$ for all).

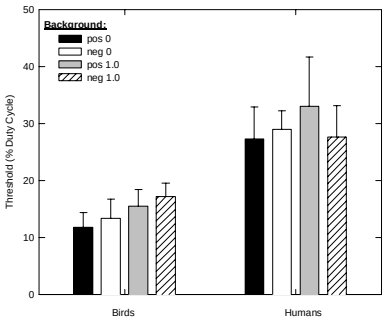


Fig. 3. Threshold duty cycles for discrimination

In order to detect that the energy in a waveform is not all contained within the one small interval corresponding to the peak of the cosine-phase wave, energy must be spread out over at least 3-4 ms for humans. For birds, the energy distribution needs only to be 1-2 ms. Further, about the

same difference in energy distribution is necessary to discriminate a target stimulus from a waveform with a very flat envelope (±1-scalars), with waveform energy across the entire period.

3.2 CM and CAP responses to scaled stimuli

As has been reported previously (e.g., Dooling *et al* 2002), the cochlear microphonic responses to these harmonic complexes follows the stimulus waveforms very precisely. Consistent with our previous data, the CM response in

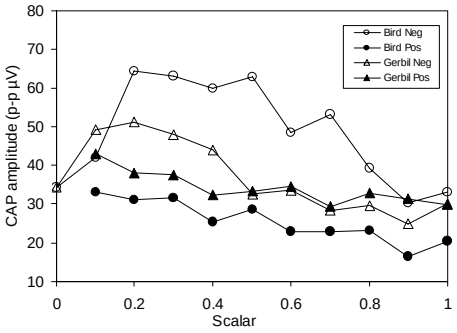


Fig. 4. Compound action potentials as a function of scalar

gerbils was much higher (40-60 µV) than in birds (5-10 µV). The CM rms amplitude showed only a small systematic increase from 0.1 to 1.0 scalars in gerbils and birds, and little difference in CM amplitude between the positive and negative scaled stimuli. The peak-to-peak amplitudes of the CAP responses to these stimuli are shown in Fig. 4. Compared to the CM, the overall amplitude of the CAPs are more similar between birds and gerbils, but there are systematic differences between the species in the pattern of responses across positive and negative scalars. The CAP amplitudes of the negative scalars in birds are considerably higher (40-70 uV) than the response to positive stimuli, which are quite low (20-30 uV). Gerbil responses are intermediate, with scalars less than |0.5| showing a larger response to the negative (40-50 uV) than to the positive (30-40 uV) stimuli. These data show a more highly synchronized neural response in birds to the negative scalar stimuli than to the positive, as well as greater synchronization in birds than in gerbils to the negative stimuli. The smaller differences between responses to positive and negative waveforms in gerbils is reminiscent of our earlier findings describing smaller positive/negative differences in Schroeder-phase stimuli with different fundamental frequencies in gerbils compared with birds. These physiological differences between birds and gerbils corresponded well with the enhanced behavioral abilities of birds compared with humans to discriminate among Schroeder-phase stimuli with increasing fundamental frequency (Dooling *et al*. 2002). In the present experiment on scalar discrimination the enhanced behavioral sensitivity of birds compared with humans remains. However, in

comparing the CM and the CAP responses of birds and gerbils, there is no clue either in the CM or the pattern of CAP amplitudes to these same stimuli that would provide insight into the physiological basis of this enhanced sensitivity to temporal fine structure in birds compared to mammals.

4 Discussion

The present research is an extension of earlier work showing that birds could discriminate between positive and negative Schroeder-phase stimuli at higher fundamental frequencies than humans and, correspondingly, that birds showed greater differences in CAP responses to positive and negative Schroeder-phase stimuli than did gerbils (Dooling *et al.* 2002). Here we approached the temporal resolution question from a different angle, asking whether there are species differences in sensitivity to temporal on/off characteristics of the stimuli or to the rates of change of instantaneous frequency over a given time period.

The behavioral discrimination experiment shows that birds require very little difference within the fine structure of the temporal waveforms of harmonic complexes in order to discriminate them. Recall that these stimuli are identical in their long-term spectra, and in other characteristics such as number of components and fundamental frequency. By scaling the phase selections, the only change within the waveforms involves the distribution of energy across the periods. With this converging discrimination task, we demonstrate again that whereas humans may require temporal discriminanda to differ by around 3 ms or more, birds can discriminate fine structure more on the order of 1-2 ms. We have argued earlier that the roots of these temporal discriminations must lie within the encoding provided by the hair cells or auditory nerves in the cochlea – an encoded discrimination of some sort that is preserved through the brainstem nuclei and the cortex in the ascending auditory pathway in order to produce the behavioral response. This argument was supported by our earlier study, in which there appeared to be a relationship between the bandwidths of the peripheral auditory system in mammals (humans) and the highest fundamental frequency that would support a discrimination between the positive and negative Schroeder-phase stimuli. The data from birds were somewhat puzzling, however, since a similar relationship did not emerge.

Here, the salient pattern found in the discrimination data, of better time resolution in birds compared with humans, is not obvious in the physiological data comparing birds and gerbils. Instead, the primary feature is that there are rather large differences in CAP amplitude between the negative and positive stimuli. Given that the magnitude of the CAP is determined by the degree of synchronization of neural firing of neurons in the auditory nerve, it would appear that these amplitudes reflect a greater degree of synchronization in the negative scaled stimuli than in the positive scaled stimuli, and a higher degree of synchronization for the scalars closer to the 0-scalar than the more dispersed |1|-scalars, as the amplitudes are usually higher for the lower-valued scalar stimuli. These observations are consistent with a synchronization of firing when the

stimulus frequencies appear within the periods from low to high (i.e., for negative-scalar stimuli), and when all frequencies occur over a very short duration, as for the lower-valued scalars. Dau, Wegner, and Kollmeier (2000) attempted to synchronize neural firing by optimizing a chirp stimulus for the human cochlea for use in auditory brainstem response measurement. They described their optimized chirp as gliding in frequency from low to high, as the negative scalars do here, and with a duration that is meant to reflect the traveling wave characteristics in the human ear. That duration would no doubt be different in the bird ears measured here, and probably also in gerbils, because the size of the cochlear structures among all these species differs considerably. Nonetheless, the traveling wave moves from high to low frequencies in all these species, and therefore, to compensate for the travel time, certainly a negative scalar would be necessary. However, other parameters of the traveling wave that vary among vertebrate groups (e.g., velocity) will also affect the response.

5 Conclusions

The present study confirms that birds show an enhanced ability to discriminate temporal fine structure in harmonic complex stimuli when compared with human abilities. The physiological basis of this enhanced performance is not seen in the CM or the CAP of either birds or gerbils (our human surrogate). Whatever the mechanisms underlying these species differences in temporal resolving power, they are not well-represented in these global electrophysiological measures of waveform following (CM) or of neural synchronization (amplitude of CAP).

Acknowledgments

This work was supported by NIH Grants DC-00198 to RJD and DC-00626 to MRL. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

References

Dau, T., Wegner, O., Mellert, V., Kollmeier, B. (2002) Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion. J. Acoust. Soc. Am. 107, 1530-1540.

Dooling, R. J., Leek, M.R., Gleich, O., and Dent, M.L. (2002) Auditory temporal resolution in birds: Discrimination of harmonic complexes. J. Acoust. Soc. Am. 112, 748-759.

Dooling, R. J. and Okanoya, K. (1995) The Method of Constant Stimuli in testing auditory sensitivity in small birds. In G. M. Klump, R. J. Dooling, R. R. Fay and W. C. Stebbins (Eds), *Methods in Comparative Psychoacoustics* Birkhaeuser Verlag, Basel, pp.161-169.

Leek, M.R., Dent, M.L. and Dooling, R.J. (2000) Masking by harmonic complexes in budgerigars (*Melopsittacus undulatus*). J. Acoust. Soc. Am. 107, 1737-1744.

Schroeder, M. R. (1970) Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. IEEE Trans. Inf. Theory IT-16, 85-89.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Dependence of binaural and cochlear "best delays" on characteristic frequency

Philip X. Joris, Marcel van der Heijden, Dries Louage, Bram Van de Sande, and Cindy Van Kerckhoven

Laboratory of Auditory Neurophysiology, K.U.Leuven, Philip.Joris@med.kuleuven.ac.be

1 Introduction

Since its inception (Jeffress 1948), Jeffress' place model of sound localization has dominated binaural thinking but has recently come under increasing criticism. More specifically, the existence and relevance of neural delay lines has been questioned. One of the key observations is the lack of delays, as measured physiologically in the inferior colliculus (IC), at high CFs, and the presence of large delays, far beyond the physiological range, at low CFs (McAlpine, Jiang and Palmer 1996; van der Heijden and Trahiotis 1999; McAlpine, Jiang and Palmer 2001). If delays are created by neural delay lines, these findings would imply longer delay lines at low CFs than at high CFs. Alternatively, the delays may reflect other processes, and over the years several alternatives have been formulated: cochlear (Schroeder 1977; Shamma 1989; Bonham and Lewis 1999), dendritic (Brew 1998), and inhibitory (Brand, Behrend, Marquardt, McAlpine and Grothe 2002). The multitude of alternatives reflects the fact that extremely small interaural time delays (ITDs) can be detected behaviorally and that many processes intervene between acoustic stimulation and binaural interaction, which have time courses that are slow relative to these small ITDs.

Although cochlear delays are one of the earliest alternatives proposed (Schroeder 1977), they have received little physiological attention. The only physiological test to date (Peña, Viete, Funabiki, Saberi and Konishi 2001), in binaural neurons of nc. laminaris in the barn owl, concluded that there was no need to invoke non-axonal delays. Schroeder's suggestion was quantitatively tested with cochlear models, first by Shamma (1989), who coined the term "stereausis" in analogy with stereopsis, and subsequently by Bonham and Lewis (1999). The modelling suggested that small CF mismatches should have significant effects on binaural ITD-tuning.

We reexamine this issue by cross-correlating responses obtained from cat auditory nerve (AN) fibers to broadband noise. Our general strategy is to obtain responses of monaural (AN) and binaural (IC) neurons to the same stimuli, then process the monaural spike trains through the simplest conceivable coincidence

detector, and compare this output with the binaural responses. In this paper, we focus on only one feature: the delay at which a peak is obtained in AN cross-correlograms and in IC noise-delay functions, i.e. the response of binaural cells to broadband noise with varying interaural time difference (ITD).

2 Methods

Recordings were obtained from the auditory nerve of cats, using standard methods. Spikes were timed with 1 μs resolution with a peak-picker (Carney Accutrig). For each fiber, we obtained: 1) a threshold tuning curve, 2) the response to many repetitions of a standard pseudorandom noise token, henceforth referred to as A+ (bandwidth 0.1-8 kHz; duration 1s, interstimulus interval 200 ms), and 3) the response to the polarity-inverted version of this noise token, henceforth referred to as A- (same stimulus parameters). The stimulus level was 70 dB SPL, but if time allowed the sequence was repeated at several levels.

3 Analysis

To quantify temporal structure, correlograms were constructed by comparing spike times (Joris 2001). The general scheme is illustrated in Fig. 1. Two sets of responses are available, consisting of multiple responses to conditions x and y. As explained below, these conditions can differ in terms of the stimuli used (e.g. A+ vs. A-), or in being obtained from different fibers (cell x vs. cell y). We refer to correlograms within a fiber as autocorrelograms (AC) and to correlograms across fibers as crosscorrelograms (CC).

Within fibers, two kinds of autocorrelogram were calculated. First, a shuffled autocorrelogram (SAC) was computed for the response to a standard noise token A+ by tallying all intervals between a spike in repetition n and all other spikes in all other repetitions, and repeating this for all spikes. Similarly, a SAC was also calculated for the response to the inverted noise token (A-). The two SACs (to A+ and A-) were then averaged. Second, a cross-stimulus correlogram (XAC) was computed by tallying all intervals between spikes in response to A+ with all spikes in response to A-. To study the temporal relationships *between* fibers, similar correlograms were constructed: a same-stimulus CC (SCC), and a cross-stimulus CC (XCC).

These correlograms are identical to the output of a binaural coincidence detector which would compare the timings of two nerve fibers over a narrow temporal window (equal to the binning window of the correlograms, 50 μs). Thus, besides being a handy analytical tool to study temporal information in response to broadband stimuli, these correlograms are also a natural display to compare peripheral, monaural responses with binaural responses that are thought to result from a coincidence process. Here, we view the cross-correlograms as predictions of binaural responses that would be obtained if coincidence detectors received inputs from the two ears that differ in CF.

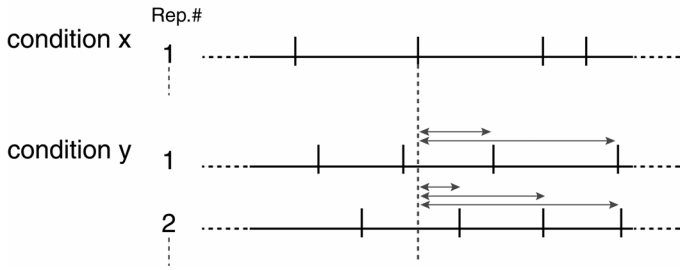


Fig. 1. Construction of correlograms from responses to multiple repetitions (Rep.#) of two conditions. In the simplest case, conditions x and y are identical (shuffled autocorrelogram, SAC; same fiber and same stimulus). The conditions can differ in the stimulus used (cross-stimulus autocorrelogram, XAC; e.g. noise token A+ in condition x and noise token A- in condition y), or in the fiber from which the responses are obtained (same-stimulus cross-correlogram, SCC; same stimulus presented to cell x and cell y), or both (cross-stimulus cross-correlogram, XCC; e.g. response to A+ from fiber x, response to A- from fiber y). In all cases, spike times are compared by tallying all intervals between all spikes in condition x and all spikes in condition y.

For a proper comparison between monaural and binaural data, we have to assign a "CF" to the cross-correlograms. Rather than using an average of the two original CFs of the fibers from which the CCs were calculated, we used the periodicity of the CC pattern itself. Subtraction of responses to the same stimuli (SAC or SCC) from the responses to anticorrelated stimuli (XAC or XCC) removes components that are common in the responses to correlated and anti-correlated stimuli, so that fine-structure is better revealed (Joris, 2001). We refer to this difference correlogram as the DIFCOR, and to the peak in the Fourier spectrum of the DIFCOR as the dominant frequency (DF). The DF is highly correlated to CF, but it is a more stable metric because it is based on a suprathreshold response and takes all spikes into account. Moreover, it allows us to compare binaural, monaural within-fiber, and monaural across-fiber data. The DF was used to estimate the place of cochlear innervation by using an empirical cochlear distance formula (Greenwood 1990).

4 Results

Figure 2A shows correlograms for fibers tuned at about 1 kHz. The thick line is the SAC for a 964 Hz fiber. SCCs between this fiber (the reference fiber) and 6 comparison fibers of neighbouring DF are shown as thin lines. The correlograms are positioned according to their DF: the intercept of the correlogram with the y-axis indicates estimated cochlear position (left ordinate) and difference in octaves between the DF of the reference fiber and that of the comparison fiber (right ordinate). The dots indicate the position of the maximum in the correlogram. Positive delays indicate a longer delay for the fiber tuned to the lower DF. Thus, the observed shifts in CC maxima are in the expected direction: increasing delay for

fibers tuned to progressively lower DFs. Within the range shown, there is little decrease in the height of the maximum, and the displacement of the peak remains smaller than 400 μs.

Figure 3 shows a similar graph for fibers from the same animal tuned to lower DFs. Although the pattern of shift in the crosscorrelograms is similar to that of Fig. 2, the shifts are larger. Differences in tuning as small as 0.1 octave result in delays that are large relative to the physiological ITD range. Interestingly, in addition to the shifts in the expected direction, there are shifts in the opposite direction at large ΔDFs.

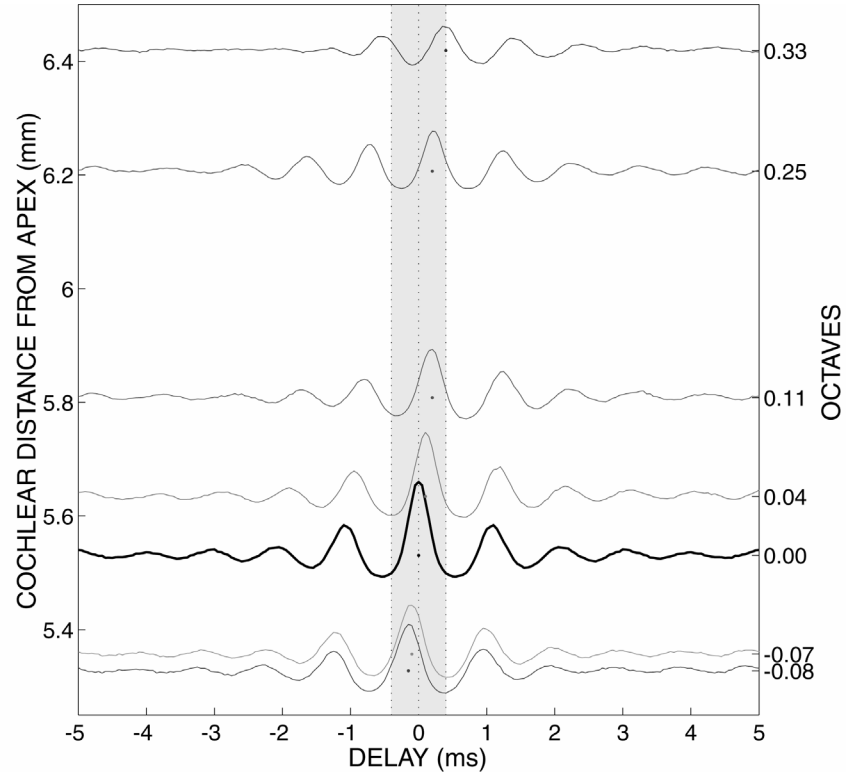


Fig. 2. Shift in cross-correlation patterns due to CF differences. For each SAC (heavy line) and SCC (thin lines), the asymptotic DC value was subtracted and the correlogram positioned according to its dominant frequency (DF). Fibers had DFs between 913 and 1213 Hz. The reference fiber (heavy line) had a DF of 964 Hz. The shaded central band indicates the approximate physiological range of interaural time differences (ITDs) for the cat (± 400 μs)

Values of peak delays as a function of DF from 4 animals are compiled in Fig. 4B. Pairs of fibers were only included if their DFs were within 0.3 oct, and if the peak of their SAC was larger than 0.5 spikes/sec. Different symbols indicate groups of different ΔDFs. At low DFs the range of peak delays is larger than at high DFs.

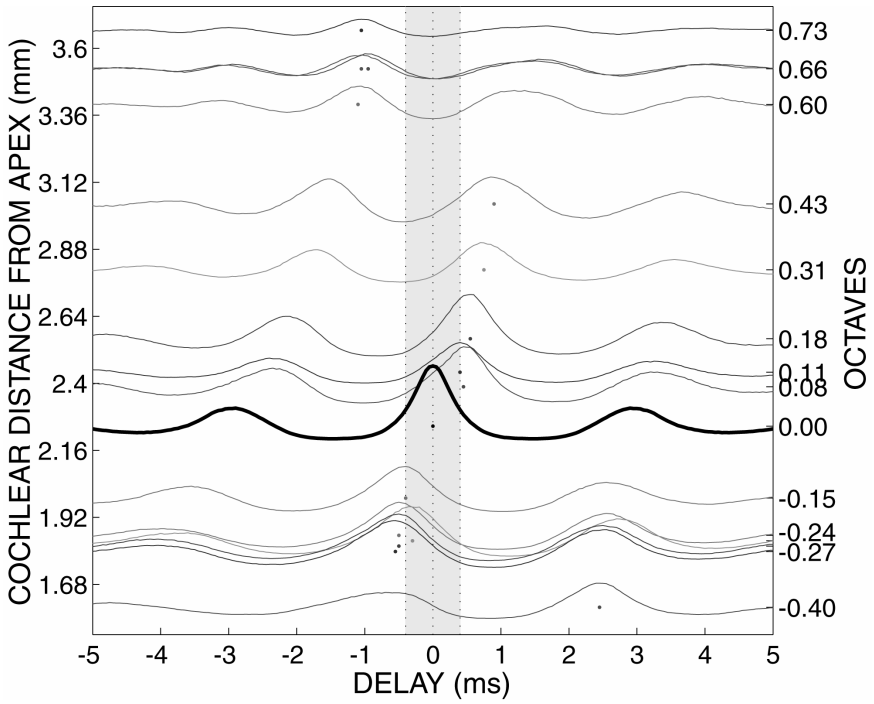


Fig. 3. Example cross-correlation patterns for fibers tuned to lower frequencies. DF of the reference fiber is 339 Hz. Range of DFs is between 256 and 562 Hz.

For comparison, best-delay IC data are shown in Fig. 4A. Noise-delay functions to correlated and anti-correlated broadband noise were available for ~ 200 IC cells (Joris 2001), which were further analysed with the procedure also used for the AN responses (calculation of DIFCOR from which DF and best delay were obtained). The IC distribution shows a wide range of best delays at low DFs, and a small range at high DFs.

5 Discussion

We developed an analysis to make predictions of the patterns of ITD-sensitivity that would be expected from an elementary coincidence detector operating on auditory-nerve inputs. Our main finding is that cross-correlograms of AN fibers with mismatched CFs show features that are similar to noise-delay functions of binaural neurons in the inferior colliculus. Here we focussed on only one datapoint of these functions and show that the delay at which auditory nerve fibers of differing CF are maximally correlated varies systematically with ΔCF. Moreover, the dependence of this delay on CF is very similar to the dependence of best ITD on CF in the IC. This

similarity suggests a view of binaural processing that differs from current models, namely that asymmetries in innervation between ipsi- and contralateral inputs have a dominant role in the creation of binaural delays.

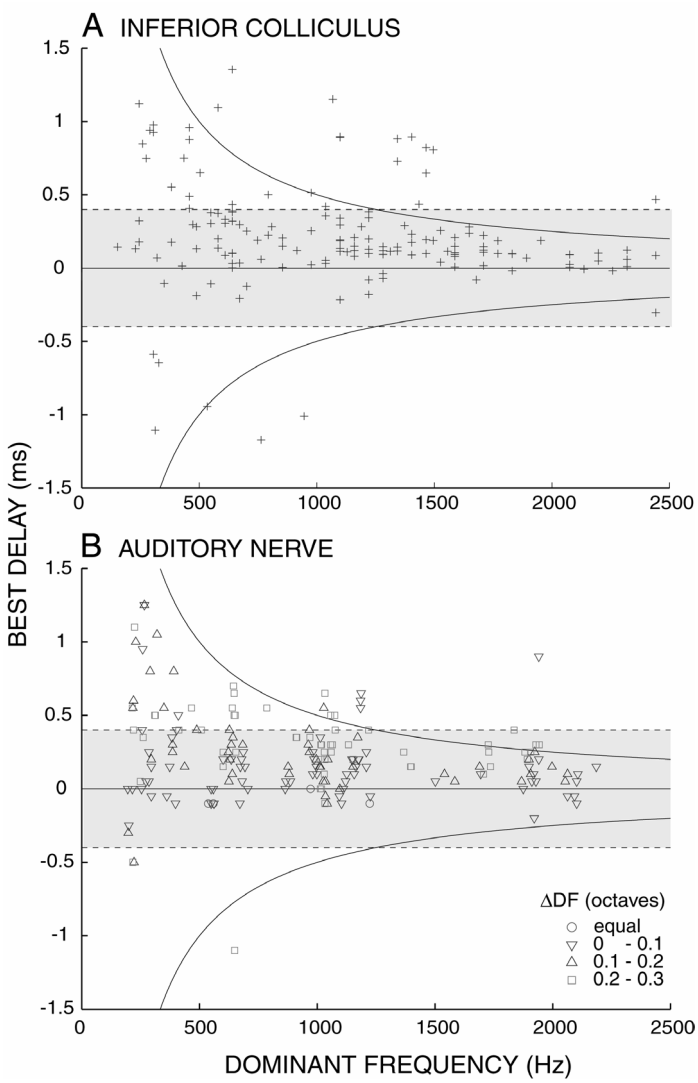


Fig. 4. Comparison of best delays in IC (A) and AN (B) (see text). The hyperbolic lines indicate the width of one DF period; the dashed lines and shading indicate the approximate width of the physiological range of ITDs for the cat.

Attention has repeatedly been drawn to the presence of delays outside of the physiological range in small mammals (McAlpine et al. 2001) and in humans (van der Heijden and Trahiotis 1999). We speculate that a broad range of best ITDs can

not be avoided at low CFs, because cochlear phase delays at these frequencies are large. Exact matching of monaural afferents to the binaural coincidence detectors may be unattainable in terms of precision in wiring, or may be undesirable because it restrains the strategies available to the central processor (cf. Carney, Heinz, Evilsizer, Gilkey and Colburn 2002).

Acknowledgements

Supported by the Fund for Scientific Research - Flanders (G.0083.02) and Research Fund K.U.Leuven (OT/10/42). We thank Laurel Carney for the Accutrig schematic.

References

Bonham, B. H. and Lewis, E.R. (1999) Localization by interaural time difference (ITD): effects of interaural frequency mismatch. J. Acoust. Soc. Am. 106, 281-290.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D. and Grothe, B. (2002) Precise inhibition is essential for microsecond interaural time difference coding. Nature 417, 543-547.

Brew, H. M. Modeling of interaural time difference detection by neurons of mammalian superior olivary nucleus. Association for Research in Otolaryngology Abstracts 21, 680. 1998.

Carney, L.H., Heinz, M.G., Evilsizer, M.E., Gilkey, R.H. and Colburn, H.S. (2002) Auditory phase opponency: a temporal model for masked detection at low frequencies. Acta Acustica united with Acustica 88, 334-346.

Greenwood, D. D. (1990) A cochlear frequency-position function for several species 29 years later. J. Acoust. Soc. Am. 87, 2592-2605.

Jeffress, L. A. (1948) A Place Theory of Sound Localization. J. Comp. Physiol. Psychol. 41, 35-39.

Joris, P. X. (2001) Sensitivity of inferior colliculus neurons to interaural time differences of broadband signals: comparison with auditory nerve firing patterns. In: D.J. Breebaart, A.J.M. Houtsma, V.F. Prijs, and R. Schoonhoven (Eds.) *Physiological and Psychophysical Bases of Auditory Function*, Shaker Publishing BV, Maastricht, pp. 177-183.

McAlpine, D., Jiang, D. and Palmer, A. (1996) Interaural delay sensitivity and the classification of low best-frequency binaural responses in the inferior colliculus of the guinea pig. Hear. Res. 97, 136-152.

McAlpine, D., Jiang, D. and Palmer, A. (2001) A neural code for low-frequency sound localization in mammals. Nature Neuroscience 4, 396-401.

Peña, J. L., Viete, S., Funabiki, K., Saberi, K. and Konishi, M. (2001) Cochlear and neural delays for coincidence detection in owls. J. Neurosc. 21, 9455-9459.

Schroeder, M. R. (1977) New viewpoints in binaural interactions. In: E.F. Evans, J.P. Wilson (Eds.) *Psychophysics and Physiology of Hearing*, Academic Press, New York, pp. 455-467.

Shamma, S. A. (1989) Stereausis: binaural processing without neural delays. J. Acoust. Soc. Am. 86, 989-1006.

van der Heijden, M. and Trahiotis, C. (1999) Masking with interaurally delayed stimuli: the use of "internal" delays in binaural detection. J. Acoust. Soc. Am. 105, 388-399.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

The enigma of cortical responses: Slow yet precise

Mounya Elhilali, David J. Klein, Jonathan B. Fritz, Jonathan Z. Simon, and Shihab A. Shamma

Institute for Systems Research & Department of Electrical and Computer Engineering, University of Maryland, College Park MD
{mounya,djklein,ripple,jzsimon,sas}@isr.umd.edu

1 Introduction

There is a fundamental paradox lurking in the characterization of cortical response dynamics. On the one hand, it has long been accepted that cortical cells are sluggish and fail to follow sustained repetitive stimuli at rates much beyond 20 Hz (Kowalski, Depireux, and Shamma 1996; Miller, Escabí, Read, and Schreiner 2002). On the other hand, numerous studies have demonstrated a remarkable temporal precision of spike occurrences that are locked to stimulus onsets and other transients, and have considered it functionally significant (Bair and Koch 1996; Heil 1997).

These two phenomena have generally been studied separately using different stimuli that tend to highlight one phenomenon or the other; e.g., AM tones and noise, ripples, and click trains versus tone onsets and dynamic dots (Bair *et al.* 1996; Heil 1997). It is, however, possible to demonstrate the *coexistence* of these two response properties, and explore their limits and characteristics with stimuli that combine *both* repetitive and transient aspects. In this report, we describe how ripples (a broadband *frozen* noise or a harmonic series with various spectro-temporally modulated envelopes) can be used to elicit responses phase-locked both to the modulation envelopes and to the "texture" of the carrier. By independently manipulating these two aspects of the stimulus, it is possible to explore (1) the dependence of the precise firings on the nature of the stimulus, (2) the mechanisms that may give rise to these finely-structured responses, and (3) their functional significance.

2 Methods

Data were collected from extra-cellular cortical recordings in a total of 8 domestic ferrets (*Mostela putorius*). Five were in awake state, and the remainder were ketamine anesthetized. Details of the surgery are as in (Kowalski *et al.* 1996).

Stimuli included various combinations of moving ripples that last 3 seconds. Ripples are broadband complex sounds with periodically modulated spectral envelopes, explained in more detail in (Kowalski *et al.* 1996). We used specific combinations of ripples referred to as TORCs (Temporally-Orthogonal Ripple Combinations) to characterize the Spectro-Temporal Receptive Fields (STRFs) of cortical neurons. A TORC typically consists of 501 random-phase tones equally-spaced along the tonotopic axis, and spanning a range of 5 octaves. These tones form an instance of *frozen* noise, whose envelope is modulated by 30 different spectro-temporal waveforms with rates up to 24 Hz and spectral densities up to 1.4 cycle/octave, as described in detail in (Klein, Depireux, Simon, and Shamma 2000).

Hence, ripple stimuli have two distinct aspects, which are better described by the cochlear-like spectrogram shown in Fig. 1: (1) A prescribed spectro-temporal envelope (top trace in right panel) which allows us to estimate rapidly and accurately the STRF (using a reverse correlation technique (Klein *et al.* 2000)); and (2) A *Fine-structure* that carries the envelope (thick black trace in right middle panel). It is created by the interaction between the tones and can be extracted by a Hilbert transform. The fine-structure waveform depends solely on the carrier tone characteristics (frequencies, phases and amplitudes), and is limited in bandwidth to that of the cochlear filter, becoming progressively broader at higher filter frequencies. An additional variant of the TORC stimulus was used. It consisted of harmonic-TORCs whose spectro-temporal envelopes were carried by sets harmonically-spaced tones with fundamental frequencies between 25 and 200 Hz. In all other aspects, the harmonic- and regular-TORCs shared identical spectro-temporal envelope parameters.

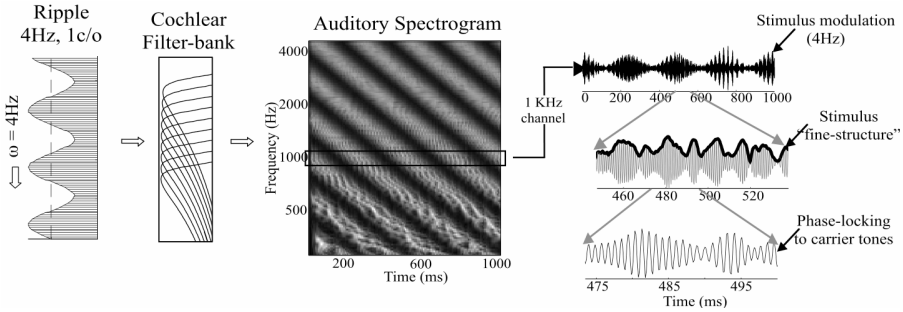


Fig. 1. Schematic of stimulus envelope and fine structure. **Left:** A ripple stimulus (4Hz and 1 c/o) is given as input to a filter-bank. **Middle:** The filters outputs show an overall pattern of a 4Hz drifting spectrogram. **Right:** The output of the 1KHz channel reveals the 4Hz envelope modulating a more dense fine structure carrier.

Neural responses to a series of TORCs are shown in the rasters of Figure 2. To quantify the precision of spiking, we computed the average cross-correlation of spike trains of different stimulus presentations (Fig. 2(B)), and then fitted it to a model of Poisson point-process cross-correlations that includes parameters to account for timing-jitter and spike deletion (Fig. 3(A)). Specifically, we assumed a Gaussian spread of the correlation peak whose variance σ represents the timing

jitter and scale α ($0<\alpha<1$) represents spike deletion between one trial and another. Combining these two parameters together with the rate of the process λ, we obtain:

$$R(\tau) = \lambda^2 + \frac{\alpha\lambda}{\sigma\sqrt{2\pi}} e^{-\tau^2/\sigma^2} \quad (1)$$

Finally, we employed the reverse correlation technique (Klein *et al.* 2000) to measure: (1) the usual STRF of the unit with respect to the spectro-temporal envelopes of the TORCs, (2) the $STRF^C$ with respect to the *complete* cochlear filter-bank output (i.e., including both the envelopes and fine-structure); and (3) the $STRF^F$ with respect to the fine-structure only (all shown Fig. 2(C)). In the last case, we averaged over all cochlear filter-bank outputs of the TORC stimuli to null out the spectro-temporal envelope and preserved only the fine-structure of the spectrogram to construct $STRF^F$.

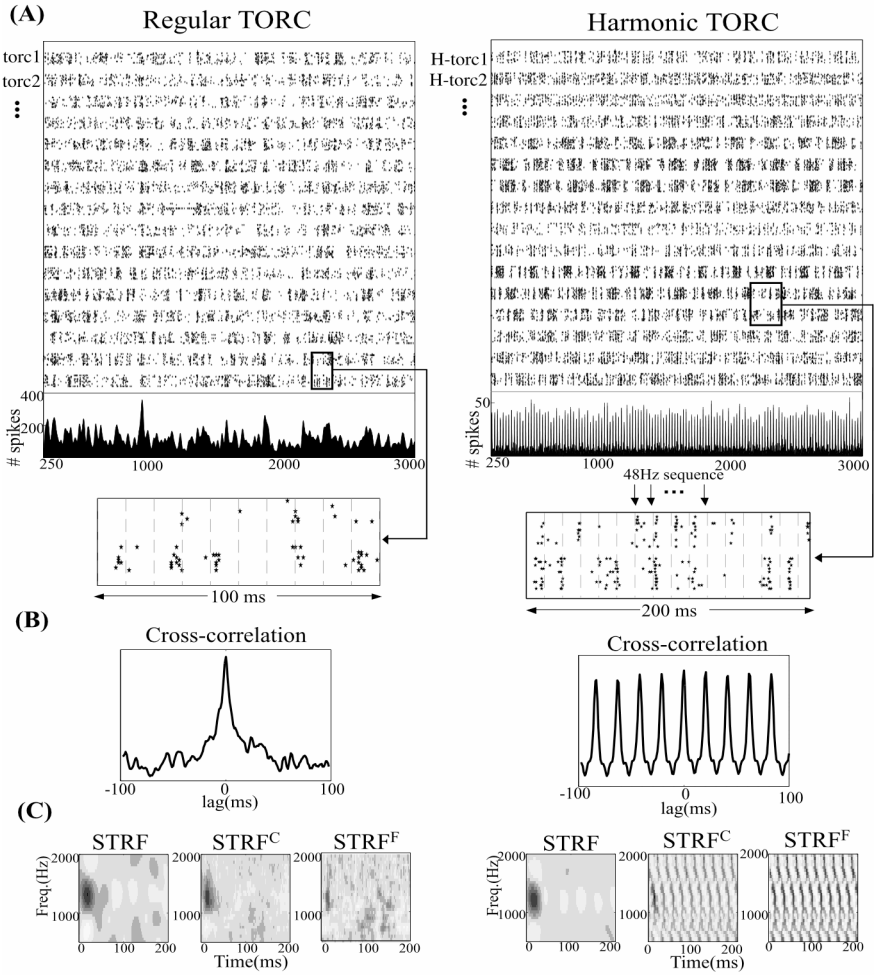


Fig. 2. Data analysis using regular (left panels), and harmonic TORCs (right panels).

3 Results

Data analyzed here were based on a total of 680 units (50% from awake ferrets). Figure 2(A) illustrates the nature of the precise spiking observed in the raster of TORC and harmonic-TORC responses. Specifically, this unit phase-locks to the fine-structure common to all the TORC stimuli with spikes that appear vertically aligned. To highlight this property, responses to all TORCs are collapsed to generate the PST histograms shown below each raster. Both histograms display strong and precise firing episodes (peaks) at numerous instants throughout the extended duration of the stimuli (3 seconds). In the case of the 48 Hz harmonic-TORC, the peaks occur regularly, reflecting the periodicity of the fine-structure.

To assess the degree of precision in the phase-locked responses, we computed the averaged cross-correlation among all TORC responses as defined by Eq 1 above. The resultant correlation functions shown in Fig. 2(B) display a sharp peak at zero lag (width of approximately 2-3 milliseconds) due to the high precision of firing from one response to another. The correlation function of the harmonic-TORC responses is periodic (right panel) demonstrating the precise phase-locking to the 48 Hz periodicity of the fine-structure. Using the Poisson model of unit responses, we estimated the distribution of σ, α, and λ for all units as shown in Fig. 3(B). Over 50% of all units exhibit relatively precise locking to the fine structure with σ<10ms. The skewness of the α distribution towards 0 suggests that spike deletion is a common property among most responses, partly because the different TORC envelopes are uncorrelated and hence spikes are suppressed differently from one TORC response to another (see discussion later). Finally, we have found that precision of spiking (σ distribution) is very similar in both the awake and the anesthetized ferret.

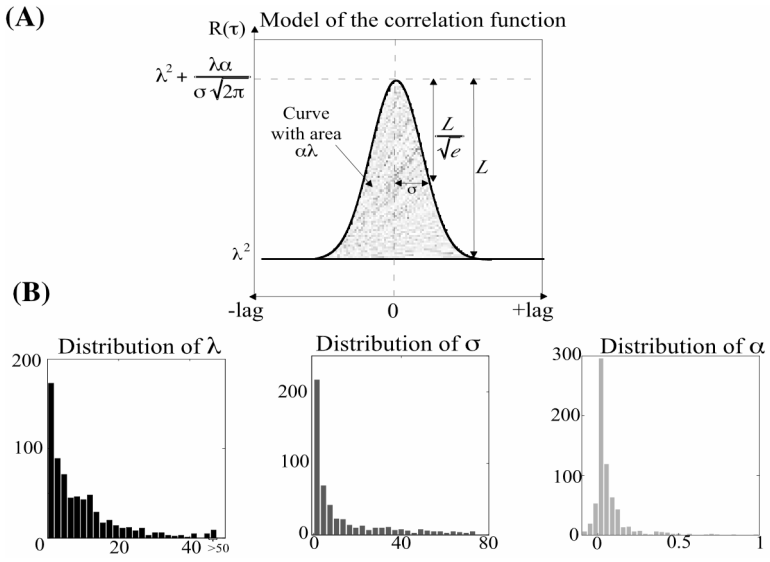


Fig. 3. (A) Model of spike correlation. **(B)** Population statistics for entire data set (N=680).

The precision and speed of phase-locking is most readily seen in the periodic fine-structure of the responses to harmonic-TORCs. We have observed phase-locked responses over a wide range of fundamental frequencies up to 200Hz, and further testing could shed more light on the upper limit of this locking, and the possible tuning of neuronal responses to different fundamentals.

The fine-structure of cortical responses at a given BF is strongly related to the "Hilbert" envelope of the cochlear filter responses near the same CF. To demonstrate this relationship, we computed the correlation between the PST histogram of each TORC response and the averaged Hilbert envelopes of the cochlear filter-bank responses to the TORCs. The results are shown in the panels labeled $STRF^F$ in Fig. 2(C). In the case of regular TORCs, the correlation maximum occurs at the cochlear CF that corresponds to the BF of the cell. It has a latency of approximately 15 ms and is temporally compact indicating precise and rapid locking to the cochlear output at that CF. This finding suggests that relatively fast temporal modulations in auditory-nerve responses are preserved through four or more synapses all the way up to the cortex. We have found correlation functions such as this in about 66% of the 340 cells that exhibited precise firings (σ<10ms). The absence of this correlation in otherwise precisely firing cells may be due to substantial convergence of cochlear channels (e.g., in broadly tuned cells), or other more elaborate linear or nonlinear transformations that alter the cochlear envelopes prior to the cortical stage. Since the auto-correlation of the cochlear modulations is concentrated around zero-lag, the correlation functions in Fig. 2(C) can also be interpreted as the effective STRF of cortical cells to these modulations (denoted by $STRF^F$ in methods).

The relation between the STRF, $STRF^C$ and $STRF^F$ is illustrated in Fig. 4. The $STRF^F$ is compared to the *regular* STRF computed from the TORC envelopes only (i.e., disregarding the fine structure; see methods), and to the $STRF^C$ computed from the cochlear outputs to each TORC (i.e., taking into account responses to TORC envelopes *and* filter-bank output fine-structure). The four examples shown illustrate the wide range of response variability observed in our experimental data. For instance, while all STRF, $STRF^C$, and $STRF^F$ of a given neuron share roughly the same BF, there is a drastic difference between the slow dynamics of the STRF and the rapid onsets of the $STRF^F$, as in units A and B. In both these units, the $STRF^C$ is intermediate, in that it combines features of the STRF and $STRF^F$. In some cases as in unit C, the $STRF^F$ is very weak or absent leading the $STRF^C$ to resemble closely the STRF. Finally, in many cases as in unit D, the STRF and $STRF^F$ may not resemble each other closely suggesting substantial transformation of processing at the cortical level.

The STRFs measured with harmonic TORCs are often virtually identical to those obtained with regular TORCs since both stimuli share the same envelope parameters (e.g. Fig. 2(C)). However, since the carrier tones of the harmonic-TORC constitute a harmonic series, the response fine-structure to the harmonic TORCs is (as expected) limited to one periodicity regardless of CF (that of the fundamental). Consequently, the fine-structure of the cortical response is well correlated with cochlear filter-bank envelopes over all CFs, and does not reflect the STRF of the cell in a meaningful way (e.g., the striped $STRF^F$ in Fig. 2(C)).

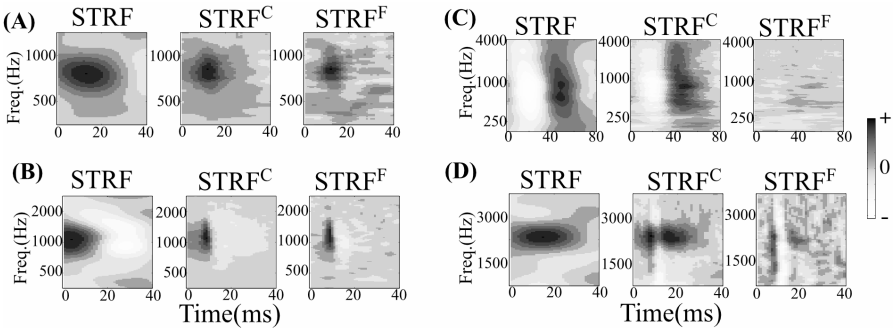


Fig. 4. Examples of STRFs. Each triplet corresponds to the STRF, $STRF^C$, and $STRF^F$ of the same neuron.

So what is the relationship between cortical cells' fine-structure responses ($STRF^F$) and TORC-following responses (STRF)? And why do cortical cells phase-lock well to fast cochlear envelopes (likely more than 200 Hz), yet are incapable of following envelope modulations much beyond 20 Hz?

To elucidate this issue, we examined the hypothesis that the TORC-envelope acts as a gain that *gates* the responses to the fast underlying cochlear modulations. To test this hypothesis, we compared unit responses to those predicted from its STRF and $STRF^F$. Figure 5 illustrates the approach for the unit already discussed in Fig. 2. Using regular TORCs, we measured the unit's STRF and $STRF^F$ (Fig. 2(C), left panel) and then used them to predict the responses to two 48Hz harmonic-TORC stimuli. The results are depicted by three curves in each panel. The *solid black* line is a smoothed period histogram of the actual response to this harmonic-TORC. The *dashed gray* is the predicted response based only on the unit's STRF. This curve captures the broad slow fluctuations in the response due to the TORC envelope, and completely ignores the response fine-structure. The $STRF^F$ predictions are not shown but they consist simply of a train of 48 Hz peaks whose locations are indicated by the arrows in the figure. The *solid gray* line is the *product* of the STRF and $STRF^F$ predictions. It tracks the actual response fairly well, giving support to the hypothesis that the response is essentially a *modulated fine-structure*. That is, the fine-structure is only visible when the envelope of the response fluctuates sufficiently strongly to reveal it; otherwise it is suppressed. This conjecture is completely consistent with the known effects of *synaptic depression* in thalamo-cortical pathway (Markam and Tsodyks 1996), and may explain the paradox of a sluggish, yet precise cortical response.

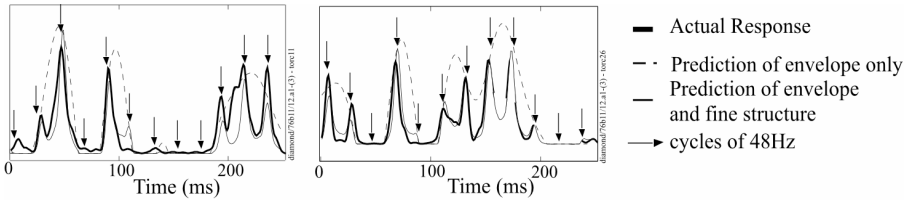


Fig. 5. Predictions of the responses of cortical neurons to a 48Hz harmonic TORC.

4 Discussion

We have demonstrated that cortical cells can phase-lock to cochlear envelope modulations with millisecond accuracy, with rates extending up to 200 Hz, and over sustained periods of time. While these response properties are common in the earlier thalamic and inferior collicular stages, they become highly constrained in the cortex. It becomes necessary to modulate the overall (spectro-temporal) envelope of the stimulus in order to reveal them. For instance, it is difficult to get rapid phase-locked responses to *sustained* simple tones, AM tones or complex tones. Rather, precise and phase-locked firings are largely confined to the onset of stimuli where the envelope rises rapidly (Heil 1997; Wallace, Shackleton, and Palmer 2002). These findings are consistent with the known properties of synaptic depression in thalamo-cortical synapses (Markram *et al.* 1996). To overcome its effects, it is necessary to modulate the stimulus strength (e.g., to turn it off and on) in order to allow the synapse to recover. The time-constant of this recovery is rather slow and explains why cortical responses in general are sluggish (<20 Hz) in following repetitive stimuli (Eggermont 2002). Therefore, allowing for periodic recovery, a non-depressed synapse is capable of conveying fast modulations and eliciting precise spiking over sustained periods, as is the case with the TORC stimuli.

Acknowledgment

This work is supported by the Office of Naval Research (grant N00014-97-1-0501), NIDCD (training grant DC00046-01), and NIH (grant DC05019-01A1).

References

Bair, W. and Koch, C. (1996) Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Comput. 8, 1185-1202.

Eggermont, J.J. (2002) Temporal modulation transfer functions in cat primary auditory cortex: Separating stimulus effects from neural mechanisms. J. Neurophysiol. 87, 305-321.

Heil, P. (1997) Auditory cortical onset responses revisited. I. First-spike timing. J. Neurophysiol. 77, 2616-2641.

Klein, D.J., Depireux, D.A., Simon, J.Z. and Shamma, S.A. (2000) Robust spectro-temporal reverse correlation for the auditory system: optimizing stimulus design. J. Comput. Neurosci. 9, 85-111.

Kowalski, N., Depireux, D.A. and Shamma, S.A. (1996) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J. Neurophysiol. 76, 3503-3523.

Markram, H. and Tsodyks, M. (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature. 382, 807-810.

Miller, L.M., Escabí, M.A., Read, H.L. and Schreiner, C.E. (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. 87, 516-527.

Wallace, M.N., Shackleton T.M. and Palmer, A.R. (2002) Phase-locked responses to pure tones in primary auditory cortex. Hear. Res. 172, 160-171.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Learning-induced sensory plasticity: Rate code, temporal code, or both?

Jean-Marc Edeline

NAMC, UMR CNRS 8620, Orsay. jean-marc.edeline@ibaic.u-psud.fr

1 Introduction

Findings relative to experience-induced sensory plasticity are systematically described based on rate coding. This is clearly a contrast compared with the long tradition of temporal coding in the auditory system. One can thus wonder if descriptions of plasticity in terms of firing rate are appropriate to unravel how experience changes the sensory code in the auditory system of an adult animal. After a brief overview of these domains, it will be suggested that considering the temporal aspects of neuronal discharges will probably help to characterize experience-induced plasticity and to elucidate its mechanisms.

2 Rate code descriptions of experience-induced plasticity

Over the last 15 years, many studies have documented that physiological reorganizations take place in the thalamo-cortical system of adult animals after behavioral training. Two types of reorganizations were described. At the single unit level, the receptive fields (RF) of cortical and thalamic neurons were shown to display selective shifts to the frequency of a significant stimulus. These effects were observed in the secondary and primary auditory cortical fields as well as at the thalamic level (for reviews see Edeline 1999; Weinberger 1998). At the system level, enlargements of tonotopic maps were described in favor of the stimulus that was used during behavioral training (Recanzone, Schreiner, and Merzenich 1993). Although one can intuitively accept the idea that map reorganizations stem from selective receptive field modifications of individual neurons (Fig. 1), it is important to stress that the conditions of induction, and of expression, of RF and map reorganization differ notably. Firstly, selective RF modifications emerged after a few (15-30) training trials. In contrast, map reorganizations were reported after 2-3 months of training in a perceptual task involving hundreds of daily trials. Secondly, RF modifications were usually obtained in awake animals shortly after completion of the training protocol, whereas map reorganizations were obtained from deeply

anesthetized animals 24-48h after completion of extensive training. Thirdly, RF modifications were not correlated with the magnitude of the conditioned behavioral responses (Edeline and Weinberger 1991; 1993). In contrast, the extent of map reorganizations was found to correlate with behavioral performance (Recanzone *et al*. 1993; Rutkowski, Than and Weinberger 2002). Fourthly, RF modifications were obtained after a simple task in which a single tone predicted the occurrence of an aversive or appetitive reward. In these situations, the animal simply had to detect the correlation between the occurrence of two events: a tone and a reward. In contrast, map reorganizations were reported after a training that required the animal to be attentive to the auditory stream and to detect small frequency differences. No new relationship had to be learned in this situation: the animal learned how to make use of activation of slightly different sets of neuronal populations to determine when it was time to produce a behavioral response. Thus, as recently suggested (Weinberger 2003), RF modifications seem to index rapidly acquired associative memory, whereas map reorganizations seem to index perceptual learning.

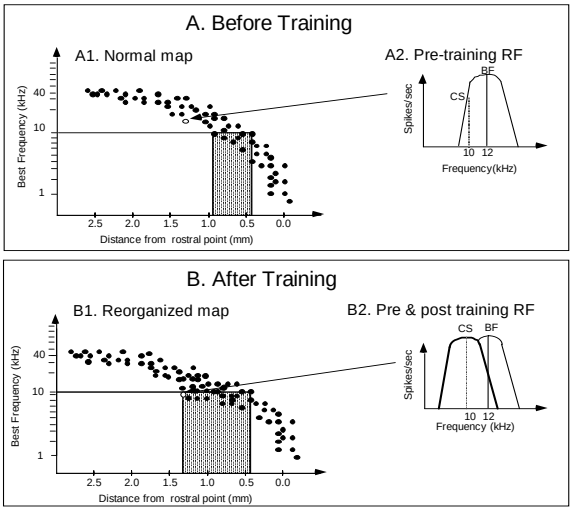


Fig. 1. Potential relationship between RF changes and map reorganizations (modified from Edeline 1999).

Although learning-induced RFs and map reorganizations might result from different processes, when the action of permissive factors such as neuromodulators was investigated, their physiological consequences were found to be similar. For example, stimulation of the nucleus basalis magnocellularis, the unique source of cortical acetylcholine, was able to promote selective RF modifications (Bakin and Weinberger 1996) and selective map reorganizations (Kilgard and Merzenich 1998), thus reinforcing the belief that a continuum exists between RF and map reorganizations.

3 Aspects of temporal coding in the auditory system

It has long been proposed that temporal aspects of neuronal responses code for parameters of auditory stimuli. In fact, there are many different ways to relate temporal aspects of neuronal discharges and acoustic parameters. The most obvious example illustrating the importance of neuronal timing is the selectivity for interaural time differences (ITD): difference in the time of arrival of a sound at the two ears is one of the cues allowing neurons of lower auditory stations (e.g. in the medial superior olive) to be sensitive to stimulus location. In this case, it is the time of occurrence of inputs converging onto a given cell that allows neuronal selectivity to be expressed... in terms of firing rate. If we now consider the output of a given cell, the temporal aspects of spike trains can also code for stimulus location. For example, studies performed in auditory cortex provided strong evidence in favor of a code based on the temporal occurrence of action potentials rather than on the rate of discharge of individual neurons (Middlebrooks, Clock, Xu, and Green 1994). Another aspect of temporal coding concerns the short time-scale coordinations of neuronal discharges. In auditory cortex, neurons can exhibit precise neuronal coordinations (as measured by cross-correlograms) that are selective for a stimulus location or stimulus movement (Ahissar, Ahissar, Bergman and Vaadia, 1992). Sound localization is not the only parameter for which the temporal characteristics of neuronal dischages code for sensory information. The fact that for low frequency auditory stimuli discharges of the VIII nerve fibers were phase-locked with the sounds has largely contributed to the notion that acoustic stimuli can be coded by temporal aspects Also, it was shown from the early 1960s that latency, and latency variability, are more reliable to code a given sound frequency than the neuron's firing rate (Hind, Goldberg, Greenwood and Rose 1963). More recently, the short time-scale interactions between neuronal discharges were found to be related with sound frequency independently of the firing rate (DeCharms and Merzenich, 1996). Thus, temporal aspects of neuronal discharges seem to be able to code for various, if not for all, aspects of acoustic stimuli.

4 Temporal coding is altered in anesthetized animals

Up to now, most of the *in vivo* electrophysiological recordings obtained in the auditory system come from anesthetized preparations, which raises problems because anesthetics strongly affect the temporal characteristics of neuronal responses. For example, the response latency and its variability are affected, being either increased or decreased depending on the type of anesthetic and the locus of recording in the auditory system. Also, the percentage of high-frequency bursts is largely increased by anesthetics, not only in spontaneous but also in evoked activity (Massaux and Edeline in press). As bursts are groups of action potentials with short inter-spike intervals, they are probably more efficient than single spikes at making post-synaptic cells discharge (Swadlow and Gusev 2001), and as such, are suspected to promote the detection of sensory stimuli (Sherman 2001) as well as the occurrence of neuronal plasticity (Steriade and Timofeev 2003). Finally, both

rhythmic activities (large scale synchronizations) and between-cell coordination totally differ in drugged and in undrugged animals (Cotillon-Williams and Edeline 2003). This suggests that many indices that are essential for a temporal code to operate are affected by anesthetics. Therefore, building hypotheses, or models, based on data obtained under anesthesia can lead to conclusions that do not apply to auditory processing in undrugged animals.

5 Progressing toward integrating temporal code in plasticity

Integrating the temporal aspects of neuronal discharges in the field of learning-induced plasticity is of importance for two reasons. First, results from psychoacoustic experiments in human subjects point out that temporal aspects of auditory stimuli are essential to perceive the characteristics of complex signals (Merzenich, Jenkins, Johnston, Schreiner, Miller and Tallal 1996). It is, therefore, quite surprizing that none of the studies describing experience-induced reorganizations in the auditory system have looked at temporal aspects of neuronal discharges. Second, the development of imaging techniques allows visualization of reorganizations in adult sensory cortices of human subjets after behavioral training. For example, studies using PET imaging revealed specific changes in auditory cortical activation after classical conditioning (Morris, Friston and Dolan 1998). However, the obvious pitfall of imaging techniques, such as PET or fMRI, is their lack of temporal resolution. From this perspective, single unit recordings in awake animals still have a considerable advantage.

In fact, a careful examination of some of the data published in the field of learning-induced plasticity reveals that temporal aspects of neuronal discharges can be modified. For example, studies describing increased auditory evoked responses during behavioral training have reported decreases in response latencies (McEchron, Green, Winters, Nolen, Schneiderman and McCabe 1996; Quirk, Repa and Ledoux 1995; Quirk, Armony and Ledoux 1997). Thus, descriptions based only on firing rate might miss important reorganizations emerging from the temporal structure of neuronal discharges. This point is illustrated by Fig. 2: in auditory cortex, a brief increase in noradrenaline (NA) concentration just before tone presentation produces a decrease in evoked response, but it also concentrates the response in time, in such a way that both the latency and its variability are decreased. The question is then: is it more important to consider that the actual number of spikes falling in an arbitrary time window is decreased, or to consider that the occurrence of short-latency, high-frequency bursts of action potentials propagates information about a stimulus more efficiently in a multilayer network? The ideal answer is, of course, that we should consider that these two effects occur simultaneously and both participate to code acoustic information.

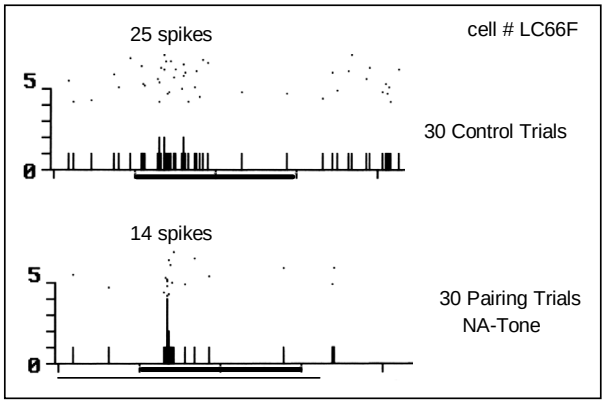


Fig. 2. Responses evoked in the auditory cortex without or with a brief pulse of noradrenaline (from Manunta and Edeline, unpublished data).

In general, rate coding is much easier to understand than the multiple facets by which temporal aspects of neuronal discharges can potentially code for the physical parameters of acoustic stimuli. Thus, it was legitimate to first describe learning-induced, and more generally, experience-induced sensory plasticity in terms of changes in firing rate, i.e., based on a rate code. However, after this first level of description, oversimplifications should be abandoned to better understand the richness and complexity of the neural code and to elucidate how sensory experience can modify this code. A large avenue of research is open to document how the temporal code used in the auditory system of an awake animal is sculpted by sensory experience.

References

Ahissar, M., Ahissar, E., Bergman, H. and Vaadia E. (1992) Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. J. Neurophysiol. 67, 203-215.

Bakin, J.S. and Weinberger, N.W. (1996) Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. Proc. Natl. Acad. Sci. USA 93, 11219-11224.

Cotillon-Williams N. and Edeline J-M. (2003) Evoked oscillations in the thalamo-cortical auditory system are present in anesthetized but not in unanesthetized rats. J. Neurophysiol. 89, 1968-1984.

deCharms, R. C. and Merzenich, M. M. (1996) Primary cortical representation of sounds by the coordination of action-potential timing. Nature 381, 610-613

Edeline, J-M. (1999) Learning-induced physiological plasticity in the thalamo-cortical sensory system: A critical evalutation of receptive field plasticity and maps changes and their potential mechanisms. Prog. Neurobiol. 57, 165-224.

Edeline, J-M. and Weinberger, N.W. (1991) Subcortical adaptive filtering in the auditory system : associative receptive field plasticity in the dorsal medial geniculate body. Behav. Neurosci. 105, 154-175.

Edeline, J-M. and Weinberger, N.W. (1993) Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. Behav. Neurosci. 107, 82-103.

Hind, J.E., Goldberg, J.M., Greenwood, D.D. and Rose, J.E. (1963) Some discharge characteristics of single neurons in the inferior colliculus of the cat. II. Timing of discharges and observations on binaural stimulation. J. Neurophysiol. 26, 321-341.

Kilgard, M.P. and Merzenich, M.M. (1998) Cortical map reorganization enabled by nucleus basalis activity. Science 279, 1714-1718.

Massaux A. and Edeline J-M. (in press) Bursts in the medial geniculate body: a comparison between anesthetized and unanesthetized states in guinea-pig. Exp. Brain Res.

McEchron, M.D., Green, E.J., Winters, R.W., Nolen, T.G., Schneiderman, N.and McCabe, P.M. (1996) Changes of synaptic efficacy in the medial geniculate nucleus as a result of auditory classical conditioning. J. Neurosci. 16, 1273-1283.

Merzenich, M.M, Jenkins, W.M, Johnston, P. Schreiner, C., Miller, S.L. and Tallal, P. (1996) Temporal processing deficits of language-learning impaired children ameliorated by training. Science 271, 77-81.

Middlebrooks, J.C, Clock, A.E., Xu, L. and Green D.M (1994). A panoramic code for sound location by cortical neurons. Science 264, 842-844.

Morris, J.S., Friston, K.J. and Dolan R.J. (1998) Experience-dependent modulation of tonotopic neural responses in human auditory cortex; Proc. R. Soc. Lond B 265, 649-657.

Quirk, G.J., Repa, C. and LeDoux, J.E.(1995) Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat Neuron 15, 1029-1039.

Quirk, G.J., Armony, J.L. and LeDoux, J.E. (1997) Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. Neuron 19, 613-624.

Rutkowski, R.G, Than, K.H. and Weinberger, N.M. (2002) Evidence for area of frequency representation encoding acquired stimulus importance in rat primary auditory cortex. Soc. Neurosci. Abstr. 80.3

Recanzone G.H., Schreiner, , C.E. and Merzenich, M.M. (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J. Neurosci. 13, 87-103.

Sherman, S.M. (2001) Tonic and burst firing: dual modes of thalamocortical relay. Trends in Neurosci. 24, 122-126.

Steriade, M. and Timofeev, I. (2003) Neuronal plasticity in thalamocortical networks during sleep and waking oscillations. Neuron 37, 563-576.

Swadlow, H.A and Gusev A.G. (2001) The impact of 'bursting' thalamic impulses at neocortical synapse. Nature Neuroscience 4, 402-408.

Weinberger N.W. (1998) Physiological Memory in Primary Auditory Cortex: Characteristics and Mechanisms. Neurobiol Learn Mem. 70, 226-251.

Weinberger N.W. (2003) Experience-dependent response plasticity in the auditory cortex: Issues, characteristics mechanisms and functions. In: T. Parks (Ed), *Handbook of Auditory Research* Springer-Verlag, New York, pp 385-430.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Synaptic dynamics and intensity coding in the cochlear nucleus

Katrina M. MacLeod and Catherine E. Carr

Department of Biology, University of Maryland, College Park, macleod@glue.umd.edu, cecarr@umd.edu

1 Introduction

Sound localization depends on two binaural cues: interaural timing differences (ITD) and interaural level, or intensity, differences (ILD). In one avian species, these cues have been shown to be encoded separately and in two parallel streams, beginning with the two divisions of the cochlear nuclei, nucleus magnocellularis (NM) for timing and nucleus angularis (NA) for intensity (Fig. 1)(Knudsen and Konishi 1978; Parks and Rubel 1978; Moiseff and Konishi 1981; Sullivan and Konishi 1984; Takahashi, Moiseff, and Konishi 1984). Both nuclei receive direct input from the auditory nerve, yet extract different information from the afferent spike trains.

How this occurs depends in large part on the nature of the auditory nerve-cochlear nucleus synaptic connection. Large, calyceal synapses are the hallmark of auditory brainstem neurons which are involved in encoding sound stimuli through the precise timing of action potentials (Oertel 1999; Trussell 1999). These highly specialized synapses are found in the avian NM, and in the mammalian ventral cochlear nucleus, medial nucleus of the trapezoid body, and ventral nucleus of the lateral leminscus. Synaptic specializations ensure reliable, high frequency transmission and tight phase-locking. The high release probability associated with reliability comes at a cost: all calyceal synapses experience short-term synaptic depression through vesicle depletion and receptor desensitization (Zhang and Trussell 1994; Schneggenburger, Sakaba, and Neher 2002; von Gersdorff and Borst 2002).

The auditory nerve axon that terminates in a calyceal synapse in NM also bifurcates to send a branch to NA where it forms ordinary, bouton-like synapses (Parks *et al.* 1978; Carr and Boudreau 1991). Synaptic function at non-calyceal auditory brainstem connections is less well understood, and almost no data exists for auditory nerve-NA connections. Short term synaptic plasticity, by dynamically modulating synaptic strength, may be an important factor in gating the auditory information transmitted to each nucleus. We therefore investigated the properties of the auditory nerve-NA synapses in the avian brainstem. This paper provides eviden-

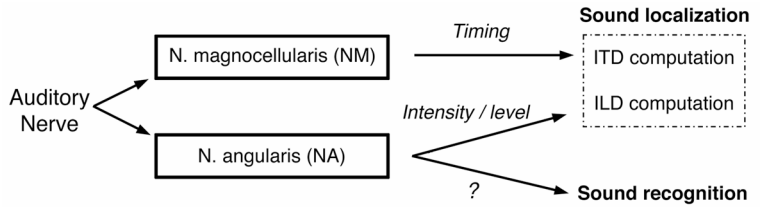


Fig. 1. Parallel pathways in the avian auditory brainstem. ITD, interaural time difference, ILD, interaural level difference.

ce that the cochlear nuclei express dramatically different short-term synaptic plasticity, and that the plasticity found in NA may contribute to its ability to encode level information for sound localization.

2 Methods

Auditory brainstem slices were made from chicken embryos aged E16-18, and whole cell recordings were made from NA and NM neurons as previously described (Soares, Chitwood, Hyson, and Carr 2002). Artificial cerebrospinal fluid contained 75-100 μM APV ([+]-2-amino-5-phosphonopentanoic acid), 20 μM bicuculline, and 0.1μM strychnine to block NMDA, GABA, and glycine receptors. A metal stimulating electrode was placed at the margin of NA or NM where 8th nerve fibers enter the nucleus. Data are pooled from experiments using both current clamp and voltage clamp (V_{hold} ~-60mV). Data was recorded with Axopatch 200B amplifier, filtered at 5-10kHz, and digitally acquired at 20-30kHz. Paired pulse amplitudes were calculated as the ratio of $EPSC_2$ / $EPSC_1$. Steady state relative amplitudes were calculated as the average of $EPSC_{7\text{-}10}$ divided by $EPSC_1$. The 'synaptic transfer function' (Markram, Wang, and Tsodyks 1998) is simply calculated as the product of the steady-state relative amplitude and stimulation frequency and represents the frequency-dependent net postsynaptic effect.

To compare our NA experimental data with the known properties of depressing synapses, we modeled a simple depressing synapse based on a previously published model of depression in neocortex (Varela, Sen, Gibson, Abbott, and Nelson 1997). If spikes occur at times $t_1, t_2 \ldots t_n$, and $EPSC_0$ is initial EPSC amplitude after long period of no activity, then the amplitude of $EPSC_n$ is calculated as:

$$EPSC_n = EPSC_0 \bullet D(t_n^-) \quad (1)$$

D is a time-varying depression factor bounded by 0 and 1, initially equal to 1. When a spike occurs D is multiplied by d, the amplitude of depression:

$$D(t_n^+) = D(t_n^-) \bullet d \quad (2)$$

where $D(t_n^-)$ indicates D just before the spike at time t_n, $D(t_n^-)$, just after. D(t) then decays exponentially back up to 1 with a time constant τ_d :

$$dD / dt = (1\text{-}D(t)) / \tau_d \quad (3)$$

We used parameters τ_d = 75 ms and d = 0.35 to represent depression in NM (Brenowitz and Trussell 2001).

3 Synaptic function in nucleus angularis

3.1 Nucleus angularis synapses show facilitation and depression

To understand how auditory information is transmitted from the auditory nerve to the cochlear nuclei, we studied the synaptic responses of nucleus angularis neurons to electrical stimulation of the 8th nerve in chick auditory brainstem slices *in vitro* (Fig. 2A). Auditory nerve afferents make glutamatergic, excitatory connections onto their postsynaptic targets in NA (MacLeod and Carr 2002). We recorded AMPA-receptor mediated responses to trains of stimuli at a constant frequency (3-10 pulses, 5-333 Hz, 1-7 frequencies tested per cell, n=19 cells).

Unlike the calyceal synapses, short term plasticity at the auditory nerve-NA synapses is composed of a mixture of depression and facilitation, the balance of which is variable across cells. The response of one NA neuron is shown in Fig. 3A. During a 200 Hz train, excitatory postsynaptic currents (EPSCs) showed little depression in their amplitude, but instead were slightly facilitating. However, a recovery stimulus showed depression relative to both the initial EPSC, and to the last EPSC in the train, suggesting that facilitation may mask an underlying depression. In addition, paired pulse facilitation occurred for at least one frequency in more than 60% of NA neurons (10/16 cells tested at 20 Hz or higher), suggesting that facilitation is a common feature in NA.

To better quantify these plasticity effects, we analyzed the frequency-dependence of the steady state relative amplitude (see Methods). Individual NA neuron response curves had a complex relationship with frequency, and this relationship varied across cells (n=7, Fig. 3C). Three types of relative amplitude-

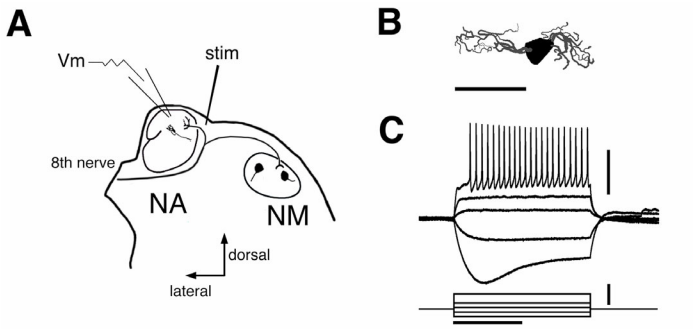


Fig. 2. Brainstem circuit, NA morphology and intrinsic physiology**.** A) Chick brainstem slice and experimental protocol. NA, nucleus angularis, NM, nucleus magnocellularis. B) Neurolucida reconstruction of a biocytin-labeled planar NA neuron. Scale bar: 50μm. C) Voltage response of neuron in B. Scale: top, 40 mV; bottom, 400 pA; horizontal, 200 ms.

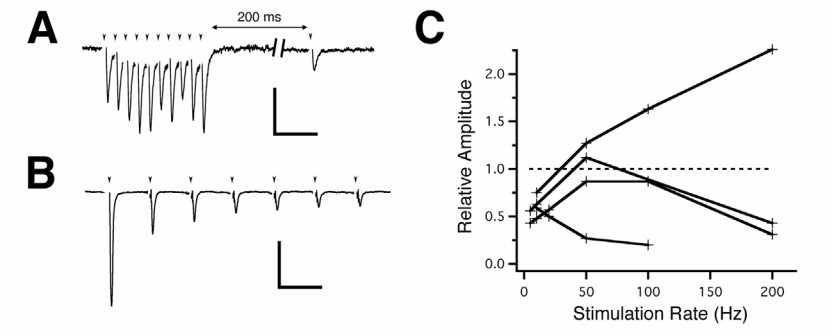


Fig. 3. Excitatory postsynaptic current response electrical stimulation of the 8th nerve. A) NA, 200 Hz train, plus recovery pulse. Scale: 20 pA, 50 ms. B) NM, 50 Hz train. Scale: 400 pA, 50 ms. Stimulation at arrowheads. C) Steady state relative amplitude versus frequency plots for 4 individual NA neurons. Values >1 equals facilitation, <1, depression.

frequency curves were observed: 1) increasing net facilitation with frequency (n=1), 2) increasing depression with frequency (n=1), and 3) a nonmonotonic relationship (n=5), in which low and high frequencies were depressing, but mid-range frequencies showed less net depression or net facilitation.

The NA population on average had a non-monotonic relationship of steady state relative amplitude with frequency (Fig. 4A, open squares; n=16 cells, n=49 cell-frequency combinations), similar to some individual cells (Fig. 3C). At lower frequencies, all cells showed steady state depression (5 Hz: 0.54±0.07, 10 Hz: 0.57±0.15, mean±s.d.). With increasing frequency, different cells showed different degrees of depression or facilitation (range: 0.27 to 2.26 relative amplitude; 50 Hz: 0.79±0.39, 100 Hz: 0.74 ±0.48, 200 Hz: 0.86±0.71, mean±s.d.). Two neurons tested at 333 Hz showed depression. As a population, as for some individual neurons, a balance between facilitation and depression occurs, resulting in little net depression in the range of 50-200 Hz.

In contrast, nucleus magnocellularis synapses strongly depress (Fig. 3B), become increasingly depressed with stimulation frequency (Fig. 4A, filled squares) as has been described in some detail (Hackett, Jackson, and Rubel 1982; Zhang *et al.* 1994; Brenowitz *et al.* 2001). To compare frequency-dependent transmission in NA and NM synapses, we modeled synaptic responses in NM with a simple, one-component depletion model, based on previously published depression models (Varela *et al.* 1997; Brenowitz *et al.* 2001). The result, plotted in Fig. 4A, demonstrates the qualitative and quantitative difference in synaptic transmission at NA and NM synapses.

3.2 A linear synaptic transfer function for nucleus angularis

To try to understand how synaptic plasticity affects information transmission in the cochlear nucleus, we analyzed the synaptic transfer function for NA synapses and model NM synapses (see Methods). Mathematical analysis of the behavior of a purely depressing synapse has shown that at higher frequencies, the steady state

relative amplitude is proportional to the inverse of the stimulus frequency (amplitude ~ 1/f) (Abbott, Varela, Sen, and Nelson 1997; Tsodyks and Markram 1997). We calculated the transfer function (relative amplitude • frequency) for an NM synapse based on a simple depletion model (see Methods; Fig. 4B, filled squares). The transfer function for a depressing synapse is strongly sublinear at low frequencies and saturates (Fig. 4B). The synaptic transfer function calculated for the NA population, however, is well fit up to 200 Hz by a straight line with a slope close to 1 (Fig. 4B, open squares; $R^2 = 0.996$, slope = 0.85). This implies an essentially linear transfer function in NA: a 2-fold increase in the afferent firing rate leads to a 2-fold increase in postsynaptic current impinging on the target neuron. Thus, the mode of information transfer is radically different for these two nuclei.

4 Discussion

The nature of the synaptic contact of an afferent onto its postsynaptic target has important consequences for the information that is passed up the neuronal hierarchy. It has become increasingly clear that the same axon can have different synaptic properties depending on the identities of the postsynaptic targets (Markram *et al.* 1998; Reyes, Lujan, Burnashev, Somogyi, and Sakmann 1998), as is evident for the auditory nerve in the morphological contrast between giant calyceal and ordinary bouton-like synapses in the cochlear nuclei.

We now provide evidence for large, target-specific differences in function at the auditory nerve-cochlear nucleus synapses, and that these differences may be important for differential processing of sensory information. We propose that the short term plasticity properties of these NA synapses may allow the linear transfer of intensity as encoded by the linear slope (in dB) of the auditory nerve rate-intensity function (Salvi, Saunders, Powers, and Boettcher 1992). These data, then, are consistent with the role of NA in encoding sound intensity for binaural comparison, where intensity information from each ear must be maintained. Lesion studies have shown that NA is required for ILD coding, although the exact nature of the intensity coding is not yet clear (Takahashi *et al.* 1984).

In contrast, depressing synapses would lead to an automatic gain control and loss of intensity information. While it is not clear what role depression plays in NM, such a gain control mechanism in nucleus laminaris may confer effective intensity-invariant coincidence detection (Kuba, Koyano, and Ohmori 2002; Cook, Schwindt, Grande, and Spain 2003).

Nucleus angularis is likely to be involved in non-localization auditory function as well (Fig. 1). It is tempting to hypothesize that the variability in the short term plasticity for individual NA neurons, which implies differences in their synaptic transfer functions, might be related to different processing streams within NA. It is therefore of interest to determine if the heterogeneity in short term plasticity correlates other known heterogeneities found in NA. Morphological and *in vitro* physiological studies suggests there are 4-5 cell types in NA (Häusler, Sullivan, Soares, and Carr 2000; Soares and Carr 2001; Soares *et al.* 2002). Multiple NA response types are also found *in vivo* with sound stimulation (Sachs and Sinnott 1978; Sullivan 1985; Warchol and Dallos 1990; Köppl and Carr 2002).

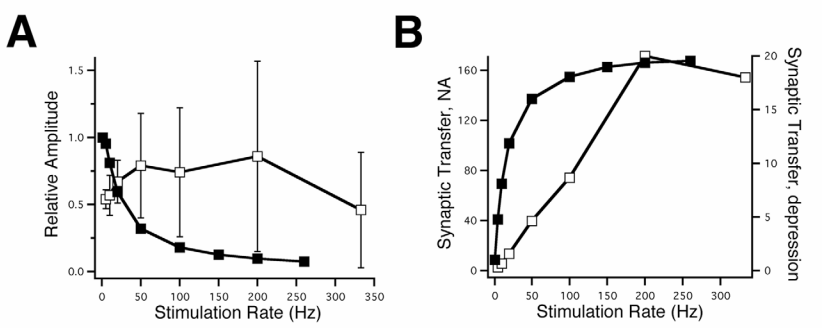


Fig. 4. Comparison of steady state synaptic transmission in NA versus a model depressing synapse. A) Steady state relative amplitude versus train frequency. B) Synaptic transfer function (relative amplitude • rate) versus frequency. Open symbols, NA, filled symbols, depressing model synapse.

Dynamic synapses may confer interesting properties to the processing of dynamic stimuli. While short term depression acts as a gain control at steady state, depressing synapses also respond briskly to *changes* in the input rate, before depressing again to its 1/f value (Abbott *et al.* 1997; Tsodyks *et al.* 1997). Thus, depressing synapses in both NA or in the timing pathway could encode *change* in intensity. These results suggest that dynamic synaptic properties will be an important component in studying how NA neurons encode auditory information.

Acknowledgements

Supported by NIH DC00436 to CEC and NRSA/NINDS fellowship to KMM. We thank Timothy Horiuchi for assistance with computational aspects of this work.

References

Abbott, L.F., Varela, J.A., Sen, K., and Nelson, S.B. (1997) Synaptic depression and cortical gain control. Science 275, 220-224.

Brenowitz, S. and Trussell, L.O. (2001) Minimizing synaptic depression by control of release probability. J Neurosci 21, 1857-67.

Carr, C.E. and Boudreau, R.E. (1991) Central projections of auditory nerve fibers in the barn owl. J Comp Neurol 314, 306-18.

Cook, D.L., Schwindt, P.C., Grande, L.A., and Spain, W.J. (2003) Synaptic depression in the localization of sound. Nature 421, 66-70.

Hackett, J., Jackson, H., and Rubel, E. (1982) Synaptic excitation of the second and third order auditory neurons in the avian brain stem. Neuroscience 7, 1455-1469.

Häusler, U.H.L., Sullivan, W.E., Soares, D., and Carr, C.E. (2000) A morphological study of the cochlear nuclei of the pigeon (*Columba livia*). Brain Behav Evol 54, 290-302.

Knudsen, E.I. and Konishi, M. (1978) A neural map of auditory space in the owl. Science 200, 795-7.

Köppl, C. and Carr, C.E. (2002) Computational diversity in the cochlear nucleus angularis of the barn owl. J Neurophysiol.

Kuba, H., Koyano, K., and Ohmori, H. (2002) Synaptic depression improves coincidence detection in the nucleus laminaris in brainstem slices of the chick embryo. Eur J Neurosci 15, 984-990.

MacLeod, K.M. and Carr, C.E. (2002) Synaptic physiology of the avian cochlear nucleus angularis., Soc Neurosci Abstracts, Orlando. pp. 844.3.

Markram, H., Wang, Y., and Tsodyks, M. (1998) Differential signaling via the same axon of neocortical pyramidal neurons. Proc Natl Acad Sci USA 95, 5323-5328.

Moiseff, A. and Konishi, M. (1981) Neuronal and behavioral sensitivity to binaural time differences in the owl. J Neurosci 1, 40-8.

Oertel, D. (1999) The role of timing in the brain stem audtiory nuclei of vertebrates. Annu Rev Physiol 61, 497-519.

Parks, T.N. and Rubel, E.W. (1978) Organization and development of the brain stem auditory nuclei of the chicken: primary afferent projections. J Comp Neurol 180, 439-448.

Reyes, A., Lujan, R., Burnashev, N., Somogyi, P., and Sakmann, B. (1998) Target-cell-specific facilitation and depression in neocortical circuits. Nature Neurosci 1, 279-285.

Sachs, M.B. and Sinnott, J.M. (1978) Responses to tones of single cells in nucleus magnocellularis and nucleus angularis of the redwing blackbird (*Agelaius phoeniceus*). J of Comp Physiol A 126, 347-361.

Salvi, R.J., Saunders, S.S., Powers, N.L., and Boettcher, F.A. (1992) Discharge patterns of cochlear ganglion neurons in the chicken. J Comp Physiol [A] 170, 227-41.

Schneggenburger, R., Sakaba, T., and Neher, E. (2002) Vesicle pools and short-term synaptic depression: lessons from a large synapse. Trends Neurosci 25, 206-212.

Soares, D. and Carr, C.E. (2001) The cytoarchitecture of the nucleus angularis of the barn owl (*Tyto alba*). J Comp Neurol 429, 192-205.

Soares, D., Chitwood, R.A., Hyson, R.L., and Carr, C.E. (2002) Intrinsic neuronal properties of the chick nucleus angularis. J Neurophysiol 88, 152-162.

Sullivan, W.E. (1985) Classification of response patterns in cochlear nucleus in the barn owl: correlation with functional response properties. J Neurophysiol 53, 201-216.

Sullivan, W.E. and Konishi, M. (1984) Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl. J Neurosci 4, 1787-99.

Takahashi, T., Moiseff, A., and Konishi, M. (1984) Time and intensity cues are processed independently in the auditory system of the owl. J Neurosci 4, 1781-6.

Trussell, L.O. (1999) Synaptic mechanisms for coding timing in auditory neurons. Annu Rev Physiol 61, 477-96.

Tsodyks, M.V. and Markram, H. (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc Natl Acad Sci, USA 94, 719-723.

Varela, J.A., Sen, K., Gibson, J., Abbott, L.F., and Nelson, S.B. (1997) A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. Journal of Neuroscience 17, 7926-7940.

von Gersdorff, H. and Borst, J.G. (2002) Short-term plasticity at the calyx of held. Nat Rev Neurosci 3, 53-64.

Warchol, M.E. and Dallos, P. (1990) Neural coding in the chick cochlear nucleus. J Comp Physiol [A] 166, 721-34.

Zhang, S. and Trussell, L.O. (1994) Voltage clamp analysis of excitatory synaptic transmission in the avian nucleus magnocellularis. J Physiol 480, 123-136.

ISH 2003. Do not distribute. To appear in Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L. (Eds). *Auditory signal processing: Physiology, psychoacoustics, and models.*

Learning and generalization on five basic auditory discrimination tasks as assessed by threshold changes

Beverly A. Wright and Matthew B. Fitzgerald

Department of Communication Sciences and Disorders and Institute for Neuroscience, Northwestern University, {b-wright, m-fitz}@northwestern.edu

1 Introduction

Relatively little is known about how practice influences the performance of human adults on basic auditory discrimination tasks. We are interested in this issue because auditory learning provides a window into the mechanisms underlying performance on the trained task, and into the learning process itself (Wright 2001). Among other benefits, a greater understanding of these issues will help guide the search for the physiological substrates of learning, and aid the development of perceptual training schemes.

With these motivations, we examined learning using simple, pure-tone stimuli, on five basic auditory discrimination tasks: frequency, intensity, interaural-time-difference (ITD), interaural-level-difference (ILD), and duration. Our overarching questions were (1) can listeners improve their ability to discriminate stimuli along each of these dimensions with practice, and if so, (2) does this learning generalize to the trained discrimination performed with untrained stimuli? Because we used the same basic format for all five experiments, any differences in the learning patterns across these trained discriminations likely reflect differences in the plasticity of the underlying mechanisms. Here we report the influence of training on discrimination thresholds assessed by comparing the mean proportional improvements on trained and untrained conditions between listeners who were, and those who were not, given multiple-hour practice on a single discrimination condition. Learning on these five tasks followed one of two general patterns. For ITD and intensity discrimination, additional practice did not lead to greater learning than that seen in untrained listeners. In contrast, for ILD, duration, and frequency discrimination, such practice yielded greater learning, but only on a subset of conditions.

2 General method

The format was the same for each of the five experiments (Wright 2001). At the beginning of each experiment, we gave a group of naïve listeners a pretest during which we measured their discrimination thresholds for tonal stimuli on six conditions. We collected five threshold estimates per condition (300 trials) over the course of a single ~2.5 hour session. We then divided the listeners into two groups. One group, referred to as trained listeners, received training on one of the conditions from the pretest. This training consisted of 12-15 threshold measurements per day (720-900 trials, ~ 1 hour), for 6 to 10 days. The other group, referred to as control listeners, received no training. Finally, at the end of the training phase, we retested all listeners on a posttest that employed the same conditions as the pretest. We randomized the condition order in the pre- and posttests across listeners, but each listener received the conditions in the same order for both of these tests. All stimuli were digitally generated and presented over headphones.

We measured the discrimination thresholds using an adaptive, two-interval, forced-choice procedure. We adjusted the signal level within each 60-trial block using the three-down/one-up rule to estimate the 79% correct point on the psychometric function (Levitt 1971). All listeners received visual feedback as to whether each of their responses was right or wrong throughout the entire experiment.

Here we assessed learning based only on the mean changes in threshold values, computed as the proportional improvement for each listener: [(pretest threshold - posttest threshold)/pretest threshold]. The advantage of this calculation is that it normalizes data for which starting values vary across listeners and/or conditions. Note that this measure emphasizes the amount of change relative to the starting value, rather than the absolute magnitude of threshold change.

We analyzed the proportional-improvement scores for each experiment using a 2 group (trained vs. control) x *n* condition analysis of variance, in which each included condition (n= 3 to 5) employed the same discrimination task with a different standard stimulus (described for each experiment, below). We did not use repeated measures on condition in these analyses because the number of listeners sometimes differed across conditions. When there was a significant group x condition interaction, we used *t*-tests to compare the scores of the trained and control listeners separately for each condition. If the proportional-improvement score of the trained listeners was greater than that of controls on the trained condition, we concluded that the trained listeners learned on the trained condition during the training phase. If that score was greater for the trained than control listeners on an untrained condition, we concluded that the trained listeners generalized their training-phase learning to that condition. We used an alpha value of 0.05 for all analyses.

3 Results

From this analysis, two general learning patterns emerged across the five trained discriminations. In one, observed for ITD and intensity discrimination, both the trained and control listeners learned, but the trained listeners learned no more than controls. In the other, observed for ILD, duration, and frequency discrimination, the trained listeners learned more than controls and generalized their learning to some conditions, but not others, in a pattern unique to each trained discrimination.

3.1 Trained learning equal to control learning

Trained listeners learned no more than controls on both ITD and intensity discrimination. In the ITD-training experiment (Fig. 1; Wright and Fitzgerald 2001), trained listeners practiced discriminating the lateral position of a standard stimulus of 300-ms, 0.5-kHz tones presented to both ears at 70 dB SPL with an ITD of 0 μs from a signal stimulus that differed from the standard only in that the ITD favored the right ear. In the untrained conditions, the standard differed from that in the trained condition either only in location (150-μs ITD vs. 0-μs ITD), frequency (1 kHz vs. 0.5 kHz), or interaural cue (ILD vs. ITD), or in both the frequency and cue (4 kHz, ILD vs. 0.5 kHz, ITD). The mean proportional-improvement scores did not differ between the trained and control listeners (main effect for group, p=0.095; group x condition interaction, p=0.239). However, listeners did learn, because, on each condition, the combined scores of both groups were always greater than zero (based on 95%-confidence intervals).

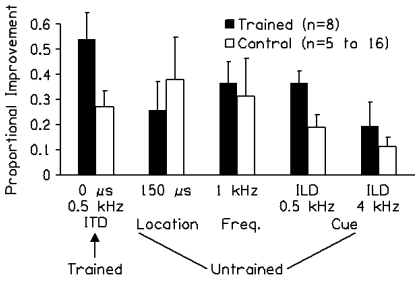


Fig. 1. Threshold proportional-improvement scores for ITD discrimination for the trained (left-most) and untrained conditions. Results are shown for the trained (black bars) and control (white bars) listeners. Error bars indicate one standard error of the mean. (Data from Wright and Fitzgerald 2001.)

In the intensity-training experiment (Fig. 2; Abrams and Wright, unpublished), the trained condition employed a standard of two 15-ms, 1-kHz, tone pips whose onsets were separated by 100 ms, presented to both ears with an ITD of 0 μs at 76 dB SPL. The listener's task was to discriminate this standard from a signal of greater intensity. We chose this particular standard because we previously had used a very similar one to train both duration and frequency discrimination (see below). Here, the untrained conditions differed from the trained one only in the standard stimulus level (46 or 91 dB SPL vs. 76 dB SPL), frequency (4 kHz vs. 1 kHz), duration (temporal interval between tone pips, 50 ms vs. 100 ms) or location (200 μs vs. 0 μs). As for ITD discrimination, the mean proportional-improvement scores did not differ between the trained and control listeners (main effect for group,

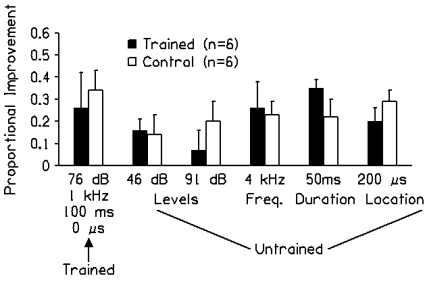


Fig. 2. Same as Fig. 1, but for intensity discrimination (Abrams and Wright, unpublished).

p=0.684; group x condition interaction, p=0.694), but listeners did learn, because, their scores combined across groups were greater than zero on each condition (based on 95%-confidence intervals). Thus, for both ITD and intensity discrimination, 6-10 hours of training yielded no greater reduction in threshold than did exposure to only the 2.5-hour pre- and posttests. For each discrimination task, this learning was relatively uniform for all stimuli.

3.2 Trained learning greater than control learning

In contrast to the learning patterns for ITD and intensity discrimination, trained listeners learned more than controls on ILD, duration, and frequency discrimination and generalized their learning only to task-dependent subsets of untrained conditions. In the ILD-training experiment (Fig. 3; Wright and Fitzgerald 2001) the standard in the trained condition consisted of 300-ms, 4-kHz tones presented to both ears at 70 dB SPL, and the signal differed from the standard only in that the ILD favored the right ear. The untrained conditions differed from the trained one only in the standard location (6-dB ILD vs. 0-dB ILD) or frequency (0.5 kHz or 6 kHz vs. 4 kHz), or in both the frequency and cue (0.5 kHz, ITD vs. 4 kHz, ILD). Here, the control listeners improved on all conditions (based on 95%-confidence intervals). However, the trained listeners learned more than controls (group x condition interaction, p=0.008), but only on the trained condition (p <0.001) and the untrained location (6-dB ILD; p= 0.036) and not on the untrained frequencies or the ITD cue.

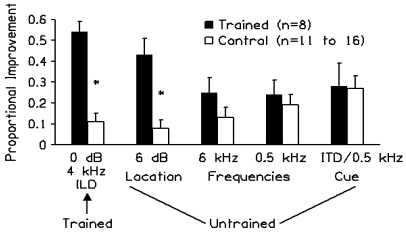


Fig. 3. Same as Fig. 1, but for ILD discrimination. (Data from Wright and Fitzgerald 2001.)

We examined learning on duration discrimination (Fig. 4; Wright, Buonomano, Mahncke, and Merzenich 1997; Wright 1998) using a monaural presentation of essentially the same standard as in the intensity-training experiment (described above). The listener's task was to discriminate the standard from a signal in which the two tone pips were separated by a longer temporal interval. The untrained conditions differed from the trained one only in the standard frequency (4 kHz vs. 1 kHz) or duration (200 or 500 ms vs. 100 ms). In this case, the controls did not improve on any condition (based on 95%-confidence intervals). The trained listeners learned significantly more than controls (group x condition interaction, p=0.026) on the trained condition (p=0.002) and the untrained frequency (4 kHz, p<0.001), but not on the untrained durations.

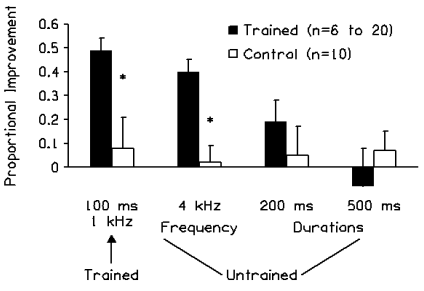


Fig. 4. Same as Fig. 1, but for duration discrimination (Some data from Wright et al. 1997 and Wright 1998).

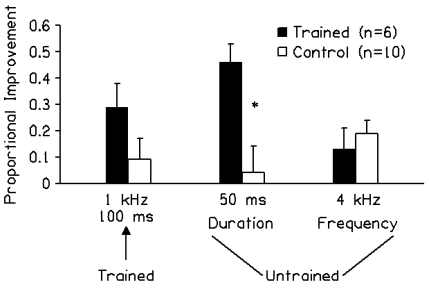


Fig. 5. Same as Fig. 1, but for frequency discrimination (Data from Wright 1998)

Finally, to examine learning on frequency discrimination (Fig. 5; Wright 1998), we trained listeners with the same standard as in the duration-training experiment but required them to discriminate that standard from a signal of lower frequency. In the two untrained frequency-discrimination conditions, the standard differed from that in the trained condition only in its duration (temporal interval between tone pips, 50 ms vs. 100 ms) or frequency (4 kHz vs. 1 kHz). Here, controls improved for the 4-kHz, but not the two 1-kHz stimuli (based on 95%-confidence intervals). The trained listeners improved more than controls (group x condition interaction, $p=0.024$), but, surprisingly, not on the trained condition ($p = 0.129$). However, there appears to have been training-induced learning, because the trained listeners learned more than controls on the untrained duration ($p=0.01$). This learning did not generalize to the untrained frequency.

Thus, for ILD, duration, and frequency discrimination, 9-10 hours of training yielded greater threshold improvements than did exposure to only the 2.5-hour pre- and posttests. Further, this training-induced learning generalized to some untrained stimuli but not others, and did so in a pattern unique to each trained discrimination task. For ILD discrimination, learning generalized to an untrained location, but was specific to the trained frequency and cue. For duration discrimination, learning generalized to an untrained frequency, but was specific to the trained duration. Finally, for frequency discrimination, learning generalized to an untrained duration, but was specific to the trained frequency.

4 Discussion

From the present results, it appears that improvements on basic auditory discrimination tasks result from two different types of learning, that this learning affects task-specific processing mechanisms, and that these mechanisms are not equally malleable. The threshold improvements on ITD and intensity discrimination may reflect primarily conceptual or procedural learning. In both instances, control listeners improved on all conditions with exposure only to the

pre- and posttests. We, and others, have proposed that such rapid and general learning reflects the acquisition of the general procedures needed to perform the task, and as such does not result from fundamental changes in stimulus processing (Recanzone, Schreiner, and Merzenich 1993; Robinson and Summerfield 1996; Wright and Fitzgerald 2001). By this account, learning on ITD and intensity discrimination is primarily procedural, because additional training did not benefit listeners on these tasks. This lack of training-phase learning may have occurred either because these discriminations are already over-learned, or because the mechanisms that govern them are relatively inflexible.

In contrast, the threshold improvements in ILD, duration, and frequency discrimination may largely result from perceptual or stimulus learning. In these cases, control listeners improved on some tasks, suggesting procedural learning. However, additional training always resulted in greater improvements on a subset of stimuli. Such slow and stimulus-specific learning may reflect fundamental changes in stimulus processing (Recanzone, Schreiner, and Merzenich 1993; Robinson and Summerfield 1996; Wright and Fitzgerald 2001). By this view, the different generalization patterns for the different tasks indicate both that training affected separate mechanisms for the different discriminations, and that these mechanisms are organized in different ways. Training modified a mechanism that, (1) for ILD discrimination, processes multiple locations, but only the ILD cue with the trained frequency, (2) for duration discrimination, processes multiple frequencies, but only the trained duration, and (3) for frequency discrimination, processes multiple durations, but only the trained frequency. Others have reported results for auditory duration (Karmarkar and Buonomano 2003) and frequency (Delhommeau, Micheyl, Jouvent, and Collet 2002; Demany and Semal 2002; Irvine, Martin, Klinkeit, and Smith 2000) discrimination that are consistent with this interpretation. Further, training-induced learning on duration discrimination has also been shown to be specific to the trained duration in the somatosensory system (Nagarajan, Blake, Wright, Byl, and Merzenich 1998), and to generalize from the somatosensory to auditory (Nagarajan et al. 1998), and from the auditory to motor (Meegan, Aslin, and Jacobs 2000) systems, suggesting that training on duration discrimination affects a multi-system timing mechanism that processes different durations separately.

In summary, a systematic examination of learning on five basic auditory discrimination tasks, assessed by threshold reductions, revealed two different learning patterns. Improvements on ITD and intensity discrimination were rapid and general, and therefore appeared to reflect procedural learning. Improvements on ILD, duration, and frequency discrimination were slow and stimulus specific, and, as such, appeared to arise from stimulus learning. Overall, the present results indicate that learning affected a different mechanism for each trained task, and that these mechanisms differ in their plasticity as well as their organization.

Acknowledgments

Julia Mossbridge, Jeanette Ortiz, and Yuxuan Zhang provided helpful comments on an earlier draft of this chapter. Chris Stewart prepared the figures. This work was supported by NIH/NIDCD (R01 DC 04453 and F31 DC05093).

References

Abrams, D. A. and Wright, B.A. (unpublished data).

Delhommeau, K., Micheyl, C., Jouvent, R., and Collet, L. (2002) Transfer of learning across duration and ears in auditory frequency discrimination. Perception and Psychophysics, 64, 426-436.

Demany, L. and Semal, C. (2002) Learning to perceive pitch differences. J. Acoust. Soc. Am. 111, 1377-1388.

Irvine, D.R.F., Martin, R.L., Klimkeit, E., and Smith, R. (2000) Specificity of perceptual learning in a frequency-discrimination task. J. Acoust. Soc. Am. 108, 2964-2968.

Karmarkar, U.R. and Buonomano, D.V. (2003) Temporal specificity of perceptual learning in an auditory discrimination task. Learning and Memory 10, 141-147.

Meegan, D.V., Aslin, R.N., and Jacobs, R.A. (2000) Motor timing learned without motor training. Nature Neuroscience 3, 860-862.

Nagarajan, S.S., Blake, D.T., Wright, B.A., Byl, N., and Merzenich, M.M. (1998) Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. J. Neurosci., 18, 1559-1570.

Recanzone, G.H., Schreiner, C.E., and Merzenich, M.M. (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J. Neurosci. 13, 87-103.

Robinson, K. and Summerfield, A.Q. (1996) Adult auditory learning and training. Ear. Hear. 17, 51-65.

Wright, B.A. (2001) Why and how we study human learning on basic auditory tasks. Audiol. Neurootol. 6, 207-210.

Wright, B.A. (1998) Generalization of auditory-discrimination learning. Assoc. Res. Otolaryngol. Abs., Abs. 413, 104. (A)

Wright, B.A., and Fitzgerald, M.B. (2001) Different patterns of human discrimination learning for two interaural cues to sound-source location. Proceed. Nat. Acad. Sciences, 98, 12307-12312.

Wright, B.A., Buonomano, D.V., Mahncke, H.W., Merzenich, M.M. (1997) Learning and generalization of auditory temporal-interval discrimination in humans. J. Neurosci., 17, 3956-3963.