Louder sounds can produce less forward masking: Effects of component phase in complex tones[a]

Hedwig Gockel[b]
CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, England

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Roy D. Patterson
CNBH, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, England

Ray Meddis
CNBH at Essex, Department of Psychology, University of Essex, Colchester CO4 3SQ, England

(Received 11 October 2002; revised 23 May 2003; accepted 30 May 2003)

The influence of the degree of envelope modulation and periodicity on the loudness and effectiveness of sounds as forward maskers was investigated. In the first experiment, listeners matched the loudness of complex tones and noise. The tones had a fundamental frequency (F0) of 62.5 or 250 Hz and were filtered into a frequency range from the 10th harmonic to 5000 Hz. The Gaussian noise was filtered in the same way. The components of the complex tones were added either in cosine phase (CPH), giving a large crest factor, or in random phase (RPH), giving a smaller crest factor. For each F0, subjects matched the loudness between all possible stimulus pairs. Six different levels of the fixed stimulus were used, ranging from about 30 dB SPL to about 80 dB SPL in 10-dB steps. Results showed that, at a given overall level, the CPH and the RPH tones were louder than the noise, and that the CPH tone was louder than the RPH tone. The difference in loudness was larger at medium than at low levels and was only slightly reduced by the addition of a noise intended to mask combination tones. The differences in loudness were slightly smaller for the higher than for the lower F0. In the second experiment, the stimuli with the lower F0s were used as forward maskers of a 20-ms sinusoid, presented at various frequencies within the spectral range of the maskers. Results showed that the CPH tone was the least effective forward masker, even though it was the loudest. The differences in effectiveness as forward maskers depended on masker level and signal frequency; in order to produce equal masking, the level of the CPH tone had to be up to 35 dB above that of the RPH tone and the noise. The implications of these results for models of loudness are discussed and a model is presented based on neural activity patterns in the auditory nerve; this predicts the general pattern of loudness matches. It is suggested that the effects observed in the experiments may have been influenced by two factors: cochlear compression and suppression.
 [DOI: 10.1121/1.1593065]

PACS numbers: 43.66.Cb, 43.66.Dc, 43.66.Nm [MRL]

I. INTRODUCTION

Most models of loudness are based on the power spectrum of the sound of interest (Fletcher and Munson, 1933; Zwicker, 1958; Zwicker and Scharf, 1965; Stevens, 1972; Moore *et al.*, 1997). For example, the loudness models presented by Zwicker and Scharf (1965) and by Moore *et al.* (1997) first derive an excitation pattern from the power spectrum of the sound. Then, using a modified power law transformation, a specific loudness pattern is calculated from the excitation pattern. The overall loudness of the sound is assumed to be proportional to the area under the specific loudness pattern. For steady sounds, these models give good predictions. However, there is evidence that for fluctuating sounds they may be inadequate. While the data are not entirely consistent across studies, recent studies suggest that, for narrowband sounds, amplitude modulation at medium rates (from 10 Hz up to about 100 Hz) results in a slight decrease of loudness relative to that of an unmodulated sound (Bauch, 1956; Fastl, 1975; Hellman, 1985; Zhang and Zeng, 1997; Moore *et al.*, 1998, 1999; Grimm *et al.*, 2002). The difference in loudness corresponds approximately to a change in level of about 1 dB. For broadband sounds the effect goes in the opposite direction; introducing amplitude modulation at medium rates increases the loudness of the sound slightly (Zhang and Zeng, 1997; Moore *et al.*, 1999;

[a]Portions of these results were presented at the symposium on "Psychoacoustics, physiology and models of the central auditory system" [Gockel *et al.*, Acta Acust. Acust. **88**, 369–377 (2002b)].

[b]Current address: MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England. Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

Grimm *et al.*, 2002). The model proposed by Glasberg and Moore (2002), which is based on the calculation of excitation patterns from the short-term spectrum of the stimulus, can predict the effect for narrowband sounds, but not for broadband sounds.

Another psychophysical measure assumed to reflect the excitation evoked in the auditory system is forward masking. The threshold of a sinusoidal signal in forward masking is assumed to be monotonically related to the excitation evoked by the masker at the signal frequency (Houtgast, 1974; Moore and Glasberg, 1983; Moore, 1997; Plack and Oxenham, 1998). As Carlyon and Datta (1997) pointed out, the excitation produced by a stimulus with a temporally fluctuating envelope is likely to be affected by fast-acting compression in the cochlea (Robles *et al.*, 1986; Ruggero, 1992; Ruggero *et al.*, 1997; Recio *et al.*, 1998). Carlyon and Datta (1997) compared the amount of forward masking produced by two harmonic tone complexes, one with components added in Schroeder-positive phase (S+), which is thought to evoke a ''peaky'' waveform on the basilar membrane (BM), and the other with components added in Schroeder-negative phase (S−), which produces a less peaky waveform. The S+ complex produced less forward masking than the S− complex of a sinusoidal signal with frequency corresponding to one of the central harmonics. This was explained by fast-acting compression which would result in lower average excitation for the peaky waveform evoked by the S+ complex than for the less peaky waveform evoked by the S− complex.

Carlyon and Datta (1997) also compared the loudness of the S+ and S− complexes, or rather of a mid-range subset of their components; the subset was turned on after the other components. The results showed that the subset was louder for the S− than for the S+ complex. Thus, both measures, loudness and ''effectiveness as a forward masker,'' were consistent with the idea that the S− complex evokes more excitation than the S+ complex.

In this paper, we present two experiments designed to further our understanding of the relationship between the loudness of *broadband* sounds and their effectiveness as forward maskers. Using broadband stimuli with the same power spectrum, but different crests factors (ratio of peak amplitude to root-mean-square, rms, amplitude), we addressed the following questions: (1) Which is louder? (2) Which produces more forward masking?

In contrast to the study of Carlyon and Datta (1997), it was the loudness of the whole sound that was compared instead of a subset of the components. We would not, however, expect this to change the direction of the effects of compression on loudness and forward masking. We have previously presented results (Gockel *et al.*, 2002b) indicating that, for bandpass filtered complex tones with a low F0 (62.5 Hz), adding the components in cosine phase (large crest factor) led to greater loudness values than adding the components in random phase (small crest factor); this appears inconsistent with the idea that a large crest factor leads to lower excitation and that this in turn leads to lower loudness. It also appears inconsistent with the results of Carlyon and Datta (1997). We also found (Gockel *et al.*, 2002b) that the random phase complex tone had slightly greater loudness than a random noise that was bandpass filtered to produce a similar excitation pattern. However, the difference was not statistically significant.

Here, we present results for more subjects to increase the statistical power. We also present results for both a low F0 (62.5 Hz) and a higher F0 (250 Hz). In a previous experiment (Gockel *et al.*, 2002a) we showed that, for the 62.5-Hz F0, a cosine-phase complex tone was a much less effective simultaneous masker of a noise (filtered into the same frequency region) than the noise was as a masker of the complex tone. This effect was much reduced when the F0 was 250 Hz. We argued that the greater effect for the lower F0 was partly caused by the greater peak factor of the waveforms evoked on the BM by the cosine-phase tone at the lower F0 (as, for a given center frequency, the number of components within the passband of the auditory filter is four times greater for the low F0 than for the high F0). Based on the arguments presented above, one might expect the peak factor to influence loudness. That expectation was assessed here.

II. EXPERIMENT 1

A. Method

1. Rationale and stimuli

The first experiment investigated the influence of the degree of envelope modulation and periodicity on loudness. Listeners matched the loudness of complex tones and noise. The stimuli were the same as those used as simultaneous maskers by Gockel *et al.* (2002a). The complex tones had a fundamental frequency (F0) of 62.5 or 250 Hz and were bandpass filtered between the 10th harmonic and 5000 Hz (3-dB down points, 100 dB/oct slope). The Gaussian noise was filtered in the same way as the complex tone with which it was to be compared in loudness. The components of the complex tones were added either in cosine phase (CPH), giving a large crest factor, or in random phase (RPH), giving a small crest factor. For each F0, subjects were presented with every possible stimulus pair, resulting in three stimulus combinations: CPH vs. RPH, CPH vs. noise, and RPH vs. noise. The long-term excitation patterns of the stimuli, calculated according to the procedure of Glasberg and Moore (1990), were essentially identical within each F0 condition (see Fig. 1 of Gockel *et al.*, 2002a). Thus, for a given F0, loudness models based on the long-term excitation pattern would not predict any differences in loudness between these stimuli. In contrast, models specifically taking into account the effects of fast-acting compression would be expected to predict that the CPH tone will be less loud than the RPH tone or the noise. This is discussed in more detail below.

In the loudness-matching procedure, one stimulus was fixed in level and the other was varied in level. The overall rms level of the fixed stimulus varied from 30 to 80 dB SPL in 10-dB steps for the 62.5-Hz F0, and from 27.6 dB SPL to 77.6 dB SPL in 10-dB steps for the 250-Hz F0. The 2.4-dB level adjustment between the two F0s was included in order to keep the level per ERB constant (Glasberg and Moore,

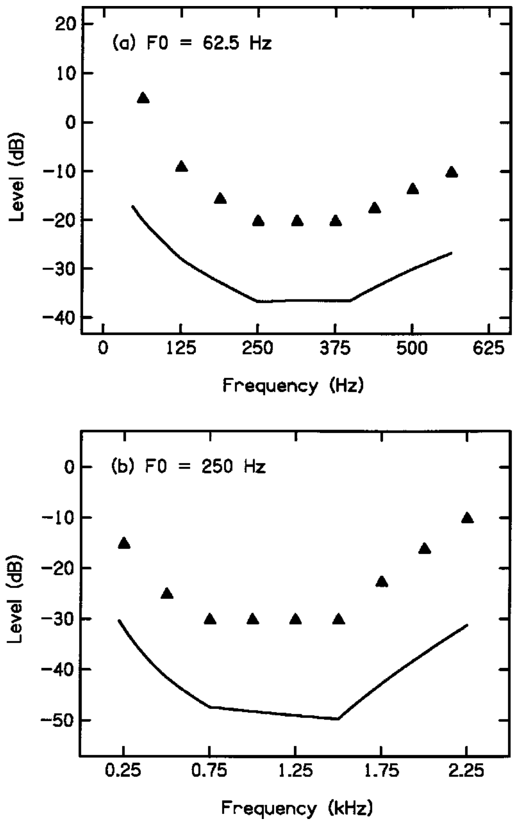

FIG. 1. Derivation of the spectral shape of the masking noise used in experiment 1 for the 62.5-Hz F0 (top) and the 250-Hz F0 (bottom). The filled upward-pointing triangles show the assumed levels of the tones needed to cancel quadratic and cubic distortion products, based on the data of Pressnitzer and Patterson (2001). The solid line shows the derived spectral density of the masking noise which was used to mask the combination tones.

1990); with this adjustment, the level per component within the passband was 6 dB lower for F0 = 62.5 Hz than for F0 = 250 Hz.

There is evidence that the strength of combination tones can depend on the phase of the primary components (Buunen *et al.*, 1974; Pressnitzer and Patterson, 2001). This could lead to differences in loudness between our stimuli. To test this possibility, in half of the conditions a noise was used to mask combination tones. Figure 1 shows the spectral shape of this noise for the 62.5-Hz F0 (top) and the 250-Hz F0 (bottom). The strongest combination tones were assumed to be quadratic distortion tones occurring at F0 and low harmonics of F0, and cubic distortion tones which are most intense for frequencies just below the lower edge of the passband of the stimulus (Goldstein, 1967; Greenwood, 1971; Zwicker and Fastl, 1999). The shape and level of the masking noise relative to the level of the components in the target tones were based partly on the measurements of Pressnitzer and Patterson (2001) of the level of a sinusoidal tone required to cancel combination tones produced by harmonic CPH tones. The filled, upward-pointing triangles in Fig. 1 show the estimated cancellation tone level relative to the level per component in our CPH tones as a function of frequency. These relative levels are applicable for moderate levels of the primary components. The continuous curve in Fig. 1 shows the derived spectrum level of the masking noise which would just mask the cancellation tones and, thus, the distortion tones. The latter derivation included the finding that, at very low frequencies, a higher signal-to-noise ratio at the output of the auditory filter is required for threshold. When present, the masking noise was continuous and at a constant level during one run; the fixed level of the masking noise was equal to that calculated to just mask the distortion products. Across runs, the noise level was kept constant relative to the level of the fixed stimulus. The noise should have substantially reduced the contribution of combination tones to loudness.

The duration of each signal was 700 ms, including 40-ms raised-cosine onset and offset ramps. The stimuli were generated digitally in advance using a sampling rate of 25 kHz. The tones were generated by adding sinusoids with frequencies ranging from F0 up to 10 kHz, while the noise was generated in the time domain by sampling from a Gaussian distribution. Bandpass filtering was then performed with a 900-tap, digital finite-impulse-response (FIR) filter with a linear phase response. For each F0, ten different realizations were produced for the RPH tone and for the Gaussian noise; one of the ten was chosen at random for each presentation. The fixed-level and variable-level signals were played through one channel of a 16-bit digital-to-analog converter (Tucker-Davis Technologies, TDT, DD1), lowpass filtered at 10 kHz (TDT FT6-2), and attenuated separately using a programmable attenuator (TDT PA4). The background noise was recorded onto CDR and played back with the test stimuli in the appropriate conditions. Stimuli were fed to a headphone buffer (TDT HB6) and presented monaurally via headphones with a diffuse-field response (AKG K 240 DF). Subjects were seated individually in an IAC double-walled, sound-attenuating booth.

2. Procedure

The two sounds (X and Y) which had to be matched in loudness were presented monaurally in regular alternation and were separated by a 200-ms silent interval. Synchronously with the two sounds, two yellow LEDs on a response box came on in regular alternation. Subjects indicated which sound was louder by pressing the button underneath the LED which accompanied that sound. Within a given run, either sound X or sound Y was fixed in level; the level of the other sound was varied to match the loudness. The starting level of the variable sound was chosen randomly from a range extending from 20 dB below to 20 dB above the level of the fixed sound, except for the lowest and the highest levels of the fixed sound. For these, the range extended from 10 dB below to 10 dB above the fixed level, to avoid inaudible or uncomfortably loud sounds. When the variable level sound was judged to be louder, it was decreased in level. When the fixed level sound was judged to be louder, the variable level sound was increased in level. The attenuator setting was changed during the silent interval between stimuli. When no button was pressed, sound presentation continued without any change. The initial step size was 5 dB. After two turnpoints it was reduced to 3 dB, and after two more turnpoints to 1 dB. Subjects were encouraged to bracket the point of equal loudness several times, i.e., to go from ‘‘sound X is louder’’ to ‘‘sound Y is louder,’’ before making their final adjustment. To indicate that they were satisfied with their match, they pressed a button underneath a green LED, which stopped the run. The variable level at this point was taken as the matching level. This ''satisfied'' button press was only accepted after the final step size was reached. If it was pressed earlier, sound presentation continued without any change. In order to balance bias effects, eight matches were done varying sound X in level, and eight matches varying sound Y. In addition, for four of these eight matches the variable-level sound was presented with the first LED, and for the other four matches it was presented together with the second LED. No feedback was provided.

For a given F0, the same subjects ran in all conditions. Half of the subjects started with the masking noise present, and half without the noise. The conditions were presented in a counterbalanced order. Within a given condition, which was determined by the specific combination of the two sounds, the identity of the fixed stimulus, and the order of the two sounds, the level of the fixed stimulus was chosen randomly from the range of fixed levels to be tested. One match was obtained for each condition and level in turn before additional measurements were obtained in any other condition. To familiarize subjects with the procedure and equipment, they participated in one loudness match for each condition before the experiment proper commenced.

3. Subjects

For the 62.5-Hz F0, six subjects participated in all conditions. For the 250-Hz F0, five subjects participated in all conditions, three of whom also took part in the 62.5-Hz F0 conditions. Their ages ranged from 20 to 50 years, and their quiet thresholds were better than 15 dB HL at audiometric frequencies between 500 and 5000 Hz.

B. Results

Figures 2 and 3 show the mean results for the 62.5-Hz F0 and the 250-Hz F0, respectively. The left and right columns give the results without and with masking noise, respectively. The first, second, and third rows show the results from the three stimulus pairs, i.e., noise vs. CPH tone, RPH vs. CPH tone, and noise vs. RPH tone. The difference between the overall rms levels of the two stimuli at the point of equal loudness is plotted as a function of the level of the fixed stimulus; the direction of the difference is indicated in each panel. For example, the upper row shows the level of the noise minus the level of the CPH complex. Results are plotted separately for the case when X was varied in level and when Y was varied in level. For example, in the upper row, the open circles indicate the difference in level for matches when the CPH tone was fixed and the noise was varied, whereas the filled upward-pointing triangles show the difference when the CPH tone was varied. Each symbol shows the difference averaged across subjects and across the two orders of presentation, and is thus based on 48 matches for the 62.5-Hz F0, and on 40 matches for the 250-Hz F0. The error bars show the standard errors of the means across subjects. The solid lines show the mean for the two cases of

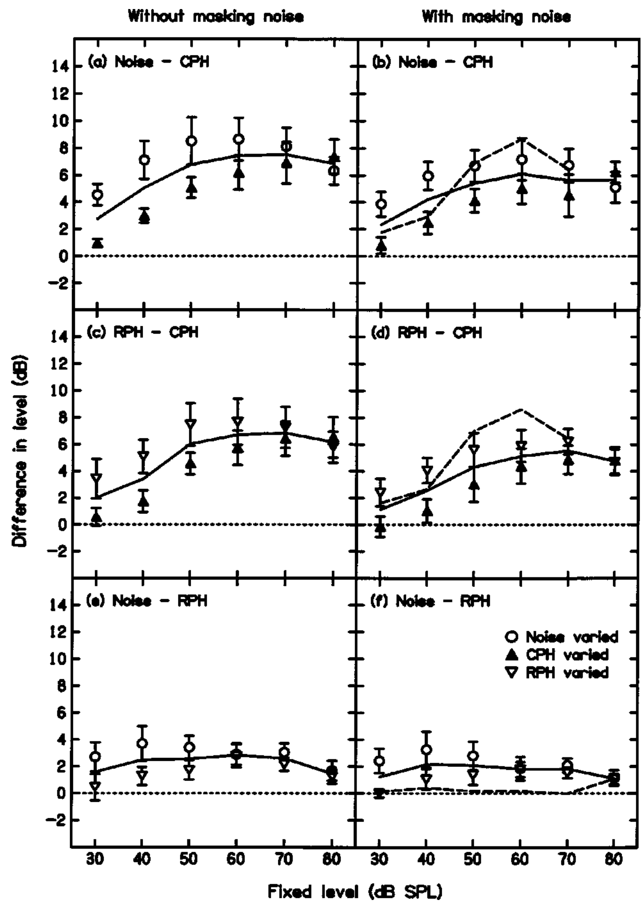

FIG. 2. Loudness results (experiment 1) for F0 = 62.5 Hz, averaged across six subjects. The left and right columns show results without and with the noise used to mask combination tones. Panels (a) and (b) (first row) show results of loudness comparisons between the Gaussian noise and the CPH tone. The difference in level (noise minus CPH tone) at the point of equal loudness is plotted. The open circles show results when the noise was varied in level; the filled upward-pointing triangles show results when the CPH tone was varied. The solid line shows the average of the two cases. Error bars show the standard error of the mean across subjects. The dashed line shows the predictions of the loudness model described in Sec. II D. Panels (c) and (d) show the results of loudness comparisons between the RPH and CPH tones. Panels (e) and (f) show the results for loudness comparisons between the Gaussian noise and the RPH tone.

the variable sound. The dashed lines in the right-hand column illustrate predictions which will be discussed later.

1. 62.5-Hz F0

Consider the results for the 62.5-Hz F0 first (Fig. 2). For each fixed level, at the point of equal loudness, the noise level was greater than the CPH tone level [Fig. 2(a)], the RPH tone had a higher level than the CPH tone [Fig. 2(c)], and the noise had a higher level than the RPH tone [Fig. 2(e)]. These effects were strongest at medium to high levels, and there was a trend for them to be slightly reduced in the presence of the masking noise [Figs. 2(b), (d) and (f)]. A clear bias effect can be seen. At nearly all levels, listeners tended to set the level of the variable sound to a level higher than that ''required for equal loudness'' (as indicated by the solid lines). A similar bias effect has been observed in previous studies on loudness matching (Stevens, 1956; Zwicker *et al.*, 1957). The bias decreased with increasing level of the

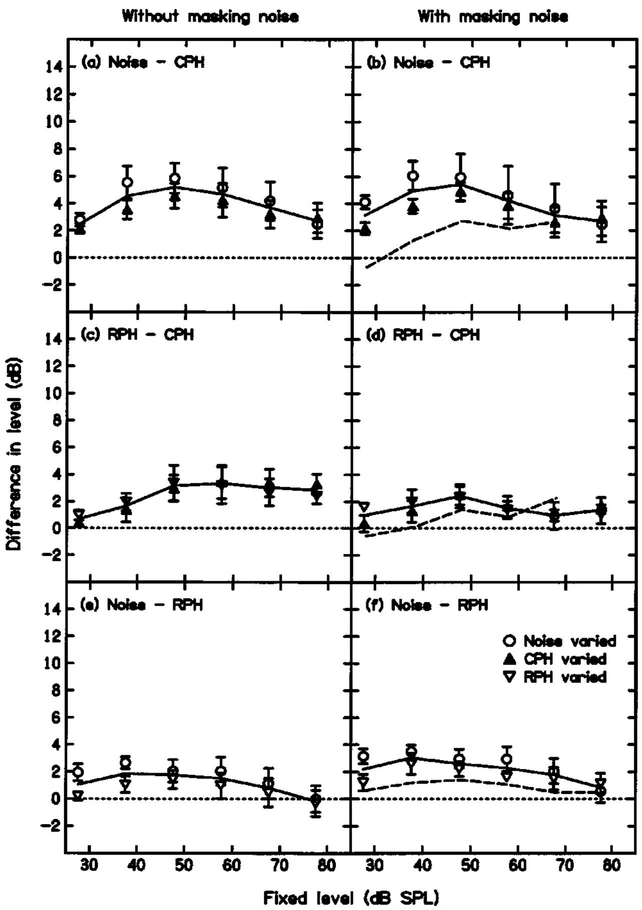

FIG. 3. As Fig. 2, but for F0 = 250 Hz.

fixed stimulus, perhaps because subjects avoided adjusting the variable sound to a high listening level. This "regression effect" has also been observed previously (Stevens, 1956; Stevens and Greenbaum, 1966; Stevens and Guirao, 1967).

Analyses of variance (ANOVAs) for repeated measures were conducted on the data for the three stimulus pairings separately. The mean value of the difference in level at the point of equal loudness for each individual subject in each condition was used as the input. Throughout this paper, the Huynh–Feldt correction was used when the condition of sphericity was not satisfied (see Howell, 1997). The results are summarized in Table I. The main effect of level was significant for the noise vs CPH comparison and for the RPH vs CPH comparison, but not for the RPH vs noise comparison. The main effect of presence or absence of the masking noise was significant only for the RPH vs CPH comparison. The main effect associated with which sound was varied was significant for the noise vs CPH and noise vs RPH comparisons, indicating that the bias was significant. The interaction between the identity of the varied sound and the fixed level was significant for all three comparisons, reflecting the reduction of the bias with increasing level of the fixed stimulus.

For the noise vs CPH comparison, the grand mean was significantly above zero ($p < 0.0012$), indicating that, at the point of equal loudness, the level of the noise was significantly greater than that of the CPH tone. For all fixed levels tested, the difference in matching level was significantly above zero ($p < 0.01$ for all levels). For the RPH vs CPH comparison, the grand mean was also significantly above zero ($p < 0.003$), indicating that, at the point of equal loudness, the level of the RPH tone was significantly greater than that of the CPH tone. The difference between matching levels was significantly above zero ($p < 0.05$) for all but the lowest fixed level tested. For the noise vs RPH comparison, the grand mean was also significantly above zero ($p < 0.05$). As the ANOVA showed no effect of level for this comparison, no *post hoc* tests were performed.

2. 250-Hz F0

Next, consider the results for the 250-Hz F0 (Fig. 3). The general pattern of the results was similar to that found for F0 = 62.5 Hz, but the differences in level at the point of equal loudness were somewhat smaller. The results of the ANOVAs are summarized in Table I. The main effect of level was significant only for the RPH vs CPH comparison. The main effect of presence or absence of the masking noise was not significant for any comparison. The main effect of identity of the varied sound was also not significant for any comparison. The interaction between the identity of the varied sound and the fixed level was significant only for the RPH vs CPH comparison. The grand mean was significantly above zero ($p < 0.05$) for all three comparisons, indicating differences in loudness at equal rms levels. Thus, as for the

TABLE I. Results of three-way ANOVAs (identity of varied sound × masker × fixed level) on level differences at the point of equal loudness measured in experiment 1. Only significant effects are listed.

	F0 = 62.5 Hz	F0 = 250 Hz
Noise vs CPH		
Varied sound	$F(1,5) = 8.1$, $p < 0.05$	⋯
Fixed level	$F(5,25) = 9.6$, $p < 0.001$	⋯
Varied sound × fixed level	$F(5,25) = 11.1$, $p < 0.001$	⋯
RPH vs CPH		
Masker	$F(1,5) = 17.1$, $p < 0.01$	⋯
Fixed level	$F(5,25) = 7.9$, $p < 0.01$	$F(5,20) = 3.7$, $p < 0.05$
Varied sound × fixed level	$F(5,25) = 4.2$, $p < 0.05$	$F(5,20) = 3.8$, $p < 0.05$
Noise vs RPH		
Varied sound	$F(1,5) = 6.6$, $p < 0.05$	⋯
Varied sound × fixed level	$F(5,25) = 3.9$, $p < 0.05$	⋯

62.5-Hz F0, the results indicate that at the point of equal loudness, the level of the noise was above that of the CPH and RPH tones and the level of the RPH tone was significantly above that of the CPH tone.

C. Discussion

At equal rms level, the CPH tone sounded louder than the noise or the RPH tone, and the RPH tone sounded louder than the noise. This was true for both F0s, although, for the 250-Hz F0, the effects were somewhat smaller than for the 62.5-Hz F0. It is not surprising that the effects of component phase are larger for the lower F0, because the number of components interacting at the output of a given auditory filter is four times greater. Thus, the envelope of the CPH tone with the lower F0 would exhibit greater modulation after auditory filtering than that of the CPH tone with the higher F0. The effects for the 62.5-Hz F0 were similar to those that we reported previously (Gockel *et al.*, 2002b), although in the former study the difference in loudness between RPH tones and noise was not significant (although it was present), whereas in the present study the difference was significant. The noise designed to mask combination tones only reduced the size of the effects slightly at the lower F0, and had no significant effect at the higher F0. Therefore, while there might be a small contribution of audible distortion products to the observed loudness differences, this cannot be the main factor involved. The effects of component phase on loudness were clear and statistically significant, but the direction of the effects was opposite to what would be predicted from BM compression. The CPH tones, the sounds with the greater crest factor (both physically and in the waveform evoked on the BM), were judged as louder than the RPH tone and the noise, both of which have lower crest factors.

The present results contrast with those of Carlyon and Datta (1997). They found the loudness of a subset of components covering the mid range of a complex tone to be greater for an S− complex than for an S+ complex. Their complex tone had an F0 of 100 Hz, with components ranging from 200 to 2000 Hz, and the central subset of components whose loudness had to be judged ranged from 900 to 1300 Hz. Thus, the 400-Hz bandwidth of the central subset of components was relatively small compared to that of our stimuli. Their results are consistent with the finding mentioned in the Introduction, that for narrowband sounds, amplitude modulation at medium rates (10–100 Hz) reduces loudness slightly, relative to that of the unmodulated sound (Zhang and Zeng, 1997; Moore *et al.*, 1998, 1999; Grimm *et al.*, 2002).

Our own results are also in the same direction as the findings mentioned in the Introduction; for broadband sounds, amplitude modulation at medium rates slightly increases loudness relative to that of the unmodulated sounds (Zhang and Zeng, 1997; Moore *et al.*, 1999; Grimm *et al.*, 2002). However, the present effect is larger than the 1–2-dB difference usually observed, especially for the lower F0. Since most earlier studies used sinusoidal amplitude modulation, one possible reason for the larger effect with the current stimuli might be the relatively large crest factor of our CPH stimuli; it was 12.4 for the 62.5-Hz F0 and 5.0 for the 250-Hz F0 (of course, the crest factor on the BM would be different from this). For broadband sounds, it appears that a large crest factor can lead to an increase in loudness. However, the loudness does not correspond to the peak value of the (physical) waveform. For the 62.5 Hz-F0, the peak value of the CPH wave is about four times greater than the peak value of the RPH wave. This corresponds to about a 12-dB difference in level, which is much larger than the observed level differences at the point of equal loudness.

Another factor which might have influenced our results is the fact that our CPH stimuli had a strong and distinct pitch corresponding to the F0 (Schouten, 1940, 1970). In contrast, for a high-frequency sinusoidal carrier modulated at a low rate, the pitch corresponding to the modulation rate is weak or absent (Ritsma, 1962; de Boer, 1976). Possibly, the existence of a strong residue pitch might increase the loudness of a sound relative to one with the same power spectrum but a weak or absent pitch. As the pitch of our CPH tones was stronger than that of the RPH tones (Lundeen and Small, 1984; Patterson, 1987; Roberts *et al.*, 2002), especially for the lower F0 (Warren and Bashford, 1981), this might have contributed to the differences in loudness at equal rms level. A problem with this account is that the subjective difference in pitch strength between the RPH tone and noise was larger than the difference in pitch strength between the CPH and RPH tones, but the mean difference in level at the point of equal loudness was larger for the latter pair of sounds. However, loudness might be a nonlinear function of pitch strength, growing rapidly with increasing pitch strength.

A third factor which might have contributed to our results is related to the annoyance of sounds. The CPH tones would have had a greater roughness than the RPH tones or the noise (Terhardt, 1974; Hellman, 1985; Zwicker and Fastl, 1999) and this might have led to greater annoyance of the CPH stimuli, and an effect on loudness. Subjects were instructed to judge loudness and ignore all other factors but they were not specifically instructed to ignore annoyance. However, it is unlikely that annoyance is the only factor influencing the effects of phase on loudness. First, several experienced listeners (including the authors), who knew about the difference between annoyance and loudness, judged the CPH stimuli to be louder than the RPH or noise stimuli at equal rms level. Second, the perceived roughness for a modulation rate of 250 Hz is substantially lower than for a modulation rate of 62.5 Hz (Zwicker and Fastl, 1999). The CPH and RPH stimuli with $F0 = 250$ Hz did not appear to differ from each other or from the noise in roughness. Thus, annoyance due to roughness is unlikely to have influenced the loudness matches for the 250-Hz F0, but effects of phase on loudness were observed.

D. Model predictions

Models of loudness perception for steady sounds generally ignore the phase spectrum of the sound (Zwicker and Scharf, 1965; Zwicker *et al.*, 1984; Moore *et al.*, 1997), and thus are unable to account for the present results. A more recent model for fluctuating sounds described by Glasberg and Moore (2002) is based on the short-term spectrum of the waveform and takes into account the effects of peripheral compression. This model predicts a *reduction* in loudness of the CPH complex relative to that of the RPH complex or the noise by about 5 to 6 phons for the 62.5-Hz F0 at medium levels. For the 250-Hz F0 a difference of about 1 phon in the same direction is predicted. For both F0s, no difference in loudness is predicted between the RPH tone and the noise. Thus, the model fails to account for our results. It is not clear whether other models for the loudness of fluctuating sounds (Zwicker, 1977; Fastl, 1993; Chalupper and Fastl, 2002; Grimm *et al.*, 2002) would do any better.

In what follows, we describe a model which correctly predicts the direction of the observed loudness differences between CPH tones and RPH tones or noise. However, the magnitudes of the effects are not accurately predicted for all conditions. The modeling was performed with the AMS/DSAM software package[1] and MATLAB.

The waveforms of the sounds used in our experiment were used as input to the model. The sampling rate was 50 kHz. The input level was varied over a wide range in 1-dB steps for each waveform. The first stage of the model is an FIR filter which simulates the combined outer and middle ear response as specified in Glasberg and Moore (2002), except that here, as our headphones had a diffuse-field response, we used a diffuse-field correction instead of the correction for frontal free-field sound incidence. The next stage was a dual-resonance nonlinear filterbank (Lopez-Poveda and Meddis, 2001) that simulates the nonlinear response of the BM at different points. We used 30 filters with characteristic frequency (CF) varying between 40 Hz and 10 kHz. The CFs were spaced according to Greenwood's (1990) frequency-place map for humans. The filter bank takes stapes motion as input, which is a linear function of the sound pressure. The output is BM velocity. The remaining stages are calculated in parallel for the 30 filter outputs. The BM velocity is lowpass filtered to simulate the displacement of the inner hair cell stereocilia, according to Eq. (1) in Sumner *et al.* (2002). The stereocilia response is converted into a receptor potential according to Eqs. (2) and (3) in Sumner *et al.* (2002). Then, the receptor potential is converted into auditory nerve (AN) spike probability (with parameters according to Table II of Sumner *et al.*). This latter stage has three parallel parts, which generate the spike probabilities for high-, medium-, and low-spontaneous rate fibers, respectively. Each of the three spike probabilities is converted into a stream of spikes, each of which is the sum of the spike activity for a population of fibers; 60, 25, and 15 fibers were used for the high-, medium-, and low-spontaneous rate fibers, respectively. The three streams of spikes were then added in each frequency channel and the simulated AN activity was summed across all 30 channels. The result is an overall neural activity pattern (NAP) for each sound as a function of time. All stages to this point were simulated with AMS/DSAM. The following calculations were done in MATLAB.

For each waveform, various measures based on the NAP were calculated. These measures were chosen as they had been suggested as correlates of loudness by previous researchers. Call a given measure (e.g., the overall mean of the NAP) M. For a given pair of sounds (e.g., CPH vs RPH with F0 = 62.5 Hz), we determined the difference in level required to give the same value of M for the two sounds, for each level of the fixed sound used in the experiment. If M is closely related to loudness, then the differences in level so determined should match those obtained in the experiment at the point of equal loudness. We tried to find a measure for which this was the case, as closely as possible, over a wide range of levels, for both F0s and for all pairs of sounds that were compared. The spike generation process in the model is stochastic and so we averaged across simulations to produce stable response values. For the CPH stimulus, ten NAPs were calculated for each input level. For each of the ten NAPs, the measures were calculated individually and then averaged. The measures which will be discussed are all based on this averaging process. For the RPH and the noise stimuli, three NAPs were calculated for each input level for each of the ten different waveforms used in the experiment, and the measures were calculated for each of them individually. The measures which will be discussed are the averages across the 30 values.

The first measures considered were the maximum value in the NAP, the average value of the NAP, the mean value of all peaks in the NAP, and the mean value of all peaks above the 80th percentile. Measures comparable to these have been used previously to model the loudness of steady and fluctuating sounds (Fastl, 1977; Zwicker, 1977; Howes, 1979; Fastl, 1993; Zwicker and Fastl, 1999). For inputs with equal rms values, the mean of the NAP for the CPH tone was below that for the RPH tone or the noise. This was true for all input levels, though the amount of the difference varied as a function of level. Thus, this measure would predict an effect on loudness in the opposite direction to that found. This indicates that, contrary to what is often assumed, loudness is not directly related to the overall activity in the auditory nerve. The two measures involving peaks in the NAP also predicted an effect in the wrong direction. The measure based on the single largest value in the NAP, which occurred close to the onset of the response, did predict an effect in the right direction. However, this is unlikely to be the basis for loudness judgments, since loudness increases with duration up to 200 ms (Scharf, 1978; Florentine *et al.*, 1996), and the size of the predicted effect was much too large.

The above measures were derived directly from the raw NAP and they do not reflect the fact that the NAP of the CPH stimulus is clearly periodic; it looks like a saw-tooth function with a period corresponding to 1/F0. Moreover, perceived loudness is presumably based on some form of running average of the NAP. To accommodate these factors, we examined the effect of asymmetric temporal smoothing of the NAP, using an operation resembling the calculation of gain in an automatic gain control system with different attack and release times. The operation is similar to the one used by Glasberg and Moore (2002) to determine short-term loudness from what they called "instantaneous loudness" (which was assumed not to be directly accessible for conscious processing). The temporal smoothing of the NAP was done in the following way. We define SM_n as the smoothed NAP at the nth sample point, and NAP_n as the raw NAP at the nth sample point.

If $NAP_n > SM_{n-1}$ (corresponding to an attack), then

$$SM_n = a\text{NAP}_n + (1-a)\text{SM}_{n-1}, \quad (1)$$

where a is a constant related to the attack time, T_a (in ms):

$$a = 1 - e^{1/(Fs \cdot T_a)} \quad (2)$$

with Fs corresponding to the number of samples per ms (50 in this case).

If $\text{NAP}_n < \text{SM}_{n-1}$ (corresponding to a release), then

$$SM_n = r\text{NAP}_n + (1-r)\text{SM}_{n-1}, \quad (3)$$

where r is a constant related to the release time, T_r (in ms):

$$r = 1 - e^{1/(Fs \cdot T_r)}. \quad (4)$$

Based on the smoothed NAPs, the same measures were calculated as before. The mean across the smoothed NAPs gave predictions in the right direction. The attack and release times that produced the best fits to our data were 3.1 and 49.49 ms, respectively. The release time happens to be identical to the one chosen by Glasberg and Moore (2002), while the attack time is about one-seventh of their value. The predictions obtained with these values are plotted as dashed lines in the right-hand columns of Figs. 2 and 3. For loudness matches involving the CPH tone at medium levels, the predicted effect is larger than the obtained effect for the low F0, but it is smaller than the obtained effect for the higher F0. No predictions are plotted for a fixed input level of 80 dB SPL for the CPH tone. The reason is that, for input levels above about 80 dB SPL, the smoothed (and the raw) NAP statistics of the RPH tone and the noise hardly increased with increasing level; to get the same mean value as for the CPH tone, the level of the RPH tone had to be about 40 dB above that of the CPH tone. This may reflect neural saturation in the model, perhaps resulting from inadequate numbers of fibers or insufficient dynamic range in the fibers with low spontaneous rates. For loudness matches between RPH tones and noise, the model failed to explain the effect observed with the lower F0; for the higher F0, there was a small effect in the right direction. The predicted effect for the higher F0 does not seem to be a consequence of the temporal smoothing, since there was a similar effect with the ''raw'' (unsmoothed) NAPs.

In summary, a model based on the ''raw'' (unsmoothed) NAPs produced results opposite in direction to the observed effects. This indicates that, contrary to what is often assumed, loudness is not directly related to the overall activity in the auditory nerve. A similar conclusion has been reached by Relkin and Doucet (1997). However, a model based on NAPs smoothed with a fast attack and slow release produced results in qualitative agreement with the data.

III. EXPERIMENT 2

The second experiment was designed to test whether the effectiveness of our stimuli as forward maskers would reflect their loudness, i.e., whether the CPH tone would be a more effective forward masker than the RPH tone and the noise. In simultaneous masking, using the same complex tones as in the present study, Gockel *et al.* (2002a) found that the CPH tone was a less efficient masker of a noise signal than the RPH tone for F0 = 62.5 Hz, while the opposite was true for F0 = 250 Hz. This was explained in terms of the temporal structure of the stimuli, which allows ''listening in the dips'' for the lower F0. Such listening in the dips is usually assumed to be irrelevant in forward masking, since the amount of forward masking reflects the average excitation or neural activity evoked by the masker over a relatively long time period (Zwicker, 1977; Carlyon and Datta, 1997). Therefore, it is not obvious that the effects of masker component phase will be the same for simultaneous masking and for forward masking. The question addressed here was whether the longer term activity evoked by the masker, as measured in forward masking, is directly related to loudness.

As mentioned in the Introduction, Carlyon and Datta (1997) measured the forward masking produced by S+ and S− complexes and found that, at least for high masker levels, the S+ complex (large peak factor on the BM) produced less forward masking than the S− complex (low peak factor on the BM). They interpreted this result in terms of the effects of fast-acting compression on the effective excitation evoked by the complexes. Their results led us to expect that the CPH tone would be a less effective forward masker than the RPH tone or noise, despite the greater loudness of the CPH tone.

A. Method

1. Stimuli

The signal was a 20-ms sinusoid with a frequency of 702, 1114, 1768, 2806, or 4454 Hz. The middle frequency, 1768 Hz, corresponded to the geometric mean of the frequency range covered by the forward masker, which extended from 625 to 5000 Hz; the other signal frequencies were offset by two semitones, and ten semitones, respectively, from the lower and upper end of the masker's frequency range. The signal had 10-ms, raised-cosine onset and offset ramps and no steady-state portion; it followed the masker without a silent gap.

Three stimuli were used as forward maskers: the CPH and RPH tones with an F0 of 62.5 Hz, and the Gaussian noise. They were filtered in the same way as in experiment 1. Their duration was 208 ms (corresponding to 13 periods), including 10-ms raised-cosine onset and offset ramps. To test whether there was an effect of time elapsed between the last peak in the masker waveform and the onset of the signal, the CPH masker began at two different points in its 16-ms period; either 0 or 12 ms after the first peak (conditions CPH 0 and CPH 12, respectively). To measure level dependence and to allow differences in masker effectiveness to be expressed as differences in effective masker level (Houtgast, 1974; Moore and Glasberg, 1983), the masker was presented at 30, 50, 68, 77, and 85 dB SPL. If compression plays a major role, then its effects should be most obvious at medium to high levels, and less clear or absent at low levels. The method of stimulus generation and the equipment were the same as for experiment 1.

2. Procedure

A two-interval two-alternative forced-choice task was used. The forward masker was presented in both intervals, which were marked by lights and separated by 300 ms of

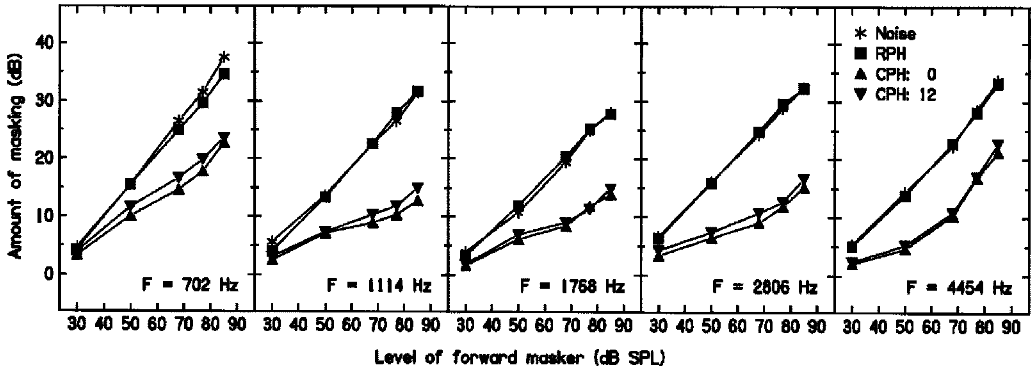

FIG. 4. Forward masking results (experiment 2), averaged across six subjects, as a function of masker level. The asterisks, squares, upward-pointing triangles, and downward-pointing triangles indicate results for the noise, the RPH tone, the CPH tone with 0-ms offset in the period, and the CPH tone with 12-ms offset in the period, respectively. The five panels show the results for the five different signal frequencies, as indicated in each panel.

silence. The subject was required to indicate which interval contained the signal. Feedback was provided following each response. The level of the signal was adjusted using a three-down one-up adaptive rule (Levitt, 1971) tracking 79% correct responses. At the beginning of each threshold measurement, the signal level was 10 dB above that of the masker, except for the lowest masker level, where it was 20 dB above the level of the masker. Initially, the signal level was increased or decreased in 8-dB steps. Following two reversals, the step size was reduced to 4 dB, and following two more reversals, it was reduced to 2 dB. Eight reversals were obtained with this final step size. The threshold was defined as the mean of the signal levels at the last eight reversals.

At least three threshold estimates were obtained for each condition from each subject. If the thresholds for a given subject and condition varied by more than 6 dB, two additional threshold estimates were obtained. The thresholds reported correspond to the mean of these three to five estimates, for each condition and subject. The total duration of a single session was about 2 h, including rest times. For a given signal frequency, one threshold was collected for each condition before the signal frequency was changed. Within a block with fixed signal frequency, one threshold was collected for each of the four maskers before the masker level was changed. Within the four runs at a given level, the identity of the masker was chosen randomly. The masker level was changed in either ascending or descending order between groups of four runs, in order to avoid having a run with a high-level masker immediately before a run with a low-level masker. The direction of level change was reversed when the next thresholds were collected for this signal frequency. The order of signal frequencies was counterbalanced across subjects. One threshold was obtained for each condition in turn, before additional measurements were obtained in any other condition. Subjects participated in at least one threshold measurement in each condition before data collection proper commenced. This took approximately 8 h. Some of the subjects showed strong learning effects, in which case their second threshold measurement in each condition was regarded as an additional practice run and the experiment proper started after that.

Absolute thresholds for each signal were measured after the main experiment was completed, using the same three-down one-up procedure as before. Four threshold estimates were obtained for each signal and each subject.

3. Subjects

Six subjects participated in all conditions, two of whom had substantial experience in other psychoacoustic experiments. One of them was the first author, and one of them was a musician. Their ages ranged from 18 to 41 years, and they all had normal hearing at audiometric frequencies between 500 and 4000 Hz.

B. Results

Figure 4 shows the masking produced by each masker as a function of masker level. Each panel shows results averaged across the six subjects for one signal frequency. The RPH tone (squares) and the noise (asterisks) produced similar amounts of masking, increasing at a steady rate from about 5 to between 27 and 36 dB with increasing masker level; the size of the increase depended on the signal frequency. The two CPH maskers (upward and inverted triangles) produced far less masking than the RPH tone or noise. The amount of masking increased from about 3 to between 14 and 23 dB with increasing masker level. The rate of increase depended on signal frequency and masker level. In summary, the CPH tones produced far less forward masking than the RPH tone or the noise, despite the fact that they were clearly louder.

To assess the statistical significance of these effects, a three-way ANOVA (masker identity×signal frequency ×masker level) for repeated measures was conducted on the amount of masking. The main effect of masker identity was highly significant [$F(3,15)=330.6$, $p<0.001$], as were the main effects of signal frequency [$F(4,20)=7.2$, $p<0.001$] and of level [$F(4,20)=280.4$, $p<0.001$]. There was also a significant interaction between masker identity and signal frequency [$F(12,60)=3.4$, $p<0.05$], indicating that the relative amount of masking produced by the different maskers depended on the signal frequency. The interaction between level and signal frequency was highly significant [$F(16,80)=11.6$, $p<0.001$], indicating that the increase in amount of masking with level depended on signal frequency. Furthermore, the interaction between masker identity and masker level was highly significant [$F(12,60)=81.6$, p

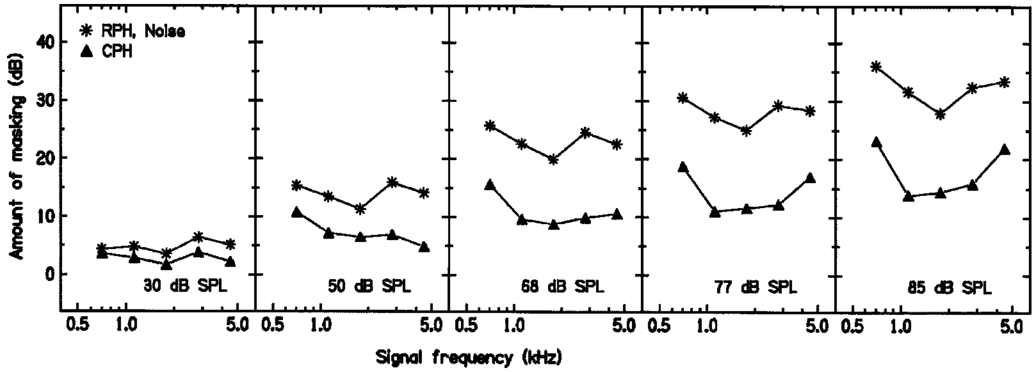

FIG. 5. Mean amount of forward masking, plotted as a function of the frequency of the signal. The asterisks show results averaged for the RPH-tone and noise maskers; the upward-pointing triangles show results averaged for the two CPH-tone maskers. The five panels show the results for the five different masker levels, as indicated in each panel.

<0.001], showing that the increase in amount of masking with increasing level differed across maskers. Finally, the three-way interaction was significant [$F(48,240)=2.9$, $p=0.001$].

Post hoc paired comparisons based on Fisher's least significant difference procedure were made across the different masker conditions. As they showed no significant difference between the RPH tone and noise masker, the data were averaged across these two masker conditions for subsequent analyses. Averaged across level and signal frequency, the amount of masking produced by the CPH 0 masker (10.1 dB) was slightly smaller than that produced by the CPH 12 masker (11.1 dB). Thus, the time between the last envelope peak in the masker's waveform and the beginning of the signal, being shorter for the CPH 12 condition, did have a small but significant effect ($p<0.001$). The fact that the difference was very small is consistent with the idea that forward masking depends upon the average effect of the masker over a relatively long time interval. Since thresholds for the CPH 0 and CPH 12 maskers were almost the same, the data for these two conditions were averaged for subsequent analyses.

To enable comparison of the masking effects across the different signal frequencies, Fig. 5 shows the amount of masking, averaged across RPH tone and noise maskers (asterisks), and averaged across the two CPH tone maskers (upward-pointing triangles) as a function of signal frequency; each panel shows the data for one masker level. Even for the lowest masker level of 30 dB SPL (left-most panel), the amount of masking produced by the RPH tone and noise masker was higher than for the CPH tone maskers, albeit only by about 2 dB. This difference was, however, significant; a two-way ANOVA (masker identity $\times$signal frequency) for the masker level of 30 dB SPL showed a highly significant main effect of masker identity [$F(1,5)=112.9$, $p<0.001$]. The interaction between masker identity and signal frequency was also significant [$F(4,20)=6.1$, $p<0.01$]. With increasing masker level (from left to right panels) the difference in masking between the RPH/noise masker and the CPH maskers increased markedly. At the highest masker level of 85 dB SPL, the RPH and noise masker produced about 32 dB of masking (averaged across signal frequency), while the CPH maskers produced about 17.8 dB of masking. The effect of masker identity was largest at 77 and 85 dB SPL for the signal frequencies of 1114 and 2806 Hz (the frequencies adjacent to the center frequency), amounting to 16–18 dB.

For the CPH masker, the masking pattern became more bowl shaped with increasing masker level, i.e., at high levels the amount of masking was greater for the outermost signal frequencies than for the three mid-frequency signals. For the RPH tone and the noise masker, it was a bent V-shaped pattern, with a clear minimum at the center frequency and a "bump" at 2806 Hz. The increased masking at 2806 Hz may be related to the peak in the response of the headphones around 3000 Hz. This peak is reflected in the long-term excitation pattern of the masker, as shown in Fig. 1(a) of Gockel *et al.* (2002a). Perhaps this peak was enhanced by a release from lateral suppression leading to increased forward masking (Houtgast, 1974; Moore and Glasberg, 1983). If this is the case, it remains unclear why no "bump" occurred for the CPH stimuli. See below for further discussion of the effects of suppression.

To express the differences in amount of masking as differences in effective level of the maskers (Houtgast, 1974; Moore and Glasberg, 1983), we estimated the increase in level of the CPH masker necessary for it to produce the same amount of masking as the RPH tone and noise masker, for the four higher levels of the CPH masker. The estimates were based on interpolation, using the short line segments connecting the data for adjacent masker levels, as in Fig. 4. Figure 6 shows these increases plotted as a function of the level of the CPH masker, with signal frequency as parameter. The CPH masker had to have a level up to 35 dB above the level of the RPH or noise masker to produce equal masking. For all levels except the lowest, the necessary increase in CPH masker level was greatest for the 2806-Hz signal frequency. The second largest difference occurred for the 1114-Hz signal frequency. This is a consequence of the V-shaped masking pattern for the RPH and noise maskers, and the bowl-shaped masking pattern of the CPH maskers. So, interestingly, the effectiveness of the CPH maskers was reduced most relative to that of the RPH or noise maskers at the two signal frequencies adjacent to the center frequency, rather than at the center frequency.

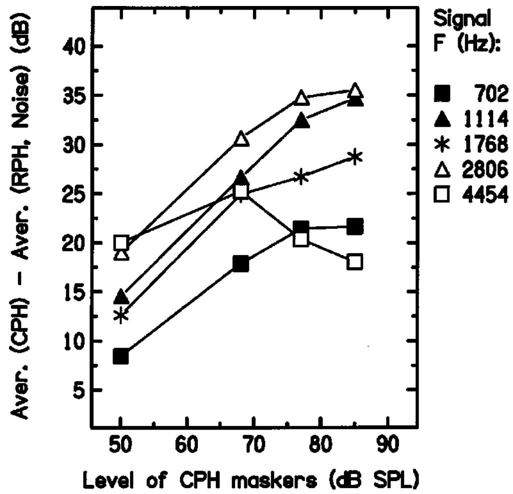

FIG. 6. Differences in effective level of the forward maskers. The difference in level of the CPH maskers and the RPH/noise maskers at the point where they produced an equal amount of forward masking is plotted as a function of the level of the CPH maskers. The five curves correspond to the five different signal frequencies, as indicated in the legend.

C. Discussion

The results of the forward-masking experiment are consistent with those of Carlyon and Datta (1997) indicating that stimuli evoking waveforms with a large crest factor on the BM (S+ or CPH tones) are less effective forward maskers than stimuli evoking waveforms with small crest factors (S− or RPH tones and noise). Carlyon and Datta explained this effect in terms of fast-acting compression on the BM, which would result in lower average excitation for waveforms with large peak factors. However, it is remarkable that the differences in effective level of the CPH and RPH maskers could be as much as 35 dB. It is difficult to conceive how such a large effect could be explained in terms of BM compression alone.

To see why this is so, consider a hypothetical example. In what follows, for simplicity, we consider only the envelopes of the hypothetical signals, but the arguments would be similar if the waveforms were considered. Assume that a signal has an envelope that is very peaky: the envelope consists of low-amplitude portions (amplitude = 0.01 units) lasting 0.99 time units, and brief high-amplitude portions (amplitude = 1 unit) lasting 0.01 time units. The rms value of the envelope is 0.10049 units and the crest factor of the envelope is 9.95. This signal is compared with a signal of constant envelope amplitude and the same rms value. Assume that each of these two signals is passed through an instantaneous compressor with a compression ratio of 3 over the whole range of input amplitudes and which applies a gain of unity (0 dB) when the input envelope amplitude is one unit. Following compression, the low-amplitude portions of the first signal have an amplitude of 0.215, the peak amplitude stays the same (unity), and the rms value of the envelope becomes 0.236. For the second signal, the (constant) envelope amplitude is increased to 0.465. To equate the rms output values of the two signals (allowing for the compression ratio of 3), the input level of the peaky stimulus would have to be increased by 17.7 dB. This is about one-half of the largest difference in effective masker level inferred from the forward masking data. The required change in input level needed to equate the output levels of the two stimuli would be somewhat larger if the compression ratio were higher. For example, a compression ratio of 5 would lead to a required change in input level of the peaky signal of 18.9 dB. We conclude that it is unlikely that the differences in effective level of our maskers inferred from the forward masking data can be accounted for solely by the effects of fast-acting compression on the BM.

It is possible that suppression has a strong influence on the masking produced by broadband CPH maskers, as used in the present experiments. When the masker evokes a highly peaky waveform on the BM, and the peaks are approximately synchronous at the outputs of different auditory filters, the peaks at the outputs of filters tuned away from the signal frequency may suppress peaks at the output of the on-frequency filter, and this may reduce the forward masking produced by the peaks. Also, if the masker dips are synchronous across channels, then suppression will be weak during the dips, as the strength of suppression depends partly on overall level (Javel, 1981; Sellick *et al.*, 1982; Recio and Rhode, 2000). If suppression is at its strongest when the masker itself is at its highest short-term level, and is weakest when the masker is at its lowest short-term level, this could produce a strong reduction of the effective level of a peaky masker, and thus contribute to differences in effectiveness of the CPH tones and RPH tones/noise as forward maskers.

Suppression within a broadband sound occurs both from low frequencies towards higher ones and vice versa, although the former effect occurs mainly for medium to high overall levels (Houtgast, 1974; Javel, 1981; Ruggero *et al.*, 1992). Suppression will be strongest in a given frequency region when there is suppression from both lower and higher regions, i.e., when the frequency is well within the spectral range of the sound. Towards the spectral edge of the sound, there is effectively a release from suppression, which results in enhancement of the edge (Houtgast, 1972, 1974). This can account for our finding that the amount of masking was more for the signal frequencies towards the edge of the spectral range of the maskers than for the signal frequencies well within that spectral range.

IV. GENERAL DISCUSSION

At equal rms level, the CPH maskers clearly produced less forward masking than the RPH or noise maskers, even though the former were louder than the latter. Thus, the loudness of the stimuli and their effectiveness as forward maskers cannot be based on identical processes. It is generally assumed that both the loudness of a sound and the effectiveness of that sound as a forward masker are related to the excitation evoked by the sound in the peripheral auditory system. Our results suggest that one or both of these assumptions is incorrect.

The model based on neural activity patterns (NAPs), introduced in Sec. II D, was able to predict some aspects of the loudness-matching data. To achieve this, we assumed that loudness is based on the output of a smoothing process applied to the NAP. The smoothing was implemented using a sliding temporal integrator analogous to an automatic gain control system, with a fast attack time and slower decay time. Effectively, this system responds strongly to the peaks in the NAP evoked by the CPH tone, thus increasing the overall response to that tone relative to the response evoked by the RPH tone or noise.

It is usually assumed that forward masking is related to the excitation or neural activity evoked by the masker, averaged or smoothed over a time interval of a few tens of milliseconds (Zwicker, 1977; Carlyon and Datta, 1997; Zwicker and Fastl, 1999). In fact, forward masking is often modeled using the concept of a sliding temporal integrator or window (Moore *et al.*, 1988; Plack and Moore, 1990; Moore and Oxenham, 1998; Plack and Oxenham, 1998). The temporal integrator in these models is asymmetric in time, but is usually assumed to be linear. Such models can correctly predict that CPH stimuli produce less forward masking than RPH or noise stimuli, provided that they include a compressive nonlinearity prior to the temporal integrator. However, a linear temporal integrator would not account for our loudness data. To model these data, we had to assume a nonlinear smoothing process, which behaves differently depending on whether the momentary input is increasing or decreasing. This mechanism emphasizes the peaks in the input (in our model the peaks in the NAP), and this leads to a greater output for the CPH than for the RPH or noise stimuli.

V. SUMMARY AND CONCLUSIONS

In experiment 1, we obtained loudness matches between complex tones with components added in cosine phase (CPH) and random phase (RPH), and between those complex tones and noise. Two fundamental frequencies (F0s) were used: 62.5 and 250 Hz. The stimuli were bandpass filtered so that only unresolved harmonics were present for the tones. For a given F0, excitation patterns calculated from the power spectrum were essentially identical for all stimuli. The results showed the following.

(1) For a given overall level, the CPH tone was louder than the RPH tone and the noise, and the RPH tone was louder than the noise. At the point of equal loudness, the level of the noise was greater than that of the CPH tone by up to 7 dB.
(2) The effects were greatest for medium to high levels.
(3) The effects were only slightly reduced by the addition of a background noise intended to mask combination tones.
(4) The effects were greater for F0 = 62.5 Hz than for F0 = 250 Hz.

The general pattern of the results could be accounted for using a model based on simulated neural activity patterns (NAPs), using a temporal smoothing mechanism resembling the operation of an automatic gain control system with a fast attack time and slow decay time. However, the model failed to predict the observed difference in loudness of the RPH-tone and noise for F0 = 62.5 Hz.

Experiment 2 compared the effectiveness of the same sounds (F0 = 62.5 Hz only) as forward maskers, using several signal frequencies within the spectral range of the maskers. The results showed the following.

(1) The noise and RPH tone produced similar amounts of forward masking and both produced more masking than the CPH tone.
(2) The results were only slightly affected by the time during the period of the CPH masker when that masker was turned off.
(3) The rate of growth of forward masking was markedly greater for the noise and RPH tone than for the CPH tone.
(4) For high masker levels, the amount of masking was greater for signal frequencies close to the spectral edges of the masker than for signal frequencies well within the spectral range of the masker.
(5) For high masker levels, the CPH masker needed to have a level as much as 35 dB above that of the RPH or noise maskers to produce equal amounts of masking.

Taken together, the results of the two experiments indicate that, for a given overall level, peaky sounds (CPH tones) are louder and produce less forward masking than less peaky sounds (RPH tones or noise). Thus, it cannot be the case that loudness and forward masking are determined by the same peripheral processes, such as the level of excitation or neural activity. We suggest that the representation of these sounds in the auditory system may be affected by at least two processes: fast-acting cochlear compression and suppression.

ACKNOWLEDGMENTS

This work was supported by the MRC (Grant No. G9900362). We thank Brian Glasberg for calculating the predictions of the Glasberg and Moore loudness model for our stimuli. We also thank Marjorie Leek, Rhona Hellman, and an anonymous reviewer for helpful comments on an earlier version of this paper.

[1]The AMS/DSAM software is available over the Internet at http://www.essex.ac.uk/psychology/hearinglab/dsam/index.htm

Bauch, H. (**1956**). ‘‘Die Bedeutung der Frequenzgruppe für die Lautheit von Klängen,’’ Acustica **6**, 40–45.
Buunen, T. J. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (**1974**). ‘‘Phase effects in a three-component signal,’’ J. Acoust. Soc. Am. **55**, 297–303.
Carlyon, R. P., and Datta, A. J. (**1997**). ‘‘Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,’’ J. Acoust. Soc. Am. **101**, 3636–3647.
Chalupper, J., and Fastl, H. (**2002**). ‘‘Dynamic loudness model (DLM) for normal and hearing-impaired listeners,’’ Acust. Acta Acust. **88**, 378–386.
de Boer, E. (**1976**). ‘‘On the ‘residue’ and auditory pitch perception,’’ in *Handbook of Sensory Physiology, Vol. 3*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin).
Fastl, H. (**1975**). ‘‘Loudness and masking patterns of narrow noise bands,’’ Acustica **33**, 266–271.
Fastl, H. (**1977**). ‘‘Loudness of running speech,’’ J. Audiol. Tech. **16**, 2–13.
Fastl, H. (**1993**). ‘‘Loudness evaluation by subjects and by a loudness meter,’’ in *Sensory Research-Multimodal Perspectives*, edited by R. T. Verrillo (Erlbaum, Hillsdale, NJ).
Fletcher, H., and Munson, W. A. (**1933**). ‘‘Loudness, its definition, measurement and calculation,’’ J. Acoust. Soc. Am. **5**, 82–108.
Florentine, M., Buus, S., and Poulsen, T. (**1996**). ‘‘Temporal integration of loudness as a function of level,’’ J. Acoust. Soc. Am. **99**, 1633–1644.

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. (**2002**). ''A model of loudness applicable to time-varying sounds,'' J. Audio Eng. Soc. **50**, 331–342.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002a**). ''Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression,'' J. Acoust. Soc. Am. **111**, 2759–2770.
Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2002b**). ''Influence of component phase on the loudness of complex tones,'' Acust. Acta Acust. **88**, 369–377.
Goldstein, J. L. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.
Greenwood, D. D. (**1971**). ''Aural combination tones and auditory masking,'' J. Acoust. Soc. Am. **50**, 502–543.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Grimm, G., Hohmann, V., and Verhey, J. L. (**2002**). ''Loudness of fluctuating sounds,'' Acust. Acta Acust. **88**, 359–368.
Hellman, R. P. (**1985**). ''Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness,'' J. Acoust. Soc. Am. **77**, 1497–1504.
Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.
Houtgast, T. (**1974**). ''Lateral suppression in hearing,'' Ph.D. thesis, Free University of Amsterdam.
Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, Belmont, CA).
Howes, W. L. (**1979**). *Overall Loudness of Steady Sounds According to Theory and Experiment* (National Aeronautics and Space Administration, Washington, DC).
Javel, E. (**1981**). ''Suppression of auditory nerve responses: I. Temporal analysis, intensity effects and suppression contours,'' J. Acoust. Soc. Am. **69**, 1735–1745.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.
Lundeen, C., and Small, A. M. (**1984**). ''The influence of temporal cues on the strength of periodicity pitches,'' J. Acoust. Soc. Am. **75**, 1578–1587.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego).
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Masking patterns of synthetic vowels in simultaneous and forward masking,'' J. Acoust. Soc. Am. **73**, 906–917.
Moore, B. C. J., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
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., Launer, S., Vickers, D., and Baer, T. (**1998**). ''Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform and overall level,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London).
Moore, B. C. J., Vickers, D. A., Baer, T., and Launer, S. (**1999**). ''Factors affecting the loudness of modulated sounds,'' J. Acoust. Soc. Am. **105**, 2757–2772.
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Pressnitzer, D., and Patterson, R. D. (**2001**). ''Distortion products and the pitch of harmonic complex tones,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht).
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
Relkin, E. M., and Doucet, J. R. (**1997**). ''Is loudness simply proportional to the auditory nerve spike count?'' J. Acoust. Soc. Am. **191**, 2735–2740.
Ritsma, R. J. (**1962**). ''Existence region of the tonal residue. I,'' J. Acoust. Soc. Am. **34**, 1224–1229.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). ''Primitive stream segregation of tone sequences without differences in fundamental frequency or passband,'' J. Acoust. Soc. Am. **112**, 2074–2085.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992**). ''Basilar membrane responses to two-tone and broadband stimuli,'' Philos. Trans. R. Soc. Lond. B **336**, 307–315.
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.
Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception, Volume IV. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).
Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. Kon. Akad. Wetenschap **43**, 991–999.
Schouten, J. F. (**1970**). ''The residue revisited,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands).
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Stevens, S. S. (**1956**). ''Calculation of the loudness of complex noise,'' J. Acoust. Soc. Am. **28**, 807–832.
Stevens, S. S. (**1972**). ''Perceived level of noise by Mark VII and decibels (E),'' J. Acoust. Soc. Am. **51**, 575–601.
Stevens, S. S., and Greenbaum, H. B. (**1966**). ''Regression effect in psychophysical judgment,'' Percept. Psychophys. **1**, 439–446.
Stevens, S. S., and Guirao, M. (**1967**). ''Loudness functions under inhibition,'' Percept. Psychophys. **2**, 459–465.
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.
Terhardt, E. (**1974**). ''On the perception of periodic sound fluctuations (roughness),'' Acustica **30**, 201–213.
Warren, R. M., and Bashford, Jr., J. A. (**1981**). ''Perception of acoustic iterance: pitch and infrapitch,'' Percept. Psychophys. **29**, 395–402.
Zhang, C., and Zeng, F.-G. (**1997**). ''Loudness of dynamic stimuli in acoustic and electric hearing,'' J. Acoust. Soc. Am. **102**, 2925–2934.
Zwicker, E. (**1958**). ''Über psychologische und methodische Grundlagen der Lautheit,'' Acustica **8**, 237–258.
Zwicker, E. (**1977**). ''Procedure for calculating loudness of temporally variable sounds,'' J. Acoust. Soc. Am. **62**, 675–682.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics—Facts and Models* (Springer-Verlag, Berlin).
Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.
Zwicker, E., Fastl, H., and Dallmayr, C. (**1984**). ''BASIC-Program for calculating the loudness of sounds from their 1/3-oct. band spectra according to ISO 532B,'' Acustica **55**, 63–67.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.