Behavioral/Systems/Cognitive

Masking by Inaudible Sounds and the Linearity of Temporal Summation

Christopher J. Plack,[1] **Andrew J. Oxenham,**[2] **and Vit Drga**[3]

[1]Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom, [2]Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455, and [3]School of Psychology, University of St. Andrews, St. Mary's College, St. Andrews, Fife, KY16 9JP, United Kingdom

Many natural sounds, including speech and animal vocalizations, involve rapid sequences that vary in spectrum and amplitude. Each sound within a sequence has the potential to affect the audibility of subsequent sounds in a process known as forward masking. Little is known about the neural mechanisms underlying forward masking, particularly in more realistic situations in which multiple sounds follow each other in rapid succession. A parsimonious hypothesis is that the effects of consecutive sounds combine linearly, so that the total masking effect is a simple sum of the contributions from the individual maskers. The experiment reported here tests a counterintuitive prediction of this linear-summation hypothesis, namely that a sound that itself is inaudible should, under certain circumstances, affect the audibility of subsequent sounds. The results show that, when two forward maskers are combined, the second of the two maskers can continue to produce substantial masking, even when it is completely masked by the first masker. Thus, inaudible sounds can affect the perception of subsequent sounds. A model incorporating instantaneous compression (reflecting the nonlinear response of the basilar membrane in the cochlea), followed by linear summation of the effects of the maskers, provides a good account of the data. Despite the presence of multiple sources of nonlinearity in the auditory system, masking effects by sequential sounds combine in a manner that is well captured by a time-invariant linear system.

Key words: hearing; auditory; psychophysics; temporal; adaptation; summation

Introduction

Our sensitivity to sound is impaired by previous moderate stimulation for durations up to ~200 ms. This phenomenon, known as "forward masking," is usually quantified by measuring how much the just-detectable level (or threshold) of a target sound is raised by the previous presentation of a masker sound (Zwislocki et al., 1959; Zwicker and Fastl, 1972). Forward masking plays an important role in the perception of speech and environmental sounds that fluctuate over time.

For a theory of forward masking to be useful in predicting the perception of everyday sounds, it must be able to predict how multiple forward maskers, such as a succession of speech sounds, interact with each other. The simplest possible model is one of linear summation, whereby the effects of forward maskers combine in an additive and linear way. Previous work in this area (Penner and Shiffrin, 1980; Humes and Jesteadt, 1989; Cokely and Humes, 1993; Oxenham and Moore, 1994) has produced results that are broadly consistent with the expected effects of linear summation, provided the stimuli are temporally nonoverlapping and are subject to a compressive nonlinearity, similar to

that observed in the vibration of the basilar membrane in the cochlea (Ruggero et al., 1997), before summation. Similarly, the effects of masker and signal duration in forward masking, and their interactions with overall level, have also been successfully modeled using a linear-summation model with front-end compression (Oxenham and Plack, 2000; Oxenham, 2001). Conversely, responses within the auditory system are generally far from linear, with time-dependent aspects to the responses of individual neurons evident from the auditory nerve (Smith, 1977) up to auditory cortex (Brosch and Schreiner, 2000; Ulanovsky et al., 2004).

One interesting prediction of a linear-summation model is that a masker can continue to contribute to masking even if it is inaudible. In particular, because the effect of one masker is not influenced by preceding events, it should continue to exert a masking effect, even if it is itself masked by a preceding sound. To our knowledge, there are no examples in the literature of inaudible sounds affecting subsequent perception. However, there are some examples from visual perception of subthreshold stimuli influencing the perception of subsequent stimuli. For instance, it has been shown that adaptation by a grating with a spatial frequency too high to be resolved by the visual system can nevertheless reduce observers' sensitivity to a resolvable grating at the same orientation, relative to sensitivity at the orthogonal orientation (He and MacLeod, 2001). Similarly, stimuli with undetectable flicker frequencies can cause adaptation to flicker at detectable frequencies (Shady et al., 2004).

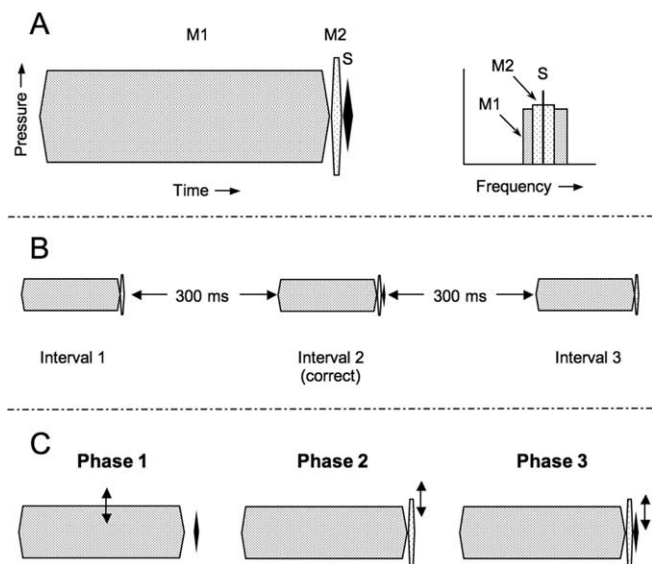The experiment presented here investigated the temporal in-

Figure 1. The stimuli and procedure used in the experiment. ***A***, A schematic illustration of the temporal and spectral characteristics of the stimuli. ***B***, The temporal presentation of the stimuli for a single trial in the three-interval task (the combined-masker conditions in phase 3 are shown). The listener had to select in which of the three intervals (chosen at random) the signal was presented. ***C***, The procedure for the three phases of the experiment. In phase 1, the level of M1 was varied to find the level needed to mask the signal. In phase 2, the level of M2 was varied to find its threshold in the presence of M1. In phase 3, the signal was varied to find its threshold in the presence of M1, M2, and M1+M2 (only the last condition is illustrated).

teractions of the masking effects of two acoustic forward maskers. The conditions were specifically selected to address the potential influence of inaudible sounds on masking as a test of the hypothesis that sequential sounds interact linearly within the auditory system.

Materials and Methods

Stimuli. Two contiguous maskers were presented shortly before a brief pure-tone signal (Fig. 1*A*). The maskers were bands of Gaussian noise. The first masker (M1) was bandpass filtered between 2800 and 5600 Hz (3 dB cutoffs, 90 dB/octave), and the second masker (M2) was bandpass filtered between 3400 and 4800 Hz. The spectral configurations were chosen so that M1, M2, and the signal would be clearly distinguishable from each other. M1 had a total duration of 200 ms, including 2 ms raised-cosine onset and offset ramps. M2 had a total duration of 6 ms, including 2 ms onset and offset ramps. The 4 kHz pure-tone signal (S) had a total duration of 4 ms, including 2 ms onset and offset ramps (no steady state). The end of M1 coincided with the start of M2. Masker gating occurred after filtering. The silent interval between the end of M2 and the start of the signal was 0 ms except for one listener (L5), for whom the interval was 10 ms. A longer gap was selected for L5 because this listener showed initially very poor detection of the signal at the short gap, suggesting that "confusion" effects (confusing the signal for the masker) may have influenced thresholds (Moore and Glasberg, 1985; Neff, 1985, 1986). A longer gap reduced this possibility and improved signal detectability.

Stimuli were generated digitally and were output by an RME (Haimhausen, Germany) Digi96/8 PAD 24-bit soundcard set at a clocking rate of 48 kHz. The headphone output of the soundcard was fed via a patch panel in the sound booth wall to Sennheiser (Wedemark, Germany) 580 headphones without filtering or amplification. Stimuli were presented to the listener's right ear.

Procedure. The experimental procedure (Fig. 1*B,C*) was similar to that used in a previous study (Plack and O'Hanlon, 2003). On each trial, listeners were presented with three observation intervals separated by 300 ms. Two intervals contained the masker(s) only, and one interval (chosen at random) contained the masker(s) plus the signal (Fig. 1*B*). Listeners

were required to select the interval containing the signal (three-alternative forced choice). The level of the masker or the signal was varied between trials using a "two-up one-down" (phase 1) or "two-down one-up" (phases 2 and 3) adaptive staircase to find the level at which the signal was just masked according to a 71% detection criterion (Levitt, 1971). In the two-up one-down procedure, the level of M1 was increased by the step size after every two consecutive correct responses and decreased by the step size after every incorrect response. In the two-down one-up procedure, the level of M2 (phase 2) or signal (phase 3) was decreased by the step size after every two consecutive correct responses and increased by the step size after every incorrect response. The step size was 4 dB for the first four "turn points" (transitions between ascending and descending level) and 2 dB thereafter. In each block of trials, 16 turn points were measured, and the threshold estimate was taken as the mean level at the last 12 turn points. At least four such estimates were made for each condition in each listener, and the results were averaged.

In phase 1 of the experiment, the signal was presented at 10 dB (low-level conditions) or at 40 dB (medium-level conditions) above its detection threshold in quiet. The level of M1 was varied to find the level required to mask the signal (Fig. 1*C*). In phase 2, M2 acted as the signal. Using the level of M1 from phase 1, the level of M2 was varied to find its masked threshold in the presence of M1. In phase 3, the level of the signal at threshold was found in the presence of M1 alone, M2 alone, and M1+M2. M1 was fixed at the level determined in phase 1. M2 was presented at levels above, equal to, and below its own masked threshold in the presence of M1.

Five normal-hearing listeners were tested. They were trained on the tasks until performance was stable. Listeners were seated in a double-walled sound-attenuating booth and made their responses via a computer keyboard. "Lights" on the computer monitor indicated the time of occurrence of the observation intervals and provided feedback as to whether the response was correct or incorrect.

For four of the five listeners (L1–L4), a parallel set of trials using a 4 kHz pure-tone masker as M2 was randomly interleaved with the noise-M2 trial blocks. The additivity of masking and modeling results for the pure-tone M2 were very similar to those with the noise M2, so only the latter are presented below.

Results

Table 1 shows the absolute thresholds of the signal and the individual results from the first two phases of the experiment. If the system were linear, then it would be expected that the difference between the levels of M1 required to mask the 10 and 40 dB sensation-level signals would be 30 dB. Although this is approximately the case for L4, the difference is greater for the other listeners and almost 60 dB for L5. Differences in level growth between a forward masker and a signal have been investigated in previous studies and can be generally well accounted for by a combination of internal noise and the effects of the compressive nonlinearity in the peripheral auditory system (Plack and Oxenham, 1998; Plack et al., 2002). Internal noise may reduce the masker level required for the low-level signal, because the internal noise contributes part of the effect required to mask the signal. This is especially true when there is compression at low levels: the more the signal is compressed, the closer in level is the internal representation of the signal to the noise floor and, hence, the greater the contribution of the noise floor to masking. A differential effect of compression on the masker and signal also leads to nonlinear growth. For example, if the masker is compressed more than the signal, then a given increase in physical masker level will require a smaller increase in physical signal level to maintain the same detectability. Because of the effects of compression, small individual differences in nonlinearity can lead to large individual differences in masked threshold.

The results from phase 3 are shown in Figure 2*A* (low-level conditions) and Figure 2*B* (medium-level conditions). Signal

Table 1. Signal absolute thresholds and the individual results from phase 1 and phase 2

	Absolute threshold for signal (dB SPL)	Phase 1: level of M1 required to mask signal (dB spectrum level)		Phase 2: threshold of M2 in presence of M1 (dB spectrum level)	
		Low level	Medium level	Low level	Medium level
L1	11.5 (0.4)	−13.0 (0.4)	37.4 (0.9)	−7.4 (0.4)	32.8 (1.1)
L2	12.3 (0.4)	−16.4 (0.4)	24.5 (1.1)	−6.9 (0.6)	32.1 (0.3)
L3	16.0 (0.4)	−4.6 (0.6)	46.9 (0.4)	−1.5 (1.2)	49.1 (0.7)
L4	22.8 (0.7)	−3.7 (0.2)	25.9 (0.5)	4.5 (0.2)	36.1 (0.4)
L5	12.2 (0.7)	−16.4 (0.4)	41.0 (0.8)	−5.3 (0.3)	50.3 (0.4)

'Low level' refers to the use of a 10 dB sensation level (absolute threshold plus 10 dB) signal in phase 1. 'Medium level' refers to the use of a 40 dB sensation level signal in phase 1. SEs are given in parentheses.
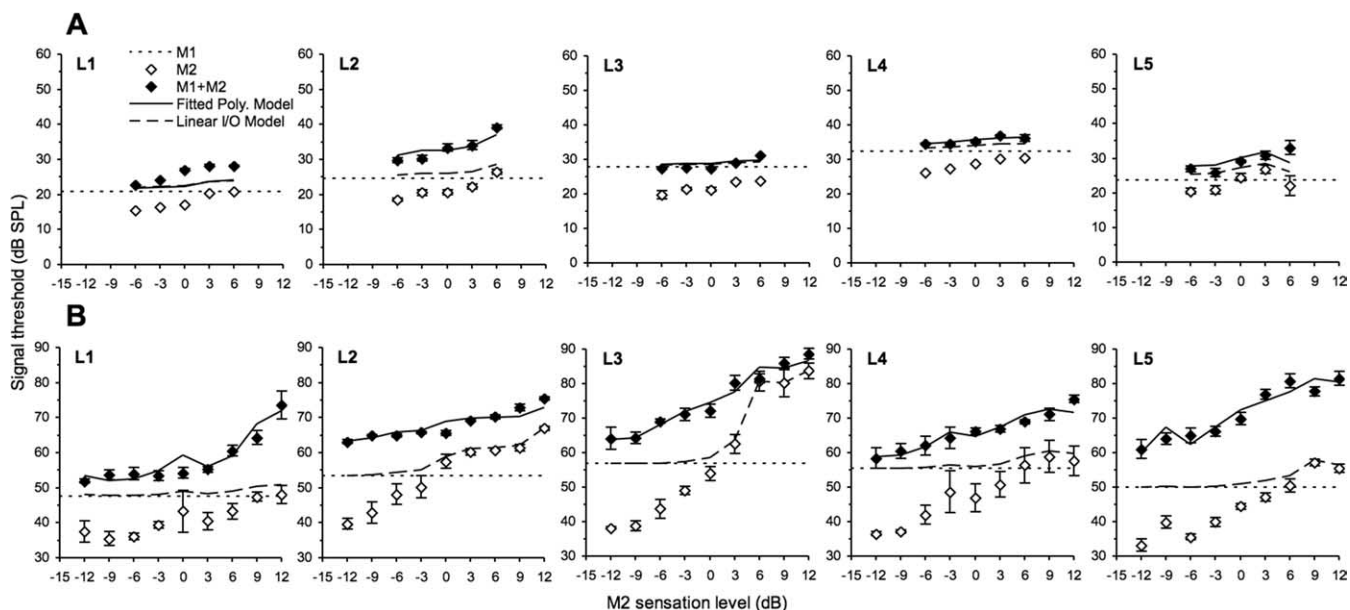

Figure 2. The individual results of phase 3 of the experiment. The results are shown separately for the low-level (**A**) and medium-level (**B**) conditions. Signal threshold is plotted as a function of the level of M2 relative to its masked threshold in the presence of M1. The horizontal dotted lines show the signal thresholds in the presence of M1 alone. The open symbols show the signal thresholds in the presence of M2 alone. The filled symbols show the signal thresholds in the combined-masker conditions (M1+M2). Error bars show SEs. Also shown are the combined-masker thresholds derived from the single-masker thresholds by a linear-summation model, using a fitted third-order polynomial to simulate the basilar-membrane input–output function (solid lines) and using a linear input–output function (dashed lines).

thresholds are plotted as a function of the level of M2 relative to its threshold in the presence of M1. A sensation level of 0 dB refers to an M2 level that was at threshold in the combined-masker condition (as determined in phase 2). In each plot, the signal threshold for M1 alone (M1 level was fixed for each overall level condition) is indicated by the horizontal dotted line. The amount of additional masking produced by M2 is indicated by the difference between the horizontal dotted line and the filled symbols.

In the medium-level conditions (Fig. 2*B*), across all levels of M2, the effect of combining the two maskers was usually much greater than their individual effects. The addition of M2 to M1 produced a substantial increase in masking even when M2 was below its own masked threshold. The increase in masking found with both maskers present was not as marked in the low-level conditions (Fig. 2*A*). Combined-masker thresholds for most listeners were close to those found for M1 alone for the lower M2 levels (and close for all M2 levels for listeners L3 and L4). Thus, the subthreshold masking effects appear to be much stronger at medium levels than at low levels, although in both cases, the effect of two maskers was usually greater than that of one alone (the exception being L3 at low levels).

Threshold here is defined (arbitrarily) as the level that gives 71% correct performance on the discrimination task. To provide a more rigorous test of whether M2 was detectable, independent
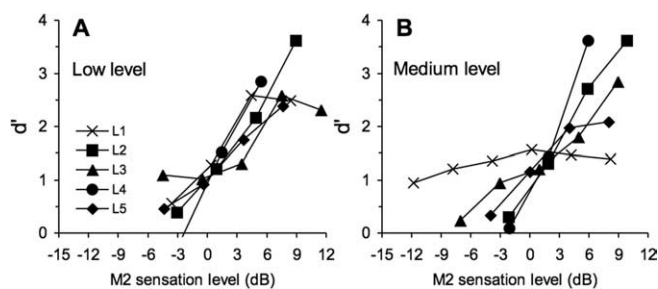

Figure 3. Individual psychometric functions derived from the adaptive tracks of phase 2 of the experiment, for the low-level (**A**) and medium-level (**B**) conditions. The detectability index, *d'*, for the detection of M2 in the presence of M1 is plotted against M2 sensation level (i.e., relative to the M2 level required for 71% correct responses). A *d'* of 0 represents chance performance (M2 is completely inaudible).

of any specific threshold criterion, an analysis of the individual trials of the adaptive tracks from phase 2 was used to construct psychometric functions for the detection of M2 in the presence of M1. Percentage correct values were converted into measures of the detectability index, *d'* (Elliot, 1964; Hacker and Ratcliff, 1979). These functions are shown in Figure 3. For the medium-level conditions, for four of the five listeners (L2, L3, L4, and L5), detection of M2 was effectively at chance (33% correct) when M2
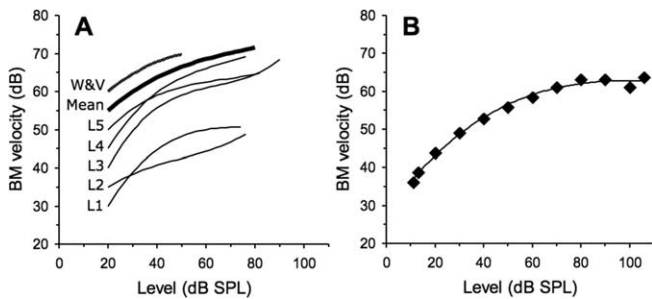
Figure 4. *A*, Third-order polynomials representing the basilar-membrane input–output function. The seven functions are those derived from the individual data, the mean data, and the data of Widin and Viemeister (1980) (W&V) and were used in the simulations described in Results. For each function, basilar-membrane velocity [expressed as $20\log_{10}$(linear velocity) in decibels] is plotted for input levels up to the maximum level constrained by the data (i.e., the maximum signal threshold). The functions have been staggered vertically for clarity (the intercept is arbitrary and is not constrained by the data). *B*, The response of the chinchilla basilar membrane at a place with a characteristic frequency of 10 kHz (Ruggero et al., 1997). Basilar-membrane velocity in decibels (relative to 1 μm/s) is plotted as a function of the level of a 10 kHz pure tone. The line shows the third-order polynomial fitted to these data, as used in the simulations described in Results.

was at a sensation level of −8 dB or less during an adaptive track. For these four listeners, linear fits to plots of d' against M2 sensation level had $d' = 0$ (chance performance) intercepts ranging from −7.8 to −2.0 dB. Thus, at sensation levels of −9 and −12 dB, M2 was completely inaudible, yet it still made a substantial contribution to masking when combined with M1 (mean threshold increase of 7.1 dB).

Linear-summation model
To test the hypothesis that the summation of the masking effects is a linear process, thresholds from phase 3 were simulated using a computational model of auditory processing. The model was similar to that used in previous studies to model forward masking and the additivity of nonsimultaneous maskers (Penner, 1980; Plack and O'Hanlon, 2003). The model assumes that the responses to the maskers and the signal are combined at some stage in the auditory system, although an advantage of the present approach is that it is not necessary to specify the nature of the combinatorial mechanism or how the temporal response declines over time. Hence, because the temporal locations of the maskers and signals in this experiment remained constant, it was not necessary to include a time parameter in the simulations. The temporal parameters of forward masking have been dealt with in previous studies (Plack and Oxenham, 1998).

The initial transformations (preprocessing) involved simulating the effects of basilar-membrane compression and hair-cell rectification. Compression was assumed to be instantaneous, applied to the intensity of the signal at the peak in the signal envelope. The input–output function was a third-order polynomial in decibel/decibel coordinates, with three parameters. In units of intensity, this becomes

$$f(x) = 10^{(a(10\,log_{10}(x))^3 + b(10\,log_{10}(x))^2 + c10\,log_{10}(x))/10}, \qquad (1)$$

where x is input intensity, and a, b, and c are the coefficients of the polynomial. (The constant or intercept in the equation is not constrained by the data and does not affect the predictions of the model.) A separate polynomial was derived for each listener and for the mean data (Fig. 4*A*). A second version of the model with no free parameters was also tested on the mean data, using a third-order polynomial fit to physiological data from the chin-

chilla cochlea (Ruggero et al., 1997) as the input–output function (Fig. 4*B*). Direct measurements from a chinchilla are a reasonable choice for modeling human performance because the frequency and dynamic range of chinchilla hearing is similar to that of humans (Heffner and Heffner, 1991). It was assumed that the form of the input–output function does not vary significantly between the 4 kHz place investigated here and the 10 kHz place investigated by Ruggero et al. Finally, a version of the model with no compression (linear input–output function) was tested.

After the simulation of compression and rectification, the responses to the stimuli (maskers and signal) were assumed to add linearly. Detection of the signal was based on the signal-to-masker ratio after preprocessing and summation, and this ratio was assumed to be constant at threshold for all conditions. This means that a measure of the masking effect can be taken as the signal intensity at masked threshold after basilar-membrane compression:

$$E = f(S), \qquad (2)$$

where E is the masking effect, and S is the signal intensity at threshold. Assuming that the effects of two maskers sum linearly,

$$E_{M1+M2} = E_{M1} + E_{M2}, \qquad (3)$$

where E_{M1}, E_{M2}, and E_{M1+M2} are the masking effects produced by M1, M2, and M1 and M2 combined. Substituting from Equation 2 and solving for S gives the following:

$$S_{M1+M2} = f^{-1}(f(S_{M1}) + f(S_{M2})), \qquad (4)$$

where S_{M1} and S_{M2} are the signal intensities at threshold in the presence of M1 and M2, respectively, and S_{M1+M2} is the signal intensity at threshold in the presence of M1 and M2 combined. Using this equation, and the simulated basilar-membrane input–output function as the function, f, the thresholds from each masker alone in phase 3 (S_{M1} and S_{M2}) were used as the input to the model, and the thresholds in the presence of both maskers (S_{M1+M2}) were predicted. In the case of the fitted polynomial basilar-membrane input–output functions (Fig. 4*A*), the parameters were selected to minimize the sum of the squared deviations of the model predictions from the thresholds in the combined-masker conditions.

For the individual data, the predicted combined thresholds (M1+M2) of the model using the fitted polynomials and the model using a linear input–output function are shown in Figure 2 as solid and dashed curves, respectively. The model incorporating a simulation of basilar-membrane compression provides a good account of the data. For each listener, a single third-order polynomial can account for both the low- and medium-level thresholds. As shown in Figure 4, there are some differences between the polynomials for the different listeners, but the overall forms of the functions are similar (note that the vertical positions of the functions are arbitrary). The compression exponents (slopes of the polynomials) in the mid-level region [40–70 dB sound pressure level (SPL)] averaged 0.20, 0.21, 0.23, 0.28, and 0.15 for L1, L2, L3, L4, and L5, respectively. These are within the range of values found in previous psychophysical and physiological studies. For example, over the same level range, the average compression exponent for the fit to the chinchilla data shown in Figure 4 was 0.23. The model assuming a linear basilar-membrane response (Fig. 2, dashed curves) produces very poor predictions, particularly for the medium-level conditions.

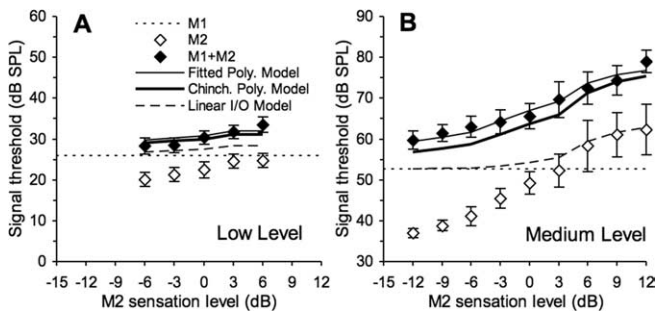The mean data from phase 3 and the predictions of the models

Figure 5. The mean results of phase 3 of the experiment for the low-level (**A**) and medium-level (**B**) conditions. Error bars show SEs across listeners. The solid lines show combined-masker thresholds derived from the single-masker thresholds by a linear-summation model, using a polynomial fitted to the present data (thin lines) and using a polynomial fitted to the chinchilla data (thick lines). Also shown are thresholds predicted by a linear-summation model using a linear initial input–output function (dashed lines).

Figure 6. The mean results of the experiment of Widin and Viemeister (1980). The figure shows the results for a fixed-level M1 and a variable-level M2 (**A**) and the results for a fixed-level M2 and a variable-level M1 (**B**). The top two solid lines in each panel show combined-masker thresholds derived from the single-masker thresholds by a linear-summation model, using a polynomial fitted to the Widin and Viemeister data (thin lines) and using a polynomial fitted to the chinchilla data (thick lines). The bottom two solid lines in each panel show the single-masker thresholds in the presence of M2 (**A**) and M1 (**B**) predicted by the two polynomials.

are shown in Figure 5. Overall, the predictions of the two models assuming a compressive basilar-membrane nonlinearity are reasonably accurate. In particular, the models predict strong effects of M2 in the medium-level conditions, even when M2 is well below its own masked threshold, and correctly predict a smaller increase in masking in the low-level conditions than in the medium-level conditions. This is because the nonlinearity in both models is less compressive at low levels than at medium levels. As expected, the model using a fitted polynomial is more accurate than the model based on the animal physiological data. The latter tends to under-predict thresholds slightly in the medium-level conditions, particularly for the lower levels of M2 (Fig. 5B). However, even this model, which has no free parameters, accounts well for the effects of combining two maskers. Again, the model assuming a linear basilar-membrane response produces a very poor fit.

Widin and Viemeister (1980) conducted an experiment broadly similar to that presented here, in which the two maskers (M1 and M2) and the signal were all 1 kHz pure tones, with 10 ms raised-cosine onset and offset ramps and no steady state. The silent interval between M1 and M2, and between M2 and the signal, was 6.5 ms. Although their reliable data (see below) did not include subthreshold levels of M2, they measured the effects of combining maskers over a range of M1 and M2 levels that provide an additional test of the linear-summation hypothesis. Their mean results are shown in Figure 6. A shows the results with a fixed M1 level and variable M2 level (Widin and Viemeister argued that the combined-masker thresholds for the lower levels of M2 were artificially low because of equipment limitations, and the lowest two values are omitted), and B shows the results with a fixed M2 level and variable M1 level. The signal thresholds are all relatively low, similar to those obtained in the low-level conditions of the present study. The dashed lines show the predictions of the linear-summation models described above, one with a fitted polynomial (Fig. 4) and one with a polynomial derived from the chinchilla data. Although the fits are not as good as those in Figure 5, the model predictions are close to the combined-masker thresholds.

As a final test of the model, predictions of the single-masker thresholds of Widin and Viemeister (i.e., the thresholds in the presence of M2 alone and M1 alone) were calculated assuming that the relative growth rates of signal and masker were determined by each of the two polynomials (fitted and chinchilla). The analysis was conducted to determine whether the growth of signal threshold with masker level is consistent with the effects of com-
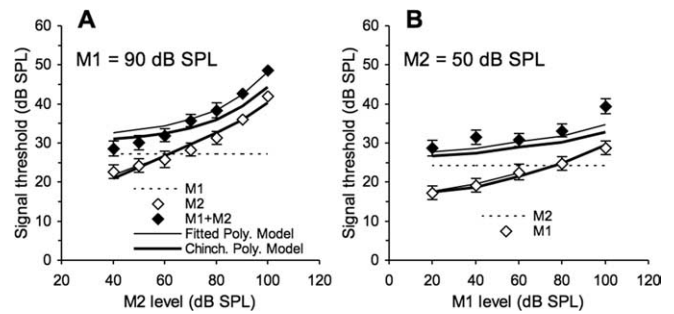
bining two maskers, on the assumption of linear summation. Signal threshold was assumed to be given by the following:

$$S = f^{-1}(k(f(M) + N_0)), \tag{5}$$

where S is the signal intensity at threshold in the presence of a masker with intensity M. f is the simulated basilar-membrane input–output function. N_0 is related to the level of an internal noise floor, assumed to limit performance for signal levels close to absolute threshold. N_0 is usually derived from the absolute threshold for the signal, by finding the value of N_0 for which the absolute threshold is equal to S when $M = 0$ (Plack and Oxenham, 1998). Because absolute threshold was not available for the Widin and Viemeister data, absolute threshold in the simulation was assumed to be equal to the lowest signal threshold measured (17.3 dB SPL). k is a measure of the efficiency of signal detection and of the temporal decay of forward masking and was assumed to be constant for a given temporal configuration of signal and masker. The value of k was varied adaptively to minimize the sum of the squared deviations of the predictions from the data. The analysis was performed independently for the M2 data (Fig. 6A) and for the M1 data (Fig. 6B). Because the fitted polynomial was only constrained for low input levels (Fig. 4A), the predictions using this function are only shown for the three lowest masker levels in each case. The two versions of the model provide a good fit to the data.

This form of analysis is possible for the data of Widin and Viemeister (1980) because their maskers and signals were all pure tones of the same frequency. However, because our data involved noises of different bandwidths, covering a wide range of frequencies, a similar simple analysis was not possible.

Discussion

The main finding of the present study is that stimuli below masked threshold can contribute substantially to decreasing the audibility of subsequent stimuli. This counterintuitive result can be explained by a simple model in which physiologically realistic compression is followed by linear summation. Linear summation implies that the contribution of inaudible maskers is the same as when they are audible. Forward masking does not reduce the masking effectiveness of stimuli, at least not over the range of levels tested here.

Related findings

Other results are also consistent with effectively linear postcochlear processing in forward masking (for review, see Plack et al., 2002). For example, a given increase in the level of a forward masker produces a much smaller increase in the signal level at threshold, for signal levels less than ~30 dB SPL [Munson and Gardner, 1950; Widin and Viemeister, 1980 (see also Fig. 6); Jesteadt et al., 1982; Moore and Glasberg, 1983]. It has been argued that the shallow, nonlinear growth of forward masking with level is a consequence of the masker level falling within the highly compressive portion of the basilar-membrane input–output function, and the lower-level signal falling within the more linear portion of the function (Plack and Oxenham, 1998; Plack et al., 2002). Hence, a given increase in masker level has a much smaller physiological effect than the same increase in signal level, resulting in a shallow masking function. A prediction of the linear-summation model is that, in the absence of basilar-membrane compression, the growth of forward masking should be linear. The prediction has been tested by measuring forward masking in listeners with moderate cochlear hearing loss. These listeners should have reduced or absent basilar-membrane compression, because direct measurements in nonhuman mammals suggest that cochlear dysfunction is associated with a linear basilar-membrane response (Ruggero and Rich, 1991; Ruggero et al., 1997). Consistent with the prediction of the linear-summation model, growth of forward masking is approximately linear for hearing-impaired listeners: a given increase in masker level produces approximately the same increase in signal level at threshold (Oxenham and Moore, 1995). Similarly, a prediction of the linear-summation model is that listeners without cochlear compression should display linear masking additivity, such that a combination of two equally effective maskers should produce a 3 dB increase in threshold compared with the individual masker conditions. This has been confirmed by measuring the effects of combining forward and backward maskers for listeners with cochlear hearing loss (Oxenham and Moore, 1995).

Because we do not know the exact form of an individual's basilar-membrane input–output function, it is hard to exclude the possibility of postcochlear nonlinearities in the context of the current model. However, the present results and previous results are at least consistent with a linear systems-analysis approach to predicting masking in everyday situations with multiple sound sources, despite the obvious nonlinearities at earlier stages in the auditory pathways and the sparse, feature-dependent representations found in auditory cortex (Nelken, 2004). Linear systems are far easier to describe and to investigate than nonlinear systems because the response to multiple inputs can be derived by summation and the contribution of each input is independent of other inputs. Our results indicate that some important aspects of hearing may be illuminated by a relatively simple formulation and provide a behavioral example of linearity, which has also been found to provide a reasonable description of specific aspects of auditory processing in the auditory nerve (Carney and Yin, 1988) and in auditory cortex (Kowalski et al., 1996a,b).

Forward masking and neural adaptation

The search for the neurophysiological basis of forward masking has focused on neural adaptation, in which the spike rate in response to a stimulus is reduced after previous stimulation. Several authors have suggested that psychophysical forward masking is a direct consequence of this process (Duifhuis, 1973; Smith, 1977, 1979; Jesteadt et al., 1982; Kidd and Feth, 1982), and it is common to describe a poststimulation reduction in neural activ-

ity as forward masking (Harris and Dallos, 1979; Shore, 1995; Brosch and Schreiner, 1997). However, a poststimulation reduction in neural activity does not necessarily imply masking. This is because a reduction in response to the signal can be accompanied by a similar reduction in the spontaneous neural activity, meaning that the presence of the signal remains easily distinguishable from its absence. Based on this type of analysis, it has been shown that adaptation in the auditory nerve is not sufficient to account for behavioral forward masking (Relkin and Turner, 1988). Hence, a more central process must be suboptimal in its use of information from the auditory nerve (Meddis and O'Mard, 2005).

Recent work has investigated the time-dependent responses of neurons in the auditory cortex. Adaptation is stronger here than in the auditory nerve and can persist for longer durations (Calford and Semple, 1995; Brosch and Schreiner, 1997; Ulanovsky et al., 2004; Wehr and Zador, 2005). However, it has not been demonstrated that the decrease in signal detectability resulting from the adaptation of cortical neurons (rather than simply the gross reduction in response) is equivalent to that measured psychophysically. The present demonstration of linear temporal summation presents an additional challenge for neural models of forward masking, particularly because cortical processing appears to be highly nonlinear (Machens et al., 2004).

It is possible to conceive of mechanisms by which neural adaptation could account for the finding that a subthreshold masker can decrease the detectability of a subsequent signal. For example, M1 could reduce the response to M2 below the threshold needed for activation at a higher stage in the auditory pathway (or below a central "noise floor"), so that M2 is inaudible. However, the attenuated representation of M2 may still be sufficient to produce adaptation of the signal at the more peripheral stage. The more specific constraint, that the "effective" contribution of M2 to masking should be unaffected by adaptation, is potentially more problematic. Our paradigm should provide a useful tool in the search for neural correlates of forward masking within the auditory pathways. In such a search, it is important to recognize the difference between masking, which implies an inability to discriminate the presence from the absence of the signal, and a reduction in response, which does not necessarily imply masking (Relkin and Turner, 1988).

References

Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77:923–943.

Brosch M, Schreiner CE (2000) Sequence sensitivity of neurons in cat primary auditory cortex. Cereb Cortex 10:1155–1167.

Calford MB, Semple MN (1995) Monaural inhibition in cat auditory cortex. J Neurophysiol 73:1876–1891.

Carney LH, Yin TCT (1988) Temporal coding of resonances by low-frequency auditory-nerve fibers - single-fiber responses and a population-model. J Neurophysiol 60:1653–1677.

Cokely CG, Humes LE (1993) Two experiments on the temporal boundaries for the nonlinear additivity of masking. J Acoust Soc Am 94:2553–2559.

Duifhuis H (1973) Consequences of peripheral frequency selectivity for nonsimultaneous masking. J Acoust Soc Am 54:1471–1488.

Elliot PB (1964) Tables of d′. In: Signal detection and recognition by human observers (Swets JA, ed), pp 651–684. New York: Wiley.

Hacker MJ, Ratcliff R (1979) A revised table of d′ for M-alternative forced choice. Percept Psychophys 26:168–170.

Harris DM, Dallos P (1979) Forward masking of auditory nerve fiber responses. J Neurophysiol 42:1083–1107.

He S, MacLeod DIA (2001) Orientation-selective adaptation and tilt aftereffect from invisible patterns. Nature 411:473–476.

Heffner RS, Heffner HE (1991) Behavioral hearing range of the chinchilla. Hear Res 52:13–16.

Humes LE, Jesteadt W (1989) Models of the additivity of masking. J Acoust Soc Am 85:1285–1294.

Jesteadt W, Bacon SP, Lehman JR (1982) Forward masking as a function of frequency, masker level, and signal delay. J Acoust Soc Am 71:950–962.

Kidd G, Feth LL (1982) Effects of masker duration in pure-tone forward masking. J Acoust Soc Am 72:1384–1386.

Kowalski N, Depireux DA, Shamma SA (1996a) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. J Neurophysiol 76:3503–3523.

Kowalski N, Depireux DA, Shamma SA (1996b) Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. J Neurophysiol 76:3524–3534.

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

Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. J Neurosci 24:1089–1100.

Meddis R, O'Mard LP (2005) A computer model of the auditory-nerve response to forward-masking stimuli. J Acoust Soc Am 117:3787–3798.

Moore BCJ, Glasberg BR (1983) Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: implications for suppression in noise. J Acoust Soc Am 73:1249–1259.

Moore BCJ, Glasberg BR (1985) The danger of using narrowband noise maskers to measure suppression. J Acoust Soc Am 77:2137–2141.

Munson WA, Gardner MB (1950) Loudness patterns—a new approach. J Acoust Soc Am 22:177–190.

Neff DL (1985) Stimulus parameters governing confusion effects in forward masking. J Acoust Soc Am 78:1966–1976.

Neff DL (1986) Confusion effects with sinusoidal and narrowband-noise forward maskers. J Acoust Soc Am 79:1519–1529.

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14:474–480.

Oxenham AJ (2001) Forward masking: adaptation or integration? J Acoust Soc Am 109:732–741.

Oxenham AJ, Moore BCJ (1994) Modeling the additivity of nonsimultaneous masking. Hear Res 80:105–118.

Oxenham AJ, Moore BCJ (1995) Additivity of masking in normally hearing and hearing-impaired subjects. J Acoust Soc Am 98:1921–1934.

Oxenham AJ, Plack CJ (2000) Effects of masker frequency and duration in forward masking: further evidence for the influence of peripheral nonlinearity. Hear Res 150:258–266.

Penner MJ (1980) The coding of intensity and the interaction of forward and backward masking. J Acoust Soc Am 67:608–616.

Penner MJ, Shiffrin RM (1980) Nonlinearities in the coding of intensity within the context of a temporal summation model. J Acoust Soc Am 67:617–627.

Plack CJ, O'Hanlon CG (2003) Forward masking additivity and auditory compression at low and high frequencies. J Assoc Res Otolaryngol 4:405–415.

Plack CJ, Oxenham AJ (1998) Basilar-membrane nonlinearity and the growth of forward masking. J Acoust Soc Am 103:1598–1608.

Plack CJ, Oxenham AJ, Drga V (2002) Linear and nonlinear processes in temporal masking. Acustica 88:348–358.

Relkin EM, Turner CW (1988) A reexamination of forward masking in the auditory nerve. J Acoust Soc Am 84:584–591.

Ruggero MA, Rich NC (1991) Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane. J Neurosci 11:1057–1067.

Ruggero MA, Rich NC, Recio A, Narayan SS, Robles L (1997) Basilar-membrane responses to tones at the base of the chinchilla cochlea. J Acoust Soc Am 101:2151–2163.

Shady S, MacLeod DI, Fisher HS (2004) Adaptation from invisible flicker. Proc Natl Acad Sci USA 101:5170–5173.

Shore SE (1995) Recovery of forward-masked responses in ventral cochlear nucleus neurons. Hear Res 82:31–43.

Smith RL (1977) Short-term adaptation in single auditory nerve fibers: some poststimulatory effects. J Neurophysiol 40:1098–1112.

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

Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. J Neurosci 24:10440–10453.

Wehr M, Zador AM (2005) Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47:437–445.

Widin GP, Viemeister NF (1980) Masker interaction in pure-tone forward masking. J Acoust Soc Am 68:475–479.

Zwicker E, Fastl H (1972) Zur Abhangigkeit der Nachverdeckung von der Störimpulsdauer. (On the dependence of forward masking on masker duration.) Acustica 26:78–82.

Zwislocki J, Pirodda E, Rubin H (1959) On some poststimulatory effects at the threshold of audibility. J Acoust Soc Am 31:9–14.