Reconciling frequency selectivity and phase effects in masking

Andrew J. Oxenham[a)]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Torsten Dau[b)]
Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

(Received 27 October 2000; revised 10 April 2001; accepted 22 June 2001)

The effects of auditory frequency selectivity and phase response on masking were studied using harmonic tone complex maskers with a 100-Hz fundamental frequency. Positive and negative Schroeder-phase complexes (m_+ and m_-), were used as maskers and the signal was a long-duration sinusoid. In the first experiment, thresholds for signal frequencies of 1 and 4 kHz were measured as a function of masker bandwidth and number of components. A large difference in thresholds between the m_+ and m_- complexes was found only when masker components were presented ipsilateral to the signal over a frequency range wider than the traditional critical band, regardless of the absolute number of components. In the second experiment, frequency selectivity was measured in harmonic tone complexes with fixed or random phases as well as in noise, using a variant of the notched-noise method with a fixed masker level. The data showed that frequency selectivity is not affected by masker type, indicating that the wide listening bandwidth suggested by the first experiment cannot be ascribed to broader effective filters in complex-tone maskers than in noise maskers. The third experiment employed a novel method of measuring frequency selectivity, which has the advantage that the overall level at the input and the output of the auditory filter remains roughly constant across all conditions. The auditory filter bandwidth measured using this method was wider than that measured in the second experiment, but may still be an underestimate, due to the effects of off-frequency listening. The data were modeled using a single-channel model with various initial filters. The main findings from the simulations were: (1) the magnitude response of the Gammatone filter is too narrow to account for the phase effects observed in the data; (2) none of the other filters currently used in auditory models can account for both frequency selectivity and phase effects in masking; (3) the Gammachirp filter can be made to provide a good account of the data by altering its phase response. The final conclusion suggests that masker phase effects can be accounted for with a single-channel model, while still remaining consistent with measures of frequency selectivity: effects that appear to involve broadband processing do not necessarily require across-channel mechanisms. [DOI: 10.1121/1.1394740]

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

I. INTRODUCTION

In a pair of papers, Smith *et al.* (1986) and Kohlrausch and Sander (1995) showed that two harmonic tone complexes with identical power spectra and very similar flat temporal envelopes could produce masked thresholds for long-duration tones that differed by as much as 20 dB. As these waveforms were based on equations proposed by Schroeder (1970), they have come to be known as Schroeder-phase complexes (Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996; Carlyon and Datta, 1997a,b; Summers and Leek, 1998). These complexes are comprised of a series of equal-amplitude harmonic tones, with the starting phase of component n assigned according to the following equation:

$$\theta_n = \pm \pi n(n-1)/N, \tag{1}$$

where N is the total number of components. Complexes derived using the "+" sign are known as Schroeder positive (m_+), while complexes derived using the "−" sign are known as Schroeder negative (m_-).[1] As pointed out by Smith *et al.* (1986), these stimuli can be thought of either as harmonic tone complexes with flat temporal envelopes, or as a series of repeating linear frequency sweeps, with the m_+ complex producing downward sweeps and the m_- complex producing upward sweeps.

These complexes are interesting stimuli for psychoacoustic experiments for a number of reasons. First, because they both have the same power spectrum but can produce very different masked thresholds, they contradict the power-spectrum model of masking (Fletcher, 1940; Patterson, 1976; Glasberg and Moore, 1990). Second, these complexes might provide some insight into the phase response of the auditory filters. In particular, it has been proposed that auditory filtering, probably already at the level of the basilar-membrane (Recio and Rhode, 2000), alters the waveforms such that m_+ complexes result in a highly modulated, or "peaky," enve-

[a)]Electronic mail: oxenham@mit.edu
[b)]Now at: Arbeitsgruppe Medizinische Physik, Universität Oldenburg, 26111 Oldenburg, Germany; electronic mail: torsten.dau@medi.physik.uni-oldenburg.de

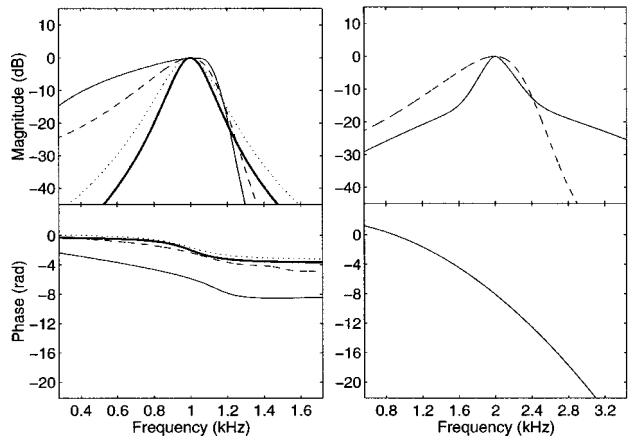

FIG. 1. The magnitude and phase responses of the filters used in the simulations are shown in upper and lower panels, respectively. The left panels show the responses of the Gammatone (heavy solid curve), 60-dB and 40-dB Gammachirp (dashed and dotted curves, respectively), and Strube's transmission-line (light solid curve) filters, all centered at 1 kHz. The upper-right panel shows the magnitude response of the fitted filter (solid curve; see text for details) and the 60-dB Gammachirp filter (dashed curve) centered at 2 kHz, and used in modeling the results from Experiment 3. The phase response shown in the lower-right panel has the same curvature as the m_- complex used in Experiment 3.

lope, while m_- complexes produce a less modulated, flatter, envelope. This in turn places certain constraints on the form of the auditory system's phase response (Kohlrausch and Sander, 1995). Certain models of basilar-membrane filtering (e.g., Strube, 1985; Giguère and Woodland, 1994a) support this view. For instance, Kohlrausch and Sander (1995) show in their Fig. 18 that the output of Strube's (1985) model is more highly modulated in response to an m_+ complex than to an m_- complex. This is in contrast to the response of a Gammatone filter (Patterson *et al.*, 1995), which shows essentially the same degree of modulation in response to both (Kohlrausch and Sander, 1995, Fig. 17). One reason for the qualitative success of BM models in accounting for the perceptual difference between m_+ and m_- complexes is that the phase response of these models has a negative curvature throughout most of the passband. In the case of the m_+ complex, this negative curvature compensates for, or ''cancels,'' the positive curvature of the complex, leading to a filtered waveform in which the starting phases of all the components come close to coinciding, giving a peaky envelope. In contrast the Gammatone filter, although generally very successful in many psychoacoustic applications, has a curvature that changes from being negative to positive at the center frequency (CF), meaning that it can never completely compensate for the curvature of the m_+ complex (Kohlrausch and Sander, 1995). The lower left panel of Fig. 1 illustrates this property by plotting the phase response of a segment of the (linear) Strube transmission-line model with a CF of 1 kHz (solid light curve), compared with the response of a fourth-order Gammatone filter with the same CF (solid heavy curve). The other two curves (dotted and dashed) are from Gammachirp filters (Irino and Patterson, 1997), discussed later.

While BM models provide a satisfying qualitative account, one troubling aspect is that their frequency selectivity is much poorer than that normally associated with the human auditory system. To illustrate this, the upper-left panel of Fig. 1 shows the magnitude responses of the respective filters. Again, the transmission-line and Gammatone filters are denoted by the light and heavy solid lines, respectively. The frequency selectivity of the Giguère and Woodland's (1994a, b) nonlinear model at medium and high levels is similar to that of the Strube (1985) transmission-line filter. While it has been speculated that the phase response of the BM filters, rather than the magnitude response, is primarily responsible for the good description of the m_+/m_- difference by these models (Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996), this has not been demonstrated. An indication that broader frequency selectivity may be necessary to account for large m_+/m_- differences can be found in the data of Carlyon and Datta (1997b). They found that thresholds were affected by masker components that fell outside what is normally considered the passband of the auditory filter. For instance, thresholds in their m_+ condition increased when components half an octave and more away from the signal frequency of 1.1 kHz were removed, even though the phase and amplitude of the more central components were unchanged. Using a model (Giguère and Woodland, 1994a) with much broader frequency selectivity than the Gammatone filter, Carlyon and Datta were able to provide a reasonable description of their results. It seems unlikely that a similarly good description would have been possible with a Gammatone filter, even if its phase response had been altered to provide negative curvature throughout the passband.

Carlyon and Datta (1997b) attributed the effect of the masker bandwidth to the absolute number of components present, as this limits the maximum peakiness of the envelope. However, it is possible that the limiting factor was not the number of components, but rather the bandwidth of the masker, relative to the signal frequency. It may be, for instance, that the presence of ''off-frequency'' components is necessary to elicit a large m_+/m_- difference, as suggested by Buss *et al.* (1998). This explanation is more likely if the effects of nonlinear interactions, such as suppression (Oxenham and Plack, 1998), or across-frequency mechanisms, such as in comodulation masking release (CMR) (Hall *et al.*, 1984), play a role in determining thresholds. The role of such mechanisms could explain why it is necessary when using a single-channel model to assume broader frequency selectivity than is necessary in most other situations. No attempts have been made so far to reconcile frequency selectivity and seemingly broadband phase effects within the same model.

The first experiment studies the effect of masker bandwidth on thresholds in m_+ and m_- maskers. By using two signal frequencies (1 and 4 kHz), it was possible to dissociate the effects of an absolute number of components and relative bandwidth. In some conditions, by presenting the off-frequency components to the ear contralateral to the signal, we were able to examine the possible role of higher-level, across-channel mechanisms in determining thresholds.

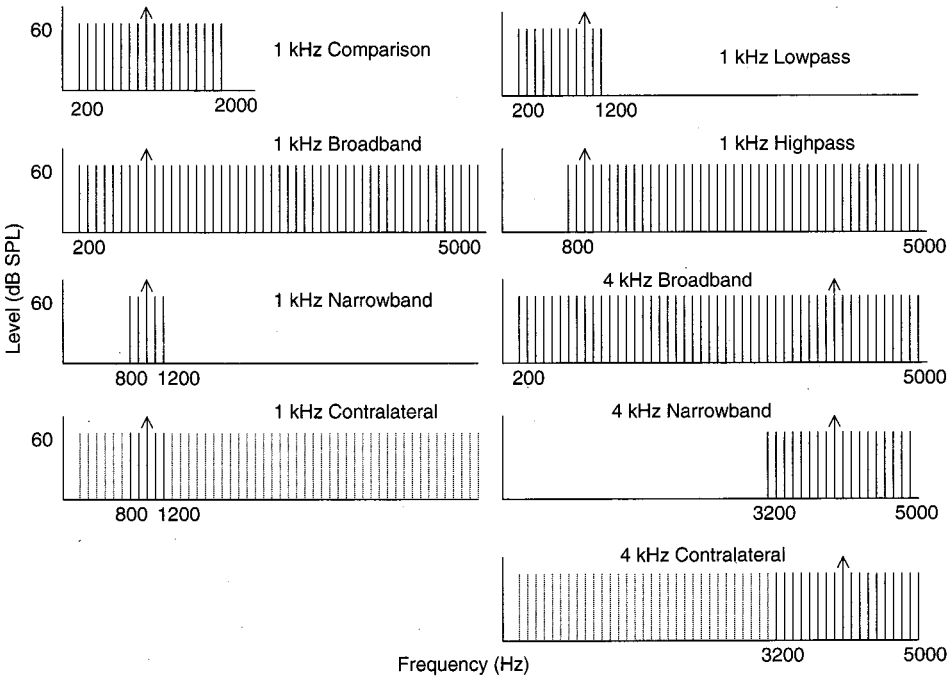

FIG. 2. Schematic diagram of the stimuli used in Experiment 1. Dashed lines denote components presented contralaterally.

II. EXPERIMENT 1. EFFECTS OF NUMBER OF COMPONENTS AND MASKER BANDWIDTH

A. Stimuli and conditions

The masker was a 320-ms (total duration) harmonic tone complex with a fundamental frequency (f_0) of 100 Hz, gated on and off with 10-ms raised cosine ramps. The level of the masker was set to 60 dB SPL per component. The sinusoidal signal was temporally centered within the masker and had a total duration of 260 ms, gated with 30-ms raised-cosine ramps. The signal frequency (f_s) was either 1 or 4 kHz. In all but one condition, the masker was derived from a harmonic tone complex with equal-amplitude frequency components from 200 to 5000 Hz, with the phase of each masker component selected based on Eq. (1) with $N=49$. One comparison condition was run using a 1-kHz signal, in which the masker was derived using components from 200 to 2000 Hz ($N=19$). For every condition, signal thresholds were measured in both m_+ and m_- configurations. The signal was always added in phase with the masker component at the signal frequency. Figure 2 shows a schematic diagram of the power spectra of the conditions tested. In certain conditions some of the masker components were eliminated, but the starting phases of the remaining components were not changed. The dotted lines indicate components presented to the ear contralateral to the signal.

The 1-kHz narrowband condition is similar to one tested by Carlyon and Datta (1997b). Based on their results, we expect the m_+/m_- difference to be reduced relative to the broadband condition. The 4-kHz narrowband masker has about the same relative bandwidth as the 1-kHz narrowband masker, but the same number of components as in the 1-kHz comparison condition. If, as suggested by Carlyon and Datta (1997b), the number of components is critical, then the 19 components of the 4-kHz narrowband masker should be sufficient to produce a large m_+/m_- difference. On the other hand, if a relatively wide masker bandwidth is necessary regardless of the number of components, then the m_+/m_- difference should be small in the 4-kHz narrowband condition.

The two contralateral conditions present the narrowband masker and the signal to one ear, and the remaining components from the broadband masker to the other ear. If higher-level across-channel processing is partly responsible for the m_+/m_- difference, then we might expect the contralateral components to have some effect. On the other hand, if the difference is mediated solely by peripheral mechanisms, then the contralateral components would not be expected to have any effect on thresholds.

The 1-kHz highpass and lowpass conditions test which range of components contribute more to the m_+/m_- difference. The 1-kHz comparison condition was included to provide a direct comparison with previous studies, many of which have used very similar parameters (Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a), and with the 4-kHz narrowband condition, which used the same number of components.

In pilot tests, it was noticed that distortion products could be heard in the 4-kHz narrowband condition and that they were modulated by the presence of the signal. Because of this, a lowpass pink noise (2.5-kHz cutoff frequency) was added at a level of 38 dB SPL per 1/3 octave band to all the 4-kHz conditions in order to mask any distortion products. The stimuli were generated digitally and played out using a LynxOne sound card at a sampling rate of 32 kHz. The stimuli were passed through a programmable attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented to the listener via Sennheiser HD 580 headphones. The stimuli (except for the contralateral masker components) were presented to the listener's left ear. Listeners were seated in a double-walled sound-attenuating booth.

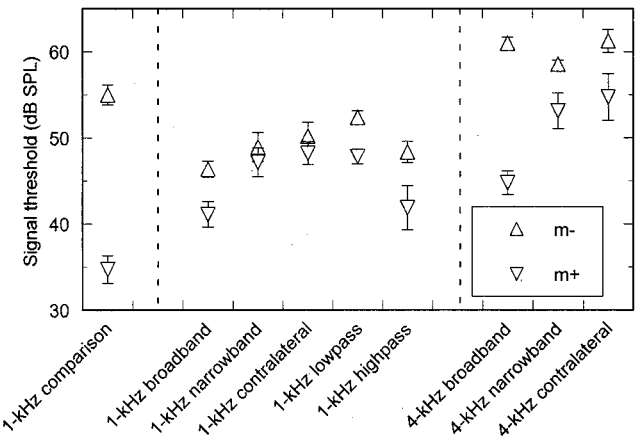

FIG. 3. Mean data from Experiment 1. Error bars denote ±1 standard error of the mean.

B. Procedure

An adaptive three-interval three-alternative forced-choice (3AFC) procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). Each interval in a trial was separated by an interstimulus interval (ISI) of 500 ms. The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size was 5 dB, which was reduced to 2 dB after the first 2 reversals. The threshold was defined as the mean of the remaining 8 reversals. Four threshold estimates were obtained from each listener in each condition.

C. Subjects

Four listeners participated as subjects in this experiment. Two were the authors, one was a graduate student in the first author's laboratory, and the other was paid for her services. The age of the subjects ranged from 24 to 34 years. All had thresholds of less than 15 dB HL at octave frequencies between 250 and 8000 Hz. The two less experienced subjects received at least 6 h training before data were collected.

D. Results

The trends of the four listeners were very similar, and so only the mean data from Experiment 1 are shown in Fig. 3. Thresholds in the presence of the m_+ and m_- maskers are shown as downward-pointing and upward-pointing triangles, respectively. The error bars denote ±1 standard error of the mean across subjects.

In the comparison condition, the difference between the two masker types is about 20 dB, which is consistent with previous studies using similar stimulus parameters (Kohlrausch and Sander, 1995; Summers and Leek, 1998). The difference in the 1-kHz broadband condition is much smaller (5.3 dB). The increase in thresholds for the m_+ complex, relative to the comparison condition, is probably in part because the sweep rate (the change in group delay as a function of frequency, or phase curvature) of the m_+ complex does not match the curvature of the auditory filters at 1 kHz as well as does the reference condition. The Schroeder-phase complex produces a linear frequency sweep, meaning that the phase curvature ($d^2\theta/df^2$) is constant and independent of frequency (f). The curvature of the m_+ complex is determined by f_0 and the number of components (N) (Kohlrausch and Sander, 1995):

$$\frac{d^2\theta}{df^2} = \frac{2\pi}{Nf_0^2}. \tag{2}$$

Increasing the number of components from 19 to 49 therefore leads to a decrease in the curvature by a factor of more than 2.5.

The *decrease* in thresholds for the m_- 1-kHz broadband condition, relative to the comparison condition, probably cannot be explained in these terms. Instead, it is helpful to consider that the proportion of time within each masker cycle in which the instantaneous frequency falls within the bandpass region of the auditory filter centered at 1 kHz is reduced when the bandwidth of the stimulus is increased. This means that, regardless of the local curvature of the complex, the frequency sweep will excite the auditory filter for a smaller proportion of the period, leading to a modulated envelope for the m_- complex, and hence to lower thresholds.

Unfortunately, the 1-kHz broadband condition does not provide a very good baseline condition, as the m_+/m_- difference is already rather small. Nevertheless, as expected, the m_+/m_- difference for the 1-kHz narrowband condition is even smaller. The m_+/m_- difference for the 1-kHz contralateral condition is about the same as for the narrowband condition, suggesting that the contralateral components did not influence performance. The 1-kHz lowpass condition shows thresholds that are elevated with respect to the broadband condition for both m_+ and m_- complexes, while the highpass results are very similar to those of the broadband condition.

The 4-kHz broadband condition showed a substantial m_+/m_- difference of about 16 dB. This is similar to the difference found by Summers and Leek (1998) for a similar stimulus configuration. The m_+/m_- difference for the 4-kHz narrowband condition was much smaller (5.4 dB), despite the fact that the number of components was the same as in the 1-kHz comparison condition. This suggests that masker components interacting over a relatively wide frequency region may be required to produce a large m_+/m_- difference. As at 1 kHz, the 4-kHz contralateral condition does not appear to differ from the 4-kHz narrowband condition, again suggesting no influence of the contralateral masker components.

In summary, a large m_+/m_- difference requires the presence of masker components over a frequency range wider than a traditional critical band (e.g., Scharf, 1970). This makes it seem intuitively unlikely that single-channel models, with sufficient frequency selectivity to predict other psychoacoustic data, could be used to predict these differences. However, as intuition can be misleading in such cases, quantitative modeling was used to test the ability of current models to predict the data. The following section uses a va-

riety of filter models, combined with the detection model of Dau *et al.* (1997a), in an attempt to describe the data from Experiment 1.

III. MODEL PREDICTIONS FOR EXPERIMENT 1

A. Description of the model

The model used here is based on the single-channel version of the model described by Dau *et al.* (1997a, b). This model has been successful in predicting a variety of simultaneous- and forward-masking data, as well as modulation-detection data. It is assumed that the signal is detected by a filter centered at the signal frequency and only the output from that filter is processed further. Briefly, the filtered stimulus is half-wave rectified and passed through a series of feedback loops, which simulate certain aspects of auditory-nerve adaptation. The output from the adaptation loops is passed through a bank of modulation filters. The output from the modulation filterbank is compared with an internal template of the signal, and decisions are derived using a version of an optimal detector. The model's ability to describe detection in the presence of temporally fluctuating maskers, and the ability of the detector to make use of information in the masker valleys, make it a good candidate for predicting the present data. All model parameters were identical to the ones used by Dau *et al.* (1997a).

A number of different front-end filters were used with the model. The first was the linear transmission-line model of Strube (1985). This was used in the first version of the model (Dau *et al.*, 1996a,b) and a nonlinear version (Strube, 1986) has also been used to predict data on vowel identification using Schroeder-phase stimuli (Alcántara *et al.*, 1996).[2] The second was a fourth-order Gammatone filter, which has been used in numerous models of auditory processing (e.g., Meddis and Hewitt, 1991; Dau *et al.*, 1997a; Ellis, 1999; Godsmark and Brown, 1999). Third, two versions of the Gammachirp filter (Irino and Patterson, 1997) were tested. The filter parameters were taken from Table II of Irino and Patterson (1997), averaged across all subjects at each center frequency. The two versions used parameters corresponding to signal levels of 40 and 60 dB SPL, which lie near the extremes of the signal levels measured in Experiment 1. The magnitude and phase responses of all four filters at 1 kHz are shown in the left-hand panels of Fig. 1. Finally, a filter was tested which combined the phase response of the Strube filter with the magnitude response of the Gammatone filter. This was done to test the idea that the phase response, rather than the magnitude, was critical in being able to predict the m_+/m_- difference (Kohlrausch and Sander, 1995). The 3-dB bandwidths and the equivalent rectangular bandwidths (ERBs) of these filters are given in Table I.

Simulations were run separately for each filter type. The stimuli were calculated in the same way as in the actual experiment and the model acted as a subject, being presented with three intervals, one of which contained the signal. The adaptive procedure was also the same as that used in the actual experiment. The predictions are the mean of four threshold estimates for each condition. Predictions were not made for the contralateral conditions, as the results would have been identical to those in the narrowband conditions.

TABLE I. Description of filters used in the simulations of Experiment 1.

	Center frequency (Hz)	3-dB bandwidth (Hz)	ERB (Hz)
Strube transmission-line filter	1000	365	407
	2000	701	803
	4000	1217	1452
Gammachirp 60 dB	1000	214	249
	2000	516	582
	4000	959	1114
Gammachirp 40 dB	1000	153	175
	2000	374	426
	4000	701	800
Gammatone	1000	116	133
	2000	208	241
	4000	406	457

B. Model results

The predictions from the five filter versions are shown in Fig. 4. As expected based on previous studies (Kohlrausch and Sander, 1995; Alcántara *et al.*, 1996; Carlyon and Datta, 1997b), the transmission-line model is able to predict an m_+/m_- difference in the comparison condition, although it is smaller than that observed experimentally (11.5 dB compared with 20 dB). The correct trend is also observed in some of the other conditions. Some discrepancies are apparent, however. For instance, the 1-kHz lowpass and 1-kHz broadband predictions are very similar and are different from the predictions for the 1-kHz narrowband and 1-kHz highpass conditions. This is expected, given the highly asymmetric magnitude response of the filter, where much more weight is given to frequency components below the CF than above. However, these trends were not observed in the data. The predicted m_+/m_- difference of 8 dB in the 4-kHz broadband condition is again less than the observed value of 16 dB. Also, the predicted difference between the 4-kHz broad- and narrowband conditions consists mainly of a decrease in the predicted m_- threshold. This can probably be explained by the reduction in the masker energy falling within the wide filter passband as the masker bandwidth is reduced, and because the elimination of many components results in a more modulated waveform. In contrast, the data show the opposite effect, with the main difference between the 4-kHz broad- and narrowband conditions being an increase in the m_+ threshold. The pattern of results for the Gammachirp filter at 60 dB SPL is very similar to that of the transmission-line model.

Both the 40-dB Gammachirp and the Gammatone filters fail to show a substantial m_+/m_- difference or any effect of stimulus bandwidth. Because the change in the Gammachirp parameters results in a change in both the magnitude and phase characteristics of the filter, it is not possible to say which is responsible for the difference in the predictions between the two versions. The final simulation, combining the Gammatone filter's magnitude response with the

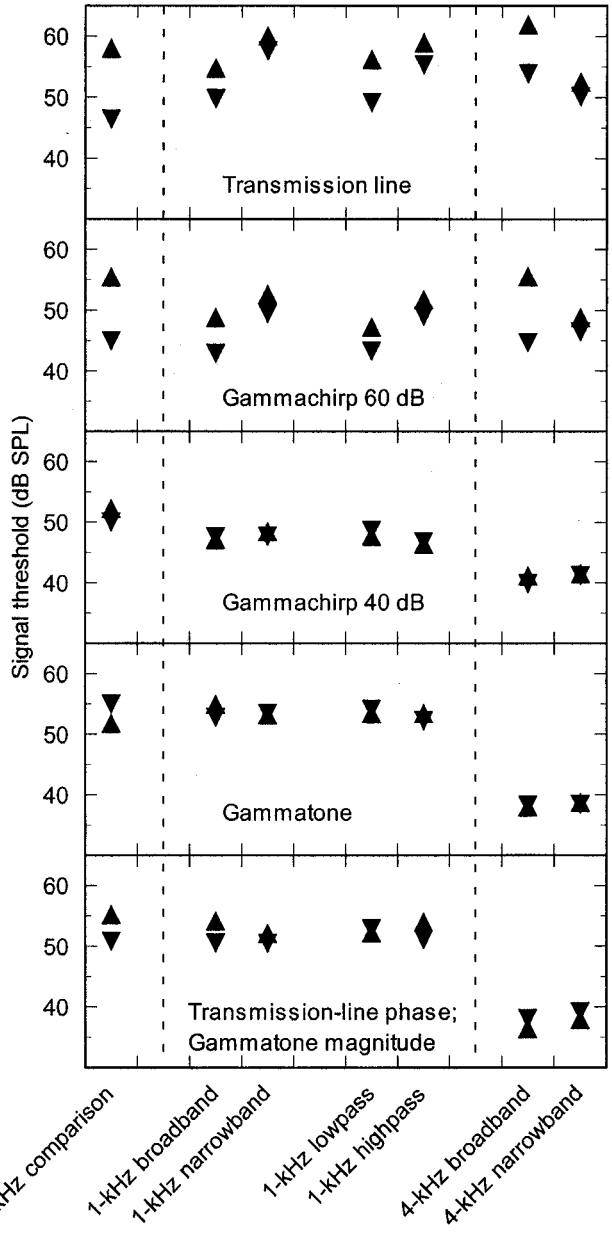

FIG. 4. Predictions of Experiment 1 using different auditory filter models. As in Fig. 3, upward- and downward-pointing triangles denote thresholds with the m_- and m_+ maskers, respectively.

transmission-line filter's phase response, answers that question. This hybrid filter shows an m_+/m_- difference of around 4.5 dB in the comparison condition, which lies in between the predictions of the transmission-line and Gammatone filters. Thus, even with a phase response that can in principle be used to predict an m_+/m_- difference, the passband of the Gammatone filter seems to be simply too narrow to allow a substantial difference to be predicted. It is interesting, however, that the small difference (3.6 dB) predicted in the 1-kHz broadband condition becomes 2 dB smaller in the 1-kHz narrowband condition. This implies that even with the frequency selectivity assumed by the Gammatone filter, components well outside its 3-dB bandwidth can still contribute to reducing thresholds. This is similar to the findings of Verhey *et al.* (1999).

C. Predictions using the temporal-window model

None of the filters tested provided very convincing fits to the data in conjunction with the model of Dau *et al.* (1997a). This could be an inherent problem with all the filters tested, or it could be a limitation of the detection model. To help address this, the data were simulated using the temporal-window model (Moore *et al.*, 1988; Plack and Moore, 1990; Oxenham and Moore, 1994; Oxenham and Plack, 2000). Stimuli were filtered using Strube's transmission-line filter, as that provided the best fits to the data in Experiment 1. The stimuli were then half-wave rectified, subjected to a power-law nonlinearity, and passed through a sliding temporal integrator, or temporal window. Threshold was determined by the maximum ratio between the window output due to the masker alone and the output due to the masker plus signal. A variety of window shapes and durations were tested in combination with a number of different power-law nonlinearities ($y=x^p$), ranging from very compressive ($p=0.1$) to energy detection ($p=2$). None of these combinations improved on the predictions of the Dau *et al.* model. This provides some indication that the failure of the model may be due to the initial filtering stage, rather than the particular post-filtering model we chose to use.

D. Discussion

A comparison of Figs. 3 and 4 shows that none of the filters can provide a completely satisfactory account of the data, at least when used in conjunction with the detection model of Dau *et al.* (1997a) or with the temporal-window model. The two broadest filters, Strube's transmission-line model and the 60-dB Gammachirp, provide the most convincing fits. This is somewhat disturbing, as the frequency selectivity associated with both filters is much broader than that exhibited by the Gammatone filter, which is generally believed to provide a good description of auditory frequency selectivity at low and medium sound pressure levels (Patterson, 1976; Glasberg and Moore, 1990; Derleth and Dau, 2000). In contrast to the speculation of Kohlrausch and Sander (1995), combining the amplitude response of the Gammatone filter with the phase response of the transmission-line model does not seem to be sufficient to produce predictions in line with the data.

There is a possibility, albeit an unattractive one, that the effective frequency selectivity of the auditory system is somehow different for harmonic tone complexes than for noise. Lentz *et al.* (1999) found similar auditory filter shapes when measuring thresholds both in noise and in tone complexes with logarithmically spaced components, although the estimated filter bandwidths for a profile-analysis task using the same complex-tone stimuli were considerably larger. To our knowledge, no direct measures of frequency selectivity have been made using harmonic tone complexes. The second experiment was designed to estimate frequency selectivity in noise and in harmonic tone complexes, with fixed or random phases, using a variant of the notched-noise method (Patterson, 1976; Glasberg and Moore, 1990).

IV. EXPERIMENT 2. FREQUENCY SELECTIVITY IN NOISE AND COMPLEX TONES

A. Methods

Many of the stimulus parameters and procedures were the same as in Experiment 1. Primarily those that differed are mentioned here. The signal frequency in this experiment was 2 kHz. This was chosen so that a masker f_0 of 100 Hz could still be used but the relative spectral notch could be manipulated in a more fine-grained way than would be possible at 1 kHz. The starting phase of the signal was random and was different in each trial. This was done because many conditions did not include a masker component at the signal frequency, and in order to make thresholds in the noise and tone complex maskers as comparable as possible. The harmonic tone complex masker was created with components from 100 Hz to 4000 Hz. In certain conditions, components around the signal frequency were eliminated. In the first part, only m_+ and Gaussian-noise maskers were tested. The noise masker had the same bandwidth as the tone complex and had a spectrum level of 40 dB SPL. This level was chosen as it was approximately equal to the average spectrum level of the tone complex (60 dB SPL per component, spaced 100 Hz apart). In the second part, three other maskers were also tested. These were an m_- complex, a cosine-phase complex, and a random-phase complex. In the random-phase complex, the phase of each component was taken from a uniform distribution on each interval of a trial. Spectral notches, placed symmetrically around the signal frequency, were introduced into the maskers. The distances between the signal frequency and the edge of the noise (or the nearest masker component) were 0.05, 0.1, 0.2, 0.3, and $0.4f_s$. For the harmonic tone complex maskers, the spectral notches were created by removing 1, 3, 5, 7, and 9 spectral components, respectively. For the noise masker, the notch was created by setting the magnitude values of the noise's Fourier transform to zero within the notch. Therefore, the steepness of the attenuation slopes at the cutoff frequencies was limited only by the gating of the maskers. The stimuli were presented diotically over Sennheiser HD 580 headphones.

An adaptive three-interval 3AFC procedure was used in conjunction with a 2-down 1-up tracking rule to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). Each interval in a trial was separated by an ISI of 500 ms. The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size was 5 dB, which was reduced to 2 dB after the first 4 reversals. The threshold was defined as the mean of the remaining 6 reversals. Three threshold estimates were obtained from each listener in each condition. Three of the four listeners from Experiment 1 (SD, TD, and AO) participated in the first part of this experiment. Only the authors (TD and AO) participated in the second part.

B. Results

The mean data from three listeners, using only the m_+ complex and the Gaussian-noise masker, are shown in the left panel of Fig. 5. Error bars denote ± 1 standard error of the mean across listeners. For the broadband maskers with no notch, the mean threshold in the m_+ condition is about 26 dB lower than in the noise condition. This difference is totally eliminated for all other conditions, and is even slightly reversed at the 0.05 deviation, where only the masker component at the signal frequency was removed. This can probably be understood in terms outlined by Duifhuis (1970, 1971). He found that a high-order harmonic in a pulse train was individually audible only if it was *not* present in the spectrum. This seemingly paradoxical result can be understood if one considers that the operation of removing a component is identical to adding one in antiphase to the original. The time waveform of such stimuli can be seen to be a pulse train with the addition of a sinusoid with a frequency of the removed component (see Fig. 5 of Duifhuis, 1970). In our case, the removal of the single component was equivalent to the addition of a 60-dB SPL sinusoidal masker at the signal frequency, which presumably had the effect of "filling up" the temporal valleys of the internal m_+ masker waveform. In the 0.05 notch condition, a tonal component at the signal frequency was heard in all intervals using the m_+ masker. As the signal had a random starting phase, the addition of the signal could result in an increase, decrease, or no change in the perceived level of that component, depending on the exact phase and level relationship between the masker component and the signal. Detecting a random level change in a tone is somewhat different from detecting the presence of an otherwise absent tone (Green, 1967), which may explain why the mean m_+ threshold for the 0.05 notch is actually higher than that for the noise.

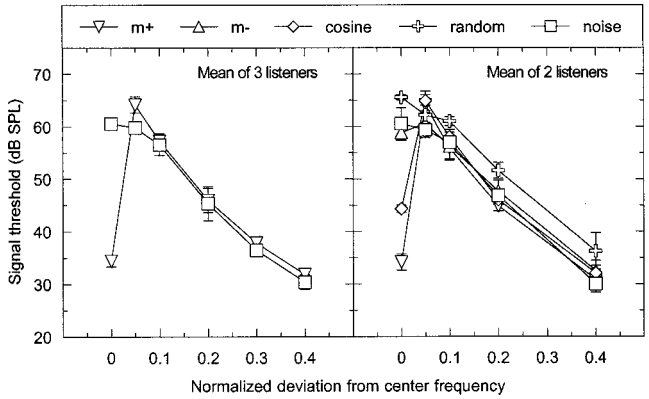

FIG. 5. Mean results from Experiment 2. The left panel shows the mean of three listeners with Gaussian noise (squares) and an m_+ tone complex (triangles) as maskers. The right panel shows the mean from a subset of two of the three listeners for a variety of different maskers (see the legend and the text for details).

Disregarding the no-notch and the 0.05 conditions, the data provide no evidence for different effective frequency selectivity in the presence of an m_+ tone complex. This finding was extended to other harmonic tone complexes, using the two authors as listeners, as can be seen in the right panel of Fig. 5. In the broadband (no-notch) condition, the m_+ complex (downward-pointing triangles) produces the least masking, with the cosine complex (diamonds) producing somewhat more masking, and all the other maskers producing similar amounts, with thresholds ranging from around 58 to 66 dB SPL. In the 0.05 notch condition, the m_+ and co-

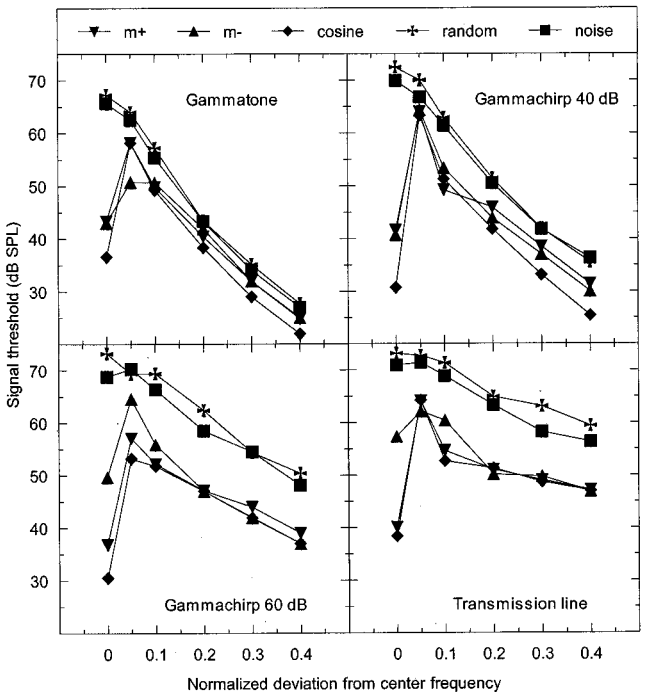

FIG. 6. Model predictions of Experiment 2.

sine maskers both produce the highest thresholds, presumably because both result in the percept of a Duifhuis pitch. Again, except for the 0 and 0.05 conditions, there is no evidence for differences in effective frequency selectivity in the presence of the different maskers. One trend, which was apparent for both listeners, is that the random-phase masker produces the most masking in almost all conditions, even though the dependence of thresholds on notch width is very similar to that found in the other conditions.

These results are reassuring: the idea that different stimuli require different filters would have been unparsimonious and difficult to justify within a physically realizable system. Nevertheless, the frequency selectivity demonstrated is rather good; increasing the total notch width from 200 Hz to 1600 Hz resulted in a mean threshold drop of about 30 dB. It seems unlikely that such good frequency selectivity could be achieved with either the transmission-line or the 60-dB Gammachirp filter. This was tested quantitatively in the following section.

C. Model predictions

The model was identical to the one described in Sec. III A. All four filters were tested again, using the stimuli from Experiment 2. This time, the Gammachirp filter parameters were taken from the 2-kHz parameters shown in Irino and Patterson (1997). The predictions are shown in Fig. 6. Considering only conditions with spectral notches, it is the Gammatone and 40-dB Gammachirp filters that provide the more convincing fits. Both show an average decrease in thresholds of around 30 to 35 dB, compared with around 20 dB for the 60-dB Gammachirp and 15 dB for the transmission-line filter. Furthermore, both of the broader filters predict much higher thresholds in the noise and random-phase conditions than in any of the other conditions. These differences are not observed in the data. None of the four filters predicts lower thresholds in the m_+ condition than in the cosine condition, while the data show about a 10-dB advantage.

The 60-dB Gammachirp filter proved to be somewhat too wide to account for the data from Experiment 2. The discrepancy between the 60-dB Gammachirp predictions and the data is probably primarily due to our method of measuring frequency selectivity, rather than a failure of the Gammachirp filter. We chose to keep the masker spectrum level constant while varying the level of the signal to achieve threshold, as has been done in most studies of frequency selectivity (Patterson, 1976; Patterson and Nimmo-Smith, 1980; Moore and Glasberg, 1983; Glasberg and Moore, 1990). In contrast, the parameters of the 60-dB Gammachirp filter (Irino and Patterson, 1997) were derived from the data of Rosen and Baker (1994), in which the signal level was held constant at 60 dB SPL, while the masker level for each notch width was varied to achieve threshold. Rosen and Baker (1994) and Rosen *et al.* (1998) have argued on both theoretical and empirical grounds that it is more appropriate to hold the signal level constant, thereby assuming that the filter shape is governed by the output level of the filter. The earlier method (e.g., Patterson and Nimmo-Smith, 1980) implicitly assumes that the filter shape is governed by the input sound level to the filter, i.e., the total sound energy, regardless of its frequency content. A recent study of Glasberg and Moore (2000) supports the conclusion of Rosen and colleagues that the filter shape is more accurately defined by keeping the signal level fixed, i.e., that the filter shape is determined by the filter output level. If the fixed-signal method provides a more accurate estimate of filter shape at a given sound level, then our results from Experiment 2 represent a composite filter shape, averaged across levels ranging from 60 to 35 dB SPL. This is expected to be considerably narrower than the ''true'' filter shape for a constant output level of 60 dB SPL, which is the basis of the 60-dB Gammachirp filter.

D. Discussion

Certain conclusions can be drawn from the data and simulations of Experiments 1 and 2. First, none of the filters correctly predicts the pattern of results in the no-notch condition. For instance, in contrast to the data, all filters resulted in lower predicted thresholds for the cosine-phase masker than for the m_+ masker. Also, whereas the data showed very little difference in thresholds between the noise masker and the m_- masker, the simulations predicted threshold differences of between 15 and 30 dB. These discrepancies suggest that the phase response of even the transmission-line or 60-dB Gammachirp filters may not be appropriate for modeling the phase response of the human auditory system. Second, the magnitude response of the Gammatone filter is too narrow to account for the phase effects observed in Experiment 1, but provides a good account of the frequency selectivity measured in Experiment 2. One interpretation of this, which is consistent with the work of Rosen and colleagues (Rosen and Baker, 1994; Rosen *et al.*, 1998), is that the ''true'' frequency selectivity in the presence of the maskers

used in Experiment 1 is considerably poorer (broader) than that implied by Experiment 2. Accordingly, the Gammatone filter with an ERB set to approximate the ERB as defined by Glasberg and Moore (1990), is probably too narrow to account for frequency selectivity with broadband stimuli at moderate sound levels of around 60 dB SPL. The considerably wider 60-dB Gammachirp filter is probably a better approximation.

V. EXPERIMENT 3. CONTRIBUTIONS OF INDIVIDUAL MASKER COMPONENTS

As discussed earlier, the assumption of the fixed-masker or fixed-signal notched-noise paradigm is that the input level or the output level of the filter, respectively, drives changes in filter shape. In fact, both methods (fixed-signal or fixed-masker level) provide a linear approximation to a nonlinear system and so it is highly unlikely that either approach is strictly correct. For instance, both approaches fail to take account of the contributions of suppression to masking (Delgutte, 1990; Moore and Vickers, 1997; Oxenham and Plack, 1998). In order to bypass the input- vs. output-level debate, it is necessary to find a method of measuring filter shape in which both the input and output levels of the filter remain roughly constant. This was achieved in Experiment 3. Noticing that removing the masker component at the signal frequency from the m_+ masker was sufficient to eliminate the m_+/m_- difference, we reasoned that the effect of removing any other single component from the masker could provide a measure of that component's effectiveness, and hence an indirect measure of its attenuation, relative to the central component. Experiment 3 therefore measured the effect of removing a single masker component from m_+ and m_- complexes as a function of the component's frequency.

A. Methods

1. Effect of component frequency

The procedure was the same as in Experiment 2, as was the signal (2 kHz) and the method of stimulus generation and presentation. The masker was a harmonic tone complex with an f_0 of 100 Hz and a level of 60 dB SPL per component. The components were in either m_+ or m_- phase, derived from a complex with components from 100 to 4000 Hz. The experiment investigated the effect of removing a single masker component as a function of the frequency separation between the masker component and the signal. For the m_+ complex, the separations tested were 0, $\pm$100, $\pm$200, $\pm$400, $\pm$800, and $\pm$1200 Hz. For the m_- complex, the effect was considerably smaller (as expected from Experiment 2) and so only separations of 0 and $\pm$400 Hz were tested. Thresholds in the broadband conditions were also tested, as were thresholds for a bandpass condition where only components between 800 and 3200 Hz (i.e., $\pm$1200 Hz) were present.

2. Effect of component level

As discussed in Sec. IV B, the absence of a given component can be treated as an addition of that component in the antiphase. The reasoning behind this experiment was that in the m_+ condition, the effect of the absence (or antiphasic addition) of a given component would decrease as the frequency separation between the component and the signal increased. This can be treated as a measure of the attenuation of a given component, relative to the component at the signal frequency. In other words, plotting the signal threshold as a function of the missing component frequency provides something akin to a masking pattern. However, the masking pattern provides only an indirect measure of filter attenuation. To derive the effective attenuation, it is necessary to know how signal thresholds change as a function of the masker component level at the signal frequency. This was done in the second part of the experiment, using the same signal and m_+ complex as in the first part, by adding a masker component at the signal frequency in antiphase to the original component at that frequency and measuring thresholds as a function of the antiphasic component level. Setting the level of the antiphasic component to be the same as that of the other masker components (0 dB) was equivalent to removing the masker component at that frequency completely. The three listeners who participated in Experiment 2 completed both parts of this experiment.

B. Results and discussion

The individual results from the second part of the experiment were fairly clear-cut. All three listeners showed a relationship between the antiphasic component level and the signal threshold that was very close to linear, at least down to a component level of -24 dB. Individual regression lines for the data down to and including -24 dB had slopes of -1.1, -0.87, and -0.97 for listeners AO, TD, and SD, respectively. The linear relationship between the signal threshold and effective on-frequency component level suggests that we can interpret changes in threshold as a function of signal–component frequency separation directly in terms of filter attenuation, down to levels of -24 dB. These results are plotted in the upper-left panel of Fig. 7. Thresholds in the presence of the m_+ and m_- complexes are shown as downward- and upward-pointing triangles, respectively. The dashed and dotted lines denote thresholds for the broadband m_+ and m_- complexes, respectively. Not shown are thresholds in the presence of the m_+ complex with components from 800 to 3200 Hz only (but with the same phases as in the broadband condition). They were not significantly different from the broadband m_+ thresholds for any listener (t-test, $p>0.05$), suggesting that components below 800 Hz or above 3200 Hz do not contribute to determining thresholds at a signal frequency of 2000 Hz.

In the m_- complex, the effect of removing a component is relatively small (about 5 dB) in the on-frequency condition, and is negligible at a frequency spacing of $\pm$400 Hz. This is presumably because the m_- waveform, once passed through the auditory filters, is not highly modulated and so there are no significant temporal valleys for the additional sinusoid to ''fill up.'' In contrast, removing the on-frequency component in the m_+ complex results in a threshold increase of more than 25 dB. The effect of removing a component decreases with increasing frequency separation, as expected, so that the effect is negligible at the largest separations of 1.2 kHz (0.6 in the normalized units of Experiment 2).

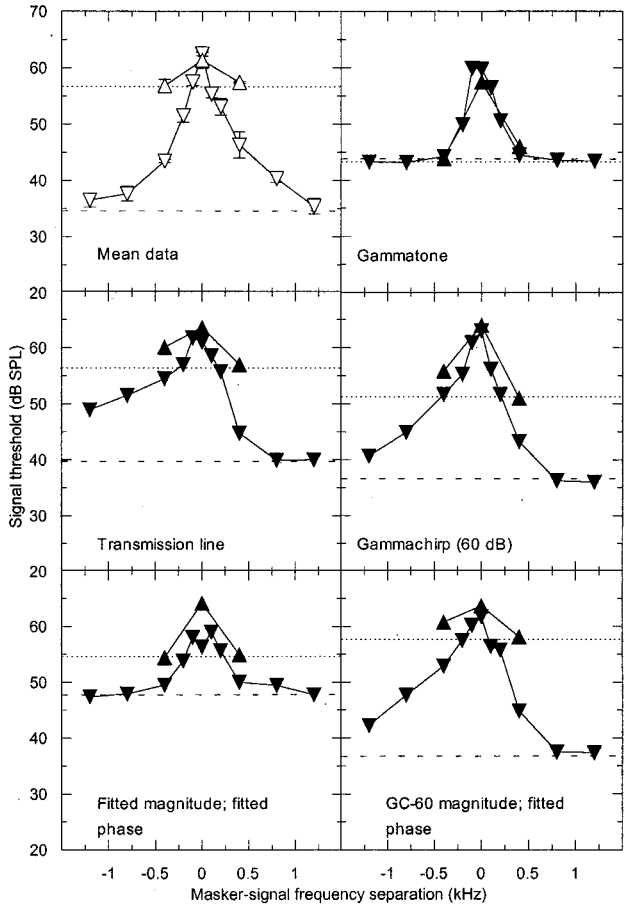

FIG. 7. Mean data (upper-left panel) and simulations from Experiment 3. Thresholds are plotted as a function of the frequency of the missing masker component. Upward- and downward-pointing triangles denote thresholds with the m_- and m_+ masker, respectively. Dotted and dashed lines denote thresholds for the m_- and m_+ maskers, respectively, with no missing components.

As a linear relationship was found between the on-frequency component level and signal threshold for attenuations up to 24 dB, the m_+ thresholds plotted in the upper-left panel of Fig. 7 can be interpreted directly in terms of auditory filter attenuation for all but the 1.2-kHz separation points. However, thresholds for the missing on-frequency component itself may require a more careful interpretation. In this case, a tonal component at the signal frequency was heard in all intervals. As discussed in Experiment 2, the addition of the signal could result in an increase, decrease, or no change, in the perceived level of that component, depending on the exact phase and level relationship between the masker component and the signal. This may explain the sharp increase in thresholds for the central data point.

Although the method used here has the advantage that the input and output level of the filter remain roughly constant, it has the disadvantage that it is not possible to rule out ''off-frequency listening'' (Patterson and Nimmo-Smith, 1980); O'Loughlin and Moore, 1981). Especially at higher signal levels, it may be that the signal is detected best in filters tuned away from the signal frequency. Because of this, the estimates of the filter bandwidth even with this new method may be somewhat too narrow. This is discussed further below.

C. Model predictions

The basic model tested here was the same as that used for Experiments 1 and 2. The filters were the transmission-line filter, the Gammatone, and a 60-dB Gammachirp, all centered at 2 kHz. The predictions are shown in the remaining panels of Fig. 7.

The transmission-line filter shows a highly asymmetric masking pattern, as would be expected from its magnitude response, which is similar to that shown for the 1-kHz transmission-line filter (Fig. 1). The decrease in thresholds above the signal frequency for the m_+ condition is similar to that observed in the data: thresholds are about 16 dB lower at the 0.4-kHz separation than at the peak in both the data and the predictions. However, the decrease in thresholds for components lower than the signal in frequency is far too shallow in the predictions: at the -1.2-kHz separation, predicted thresholds are only 12 dB lower than at the peak, compared with 26 dB in the data.

The 60-dB Gammachirp provides a reasonably good account of the frequency selectivity, thereby validating its use in Experiment 1, and supporting the idea that its failure in Experiment 2 was due to in part to a changing of the effective auditory filter shape in the experiment as the spectral notch widened. However, the tuning is still somewhat too broad on the low-frequency side. For instance, at the -0.4-kHz separation, predicted thresholds are 11 dB lower than at the peak, compared with 19 dB in the data. Some of this discrepancy may be due to off-frequency listening, not accounted for within the model.

The Gammatone filter (upper right panel) provides a reasonable account of frequency selectivity in the tip region of the filter: the decrease in m_+ thresholds is predicted to be about 16 dB for the 0.4- and -0.4-kHz separations, which is in line with the data. The small dynamic range of the predictions makes it impossible to evaluate the suitability of the Gammatone's response at wider frequency separations. As in the previous two experiments, the Gammatone fails to predict any real difference between the masked thresholds of the two complexes.

While the frequency selectivity shown by the 60-dB Gammachirp filter is reasonable, the predicted threshold difference between the broadband m_+ and m_- maskers is still too small. After these simulations had been completed, a new version of the Gammachirp filter was published (Irino and Patterson, 2001). Predictions with this new filter were in fact worse than for the original Gammachirp filter: the predicted difference in thresholds for the broadband m_+ and m_- maskers was only 7 dB, compared with 13 dB for the original Gammachirp, and 22 dB in the experimental data. As the magnitude responses of the old and new Gammachirp filters are very similar, it seems likely that the poorer predictions of the new Gammachirp filter are due to its phase response.

VI. DERIVING A NEW FILTER

The relatively good predictions using the 60-dB Gammachirp filter provide some hope that it may be possible to find a quasi-linear filter with reasonable frequency selectivity that can also be used to describe the phase properties of the

auditory system. However, even this filter exhibits frequency selectivity that is somewhat broader than that observed in the data, while predicting an m_+/m_- difference that is too small. It may be possible to ''fine-tune'' the phase and magnitude response of the filter so that predictions better match the data. In this section, we derive a magnitude response from the data of Experiment 3 and we combine it with an ''optimal'' phase response.

A. Magnitude response

The mean data from the m_+ complex in Experiment 3 were used to define the filter's magnitude response. In the second part of Experiment 3 it was found that the linear relationship between the on-frequency masker component level and the signal held for levels at and above -24 dB. Also, as discussed in Sec. IV B, the high threshold for the missing on-frequency component may be due to factors that did not play a role in any of the other conditions. For these reasons, data from the ± 1.2-kHz and the on-frequency conditions were excluded from the fit. The remaining mean data from the m_+ condition were fitted with a rounded exponential (roex) function (Patterson *et al.*, 1982). The version of the roex filter used here is known as the roex (p,w,t) filter and is described by two rounded exponential functions (one for the peak, and one for the tail) on either side of the function with a weighting factor, w, determining the relative contribution of each. It was assumed that the weighting function and the ratio between the two roex functions were the same on both sides of the filter, so the function had four free parameters, p_u and p_l to describe the slopes above and below the peak of the filter, respectively, w to describe the breakpoint, and tf, which is the factor by which the constant describing the tail of the filter is smaller than that describing the peak. Using a multi-dimensional minimization routine with a least-squares criterion, the best-fitting roex function was found. The final parameters were $p_l = 33.11$, $p_u = 26.74$, $tf = 0.224$, and $w = 0.0402$. The fit is rather good, with a rms error of only 0.78 dB, which is not surprising, given that 4 parameters were used to fit only 8 data points. The resulting ERB of the filter is 308 Hz. This value is somewhat higher than 241 Hz, as proposed by Glasberg and Moore (1990) and implemented in the Gammatone filter, but is still considerably lower than the ERB of the 60-dB Gammachirp filter (582 Hz). The derived magnitude response of the filter, based on the roex equations, is shown in the upper right panel of Fig. 1 (solid curve).

B. Phase response

With the magnitude response of the filter defined, the next step was to specify the phase response. Neither the absolute values of phase as a function of frequency, nor the slope of the function (the first derivative, corresponding to the group delay) can be determined behaviorally, the latter because delays in subsequent processing and reaction time would always swamp any delay introduced by cochlear filtering itself. The types of signal used in this study can, however, be employed to study the curvature, or second derivative, of the phase response (Kohlrausch and Sander, 1995). This is equivalent to the slope of the function relating group delay to frequency at a given CF. Given that our stimuli have constant curvature, our data cannot be sensitive to changes in curvature with frequency. It is therefore assumed that the phase curvature of the auditory filters can be approximated as being constant within the filter passband. Inspection of phase-response data from the basilar membrane (Ruggero *et al.*, 1997; Rhode and Recio, 2000) suggests that this may be a reasonable first approximation. Also, the frequency glides observed by Carney *et al.* (1999) in the impulse responses of auditory-nerve fibers were approximated by straight lines on a frequency–time plot, thereby implying a constant frequency sweep rate, or constant phase curvature. Another assumption is that the lowest predicted thresholds in the complex will occur when the phase curvature of the filter is equal in magnitude, but opposite in sign, to the phase curvature of the m_+ complex (Kohlrausch and Sander, 1995), thereby transforming the filtered m_+ complex into a complex in which all components have the same starting phase. In other words, the lowest thresholds in the m_+ condition will be predicted when the filter has a curvature equal to the m_- complex. Note that this filter curvature does not guarantee the largest predicted m_+/m_- *difference*, as this depends not only on the peakiness of the m_+ envelope, but also on the flatness of the m_- envelope. However, pilot simulations allowing the filter phase curvature to vary showed that the predicted m_+/m_- differences using the optimum phase were less than 0.5 dB larger than the differences predicted simply by using the phase curvature of the m_- complex. Therefore, the curvature of the m_- complex from Experiment 3, namely $-5\times10^{-6}\pi$ rad/Hz^2, was used to define the new filter's phase curvature. The realization of this curvature, as used in our simulations, is shown in the lower right panel of Fig. 1.

C. Predictions

The derived magnitude and phase response of the new filter were combined and an impulse response was calculated by applying an inverse Fourier transform. The new filter was then incorporated into the model of Dau *et al.* (1997a). The resulting predictions of the data from Experiment 3 are shown in the lower left panel of Fig. 7. While the frequency selectivity is in line with the data for thresholds above the asymptotic threshold (dashed and dotted lines), the model fails to describe either the large m_+/m_- difference or the large effect of adding an on-frequency antiphasic component at the signal frequency. Thus, even using a filter with a fitted frequency response and optimized phase response, the model of Dau *et al.* does not successfully predict the data.

One reason for this failure might be that our estimate of frequency selectivity is in error. For instance, as previously mentioned, off-frequency listening was not accounted for in the filter fitting procedure. As discussed in many previous studies (e.g., Patterson, 1976), this would lead to our estimate of the filter bandwidth being too narrow. Whatever the reason, it seems clear that the derived frequency selectivity is too narrow to allow the model to predict a sufficiently large m_+/m_- difference. For this reason, we returned to the 60-dB Gammachirp filter. The parameters of the Gammachirp filter (Irino and Patterson, 1997) were selected to fit

data obtained by Rosen and Baker (1994) using the notched-noise technique, which is designed to minimize off-frequency listening. As discussed earlier, the Gammachirp filter predicts somewhat broader tuning than observed in Experiment 3, but that may be primarily due to the model not taking off-frequency listening into account. The predicted m_+/m_- difference was also too small with the 60-dB Gammachirp filter, but that may be due to the filter's nonoptimal phase response. The next simulation combined the magnitude response of the 60-dB Gammachirp filter with the optimal phase response, as described above.

The predictions from combining the 60-dB Gammachirp filter's magnitude response and the fitted phase response are shown in the lower right panel of Fig. 7. The fit to the data is rather good. In particular, these are the first simulations of this study in which the m_+/m_- difference found in the data is matched well by the model. The predictions are an improvement on the predictions of the original 60-dB Gammachirp, primarily because of the higher predicted thresholds in the m-conditions. This results in an overall predicted m_+/m_- difference of about 21.7 dB, as opposed to 13.3 dB for the original 60-dB Gammachirp. A mean m_+/m_- difference of 22 dB was observed in the data. The results therefore provide the first evidence that predictions from a single-channel model may indeed be able to quantitatively match the results from such experiments, even without taking into account nonlinear effects such as suppression.

It is important to note that a single-filter model, with a bandwidth sufficiently narrow to predict traditional measures of frequency selectivity (Rosen *et al.*, 1998), was still able to predict the effects of components extending well beyond the traditional critical band. This is probably due to the highly modulated response to the m_+ masker, so that even highly attenuated components can affect the short-term power of the masker in its low-level epochs. This may explain the findings of Lin and Hartmann (1997), who reported that the strength of the Duifhuis (1970) pitch was influenced by components well outside the traditional critical band. Lin and Hartmann suggested that their results were evidence for the influence of across-channel processing. In contrast, our data and simulations suggest that their results may in fact be accounted for using a single-channel model.

At least two discrepancies between the data and the predictions remain. First, in Experiment 3 it was found that removing masking components below 800 Hz and above 3200 Hz had no effect on masked thresholds for the 2000-Hz signal. In the simulations, predicted thresholds in the m_+ masker increased by 2.5 dB when these components were removed. This suggests that the frequency selectivity assumed by the Gammachirp filter is still somewhat too broad: just as the constant masker level paradigm (Glasberg and Moore, 1990) may underestimate auditory filter bandwidth, the constant probe level paradigm (Rosen *et al.*, 1998) may overestimate it somewhat. Second, the data from Experiment 3 suggest a rather symmetric filter shape, whereas the shape of the 60-dB Gammachirp filter and its predictions are rather asymmetric. Overall, though, these discrepancies do not appear to represent fundamental failings of the model, and could probably be reduced with further fine-tuning of the filter shape.

D. Effects of peripheral compression

It has been shown by Carlyon and Datta (1997a,b) and Summers and Leek (1998) that stimulus level plays a major role in determining the m_+/m_- difference. Carlyon and Datta found that the difference was greatly reduced at low levels, an effect that could be due to narrower filters, more linear peripheral processing at low levels, a change in the filter's phase response at low levels, or a combination of these. Summers and Leek found that the m_+/m_- difference became somewhat smaller at very high levels, and was greatly reduced or even absent in listeners with sensorineural hearing loss, suggesting that cochlear nonlinearity may play an important role, and filter bandwidth (which should if anything be broader at high levels and in impaired listeners) is not the sole determining factor. The detection model used in the present study does not include any static nonlinearity, such as is observed on the basilar membrane (BM), and instead introduces time-dependent nonlinearities within the feedback loops. It is not clear to what extent these can be substituted for the effects of cochlear compression. To address this issue, further simulations were carried out using a version of the temporal-window model (Moore *et al.*, 1988), which incorporates a static compressive nonlinearity prior to temporal integration (Oxenham and Moore, 1994; Plack and Oxenham, 1998).

The temporal-window model was tested with the final model (60-dB Gammachirp magnitude response together with the ''optimal'' phase response). In Experiments 2 and 3, the signal was added in random phase to the masker. The temporal-window model in its current form is a deterministic model, and so cannot predict thresholds for randomly varying stimuli. For this reason, it was assumed that the signal could be approximated as adding, on average, in quadrature phase to the masker. This is the same as adding the intensities of the two stimuli and is basically the same procedure that was used by Oxenham *et al.* (1997). After filtering, the envelopes of the masker and signal were extracted using a Hilbert transform. The envelopes were squared to produce intensity-like quantities and then added before being subjected to a power-law compression (p). The value of p ranged from 1 (energy detection) down to 0.16, which is the value found by Oxenham and Plack (1997) at mid levels in their behavioral measure of BM compression. The stimuli were then passed through a temporal window of various shapes and durations. The smallest window was a Hanning window with a 2.5-ms half-amplitude duration. The longest was a rectangular window that spanned an entire period of the masker waveform (10 ms). Detection was assumed to occur at the instant in time when the ratio between the window output due to the masker-plus-signal and that due to the masker was maximal. This ratio was assumed to be the same for all conditions.

Predictions of Experiment 3 using the temporal-window model could be made similarly good to those of the Dau *et al.* model with the right combination of window shape and nonlinearity. The effect of nonlinearity was dependent on the

window size: the wider the temporal window, the more effect had the compression. This can be understood in terms of the differences between simultaneous and nonsimultaneous masking when using compression (Oxenham *et al.*, 1997). For very short windows, such as the 2.5-ms Hanning window, the predicted m_+/m_- difference with no missing components was larger than that observed in the data and was very similar for both the maximum ($p=0.16$) and minimum ($p=1$) compression applied (37.8 and 35.6 dB, respectively). For the 10-ms rectangular window, the predicted m_+/m_- differences were 26.8 and 0 dB for $p=0.16$ and $p=1$, respectively. The 0-dB prediction is the obvious outcome of an energy detector, integrating over a complete period of the masker, as both m_+ and m_- maskers have the same overall energy. Finally, we tested the double-exponential window employed by Plack and Oxenham (1998) and Oxenham and Plack (2000) to account for forward masking. This window, which has an equivalent rectangular duration of approximately 8 ms, predicted m_+/m_- differences of 28.4 dB and 0.2 dB for $p=0.16$ and $p=1$, respectively.

In summary, when using temporal integration windows of the shape and duration used in previous studies (Oxenham and Plack, 2000), the effects of peripheral compression are dramatic: ''normal'' compression can produce predicted m_+/m_- differences in excess of 25 dB, while linear processing produces essentially no difference. This finding provides quantitative support for the idea that differences in within-channel peripheral compression are sufficient to account for the differences in masker phase effects between normal-hearing and hearing-impaired listeners (Summers and Leek, 1998).

VII. SUMMARY

In Experiment 1 thresholds were measured in the presence of m_+ and m_- maskers at signal frequencies of 1 and 4 kHz. The two main findings were: (i) large threshold differences between m_+ and m_- maskers require a relatively wide masker bandwidth, irrespective of the number of masker components; and (ii) contralateral masker components have no effect on thresholds, suggesting no influence of across-channel mechanisms, at least beyond the level of binaural interaction.

Experiment 2 used a variant of the notched-noise technique to show that the effective frequency selectivity of the auditory system is the same for noise and harmonic complex tones. However, following Rosen and Baker (1994), it was concluded that the technique of varying the signal level while keeping the masker spectrum level constant probably underestimates auditory filter bandwidth at a given masker level.

Experiment 3 introduced a new method of estimating frequency selectivity, in which the masker level and frequency content remain roughly constant in all conditions, thereby bypassing any debate about whether filter input or output level governs filter shape. However, even this new method may underestimate the auditory filter bandwidth, as it does not take into account the possible effects of off-frequency listening.

The results from all three experiments were simulated using a detection model proposed by Dau *et al.* (1997a) with a variety of front-end filters, including the Gammatone, the Gammachirp, and the transmission-line model proposed by Strube (1985). None of these filters was able to provide a convincing fit to the data. The data and simulations suggest that none of the current models of auditory filtering successfully matches both the frequency selectivity and the phase response of the human auditory system. The Gammatone filter has a bandwidth that is too narrow to account for the present data, regardless of its assumed phase response. The transmission-line filter bandwidth was too broad to account for results from frequency–selectivity experiments. The Gammachirp filter, while having a sufficiently wide bandwidth, does not mimic the phase response of the auditory system with sufficient accuracy to predict the large m_+/m_- differences found in the data.

The data from Experiment 3 were well predicted by combining the magnitude response of the Gammachirp filter with a phase response derived from the m_- stimulus. This shows that phase effects can, in principle, be accounted for with a single-channel model. More generally, some effects that appear to require across-channel processing (Lin and Hartmann, 1997) may in fact be accounted for with an appropriate single-channel model. Good predictions were also obtained using the same filter in conjunction with the temporal-window model. Eliminating the static compression from the temporal-window model was sufficient to completely eliminate any predicted m_+/m_- masking differences. The importance of peripheral compression in these simulations may explain why hearing-impaired listeners, who are thought to have little or no peripheral compression, also show very small m_+/m_- masking differences.

Finally, as all the tested models of peripheral filtering failed to produce quantitatively accurate predictions, these data promise to provide strong constraints for testing future models of peripheral auditory processing.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD Grant No. R01 DC 03909 (A. J. O.) and by the Max Kade Foundation and the Deutsche Forschungsgemeinschaft (DFG) (T. D.). Stephan Ewert provided help with the programming. Comments by the reviewers, Larry Feth and Van Summers, greatly improved the manuscript.

[1]In fact, Smith *et al.* (1986) used the opposite designation. However, all studies since then have used the convention applied here.

[2]At levels at which a difference between m_+ and m_- thresholds is predicted, the response of Strube's nonlinear model is very similar to that of his linear version.

Alcántara, J. I., Holube, I., and Moore, B. C. J. (**1996**). ''Effects of phase and level on vowel identification: Data and predictions based on a nonlinear basilar-membrane model,'' J. Acoust. Soc. Am. **100**, 2382–2392.

Buss, E., Grose, J. H., and Hall, J. W. (**1998**). ''Fast-acting compression of auditory stimuli: A test with narrowband stimuli,'' Assoc. Res. Otolaryngol. Abs., Vol. 21, p. 798.

Carlyon, R. P., and Datta, A. J. (**1997a**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Datta, A. J. (**1997b**). ''Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components,'' J. Acoust. Soc. Am. **101**, 3648–3657.
Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). ''Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,'' J. Acoust. Soc. Am. **102**, 2906–2919.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). ''A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). ''A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements,'' J. Acoust. Soc. Am. **99**, 3623–3631.
Delgutte, B. (**1990**). ''Physiological mechanisms of psychophysical masking: Observations from auditory-nerve fibers,'' J. Acoust. Soc. Am. **87**, 791–809.
Derleth, R. P., and Dau, T. (**2000**). ''On the role of envelope fluctuation processing in spectral masking,'' J. Acoust. Soc. Am. **108**, 285–296.
Duifhuis, H. (**1970**). ''Audibility of high harmonics in a periodic pulse,'' J. Acoust. Soc. Am. **48**, 888–893.
Duifhuis, H. (**1971**). ''Audibility of high harmonics in a periodic pulse. II. Time effects,'' J. Acoust. Soc. Am. **49**, 1155–1162.
Ellis, D. P. W. (**1999**). ''Using knowledge to organize sound: The prediction-driven approach to computational auditory scene analysis and its application to speech/nonspeech mixtures,'' Speech Commun. **27**, 281–298.
Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.
Giguère, C., and Woodland, P. C. (**1994a**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.
Giguère, C., and Woodland, P. C. (**1994b**). ''A computational model of the auditory periphery for speech and hearing research. II. Descending paths,'' J. Acoust. Soc. Am. **95**, 343–349.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.
Godsmark, D., and Brown, G. J. (**1999**). ''A blackboard architecture for computational auditory scene analysis,'' Speech Commun. **27**, 351–366.
Green, D. M. (**1967**). ''Additivity of masking,'' J. Acoust. Soc. Am. **41**, 1517–1525.
Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectro-temporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.
Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.
Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.
Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.
Lentz, J. J., Richards, V. M., and Matiasek, M. R. (**1999**). ''Different auditory filter bandwidth estimates based on profile analysis, notched noise, and hybrid tasks,'' J. Acoust. Soc. Am. **106**, 2779–2792.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lin, J.-Y., and Hartmann, W. M. (**1997**). ''On the Duifhuis pitch effect,'' J. Acoust. Soc. Am. **101**, 1034–1043.
Meddis, R., and Hewitt, M. (**1991**). ''Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.
Moore, B. C. J., and Vickers, D. A. (**1997**). ''The role of spread of excitation and suppression in simultaneous masking,'' J. Acoust. Soc. Am. **102**, 2284–2290.
O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''Off-frequency listening: effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.
Oxenham, A. J., and Moore, B. C. J. (**1994**). ''Modeling the additivity of nonsimultaneous masking,'' Hear. Res. **80**, 105–118.
Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: Evidence for the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**, 3676–3687.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**, 3500–3510.
Oxenham, A. J., and Plack, C. J. (**2000**). ''Effects of masker frequency and duration in forward masking: Further evidence for the influence of peripheral nonlinearity,'' Hear. Res. **150**, 258–266.
Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., and Nimmo-Smith, I. (**1980**). ''Off-frequency listening and auditory filter asymmetry,'' J. Acoust. Soc. Am. **67**, 229–245.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
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.
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.
Rosen, S., and Baker, R. J. (**1994**). ''Characterising auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
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. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).
Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). ''Phase effects in masking related to dispersion in the inner ear,'' J. Acoust. Soc. Am. **80**, 1631–1637.
Strube, H. W. (**1985**). ''A computationally efficient basilar-membrane model,'' Acustica **58**, 207–214.
Strube, H. W. (**1986**). ''The shape of the nonlinearity generating the combination tone $2f_1-f_2$,'' J. Acoust. Soc. Am. **79**, 1511–1518.
Summers, V., and Leek, M. R. (**1998**). ''Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners,'' Hear. Res. **118**, 139–150.
Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). ''Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model,'' J. Acoust. Soc. Am. **106**, 2733–2745.