Second-order temporal modulation transfer functions

Christian Lorenzi,[a] Catherine Soares, and Thomas Vonner
Laboratoire de Psychologie Expérimentale, UMR CNRS 8581, Institut de Psychologie, Université René Descartes Paris V, 71, Av. Edouard Vaillant, 92774 Boulogne-Billancourt, France

(Received 3 October 2000; revised 15 February 2001; accepted 3 May 2001)

Detection thresholds were measured for a sinusoidal modulation applied to the modulation depth of a sinusoidally amplitude-modulated (SAM) white noise carrier as a function of the frequency of the modulation applied to the modulation depth (referred to as f'_m). The SAM noise acted therefore as a ''carrier'' stimulus of frequency f_m, and sinusoidal modulation of the SAM-noise modulation depth generated two additional components in the modulation spectrum: $f_m - f'_m$ and $f_m + f'_m$. The tracking variable was the modulation depth of the sinusoidal variation applied to the ''carrier'' modulation depth. The resulting ''second-order'' temporal modulation transfer functions (TMTFs) measured on four listeners for ''carrier'' modulation frequencies f_m of 16, 64, and 256 Hz display a low-pass segment followed by a plateau. This indicates that sensitivity to fluctuations in the strength of amplitude modulation is best for fluctuation rates f'_m below about 2–4 Hz when using broadband noise carriers. Measurements of masked modulation detection thresholds for the lower and upper modulation sideband suggest that this capacity is possibly related to the detection of a beat in the sound's temporal envelope. The results appear qualitatively consistent with the predictions of an envelope detector model consisting of a low-pass filtering stage followed by a decision stage. Unlike listeners' performance, a modulation filterbank model using Q values $\geqslant 2$ should predict that second-order modulation detection thresholds should decrease at high values of f'_m due to the spectral resolution of the modulation sidebands (in the modulation domain). This suggests that, if such modulation filters do exist, their selectivity is poor. In the latter case, the Q value of modulation filters would have to be less than 2. This estimate of modulation filter selectivity is consistent with the results of a previous study using a modulation-masking paradigm [S. D. Ewert and T. Dau, J. Acoust. Soc. Am. **108**, 1181–1196 (2000)]. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1383295]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [SPB]

I. INTRODUCTION

The characteristics of the temporal envelope of acoustic stimuli may play a crucial role in sound identification. Sensitivity to the temporal envelope is traditionally assessed by measuring a temporal modulation transfer function (TMTF), which displays the ability of listeners to detect sinusoidal amplitude modulation (SAM) as a function of the frequency of that modulation (Viemeister, 1977, 1979). Empirically, TMTFs are obtained by measuring the modulation depth, *m*, necessary to just detect the modulation of a SAM carrier as a function of the modulation frequency, f_m. Usually, TMTFs are measured with broadband noise carriers so as to preclude the use of spectral cues, since the modulation of broadband noise does not affect its (flat) long-term power spectrum. Such TMTFs are generally low pass in shape: Sensitivity to SAM is relatively independent of modulation frequency up to 50–60 Hz, and decreases progressively at higher modulation frequencies (e.g., Rodenburg, 1977; Viemeister, 1977, 1979; Bacon and Viemeister, 1985).

A traditional model used to account for the characteristics of these TMTFs is the ''linear envelope detector model.'' This model assumes that the temporal envelope of the stimuli is smoothed by a single (first-order) low-pass filter operating at a central (postcochlear) level (Viemeister, 1979; Forrest and Green, 1987; Strickland and Viemeister, 1996; Lorenzi *et al.*, 1999). Amplitude fluctuations faster than the cutoff frequency of this low-pass filter (estimated to be about 60–150 Hz) are attenuated, and this may explain why SAM detection thresholds increase at high modulation frequencies. The results of adaptation and masking experiments performed with amplitude-modulated sounds (e.g., Kay and Matthews, 1972; Green, 1976; Green and Kay, 1974; Tansley and Suffield, 1983; Houtgast, 1989; Bacon and Grantham, 1989; Yost *et al.*, 1989; Dau *et al.*, 1997a, b, 1999) have, however, suggested an alternative model, the so-called ''modulation filterbank,'' in which modulation filters, each tuned to a given modulation frequency, decompose the temporal envelope of sounds at a central level. In this approach, modulation at a given frequency is assumed to be detected by monitoring the output of a modulation filter tuned close to that frequency (e.g., Lorenzi *et al.*, 1995). In recent implementations of the modulation filterbank (Dau *et al.*, 1997a, b, 1999; Ewert and Dau, 2000), the bandwidth of modulation filters is assumed to increase with increasing center modulation frequency. The Q value of such filters was initially assumed to be equal to 2 for center modulation frequencies above 10 Hz (Dau *et al.*, 1997a, b, 1999). However, in a more recent implementation of the modulation filterbank, the Q value of such filters is assumed to be equal to 1 for center modulation frequencies up to 64 Hz (Ewert and Dau, 2000).

[a]Electronic mail: christian.lorenzi@psycho.univ-paris5.fr

 0001-4966/2001/110(2)/1030/9/$18.00

For a broadband noise carrier, the power of the intrinsic random fluctuations in the noise carrier appearing at the output of a modulation filter will therefore increase with increasing filter bandwidth (that is, with increasing filter center modulation frequency). The masking effect produced by these intrinsic random fluctuations will increase at high modulation frequencies, and this may explain why SAM detection thresholds degrade at high modulation frequencies (Dau *et al.*, 1997a, b, 1999). To date, the controversy regarding the nature of the temporal processor (a modulation filterbank or a linear envelope detector) still persists.

TMTFs measured with pure-tone carriers are quite different in shape (e.g., Zwicker, 1952; Viemeister, 1979; Fassel, 1994; Dau, 1996; Kohlrausch *et al.*, 2000). For a 5-kHz pure-tone carrier, SAM detection thresholds are constant between about 10–100 Hz. They increase in the range 100–400 Hz at a rate of 3–4 dB/oct. Finally, detection thresholds decrease abruptly above about 400 Hz (Dau, 1996; Kohlrausch *et al.*, 2000). The modulation frequency at which this rolloff occurs increases with the carrier frequency (Fassel and Püschel, 1993; Fassel, 1994; Dau, 1996; Kohlrausch *et al.*, 2000). The high-frequency region of these TMTFs can be understood easily on the basis of the auditory filter bank model (Fletcher, 1940; Patterson, 1976): (1) At high modulation frequencies, the sidebands of the modulated signal are spectrally resolved. Instead of basing decisions on the amplitude fluctuations, subjects listen to additional tones. (2) Tone-on-tone experiments (Dau, 1996; Kohlrausch *et al.*, 2000) also show that, at high modulation frequencies, SAM detection thresholds are determined by the masked threshold of the spectrally resolved lower sideband. (3) Finally, the progressive broadening of auditory filter bandwidths at high (audio) frequencies explains why the modulation frequency at which the rolloff occurs increases with the carrier frequency.

Thus, the characteristics of TMTFs measured with pure-tone carriers indirectly support the existence of a limited spectral decomposition of complex sounds within a bank of channels (the auditory filterbank). With this idea in mind, we suggest to test the existence of filters in the modulation domain by measuring ''second-order'' amplitude modulation detection thresholds, that is detection thresholds of a sinusoidal modulation applied to the modulation depth of a SAM white noise carrier (instead of unmodulated white noise or pure-tone carriers). The tracking variable would be the modulation depth of the sinusoidal variation applied to the modulation depth of the SAM noise. The effects of applying sinusoidal modulation to the depth of a SAM noise are illustrated in Fig. 1. In the top panel, the left signal corresponds to the waveform of a white noise sinusoidally amplitude modulated at a modulation frequency f_m of 16 Hz, with a constant modulation depth *m* of 50%. The right signal also corresponds to the waveform of a white noise sinusoidally amplitude modulated at $f_m = 16\,\mathrm{Hz}$; however, sinusoidal modulation has been applied to its modulation depth *m* at a frequency f'_m of 8 Hz, with a modulation depth m' of 50%. The cyclic variation in modulation depth produces a beat in its temporal envelope at a slow rate equal to f'_m. The bottom panel of Fig. 1 shows the modulation spectra of the noise

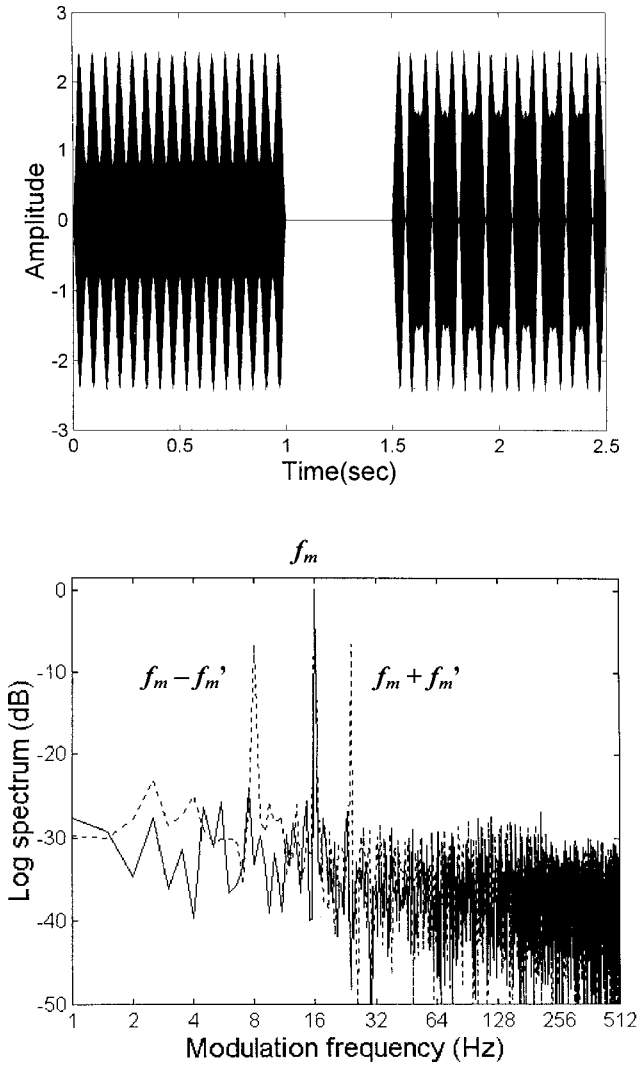

FIG. 1. Top panel, left signal: Waveform of a white noise with first-order SAM ($f_m = 16$ Hz, $m = 0.5$). Top panel, right signal: Waveform of a white noise with second-order SAM ($f_m = 16$ Hz, $f'_m = 8$ Hz, $m = m' = 0.5$). Bottom panel: Modulation spectra of the noises with first-order (continuous line) and second-order (dotted line) SAM.

with first-order (solid line) and second-order (dotted line) SAM (the modulation spectra are calculated for only one realization of the noise carrier). As expected, sinusoidal modulation applied to the modulation depth of a SAM noise carrier generates sidebands at $f_m - f'_m$ and $f_m + f'_m$ in the modulation spectrum of the ''second-order'' SAM noise.

By analogy with the results obtained with TMTFs measured with pure-tone carriers, the modulation filterbank model should therefore predict the following. (1) Second-order SAM detection thresholds should decrease at high second-order modulation frequencies f'_m (because modulation sidebands should be spectrally resolved at high second-order modulation frequencies if modulation filters are *narrowly* tuned). (2) Knowing that low SAM components are better detected than high ones (as a consequence of better modulation tuning to low modulation center frequencies), second-order SAM detection thresholds should be determined by the masked threshold of the resolved lower modulation sideband [masking being considered in the modulation domain, since previous studies (e.g., Bacon and Grantham,

1989; Houtgast, 1989) have shown that the detectability of SAM is generally degraded when a masking SAM is added]. (3) The second-order modulation frequency at which the rolloff should occur in the second-order TMTF should increase with increasing ''carrier'' modulation frequency, f_m (because of the progressive broadening of modulation filters at high center modulation frequencies). By contrast, the linear envelope detector model—a sluggish mechanism that smoothes fast fluctuations—should predict that detectability of second-order SAM should degrade at high second-order modulation frequencies. A very similar method was used by Rees and Kay (1985) to test the hypothesis of selective channels in the frequency-modulation domain.

The present study reports second-order TMTFs measured at various ''carrier'' modulation frequencies in normal-hearing listeners. First-order TMTFs and masked modulation detection thresholds were also collected to evaluate the extent to which second-order modulation sensitivity relates to first-order modulation sensitivity, modulation masking, and envelope beat detection. The empirical results are finally discussed in light of the modulation filterbank and linear envelope detector models.

II. METHOD

A. Listeners

Four listeners with normal hearing (mean age=20 years; s.d.=1 year), CS, TV, SA, and SR, participated in the experiments.

B. Stimuli and procedures

All psychophysical experiments were controlled by a PC-compatible computer. All stimuli were generated using a 16-bit D/A converter at a sampling frequency of 44.1 kHz, and were delivered binaurally (i.e., diotically) via Sennheiser HD 565 earphones. Statistically independent realizations of the white noise were used in all experiments (i.e., within and between trials). Listeners were tested individually in a soundproof booth. In each task, the standard and target stimuli were presented at 75 dB SPL. Both the standard and target stimuli had a 2-s duration including 25-ms rise/fall times shaped using a raised-cosine function. The interstimulus interval was 1 s.

1. First-order TMTFs

The listener's task was to detect the presence of a SAM applied to a white noise carrier. On each trial, a standard and a target stimulus were successively presented in random order to the listener. The standard, $S(t)$, consisted of a white noise $n(t)$. The target, $T(t)$, consisted of a white noise carrier sinusoidally amplitude modulated at a given modulation frequency. The expression describing the target was

$$T(t)=c[1+m\sin(2\pi f_m t+\phi_m)]n(t), \tag{1}$$

where m is the modulation depth ($0\leqslant m\leqslant 1$), f_m is the modulation frequency (f_m was 4, 8, 16, 32, 64, 128, or 256 Hz), and ϕ_m is the starting phase of the modulation, randomized on each interval. The term c is a multiplicative compensation term (Viemeister, 1979) set such that the overall power was the same in all intervals. The expression for c is given as follows:

$$c=[1+m^2/2]^{-0.5}. \tag{2}$$

The SAM detection thresholds were obtained using an adaptive two-interval, two-alternative forced-choice (2I, 2AFC) procedure with a two-down, one-up stepping rule that estimates the modulation depth, m, necessary for 70.7% correct detection. The listener's task was to identify the interval containing the modulation. Visual feedback about the correct interval was given after each trial. The step size of m variation corresponded initially to a factor of 1.585 [4 dB in decibels ($20\log m$)]; it was reduced to 1.26 (2 dB) after the first two reversals. The mean of the last 10 reversals in a block of 16 reversals was taken as the threshold estimate for that block (in %). For each listener and each modulation frequency, thresholds presented here are based upon three estimates. The worst threshold that can be measured corresponds to a modulation depth of 1 (100% modulated noise). The closer to 0 the value of m, the better the detection threshold.

2. Second-order TMTFs

The listener's task was to detect the presence of a sinusoidal modulation applied to the modulation depth of a SAM white noise carrier. On each trial, a standard and a target stimulus were successively presented in random order to the listener. The standard, $S(t)$, consisted of a white noise $n(t)$ sinusoidally amplitude modulated at a given modulation frequency f_m, with a fixed modulation depth m of 50% ($m=0.5$). The expression describing the standard was

$$S(t)=[1+m\sin(2\pi f_m t+\phi_m)]n(t) \tag{3}$$

where ϕ_m represents the starting phase of the modulation, randomized on each interval. The target, $T(t)$, consisted of a white noise sinusoidally amplitude modulated at a given modulation frequency f_m. Modulation depth was sinusoidally amplitude modulated at a given modulation frequency f'_m. The expression describing the target was

$$T(t)=[1+[0.5+m'\sin(2\pi f'_m t+\phi'_m)]\times\sin(2\pi f_m t+\phi'_m)]n(t), \tag{4}$$

where m' is the modulation depth of modulation-depth variation ($0\leqslant m'\leqslant 0.5$), f'_m is the frequency of modulation depth variation, and ϕ'_m represents the starting phase of modulations, randomized on each interval. The overall power was the same in all intervals. f_m (the ''carrier'' modulation frequency) was 16, 64, or 256 Hz. f'_m (the ''second-order'' modulation frequency) was 1, 2, 3, 4, 5, 6, 7, 8, or 11 Hz when f_m was 16 Hz; 3, 4, 6, 8, 11, 16, 23, 32, or 45 Hz when f_m was 64 Hz; and 3, 8, 23, 32, 45, 64, 90, 128, or 181 Hz when f_m was 256 Hz.

Second-order SAM detection thresholds were obtained using an adaptive 2I, 2AFC procedure with a two-down, one-up stepping rule that estimates the second-order modulation depth, m', necessary for 70.7% correct detection. The listener's task was to identify the interval containing the *modulation of modulation*. Visual feedback about the correct interval was given after each trial. The step size of m' varia-

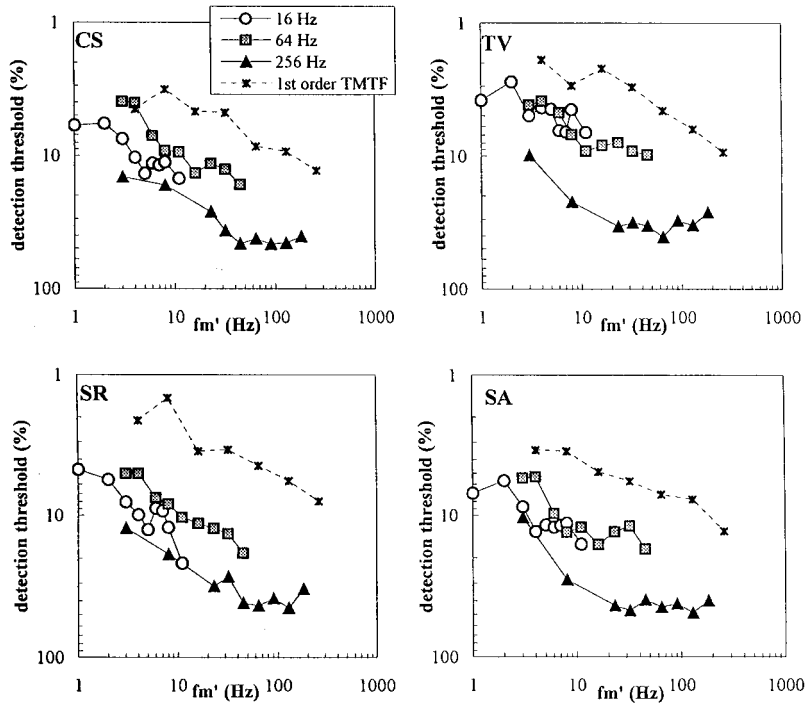

FIG. 2. Individual data for the four listeners. In each panel, the second-order TMTFs measured for $f_m = 16$ Hz (solid lines with unfilled circles), $f_m = 64$ Hz (solid lines with gray squares), and $f_m = 256$ Hz (solid lines with black triangles) are plotted along with the first-order TMTF (dotted lines with stars). In the case of second-order TMTFs, the ordinate indicates second-order modulation depth at threshold m', and the abscissa represents f'_m; in the case of first-order TMTFs, the ordinate indicates first-order modulation depth at threshold m, and the abscissa represents f_m. In both cases, modulation depth at threshold is expressed in linear units, and plotted on a logarithmic scale.

tion corresponded initially to a factor of 1.585 [4 dB in decibels ($20 \log m'$)]; this factor was reduced to 1.26 (2 dB) after the first two reversals. The mean of the last 10 reversals in a block of 16 reversals was taken as the threshold estimate for that block (in %). For each listener and experimental condition, thresholds presented here are based upon three estimates. The worst threshold that can be measured corresponds to a modulation depth m' of 0.5. The closer to 0 the value of m', the better the detection threshold.

3. Masked modulation detection thresholds

The third task was based on a classical ''modulation masking'' paradigm. On each trial, a standard and a target stimulus were successively presented in random order to the listener. The standard consisted of a white noise $n(t)$ sinusoidally amplitude modulated at the masker frequency. The expression describing the standard (i.e., the masker alone) was

$$S(t) = c[1 + m_{\text{mask}} \cos(2\pi f_{\text{mask}} t)]n(t), \tag{5}$$

where m_{mask} is the modulation depth of the masker, fixed at 0.5, f_{mask} is the masker modulation frequency, and c is a multiplicative compensation term defined below. f_{mask} was fixed at 64 Hz. The target consisted of white noise carrier $n(t)$ amplitude modulated by the sum of two sinusoidal modulators, the signal and the masker modulators. The expression describing the target (i.e., the signal-plus-masker waveform) was

$$T(t) = c[1 + m_{\text{mask}} \cos(2\pi f_{\text{mask}} t) + m_{\text{sig}} \cos(2\pi f_{\text{sig}} t)]n(t), \tag{6}$$

where m_{sig} is the modulation depth of the signal and f_{sig} is the signal modulation frequency. The masker and signal were in phase. The frequency f_{sig} was 32, 56, 61, 67, 72, or 96 Hz (the absolute spectral distance between signal and masker modulations was therefore of 3, 8, or 32 Hz, the signal modulation frequency being either below or above the masker modulation frequency). The term c is a compensation factor (Bacon and Grantham, 1989) set such that the overall power was the same in all intervals. The expression for c is given as follows:

$$c = [1 + (m_{\text{mask}}^2 + m_{\text{sig}}^2 + 2m_{\text{mask}} m_{\text{sig}})/2]^{-0.5}. \tag{7}$$

Masked modulation detection thresholds were obtained using an adaptive 2I, 2AFC procedure with a two-down, one-up stepping rule that estimates the signal modulation depth, m_{sig}, necessary for 70.7% correct detection. The listener's task was to choose the interval containing the signal modulation. Visual feedback as to the correct interval was given after each trial. The step size of m_{sig} variation corresponded initially to a factor of 1.585 (4 dB); this factor was reduced to 1.26 (2 dB) after the first two reversals. The mean of the last 10 reversals in a block of 16 reversals was taken as the threshold estimate for that block (in %). For each listener and experimental condition, thresholds presented here are based upon three estimates. The worst threshold that can be measured corresponds to a modulation depth of 0.5. The closer to 0 the value of m_{sig}, the better the detection threshold.

III. RESULTS

A. First- and second-order TMTFs

Individual and mean data for the four listeners are shown in Figs. 2 and 3, respectively. In each figure and each panel, the second-order TMTFs measured for $f_m = 16$ Hz (solid lines with unfilled circles), $f_m = 64$ Hz (solid lines with gray squares), and $f_m = 256$ Hz (solid lines with black triangles) are plotted along with the first-order TMTF (dotted lines with stars). In the case of second-order TMTFs, the

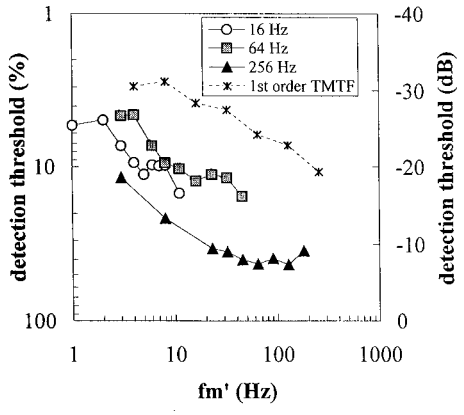

FIG. 3. Mean data for the four listeners. See Fig. 2 for legend details. The left ordinate shows modulation depth at threshold expressed in linear units ($100 \cdot m$), and plotted on a logarithmic scale. The right ordinate shows modulation depth at threshold expressed on a decibel scale ($20 \log m$).

ordinate indicates second-order modulation depth at threshold m', and the abscissa represents f'_m; in the case of first-order TMTFs, the ordinate indicates first-order modulation depth at threshold m, and the abscissa represents f_m. In Figs. 2 and 3, the left ordinate shows detection thresholds expressed in linear units ($100 \cdot m$) and plotted on a logarithmic scale; in Fig. 3 (mean data), the right ordinate shows detection thresholds expressed on a decibel scale ($20 \log m$), frequently used in previous studies. In agreement with previous studies (e.g., Viemeister, 1979; Bacon and Viemeister, 1985), first-order TMTFs display a typical low-pass characteristic (Fig. 3). Sensitivity is reduced by about 3 dB (in $20 \log m$) at $f_m = 64$ Hz; from this frequency, sensitivity decreases at a rate of about 3 dB/oct. For each ''carrier'' modulation frequency f_m, second-order TMTFs display a low-pass segment: For the lowest second-order modulation frequencies f'_m, sensitivity to second-order modulation remains constant up to 2 Hz when $f_m = 16$ Hz, and up to 4 Hz when $f_m = 64$ Hz; from these frequencies, sensitivity decreases with increasing f'_m. When $f_m = 256$ Hz, sensitivity decreases from 3 Hz (the lowest value of f'_m tested for this ''carrier'' modulation frequency). The rate of decrease estimated on the low-pass segment of the second-order TMTFs decreases when the ''carrier'' modulation frequency f_m increases: It is about 6 dB (in $20 \log m'$) per octave when $f_m = 16$ Hz, 4 dB/oct when $f_m = 64$ Hz, and 3 dB/oct when $f_m = 256$ Hz. Overall, the low-pass segment seems to be followed by a plateau: Above a given value of f'_m (of about $f_m/3$ or $f_m/4$), sensitivity remains relatively constant. However, individual second-order TMTFs (e.g., listeners CS and SR) show a noticeable notch (i.e., a loss of sensitivity) when f'_m is slightly below $f_m/2$ (this is especially noticeable when $f_m = 16$ Hz). Figures 2 and 3 also show that sensitivity measured for second-order modulation frequencies between 3–6 Hz increases when f_m increases from 16 to 64 Hz. However, overall sensitivity to second-order modulation degrades considerably for $f_m = 256$ Hz. First-order TMTFs show that modulation detection at 256 Hz is reduced by 9 dB compared to modulation detection at 16 Hz, and by 5 dB compared to modulation detection at 64 Hz. Nevertheless, the ''carrier'' modulation remains highly audible for $f_m = 256$ Hz, the detection threshold being about 10%. This suggests that the audibility of the ''carrier'' modulation is not responsible for the overall drop in sensitivity to second-order modulation when $f_m = 256$ Hz.

To assess the statistical significance of the differences in second-order modulation detection thresholds, a repeated measure analysis of variance (ANOVA) was conducted with factors f_m and f'_m. The analysis showed significant main effects of f_m [$F(2,6) = 262.8$, $p < 0.000\,01$] and f'_m [$F(8,24) = 23.3$, $p < 0.000\,01$], and significant interaction between f_m and f'_m [$F(16,48) = 9.6$, $p < 0.000\,01$]. *Post-hoc* tests (LSD) showed that, when $f_m = 16$ Hz, second-order modulation detection thresholds measured at $f'_m = 2$ Hz were not significantly different from those measured at $f'_m = 1$ Hz ($p = 0.84$), but they were significantly lower than those measured at $f'_m = 5$ Hz ($p < 0.05$); the analysis also showed that thresholds measured for $f'_m = 5$, 6, 7, 8, and 11 Hz were not significantly different from each other ($p > 0.60$). Similar patterns of results were obtained for $f_m = 64$ and 256 Hz, indicating that sensitivity degraded from $f'_m = 4$ Hz when $f_m = 64$ Hz, and from $f'_m = 3$ Hz when $f_m = 256$ Hz. *Post-hoc* tests also showed that second-order modulation detection thresholds measured for $f'_m = 4$ Hz were significantly lower when $f_m = 64$ Hz than when $f_m = 16$ Hz ($p < 0.05$); however, detection thresholds measured for $f'_m = 3$ or 6 Hz were not significantly different when $f_m = 16$ or 64 Hz. Finally, the analysis showed that detection thresholds measured for $f'_m = 3$ or 8 Hz were significantly higher when $f_m = 256$ Hz than when $f_m = 16$ or 64 Hz ($p < 0.05$).

B. Masked modulation detection thresholds

A modulation masking experiment was performed to investigate the extent to which each modulation sideband (of frequency $f_m - f'_m$ and $f_m + f'_m$) contributes to the second-order modulation detection threshold. This experimental paradigm corresponds to a transposition of that used by Kohlrausch *et al.* (2000) to the temporal-envelope domain. The masked modulation detection thresholds of the lower and the upper modulation sideband were measured separately in the presence of the ''carrier'' modulation acting as masker. Here, masker modulation frequency was fixed at 64 Hz. Measurement was performed as a function of the absolute spectral distance Δf (in the modulation spectrum) between the sideband and the carrier ($\Delta f = |f_{\text{mask}} - f_{\text{sig}}|$; Δf was 3, 8, or 32 Hz). Figure 4 shows the individual masked modulation detection thresholds plotted along with second-order modulation detection thresholds measured for a ''carrier'' modulation frequency f_m of 64 Hz. In the case of second-order TMTFs, the left and right ordinates indicate second-order modulation depth at threshold m', and the abscissa represents f'_m; in the case of masked modulation thresholds, the left and right ordinates indicate signal modulation depth at threshold m_{sig}, and the abscissa represents the absolute spectral distance between signal and masker modulation frequencies, Δf. The left ordinate shows detection thresholds expressed in linear units ($100 \cdot m_{\text{sig}}$ or $100 \cdot m'$) and plotted on a logarithmic scale, while the right ordinate

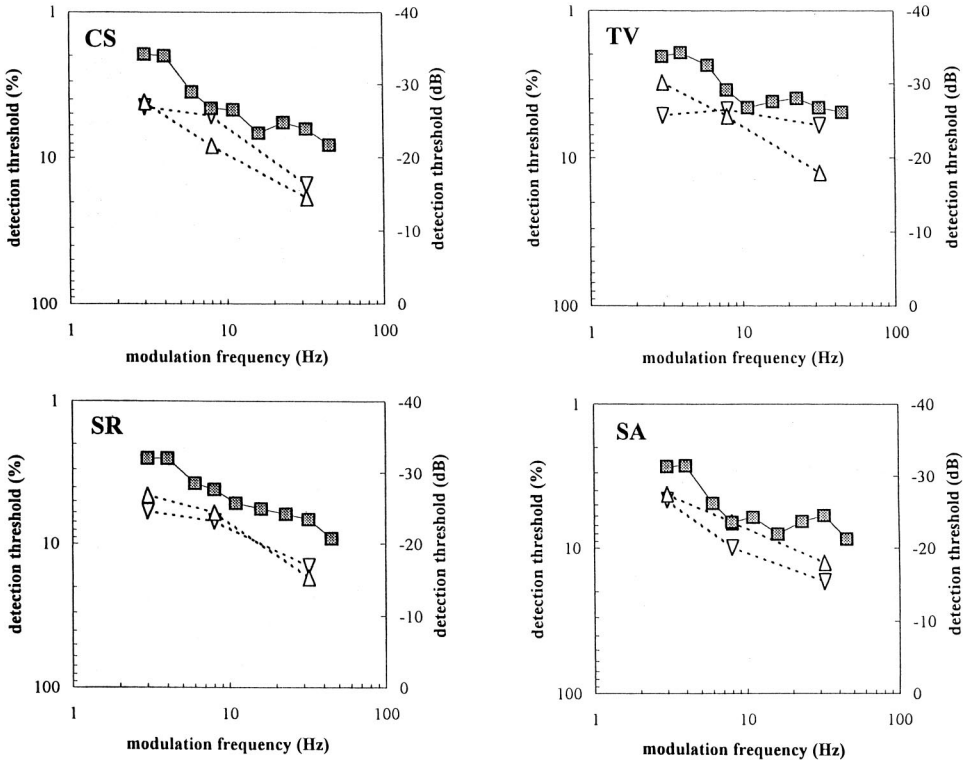

FIG. 4. Individual masked modulation detection thresholds (dotted lines with unfilled triangles) plotted along with second-order modulation detection thresholds measured for $f_m = 64$ Hz (solid lines with gray squares). In the case of second-order modulation thresholds, modulation depth at threshold (m') was transformed into modulation depth per sideband relative to the ''modulation'' carrier depth by subtracting 6 dB from $20 \log m'$. Triangles and inverted triangles represent masked detection thresholds for the upper and lower modulation sideband, respectively.

shows detection thresholds expressed on a decibel scale ($20 \log m_{\text{sig}}$ or $20 \log m'$). Triangles and inverted triangles represent masked detection thresholds for the upper and lower modulation sideband, respectively.

For a fully amplitude-modulated tone, the two sidebands are 6 dB attenuated relative to the carrier level. Therefore, in the case of second-order modulation thresholds, the depth of the modulation sidebands relative to the ''modulation'' carrier depth can be derived by dividing m' by a factor of 2 (or by subtracting 6 dB from $20 \log m'$). In Fig. 4, modulation depth at threshold (m') was therefore transformed into modulation depth per sideband relative to the ''modulation'' carrier depth, which implies a -6-dB vertical shift of the second-order modulation detection data relative to the values shown in Figs. 2 and 3.

For all listeners, masked modulation thresholds increase as the sideband modulation frequency is moved above or below the ''carrier'' modulation frequency f_m. This contrasts with previously reported modulation masking patterns, showing that modulation masking generally decreases as the modulation frequency is moved above or below the masker modulation frequency (e.g., Bacon and Grantham, 1989; Houtgast, 1989; Lorenzi *et al.*, 1997). However, the spectral distances Δf used in the present study are much smaller than what was used in these classical modulation masking studies. This inverse pattern of masking suggests that listeners are able to detect ''beats'' in the temporal envelope, which consists of a cyclic variation in the modulation depth at a slow rate f'_m. Similar effects have been reported by Strickland and Viemeister (1996), Sheft and Yost (1997), and Moore *et al.* (1999).

In order to reach the same modulation depth in a second-order modulation and a single modulation-sideband stimulus, the modulation depth per sideband for the latter has to be 6 dB higher than for the second-order modulation stimulus. The observed threshold difference of nearly 6 dB between these two conditions at small spectral distances can therefore be taken as strong evidence that envelope beat was the detection cue.

IV. GENERAL DISCUSSION

In summary, the data show the following.

(1) For each ''carrier'' modulation frequency, second-order modulation detection thresholds measured with a broadband noise carrier are lowest for (second-order) modulation frequencies below about 2–4 Hz; above 2–4 Hz, they generally increase with second-order modulation frequency up to a given second-order modulation frequency; above this frequency, second-order modulation thresholds remain roughly constant.
(2) Second-order modulation detection thresholds measured for low second-order modulation frequencies decrease when the ''carrier'' modulation increases from 16 to 64 Hz. However, they increase when the ''carrier'' modulation frequency increases to 256 Hz.
(3) Second-order modulation detection thresholds mirror the masked modulation detection thresholds of the lower and the upper modulation sidebands.

In the present experiments, the relative phase of second-order and ''carrier'' modulations was fixed [cf. Eq. (4)] because changes in relative phase affect modulation detectability (*i*) when modulations are harmonically related (e.g., Strickland and Viemeister, 1996; Lorenzi *et al.*, 1999), and (*ii*) when modulation depths are high (i.e., when modula-

tions are highly detectable; Moore and Sek, 2000). This may explain why noticeable notches or rebounds were observed in the second-order TMTFs when f'_m was close to $f_m/2$. (This was especially noticeable when $f_m = 16$ Hz.) Such local changes in detection threshold may therefore be due to the choice of a specific phase relationship between carrier and sidebands modulations. Further investigation of phase effects may help to clarify this issue.

At first sight, the results appear qualitatively (*i*) consistent with theoretical predictions of a linear envelope detector model consisting of a low-pass filtering stage and a decision stage, and (*ii*) inconsistent with predictions of a modulation filterbank model using narrowly tuned filters. Nevertheless, the current results seem consistent with predictions of a modulation filterbank model using *broadly* tuned filters. The latter point was studied by implementing a simplified version of the modulation filterbank model described by Dau *et al.* (1999) and Ewert and Dau (2000) consisting of three successive processing stages: (1) a half-wave rectifier, (2) a low-pass filter (first-order Butterworth) with a 3-dB cutoff frequency of 500 Hz, and (3) an array of overlapping linear bandpass filters (Dau *et al.*, 1996, 1997a, b) whose center frequencies range between 2 and 512 Hz. Center frequencies are spaced on a logarithmic scale from 2 to 512 Hz, and filters' density is set to 5 filters/oct. Filter bandwidths increase logarithmically over the whole range of center frequencies with constant Q values of 1, 2, or 4. The *change* produced by second-order modulation in the output of the modulation filter tuned to the lower sideband modulation frequency was calculated by subtracting the excitation evoked by the target to that evoked by the standard in this filter (i.e., the filter tuned to $f_m - f'_m$). Figure 5 shows the change in the output of this modulation filter as a function of f'_m. Simulations were performed for $f_m = 16$ Hz (left panel), 64 Hz (middle panel), and 256 Hz (right panel). Each panel shows the results of simulations for constant Q values of 1 (circles), 2 (squares), and 4 (triangles). Modulation depths of the ''carrier'' and second-order modulation (i.e., m and m', respectively) were both fixed at 50%. Figure 5 clearly shows that a modulation filter with a Q value of 1 would hardly signal the presence of second-order modulation whatever the value of f'_m (output change <1 dB); on the other hand, modulation filters with $Q \geqslant 2$ would easily signal the presence of second-order modulation (output change $\in [1-15$ dB$]$), but the pattern of change at the output of such filters predicts that detectability should *increase* with second-order modulation frequency. The present results clearly argue against a model based on finely tuned modulation bandpass filters [e.g., for $Q \geqslant 2$; Dau *et al.* (1997a, b, 1999)]. This suggests that, if such modulation filters do exist, their selectivity is poor. In the latter case, the Q value of modulation filters would have to be less than 2. This is more in line with the implementation of the modulation filterbank model proposed by Ewert and Dau (2000), in which the Q value of such filters is assumed to be equal to 1 for center modulation frequencies up to 64 Hz.

The present results also show that sensitivity to second-order modulation degrades considerably when the ''carrier'' modulation frequency increases from 64 to 256 Hz. First-order TMTFs indicate that the ''carrier'' modulation remains highly audible when $f_m = 256$ Hz, detection threshold being about 10%. This suggests that the audibility of the ''carrier'' modulation is not responsible for this overall loss in sensitivity to second-order modulation. Dau (1996) has shown that the deleterious effects of the intrinsic fluctuations of the noise carrier are greater at high modulation frequencies. Such statistical fluctuations might have affected the modulation sidebands generated by the second-order modulation, these deleterious effects being certainly strongest for the upper modulation sideband ($f_m + f'_m$). However, the way modulation sidebands might have been disrupted remains currently unexplained.

The experiment on modulation masking reveals that second-order modulation detection thresholds mirror the masked modulation detection thresholds of the lower and the upper modulation sidebands. Therefore, in contrast to the ''classical'' modulation masking patterns (e.g., Bacon and Grantham, 1989; Houtgast, 1989; Lorenzi *et al.*, 1997), the ability to detect modulation sidebands improves when sidebands and carrier modulations get closer. As for spectral masking in the audio-frequency domain, this result suggests (without demonstrating it) that the task is performed by listening to low envelope beat cues produced by the interaction

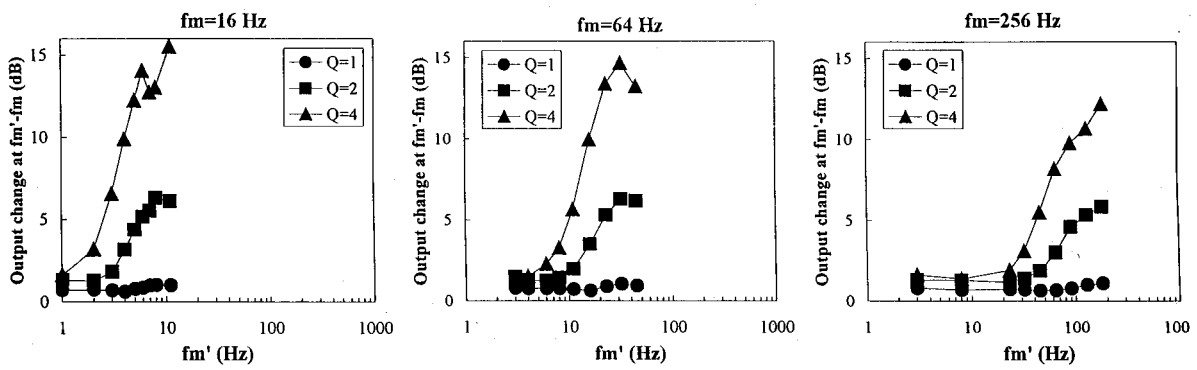

FIG. 5. Change produced by second-order modulation in the output of the modulation filter tuned to the lower sideband modulation frequency as a function of f'_m. Simulations were performed for $f_m = 16$ Hz (left panel), 64 Hz (middle panel), and 256 Hz (right panel). Each panel shows the results of simulations for constant Q values of 1 (circles), 2 (squares), and 4 (triangles). In each case, $m = m' = 0.5$.

between carrier and sideband modulation components. This is consistent with the subjective experience of subjects when listening to second-order modulation: A slow, cyclic variation in amplitude at f_m' is heard. At first sight, it appears difficult to account for this effect in terms of modulation filters, since there is no spectral energy at the envelope beat frequency in the modulation spectrum of the stimuli (see Fig. 1). However, applying a compressive nonlinearity before envelope extraction yields such a component in the modulation spectrum of stimuli with second-order SAM. This compressive nonlinearity may correspond to the fast acting compression performed by active mechanisms within the cochlea (e.g., Ruggero *et al.*, 1997; Moore and Oxenham, 1998) or to the transduction from basilar membrane vibration to neural activity (e.g., Yates, 1990; Regan and Regan, 1993). The idea that nonlinearities within the auditory system introduce distortion in the internal representation of the temporal envelope of sounds is also supported by previous electrophysiological and psychoacoustical studies using two-component modulators (Shofner *et al.*, 1996; Sheft and Yost, 1997; Moore *et al.*, 1999). For instance, the decrease in sensitivity to second-order modulation above 2–4 Hz is consistent with the results of Sheft and Yost (1997), who showed that the ''interference'' effect (i.e., the deleterious effect) of a beating two-component modulator on the detection of a probe modulator at the beat rate was greater for a 4-Hz beat rate than for a 10-Hz rate. The origin of this effect remains unclear, and further work is now needed to assess whether or not a modulation filterbank model including such nonlinearities may account for it. If detection of second-order modulation is actually based on the detection of spectral energy at the envelope beat frequency, second-order TMTFs may then be viewed as quantitative descriptions of the attenuation characteristics of low distortion components in the amplitude-modulation domain.

While sensitivity to first-order modulation is related to some aspects of *tempo* perception (for $f_m < \approx 30$ Hz), sensitivity to second-order modulation appears to be related to some aspects of *rhythm* perception, in that the sinusoidal increase and decrease in the modulation depth of SAM produce complex temporal patterns in which series of cycles of modulated noise alternate with segments of unmodulated noise at a low rate f_m'. Indeed, listeners reported performing the second-order modulation detection task by determining the interval evoking a rhythm percept instead of a regular tempo percept. Second-order TMTFs may therefore provide an interesting framework to study how perception transforms gradually from tempo to rhythm. In the same spirit, second-order TMTFs may also be of relevance to speech perception, in that our ability to detect patterns of periodicity (such as patterns of voicing) appears to be a major cue to consonant identity (e.g., Faulkner and Rosen, 1999).

ACKNOWLEDGMENTS

We thank Daniel Pressnitzer for helping us in running the modulation filterbank model and Laurent Demany, Gary Green, Tim Griffiths, Adrian Rees, Rebecca Millman, Michael Simpson, and William Woods for helpful and stimulating discussions. We also thank Sid P. Bacon, Torsten Dau, and an anonymous reviewer for helpful comments on an earlier version of this manuscript. This work was funded from the Cognitique program (MENRT).

Bacon, S. P., and Grantham, D. W. (**1989**). ''Modulation masking patterns: Effects of modulation frequency, depth and phase,'' J. Acoust. Soc. Am. **85**, 2575–2580.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects,'' Audiology **24**, 117–134.

Dau, T. (**1996**). ''Modeling auditory processing of amplitude modulation,'' Ph.D. dissertation, Oldenburg.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). ''Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrow-band 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 in modulation detection,'' J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). ''Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band carriers,'' J. Acoust. Soc. Am. **106**, 2752–2760.

Ewert, S. D., and Dau, T. (**2000**). ''Characterizing frequency selectivity for envelope fluctuations,'' J. Acoust. Soc. Am. **108**, 1181–1196.

Fassel, R. (**1994**). ''Experimente und simulationsrechnungen zur wahrnehmung von amplitudenmodulationen im menschlichen gehör,'' Ph.D. dissertation, University of Göttingen.

Fassel, R., and Püschel, D. (**1993**). ''Modulation detection and masking using deterministic and random maskers,'' in *Contributions to Psychological Acoustics* (BIS, Oldenburg, Germany), pp. 419–429.

Faulkner, A., and Rosen, S. (**1999**). ''Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception,'' J. Acoust. Soc. Am. **106**, 2063–2073.

Fletcher, H. F. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.

Forrest, T. G., and Green, D. M. (**1987**). ''Detection of partially filled gaps in noise and the temporal modulation transfer function,'' J. Acoust. Soc. Am. **82**, 1933–1943.

Green, G. G. R. (**1976**). ''Temporal aspects of audition,'' Ph.D. dissertation, Oxford.

Green, G. G. R., and Kay, R. H. (**1974**). ''Channels in the human auditory system concerned with the waveform of the modulation present in amplitude- and frequency-modulated tones,'' J. Physiol. (London) **241**, 29–30P.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

Kay, R. H., and Matthews, D. R. (**1972**). ''On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones,'' J. Physiol. (London) **225**, 657–677.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' J. Acoust. Soc. Am. **108**, 723–734.

Lorenzi, C., Berthommier, F., and Demany, L. (**1999**). ''Discrimination of amplitude-modulation phase spectrum,'' J. Acoust. Soc. Am. **105**, 2987–2990.

Lorenzi, C., Micheyl, C., and Berthommier, F. (**1995**). ''Neuronal correlates of perceptual amplitude-modulation detection,'' Hear. Res. **90**, 219–227.

Lorenzi, C., Micheyl, C., Berthommier, F., and Portalier, S. (**1997**). ''Modulation masking in listeners with sensorineural hearing loss,'' J. Speech Lang. Hear. Res. **40**, 200–207.

Moore, B. C. J., and Oxenham, A. J. (**1998**). ''Psychoacoustic consequences of compression in the peripheral auditory system,'' Psychol. Rev. **105**, 108–124.

Moore, B. C. J., and Sek, A. (**2000**). ''Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation,'' J. Acoust. Soc. Am. **108**, 2337–2344.

Moore, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). ''Modulation masking produced by beating modulators,'' J. Acoust. Soc. Am. **106**, 908–918.

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

Rees, A., and Kay, R. H. (**1985**). ''Delineation of FM rate channels in man

by detectability of a three-component modulation waveform,'' Hear. Res. **18**, 211–221.
Regan, M. P., and Regan, D. (**1993**). ''Non linear terms produced by passing amplitude-modulated sinusoids through a hair cell transducer function,'' Biol. Cybern. **69**, 439–446.
Rodenburg, M. (**1977**). ''Investigation of temporal effects with amplitude modulated signals,'' in *Psychophysics and Physiology of Hearing* (Academic, London), pp. 429–437.
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.
Sheft, S., and Yost, W. A. (**1997**). ''Modulation detection interference with two-component masker modulators,'' J. Acoust. Soc. Am. **102**, 1106–1112.
Shofner, W. P., Sheft, S., and Guzman, S. J. (**1996**). ''Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes,'' J. Acoust. Soc. Am. **99**, 3592–3605.
Strickland, E. A., and Viemeister, N. F. (**1996**). ''Cues for discrimination of envelopes,'' J. Acoust. Soc. Am. **99**, 3638–3646.
Tansley, B. W., and Suffield, J. B. (**1983**). ''Time-course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation,'' J. Acoust. Soc. Am. **74**, 765–775.
Veimeister, N. F. (**1977**). ''Temporal factors in audition: A system analysis approach,'' in *Psychophysics and Physiology of Hearing* (Academic, London), pp. 419–428.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Yates, G. K. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **50**, 145–162.
Yost, W. A., Sheft, S., and Opie, J. (**1989**). ''Modulation interference in detection and discrimination of amplitude-modulation,'' J. Acoust. Soc. Am. **86**, 2138–2147.
Zwicker, E. (**1952**). ''Die grenzen der hörbarkeit der amplitudenmodulation und der frequenzmodulation eines tones,'' Acustica **2**, 125–133.