

# Modeling interaural-delay sensitivity to frequency modulation at high frequencies

Kourosh Saberi<sup>a)</sup>

Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Massachusetts 02139

(Received 2 December 1996; revised 8 November 1997; accepted 20 January 1998)

Interaural-delay sensitivity to high-frequency ( $\geq 3$  kHz) sinusoidal-frequency-modulated (SFM) tones is examined for rates from 25 to 800 Hz and depths of  $-12$  to  $18$  dB. Comparison is made to thresholds obtained for sinusoidal-amplitude-modulated (SAM) tones for the same observers and modulation rates. Both SAM and SFM threshold-by-rate functions are U-shaped with optimum sensitivity to SFM tones occurring at higher rates ( $f_m = 200$ – $400$  Hz) compared to those for SAM tones ( $f_m = 100$ – $200$  Hz). Effects of modulation depth were examined for rates from 50 to 300 Hz. In all cases thresholds improved considerably with increasing modulation depth. It is also shown that a hybrid dichotic signal composed of an SFM tone presented to one ear and an SAM tone to the other, can perceptually fuse and be lateralized, with the contingency that both stimuli have equal modulation rates but not necessarily equal carrier frequencies. Using bandpass noise to restrict off-frequency listening, it was shown that for this stimulus, observers can use information from filters either below or above the carrier frequency. Consistent with FM-to-AM conversion from cochlear bandpass filtering, several important differences between the SAM- and SFM-tone data can be predicted from a nonstationary stochastic model of binaural interaction whose parameters are uniquely determined from the SAM-tone data. © 1998 Acoustical Society of America. [S0001-4966(98)03705-9]

PACS numbers: 43.66.Ba, 43.66.Pn, 43.66.Nm, 43.66.Lj [RHD]

## INTRODUCTION

Sensitivity to interaural delays at high frequencies is documented extensively for amplitude-modulated (AM) sounds. The early studies of AM processing investigated the dependence of sensitivity on a wide range of simple stimulus parameters such as envelope rate, carrier frequency, and modulation type (Leakey *et al.*, 1958; David *et al.*, 1959; Harris, 1960; McFadden and Pasanen, 1974, 1975, 1976; Henning, 1974a, 1980, 1981; Hafter and DeMaio, 1975; Nuetzel and Hafter, 1976, 1981; Yost *et al.*, 1971). Later investigations extended to more complex binaural phenomena such as interference (McFadden and Pasanen, 1976; Zurek, 1985; Trahiotis and Bernstein, 1990; Buell and Trahiotis, 1993; Bernstein and Trahiotis, 1995), summation of interaural information (McFadden and Pasanen, 1976; Hafter *et al.*, 1990; Buell and Trahiotis, 1993; Saberi, 1995), testing models of cross correlation (Colburn and Esquissaud, 1976; Stern *et al.*, 1988a), masking (Henning, 1974b; Bernstein and Trahiotis, 1992; Kohlrausch *et al.*, 1995), motion discrimination (Saberi and Hafter, 1996), adaptation and precedence (Hafter *et al.*, 1988; Blauert and Divenyi, 1988; Saberi, 1996), all at high frequencies.

In contrast, there are only a handful of papers on binaural sensitivity to frequency-modulated (FM) sounds (Henning, 1980; Blauert, 1981; Hukin and Darwin, 1994). The types of FM stimuli to which binaural sensitivity has been shown include phase-jittered tones (Nordmark, 1976; Blauert, 1981), square-wave modulation (Green and Kay,

1975; Green *et al.*, 1976; Kay, 1982), noise-modulated phase (Bielek, 1975), and sinusoidal-frequency-modulated (SFM) tones (Henning, 1980). In spite of their completely flat envelopes, lateralization of high-frequency FM sounds is possible because of an FM-to-AM conversion by cochlear bandpass filtering (Zwicker, 1958; Henning, 1980; Blauert, 1981). The rate, depth, envelope shape, and phase of this AM-like response vary across filters and primarily depend on: (1) the frequency separation between the filter's resonant frequency and the stimulus carrier; (2) stimulus modulation rate, phase and depth; and (3) the impulse response of the filter, including its phase spectrum. The manner in which these factors interact is complex and requires careful and in-depth consideration.

With the exception of Henning (1980), all previous binaural studies on FM at high frequencies have used nonsinusoidal modulation. These stimuli have discontinuities in the slope of the carrier waveform and, therefore, transient byproducts. Sinusoidal modulation however has smoothly varying temporal properties with a power spectrum that is more confined and symmetric about the carrier frequency, and is therefore a useful stimulus for isolating the effects of frequency modulation. The one study that has examined binaural sensitivity to SFM tones at high frequencies was conducted by Henning (1980) who used only a single modulation rate (300 Hz) and carrier frequency (3.9 kHz). Henning reported that thresholds of  $80$ – $200$   $\mu$ s for that stimulus were comparable to those observed for quasi-frequency-modulation (QFM) and slightly higher than those for SAM tones with the same rate.

The goal of the present study is to: (1) investigate binaural processing of SFM tones at high frequencies for a wide

<sup>a)</sup>Current address: 216-76 Division of Biology, Caltech, Pasadena, CA 91125, Electronic mail: kourosh@caltech.edu

range of stimulus parameters and in comparison to SAM tones, and (2) evaluate for SFM tones the predictions of a model of binaural interaction whose parameters are uniquely determined from the SAM-tone data. The experiments begin with a measurement of interaural time difference ( $\Delta$ ITD) thresholds for SFM tones at several carrier frequencies and modulation rates. Thresholds are then compared at one carrier frequency to those obtained with SAM tones using the same observers. Data and observations are also reported on a variety of other psychophysical tasks that examine the effects of modulation versus carrier delay, and modulation depth. Finally, binaural interaction between an SFM tone presented to one ear, and an SAM tone to the other ear is examined. We begin with a description of the spectral characteristics of SFM tones.

## I. FOURIER SPECTRUM OF SFM TONES

An SFM tone  $X(t) = A_c \cos[2\pi f_c t + \beta \cos(2\pi f_m t)]$  has an instantaneous frequency  $f_c + \Delta f \sin(2\pi f_m t)$  and a Fourier expansion (Inglis, 1988; Carlson, 1986) expressed as the harmonic series

$$X(t) = A_c J_0(\beta) \cos 2\pi f_c t - A_c \sum_{n=0}^{\infty} J_{2n+1}(\beta) \times \{ \cos 2\pi [f_c - (2n+1)f_m]t - \cos 2\pi [f_c + (2n+1)f_m]t \} + A_c \sum_{n=0}^{\infty} J_{2n}(\beta) \{ \cos 2\pi [f_c - 2nf_m]t + \cos 2\pi [f_c + 2nf_m]t \}, \quad (1)$$

where  $\beta = \Delta f / f_m$ , the frequency-modulation index, is the ratio of peak-frequency-deviation to modulation frequency, and

$$J_n(\beta) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(\beta \sin \lambda - n\lambda)} d\lambda \quad (2)$$

are Bessel functions of the first kind of order  $n$  (sideband number) and argument  $\beta$  which define the amplitude coefficients of the line spectra (Fig. 1). Equation (1) shows that the line spectra of SFM tones consist of a carrier frequency  $f_c$  and sidebands at  $f_c \pm nf_m$  with the *odd-order* lower sidebands inverted in phase relative to the unmodulated carrier.

As  $\beta$  increases, so does the bandwidth of the SFM tone. Usually, the effective bandwidth of this stimulus is approximated from Carson's rule which defines the 98% energy bandwidth as  $2(\beta+1)f_m$  (Cooper and McGillem, 1986; Carlson, 1986). While the bandwidth of the SFM tone increases with  $\beta$ , the amplitude  $A_c$  of the time waveform and therefore the stimulus power are constant and independent of  $\beta$ . Consequently, changing  $\beta$  must change the stimulus energy at the carrier frequency. For some values ( $\beta=2.4, 5.3$ ) there is no energy at the carrier. This property is very different from that of SAM tones. The carrier of an SFM tone contains part of the modulation information while the carrier of an SAM tone contains no information about the modulation waveform (Couch, 1987; Carlson, 1986).

The amplitude of the sidebands of an SFM tone [Eq. (2)] cannot be expressed in nonintegral form, however, they have

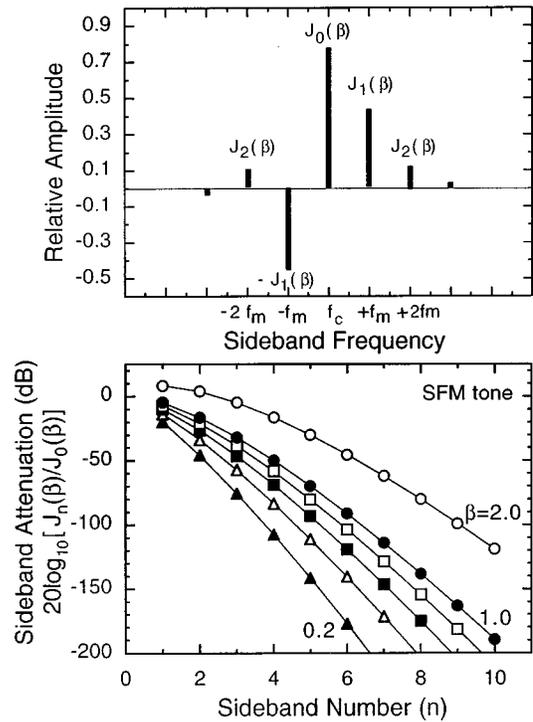


FIG. 1. Top panel shows the amplitude spectrum of an SFM tone with carrier  $f_c$ , modulation rate  $f_m$ , and a modulation index  $\beta$  of unity. The amplitudes are Bessel functions of the first kind of order  $n$  (harmonic number) and argument  $\beta$ . The lower odd harmonics are phase reversed relative to the remaining components [Eq. (1)]. The lower panel shows sideband attenuation in dB *re*: the amplitude of the carrier for  $\beta=0.2, 0.4, 0.6, 0.8, 1.0$ , and  $2.0$ .

been numerically evaluated and tabulated for many values of  $\beta$  (Abramowitz and Stegun, 1972). The lower panel of Fig. 1 shows calculations of sideband attenuation relative to carrier amplitude in decibels for various values of  $\beta=0.2, 0.4, 0.6, 0.8, 1.0, 2.0$ . Note that these functions are dependent on only one stimulus parameter,  $\beta$ . The amplitude of the sidebands drop rapidly when the sideband number exceeds the value of  $\beta$  for  $\beta > 1$ . For  $\beta \leq 1$ , the stimulus bandwidth from Carson's rule is confined to the frequency region bordered by the second sidebands. For most of the carrier frequencies and modulation rates used in the following experiments, the SFM-tone bandwidth is confined to frequencies considerably above 1.5 kHz. Nonetheless, lowpass noise and highpass filtering are used to eliminate possible low-frequency cues or intermodulation distortions.

## II. MODEL STRUCTURE

In this section a model of binaural interaction is described that is later used to examine the data obtained for both SAM and SFM stimuli. The important feature of the model is that its parameters are determined from the SAM data and predictions are subsequently made for SFM tones. Thus the extent to which the model captures performance reflects the extent to which amplitude envelope cues are used from an FM-to-AM conversion by bandpass filtering. A block diagram of the model structure is shown in Fig. 2. The stimuli are first filtered through a Gammatone filterbank (Holdsworth *et al.*, 1988). The filterbank consists of 30 filters

distributed logarithmically from 2 to 5 kHz. The Hilbert envelopes are then extracted and used to drive a nonstationary fractal point process.<sup>1</sup> The point process, which models the serial correlation of spike patterns from auditory afferents (Teich, 1989; Teich and Lowen, 1994; Lowen and Teich, 1995, 1996) is then subjected to a refractoriness model (Westerman and Smith, 1985; Carney, 1993) with a random dead-time distributed uniformly from  $R_A=0.25-0.75$  ms, and a recovery function described by the sum of two exponentials

$$Q(t) = \begin{cases} Q_{\max} \{ c_0 \exp[-(t-t_1-R_A)/s_0] \\ + c_1 \exp[-(t-t_1-R_A)/s_1] \}, & \text{for } (t-t_1) \geq R_A; \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where  $t-t_1$  is the time since the last discharge, and  $Q_{\max}$  is the maximum increase in threshold for the recovery function (normalized to unity in the current implementation),  $c_0=0.55$  and  $c_1=0.45$  are coefficients for the discharge history effect, and  $s_0=0.8$  and  $s_1=25$  ms are time constants. This first stage of the model is used to generate post-stimulus time histograms (PSTH) from 1000 sample waveform presentations; the PSTHs are then low-pass filtered to reconstruct the envelopes of the filtered waveforms. The primary effect of the point process and refractoriness was to enhance the waveform peaks, similar to an expansive nonlinearity used in psychophysical models (Colburn, 1973; Stern *et al.*, 1988a; Shear, 1987) and also reported for physiological recording in the auditory nerve in response to SAM tones (Joris and Yin, 1992). The resultant output of the point process is low-pass filtered at 150 Hz; the slope of the FIR low-pass filter is chosen to fit the SAM-tone data for modulation rates above 150 Hz.

The second stage of the model uses the reconstructed envelopes as inputs to an interaural cross-correlation function

$$\chi(t, \tau, f) = p(\tau) \int_{f_a}^{f_b} \int_{-\infty}^t X_l(t, f) X_r(t-\tau, f) \times \{ 1 + e^{-(t+\mu)/k} \}^{-1} dt df. \quad (4)$$

The running cross-correlation has a logistic decay with mean  $\mu=20$  ms and slope parameter  $k=0.002$ , and a centrality-weighting function  $p(\tau)$  that heavily weights  $\tau$  values near zero and has the form of a sixth-order Butterworth bandpass filter that is maximally flat in the passband and monotonic overall with rolloffs at  $\pm 0.6$  ms.<sup>2</sup> The cross-correlation function of Eq. (4) is discretized at a sampling rate of 20 kHz and a deviate  $\xi_{\tau, f} \sim \text{normal}(0, \sigma_\xi^2)$  is added, before frequency integration, to each delay-by-frequency sample of the  $\tau-f$  plane to model the stochastic nature of the process. The lag ( $\tau$ ) associated with the peak of the cross-correlation output,  $\chi_{\max}$ , is then used to obtain a lateral position estimate, from which the index of detectability is calculated

$$d' = E[\tau(\chi_{\max}) + \epsilon_p] / (\text{var}[\tau(\chi_{\max})] + \text{var}[\epsilon_p])^{0.5}, \quad (5)$$

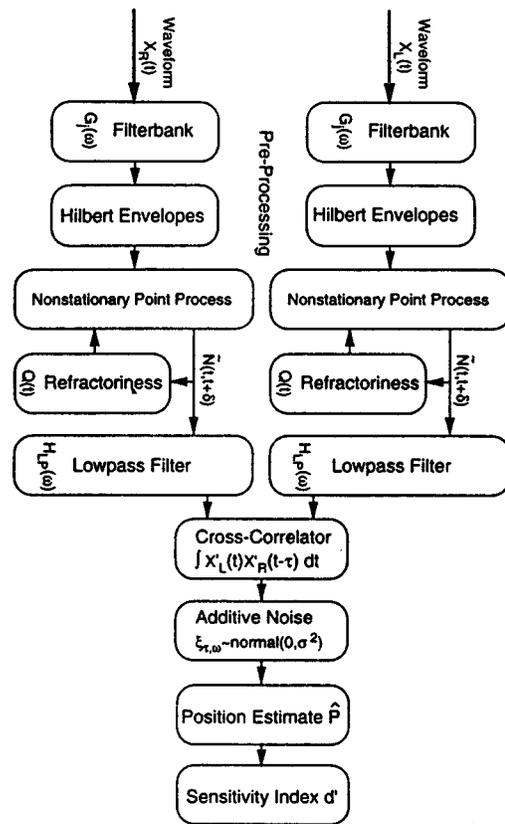


FIG. 2. Schematic diagram of the model structure.

where  $\epsilon_p \sim \text{normal}(0, \sigma_p^2)$  is a position noise added to the estimate before calculation of  $d'$  to prevent the model from perfect discrimination in those cases where the waveform envelopes have a very high peak factor. Although it is, in principle, useful to maintain this parameter ( $\epsilon$ ), in practice the position noise may be chosen to be substantially smaller than  $\xi$  when using high-frequency stimuli, and therefore  $\epsilon$  may be ignored or set to zero variance. In the model implementation the expected value of  $d'$  was estimated from computer simulations of 10 000 trials per each value of interaural delay. Between 2 and 4 values of interaural delay were used to obtain 2–4 values of  $d'$  that bracketed 0.78. A least-squares linear fit to  $\log d'$  vs  $\log$  ITD was used to obtain the interaural-delay threshold corresponding to 71% correct performance in a 2IFC task ( $d'=0.78$ ).

There are several reasons for choosing a stochastic approach. First, this approach predicts a reduction of interaural sensitivity at low modulation rates as previously reported for SAM-tone data (Nuetzel and Hafter, 1981). As the rate is reduced, the maximum slope of the envelope diminishes. As the slope of the envelope decreases, the addition of a constant-variance internal noise  $\xi$  increases the variance of estimating the peak of the cross-correlation pattern. As the modulation rate is increased, the slope of the envelope increases and the variance of the position estimate is reduced, thus a higher  $d'$ . As the rate is still further increased, the sidebands of the complex are increasingly filtered (due to increased component spacing) resulting in a reduced depth of modulation, increased variance of the position estimate, and thereby smaller  $d'$ . We next describe experiments that mea-

sure interaural-delay sensitivity to a variety of SFM stimulus conditions and compare performance with that from SAM tones. Model predictions are evaluated for several of these experimental conditions.

### III. EXPERIMENT I

#### A. Lateralization thresholds for SFM tones as a function of modulation rate and carrier frequency

##### 1. Procedure

In each interval of a two-interval, forced-choice (2IFC) task, a 400-ms dichotic SFM tone with a 20-ms cosine-squared rise-decay envelope was presented. The two intervals of the trial were separated by 300 ms. In one interval, the waveform ITD led to the left ear and in the other interval it led to the right ear (except for the gating envelopes which were not interaurally delayed). Waveform delay was achieved by appropriately phase shifting the carrier and modulation waveforms by  $\phi_c = 2\pi f_c \text{ITD}$  and  $\phi_m = 2\pi f_m \text{ITD}$ , respectively. The subject's task was to identify the order of presentation of the tones (i.e., left-leading then right, or, right-leading then left). Each run consisted of 60 trials with visual feedback after each trial. A two-down, one-up procedure (Wetherill and Levitt, 1965; Levitt, 1971) was used to adaptively track the 0.707 probability of a correct response on the observer's psychometric function. The step change in interaural delay was  $0.2 \log \mu\text{s}$  up to the fourth reversal and  $0.05 \log \mu\text{s}$  for the remaining trials. The starting  $\Delta\text{ITD}$  value was  $1300 \mu\text{s}$  ( $650 \mu\text{s}$  for each interval of the 2IFC). The first four reversals were discarded and the values of interaural delays at the remaining reversals were averaged to obtain one estimate of threshold.

Each subject completed a minimum of six runs per condition. The carrier frequencies were either 3, 4, or 8 kHz and the modulation rates were 100, 200, 250, 300, 350, 400, 450, 500, 600, and 700 Hz. The modulation index  $\beta$  was equal to unity. Each condition was defined by one value of modulation rate and one value of carrier frequency selected randomly between runs, but fixed within a run. All signals were computed before each trial of the adaptive procedure using an IBM PC and an array processor (Tucker-Davis Technologies TDT AP2). Signals were presented through 16-bit digital-to-analog converters (TDT DA2) at a sampling rate of 40 kHz and were low-pass filtered at 20 kHz. The sound pressure level for a continuous SFM tone was 60 dB SPL. Signals were highpass filtered at 1.5 kHz. In addition, continuous Gaussian noise, low-pass filtered at 1.5 kHz, was presented at a spectrum level of 32 dB to mask low-frequency intermodulation distortions (Henning, 1974a; Henning, 1980; Nuetzel and Hafter, 1981). Delays between left and right channels were checked for accuracy with a Philips dual-channel storage oscilloscope (model PM 3335). Signals were then led to a single-walled sound booth and presented to subjects through Sennheiser (HD-450) headphones.<sup>3</sup> All three subjects (EM, AN, AS) were experienced in lateralization experiments, and each was practiced on various conditions of the experiment until performance seemed to be stable. Two subjects were female and one male (EM) and their ages ranged from 18 to 22. The ear canals of

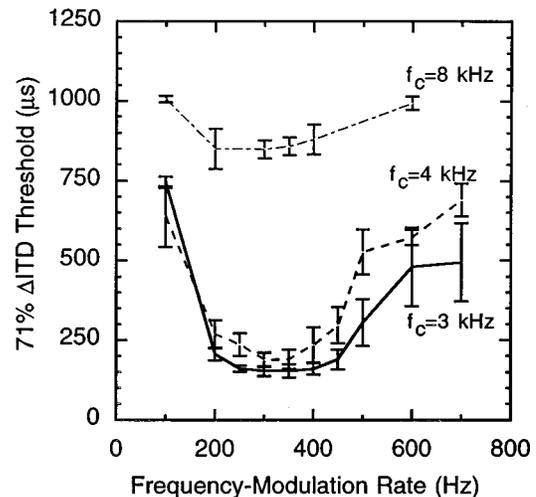


FIG. 3. Interaural-delay difference thresholds for SFM tones as a function of modulation rate. The parameter is carrier frequency. Note that each interval of the 2IFC task carried an interaural delay of  $\Delta\text{ITD}/2$ . Data are averaged for three observers and the error bars are one standard error of the mean.

all subjects were examined and cleaned at the University Infirmary. All subjects had normal hearing within 10 dB of ISO standards for frequencies between 125 and 8000 Hz as determined from a Bekesy audiometric test.

##### 2. Results

Results are shown in Fig. 3. Each line represents the averaged data for one carrier frequency. For the 8-kHz carrier, thresholds are elevated at all rates. At this carrier frequency, the signals are clearly audible and the decline in performance is not due to stimulus detectability. For the 3- and 4-kHz carriers, lowest thresholds are obtained at modulation rates of 200–400 Hz, with the 3-kHz carrier producing slightly better thresholds than the 4-kHz carrier. Best thresholds are on the order of 100–200  $\mu\text{s}$ . One observer (AN) produced low thresholds even at a rate of 700 Hz (thus the large variability). For this one subject, performance degraded to chance as the modulation rate was increased to 1 kHz (not shown). The only previously available related data is from Henning (1980) who used a 300-Hz modulation of a 3.9-kHz carrier. The threshold of 200  $\mu\text{s}$  reported by Henning is similar to the current results. Statistical analysis showed a significant effect of modulation rate for 3- and 4-kHz carriers ( $F_{9,20} = 2.66$ ,  $p < 0.05$ ; and  $F_{9,20} = 3.48$ ,  $p < 0.05$ ) but not for the 8-kHz carrier ( $F_{5,12} = 0.98$ , n.s.). Carrier-frequency effects were significantly different between the 3- and 4-kHz conditions [ $t(29) = 2.36$ ,  $p < 0.05$ ].

##### B. Effects of carrier versus modulation delay

Studies on lateralization of SAM tones at high frequencies have shown that observers are only sensitive to the interaural delay in the envelope of the stimulus and not the carrier (Henning, 1974a, 1980; Henning and Ashton, 1981; Nuetzel and Hafter, 1976, 1981). If the SFM stimulus is converted into an AM-like response at the auditory periphery, then one may expect a similar dominance of modulation compared to carrier delay. It should be noted however that

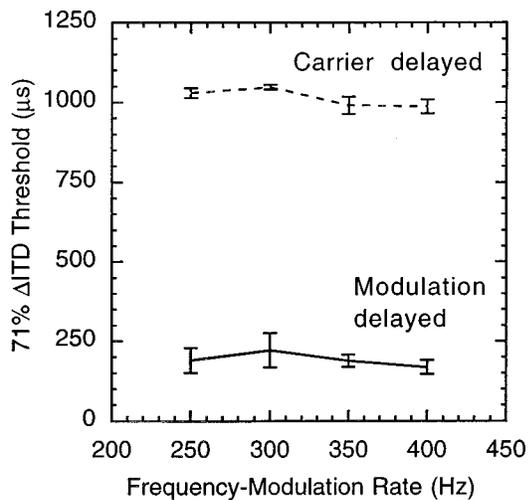


FIG. 4. The dashed line shows interaural-delay difference thresholds measured when the modulation waveforms of an SFM tone remained in phase at the two ears, and the 3-kHz carrier was interaurally phase shifted by  $2\pi f_c \text{ITD}$ . The solid line shows thresholds for the SFM tone when the modulation waveform was interaurally delayed by  $2\pi f_m \text{ITD}$  and the carrier phase was zero at the two ears. The data are averaged for three observers and the error bars are one standard error.

while phase shifting the carrier of an SAM tone produces a constant delay of the carrier, this is not true for SFM tones. A constant carrier phase shift of the SFM tone introduces a dynamic delay in the fine structure. For example, an interaural phase inversion would produce a smaller instantaneous interaural delay in the fine structure when the instantaneous frequency of the waveform is equal to the carrier frequency, compared to when it is at its minimum value  $f_c - \Delta f$ . While our results below show that these dynamic carrier effects are not important at high frequencies, they should be accounted for when low-frequency carriers are used, especially for large  $\beta$ . All subjects, procedures and stimuli for this experiment were the same as part A with the exception that for the carrier-delayed case the phase of the carrier to one ear was additionally delayed by  $\phi_c = 2\pi f_c \text{ITD}$  while  $\phi_m = 0$  [see Eq. (6b)]. In the modulation-delayed case the modulation waveform was delayed by  $\phi_m = 2\pi f_m \text{ITD}$  while  $\phi_c = 0$ . Only one carrier frequency, 3 kHz, and four modulation rates of 250, 300, 350, and 400 Hz for which low thresholds had previously been observed were used ( $\beta = 1$ ). Averaged results for the three observers are shown in Fig. 4. The dashed line represents the carrier-delayed condition and the solid line shows the modulation-delayed condition. Error bars are one standard error of the mean. There is no significant carrier-phase effect for the range of parameters examined. Thresholds for the modulation-delayed case (solid line) are comparable to the whole-waveform delayed thresholds shown in Fig. 3.

## IV. EXPERIMENT II

### A. Comparison of SFM to SAM thresholds

To what extent is the AM induced by bandpass filtering of an SFM tone comparable to that produced by an SAM tone, and to what extent can performance for the SFM stimulus be predicted from data on SAM stimuli? In this section,

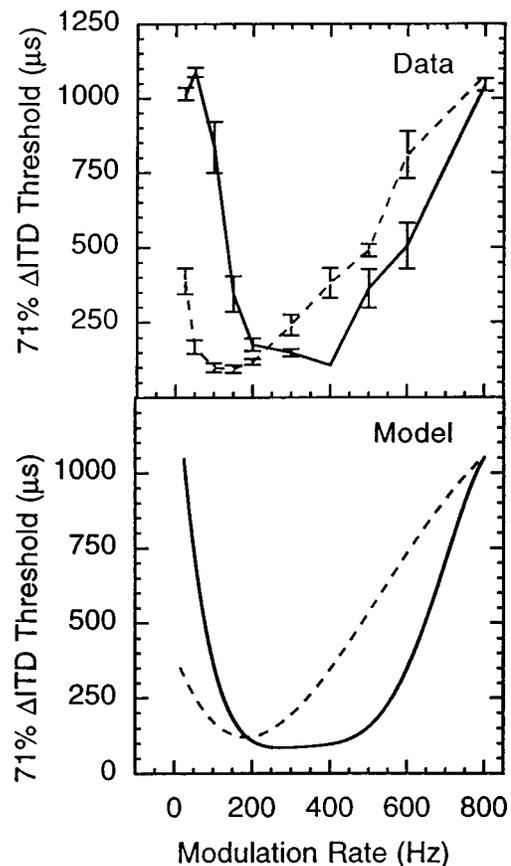


FIG. 5.  $\Delta$ ITD thresholds for SAM tones (dashed line) compared to SFM tones (solid line). The carrier frequency was 3 kHz. Data are averaged across three subjects. Error bars are one standard error of the mean. Lower panel shows model predictions.

$\Delta$ ITD thresholds are measured for a 3-kHz carrier, either frequency or amplitude modulated at rates from 25 to 800 Hz. This is a slightly larger range of rates than those used in experiment I. The same controls and procedures as before were used in this experiment, except that the amplitude of the SAM tone was adjusted by  $m^2/2$  (Viemeister, 1979) to equate the stimulus power with that of the SFM tone for which  $\beta$  was equal to one. Three new subjects were used in this experiment (JA,AW,KS).

### 1. Results

Top panel of Fig. 5 shows the results averaged for three observers. The bottom panel shows model predictions and will be discussed shortly. The dashed line shows thresholds measured for the SAM tone and the solid line, for the SFM tone. Error bars are one standard error. There is good agreement between the SFM data from experiment I (at 3 kHz) and the current data, in spite of using different observers. It is clear that the SFM data are shifted to the right by about 200 Hz compared to the SAM data. A second observation is that the same observers produced somewhat similar minimum thresholds for SFM and SAM tones, although the minimums occur at different rates (there is a slight edge for SAM thresholds). Previously published data on SAM tones has shown some variability in the minimum of this function, but generally, best sensitivity is observed anywhere from rates of

125–300 Hz (Henning, 1974a, Figs. 7, 8; Nuetzel and Hafter, 1981, Fig. 1). The data reported here show a minimum for the SAM function at a rate of 100–150 Hz for all observers, and a minimum of the SFM function that is centered around 200–400 Hz (Figs. 3, 5). There are statistically significant differences between SAM and SFM functions ( $F_{1,40}=21.52, p<0.05$ ), rate effects ( $F_{9,40}=23.42, p<0.05$ ), and interaction between condition (FM vs AM) and rate ( $F_{9,40}=13.33, p<0.05$ ).

The bottom panel shows model predictions. It is critical to note that the model parameters were determined from the SAM-tone data of experiment II. The following procedure was used to determine these parameters. The noise parameter  $\xi$  was adjusted to force the model performance to match that of the single SAM-tone datum at a modulation rate of 150 Hz (average value in Fig. 5). The low-pass FIR filter was then adjusted to best fit the slope of the increasing thresholds as the modulation rate is increased (dashed line in top panel of Fig. 5 from 200 to 800 Hz). No parameter adjustments were made for the low modulation rates below 150 Hz. The model was then presented with the SAM stimuli at all rates (low and high) and with the SFM stimuli. The final output of the model was smoothed with a polynomial function. The following are the important features of the model prediction for experiment II.

(1) An increase in  $\Delta$ ITD threshold with SAM tones is observed as the modulation rate is decreased below 150 Hz. This feature is also observed in the model and is a result of its stochastic nature and the reduction in the slope of the envelopes at the filter outputs as modulation rate is decreased (approaching a dc function as modulation rate approaches zero). Because the cross-correlation function also becomes less peaked, for a constant variance noise  $\xi$ , the position estimate becomes less reliable.

(2) At low rates of modulation, thresholds for SFM tones also increase for the same reason, however, thresholds increase much faster than those for SAM tones. This is also observed in the model behavior and is due to the following. For a constant index of frequency modulation ( $\beta = \Delta f / f_m = 1$ ) as the modulation rate ( $f_m$ ) is decreased, the peak frequency deviation must also decrease. Consequently, the instantaneous frequency of the stimulus sweeps through a progressively smaller region of an auditory filter, hence an output which will have a lesser depth of modulation. This is not true of an SAM tone whose amplitude is always fully modulated when  $m=1$ . Thus thresholds for an SFM tone increase not only because of a decrease in envelope slope (as for SAM tones) but additionally because of a reduction in peak-frequency-deviation and therefore the depth of modulation.

(3) As the modulation rate is increased above 150 Hz, thresholds increase for both SAM and SFM conditions, but they increase more quickly for the SAM stimuli (i.e., for a given modulation rate above 200 Hz, SFM stimuli produce lower thresholds). The model also shows this feature. Here, the reason is similar to that described in the previous point except in the opposite direction. As the rate increases, so does the peak-frequency-deviation. The output of a filter presented with a moderately high rate of frequency modulation

(200–500 Hz) is highly peaked since the instantaneous frequency of the stimulus sweeps rapidly through the filter's passband. The highly peaked nature of the outputs of such filters provides a cross-correlation pattern that is also highly peaked. Therefore, the position estimates have a low variance in the SFM case.

(4) The predictions for SFM conditions show a wider region of “good” performance at both high and low rates of modulation. This difference is small, but consistent. The width of this region is primarily affected by the slope of the central lowpass filter and the bandwidth of the filterbank model.

(5) Thresholds for both SAM and SFM stimuli converge at the highest rate of modulation (800 Hz), both in the data and in the model predictions. This is not surprising in either case. In the experiments, the adaptive procedure had a ceiling  $\Delta$ ITD of 1300  $\mu$ s (650 for each interval of the 2IFC task). The procedure produces a maximum threshold estimate of about 1050  $\mu$ s, even for an undetectable signal, because correct responses are recorded by chance on half the trials (i.e., the probability of two successive chance correct responses is 0.25). For this same reason, we limited the model performance to a high threshold of 1050  $\mu$ s. Both in the data and predictions, a negative curvature from this feature is observed as the modulation rate approaches 800 Hz.

## B. Thresholds as $\beta$ is varied

Interaural-delay thresholds for SAM tones improve with increasing depth of modulation (Nuetzel and Hafter, 1981). For an SFM tone, as the depth of frequency modulation increases, its instantaneous frequency sweeps through increasingly wider frequency regions and its spectrum widens. As the tone sweeps through wider frequency regions, the AM induced from bandpass filtering may also increase in depth, but not necessarily in all conditions. Because the SFM tone has an infinite number of sidebands and a constant power, increasing the depth of frequency modulation may in fact distribute the spectral energy such that components falling within a filter's passband would produce smaller envelope depths. The widening of the spectrum may also be problematic if leakage through filters with low resonant frequencies is suspected. In the current experiment, the carrier frequency was 3 kHz and three modulation rates of 50, 100, and 300 Hz were used. For the 300-Hz rate,  $\beta$  was restricted to less than 1.5. For the lowest rate of 50 Hz, a maximum value of  $\beta = 8$  was used (a peak-frequency-deviation of 400 Hz). Note that all signals were additionally highpass filtered at 1.5 kHz and low-pass noise with the same cutoff frequency was continuously presented at  $N_0 = 32$  dB.

Top panel of Fig. 6 shows the averaged data from this experiment for three observers. The error bars are one standard error of the mean. The parameter is modulation rate. Modulation depth is expressed as  $20 \log(\beta)$  to be consistent with the expression of amplitude modulation depth  $20 \log(m)$  in decibels. As before, there is an upper bound of  $\approx 1050 \mu$ s on performance measurement due to the ceiling imposed on the adaptive procedure, thus the negative curvature for the 50-Hz rate. Clearly, a significant monotonic effect of  $\beta$  is observed. It is noteworthy that thresholds at a modulation

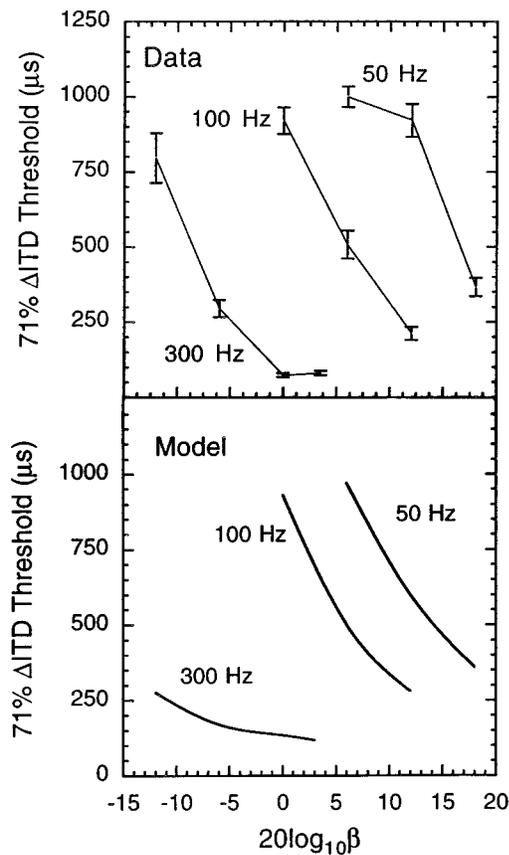


FIG. 6. Top panel shows the effects of changing the frequency-modulation depth on interaural-delay difference thresholds. The stimulus was an SFM tone with a 3-kHz carrier frequency and the parameter is modulation rate. There is an upper bound of  $\approx 1050 \mu s$  on performance measurement due to the ceiling imposed on the adaptive procedure. Data are averaged across three observers. Error bars are one standard error of the mean. Lower panel shows model predictions.

rate of 50 Hz, where virtually no sensitivity was previously observed, have noticeably improved as  $\beta$  was increased to 8. This indicates that the limiting factor is not rate *per se* but the depth of the induced AM.

The lower panel shows model predictions. Predicted thresholds improve with increasing  $\beta$  because for a constant rate of modulation the peak-frequency-deviation ( $\Delta f$ ) increases, producing envelopes at the filter outputs that have a higher peak factor. For  $f_m = 300$  Hz the model substantially underestimates the rate of change in threshold as  $\beta$  varies. The reason for this discrepancy is not clear. It should be noted that for these predictions, the noise parameter  $\xi$  was adjusted to best match the middle datum at a rate of 100 Hz (0-dB depth). If however,  $\xi$  is adjusted individually at each rate to best fit the datum at the middle value of  $\beta$  for that rate, considerably better fits are obtained even at 300 Hz (not shown).<sup>4</sup>

## V. EXPERIMENT III

### A. Binaural interaction between SFM and SAM tones

It is natural to pose the question that if an FM-to-AM conversion occurs in the auditory periphery, then would an SFM tone presented to one ear, interact with an SAM tone of identical parameters presented to the other ear? Consider a

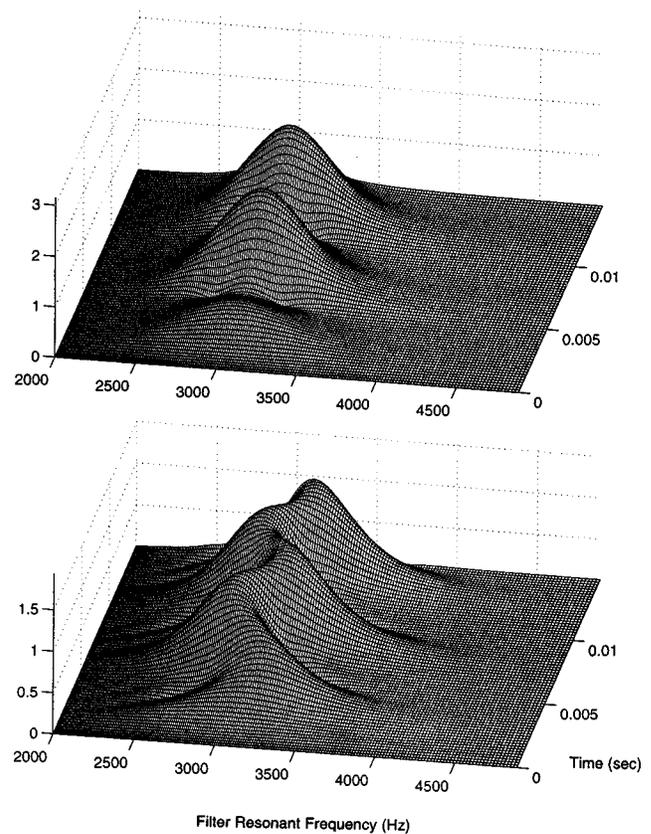


FIG. 7. Output of a Gammatone filterbank in response to an SAM tone (top panel) and an SFM tone (bottom panel). Both stimuli had a carrier frequency of 3 kHz, a modulation rate of 200 Hz, and a rise/fall time of 5 ms.

comparison of the outputs of a filterbank model in response to SFM- and SAM-tone stimulation (Fig. 7). The filterbank consisted of 300 Gammatone filters whose resonant frequencies were spaced logarithmically from 2 to 5 kHz. The top panel shows this response to an SAM tone and the bottom panel to an SFM tone, both modulated at 200 Hz and centered at 3 kHz. The SAM tone produces a modulation pattern whose peak amplitude is at the filter centered on the carrier. The rate of modulation is constant and independent of the filter placement. An SFM tone on the other hand produces a somewhat more complicated pattern. Both the depth and rate of modulation vary across filters. The output of the filter centered on the carrier produces a small AM response whose rate is twice that of the stimulus modulation rate. For the filters centered slightly away from the stimulus carrier frequency, the response is an envelope with asymmetrically occurring double-peaks during each cycle of modulation. There is, consequently, substantial spectral energy at the fundamental frequency of modulation at the outputs of even these off-frequency filters (see also Fig. 10). For filters centered further than the first sideband, the filter response is at the modulation rate, however, the response peaks occur at different times for filters centered above and below the carrier frequency. How do the different responses for different filters combine at a later processing stage? One may address some of these questions by examining the interactions that occur when an SFM tone is presented to one ear and an SAM tone to the other. The first part of the data reported here was

presented in summary form in an earlier paper (Saber and Hafter, 1995). Here individual data are presented in addition to newly collected data and model predictions.

## B. Procedure

The left- and right-ear stimuli were SAM and SFM tones, respectively,

$$X_L(t) = A'_C \sin 2\pi f_c t [1 + m \sin 2\pi f_m t], \quad (6a)$$

$$X_R(t) = A_C \sin[2\pi f_c t + \phi_c + \beta \sin(2\pi f_m t + \phi_m)]. \quad (6b)$$

Both modulator waveforms were defined in sine phase, were fully modulated  $m = \beta = 1$ , had equal modulation rates (250 Hz) and carrier frequencies (3 kHz). The SAM-tone modulation waveform remained in sine phase throughout the experiment and an interaural delay was introduced by shifting the phase of the SFM-tone modulation. Both waveforms had 20-ms cosine-squared ramps that were not interaurally delayed. The task was 2IFC with feedback. In one interval  $\phi_m = 0$ , and in the other interval  $\phi_m = 2\pi f_m$  ITD. As before, observers were instructed to indicate if the auditory image was perceived left–right or right–left. Most observers reported that for the homophasic case, the images were not symmetric about the median plane, but shifted toward the right ear which carried the SAM tone. Because this was the referent condition, the level of the SAM tone was adjusted ( $A'_C$ ) so that when the modulators were homophasic at the two ears, the intracranial image was centered. All other procedures and apparatus were the same as those described in experiment IA, except that a blocked psychophysical design with fixed stimulus parameters was used instead of the adaptive procedure and the measure of performance was proportion of correct responses in runs of 100 trials, with one run per subject per interaural delay.

## C. Results

Subjects reported that a single intracranial image was perceived whose position varied as the interaural modulation phase was varied. Figure 8 shows proportion of correct responses for three observers (filled symbol) as a function of interaural phase disparity. The open symbols are the averaged data and the solid line is model predictions and will be discussed later. When the modulation of the SFM tone was antiphase relative to the SAM phase ( $\pi$  radians or a 2-ms ITD) discrimination performance was at its maximum. As the phase difference approached a full cycle, performance dropped to chance. One observer reported that for the conditions where  $\phi_m > \pi$ , the images reversed position and she had to reverse her response pattern (i.e., the correct feedback light became correlated with a reversal of the order of the images). For this case, all observers were instructed to adopt a response strategy to maximize the number of correct responses. Such a reversal of image positions is consistent with the cyclic nature of the modulation waveform.

Figure 9 shows results of a second experiment of this type. As before, an SAM tone with a carrier frequency of 3 kHz and a modulation rate of 250 Hz was presented to the right ear. The left ear, however, received an SFM tone whose

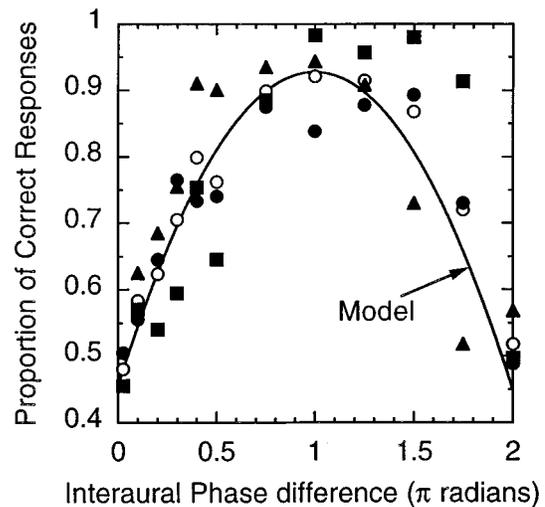


FIG. 8. Results of experiment III. An SFM tone was presented to the right ear and an SAM tone to the left ear. Proportion of correct responses are shown for three subjects (filled symbols). The task was to discriminate homophasic modulation from a delayed-modulation condition (gating envelopes were not delayed). The open symbols are the average for the three subjects and the line shows model prediction.

carrier frequency and modulation rate were parameters of the study. On each run of 100 trials, a single modulation rate and a single carrier frequency were selected for the SFM tone and percent-correct detectability of a 2-ms delay imposed on the SFM-tone modulator relative to the SAM-tone modulator were examined. Two features of these data are important. First, performance is above chance only when the modulation rate of the SFM tone is equal to that of the SAM tone. Other modulation rates do not binaurally interact with the 250-Hz SAM tone.

It is instructive to note a lack of interaction for an SFM-tone rate of 125 Hz with an SAM tone of 250 Hz. An SFM tone with a rate of 125 Hz produces an envelope with a rate of 250 Hz at the output of a filter centered on the carrier. However, this envelope, as observed from Figs. 7 and 10, has a very small depth and its effects are apparently easily overwhelmed by responses from off-frequency filters. Further support for off-frequency listening is provided by Fig. 8. If only the information from the filter centered on the stimulus carrier was used, a bimodal function may be expected with peaks at 0.5 and 1.5 rads.

A second feature of Fig. 9 is that binaural interaction is maintained for a wide range of interaural carrier-frequency disparities, as large as 1 kHz. This result is consistent with that reported for high-frequency SAM tones. Nuetzel and Hafter (1976, 1981) and Henning (1974a) have shown that some observers maintain interaural sensitivity to an SAM tone with interaural carrier-frequency disparity of 0.3–1 kHz. McFadden and Pasanan (1975) have also shown binaural beats with two-tone complexes for relatively large interaural-frequency differences between the complex to the left and right ears ( $\approx 1$  kHz). Naturally, the skirts of cochlear filters extend beyond one critical band. If we consider a Gammatone filter as an approximation, the level of a 3-kHz carrier component of an SAM tone is 30 dB attenuated at the output of a filter centered at 4 kHz, and it is about

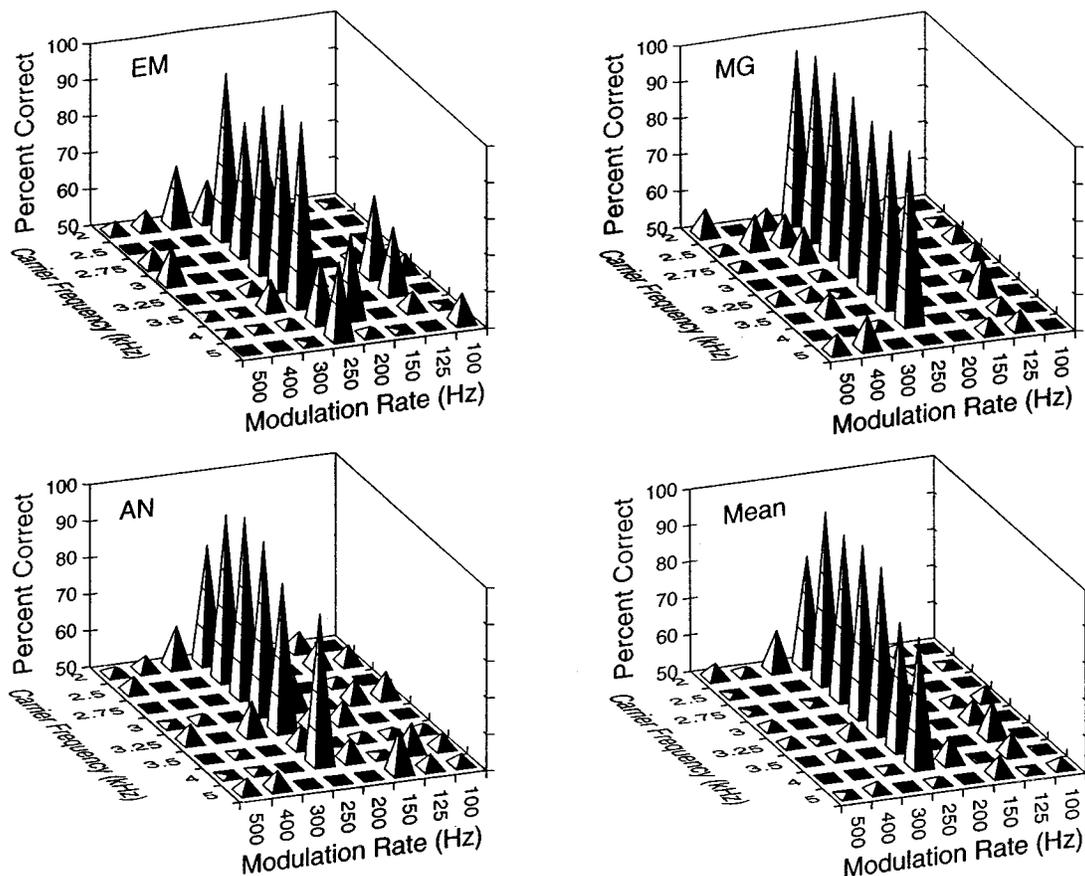


FIG. 9. Interaural-delay sensitivity for unequal carriers and modulation rates. The left ear received an SAM tone with a carrier of 3 kHz and a modulation rate of 250 Hz. The right ear received an SFM tone, the carrier and modulation rate of which were parameters ( $x$  and  $y$  coordinates). Data were collected for all 64 combinations of carrier versus rate. Each panel shows data from one observer. The lower right panel shows the average data for the three observers.

50 dB attenuated at the output of a filter centered at 2 kHz. These are only moderate attenuations and apparently insufficient to prevent binaural interactions.

Data were also collected at other modulation rates. Some preliminary listening showed that different observers showed best performance at different modulation rates, but generally rates between 100 and 250 Hz were optimal. For one observer, performance was measured at a modulation rate of 100 Hz as  $\beta$  was varied. The number of correct responses out of 100 trials per run were 100 for  $\beta=1.0$ , 92 for  $\beta=0.2$ , 88 for  $\beta=0.1$ , and 72 for  $\beta=0.05$ . Even at a modulation depth of  $-25$  dB ( $\beta=0.05$ ) where the peak-frequency-deviation is only  $\pm 5$  Hz this observer showed above chance performance. The frequency deviation in this case is too small to produce a noticeable modulation depth at the filter centered on the carrier ( $\approx 0.1$  dB if calculated from a Gammatone filter). Listening must therefore occur at more remote filters. We show further data of this type from more observers in the discussion section when considering the off-frequency process underlying this interaction.

For simplicity, the data of experiment III on binaural interaction between an SFM and SAM tone are analyzed by a statistical approach since the data were collected at a fixed modulation rate (the stochastic aspects of the model do not significantly affect this analysis). Figure 11 shows a surface plot of the SFM–SAM cross-correlation function (after bandpass filtering) for the two extreme stimulus phase con-

ditions, that is when the modulation waveforms of the two stimuli were either in-phase (left panels) or out-of-phase (right panels) relative to each other. The surface plots are a top-down view of the  $\tau$ - $f$  plane; the lighter regions represent the areas of greatest activity. The lower panels are the same as the upper panels except that a logistic frequency-weighting function was used to attenuate the activity above 3 kHz. The upper plots show nearly identical patterns that are mirror images along the frequency axis. It is easy to see that if one were to integrate across frequency, the in-phase and out-of-phase conditions would produce similar outputs and would therefore be the least discriminable conditions.<sup>5</sup> This is opposite to what was observed in the data. However, if observers were to listen off frequency either to auditory filters above or below the carrier (but not both), then these two phase conditions would be maximally discriminable as observed in the data. As the lower panels show, the inclusion of a frequency-weighting function allows discrimination between homophasic and antiphase conditions because the peak activity is at negative delays for the homophasic conditions and at positive delays for the antiphase conditions. To evaluate model predictions for this stimulus condition, the cross-correlation function was frequency-weighted prior to frequency integration as shown in Fig. 11. The position estimate ( $\hat{p}$ ) was then calculated as the interaural delay corresponding to the peak cross correlation. Index of detectability ( $d' = \hat{p}/\sigma$ ) was then determined from the position esti-

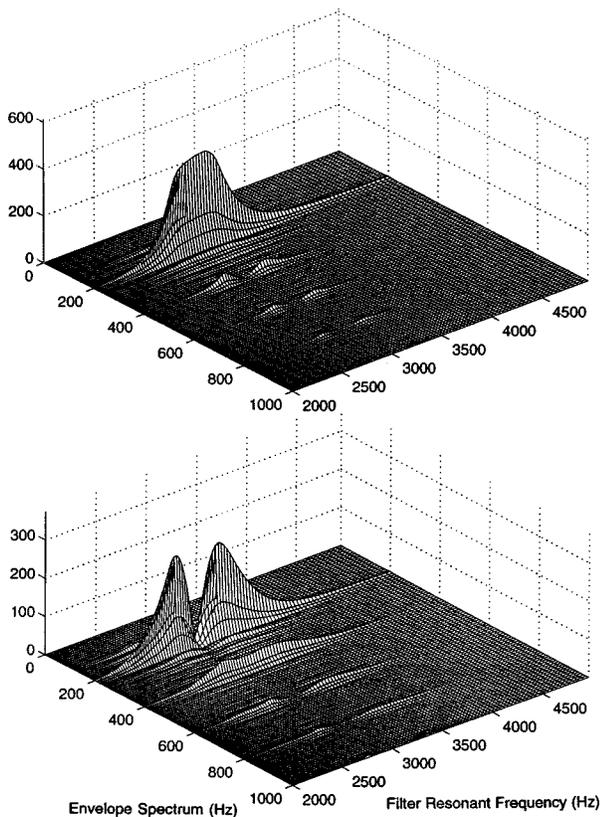


FIG. 10. Spectrum of the envelopes at the output of a Gammatone filterbank for an SAM tone (upper panel) and an SFM tone (lower panel). The stimulus was a 3-kHz carrier modulated at 200 Hz.

mate where  $\sigma$  is the single free parameter of the model. Predictions are plotted in Fig. 8 (solid line) together with the individual data (filled symbols) and averaged data (open symbols) of experiment III. Thus de-emphasizing binaural cross-correlation activity at the higher frequency regions captures the trend of the data.

To further test the validity of this analysis and the off-

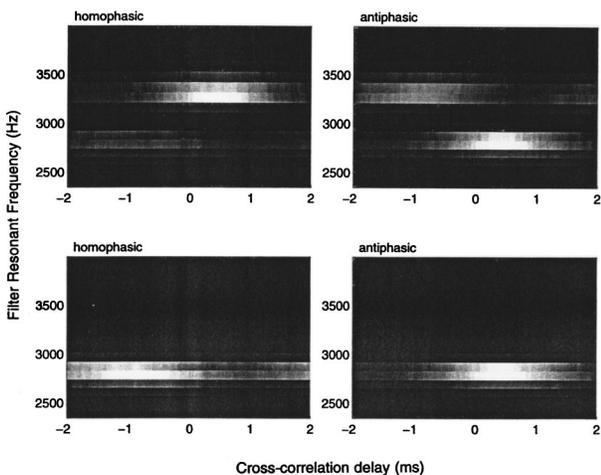


FIG. 11. Surface plots of interaural cross-correlation functions for four conditions (top-down view of the  $\tau$ - $f$  plane). The lighter regions show the areas of greatest activity. Left panels show the case where the SFM and SAM modulation waveforms were homophasic, and the right panels show antiphase conditions. The upper panels are unweighted and the lower panels are frequency weighted.

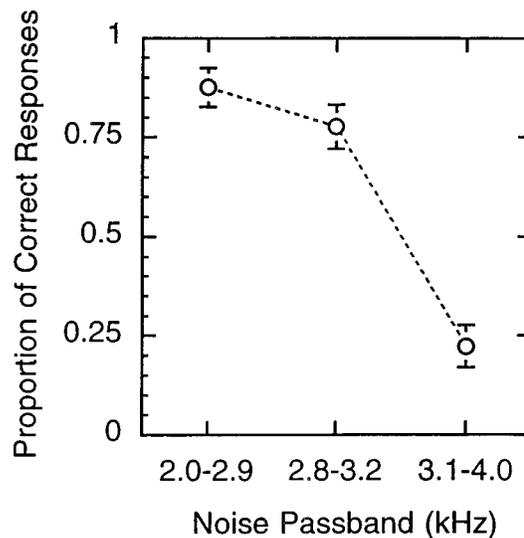


FIG. 12. Proportion of correct responses of detecting a homophasic from antiphase modulation waveform at the two ears, as a function of masking either lower, upper, or middle frequency regions of the stimuli. Chance performance is 0.5. Left ear received an SAM tone and right ear an SFM tone, both with carriers of 3 kHz and modulation rates of 250 Hz. The subject's task was to respond right if the image orders were left-right and to respond left if the orders were right-left. Correct response was arbitrarily assigned to the homophasic-antiphase, and incorrect to antiphase-homophase order of stimulus presentation in the 2IFC task. Data are averaged for three observers and error bars are one standard error.

frequency explanation, an additional control experiment was run in which filtered noise was used to mask either upper, lower, or middle frequency regions of the stimuli. The stimulus conditions and procedures were the same as described for experiment III. The modulation rate and carrier frequency of both stimuli (SFM to one ear and SAM to other) were 250 Hz and 3 kHz, respectively. The task was to discriminate a homophasic from antiphase condition. The three masking conditions included gated Gaussian noise with a passband of either 2.0–2.9 kHz, 2.8–3.2 kHz, or 3.1–4.0 kHz. No trial-by-trial feedback was provided.

Averaged data for three observers are shown in Fig. 12. Correct responses were arbitrarily assigned to one of the two orders of presentation (i.e., homophase-antiphase=correct; antiphase-homophase=incorrect), although as before, no trial-by-trial feedback was provided to the subjects. A few features of these data are worth emphasizing. First, a reversal of performance is observed when opposite frequency regions of the stimulus are masked, consistent with off-frequency listening and predicted from the cross-correlation analysis discussed above (note that chance performance is 0.5). Second, when the frequency region near the carrier is masked, observers seem to focus on the upper frequency region (i.e., similar performance for 2.0–2.9 and 2.8–3.2 kHz conditions). Third, unmasked performance (Fig. 8) is consistent with the low-frequency masker condition, suggesting that in the unmasked case, observers also attend to the frequency region above the carrier frequency, even though information below the carrier may also be available.<sup>6</sup>

## VI. DISCUSSION

We would like to address three issues in this section: (1) Why are threshold-by-rate functions different for SAM and

SFM tones; (2) What are the envelope cues that promote off-frequency listening to SFM tones; and (3) How does the phase response of auditory filters alter the envelope of the AM-like response. The data and predictions both show that interaural-delay thresholds follow U-shaped functions for both SAM and SFM tones as modulation rate is varied (Fig. 5). Why do the minima of these functions occur at different modulation rates? First, consider the stimulus properties that limit interaural-delay sensitivity for SAM tones. At low modulation rates, even though the carrier is fully modulated, performance is limited by the inability of the binaural system to track envelope synchrony for tones whose envelopes are changing very slowly (i.e., a dc effect). This is of course not a characteristic of high-frequency fibers *per se* but of envelope slope since highpass filtered clicks are easily lateralized with a precision of better than 100  $\mu$ s (Hafters *et al.*, 1983, 1988). Performance at low rates may also be affected by a reduction in the number of samples or envelope peaks that the binaural system receives for a fixed stimulus duration (McFadden and Moffitt, 1977). While we do not exclude this factor, it should be noted that the experiment on frequency-modulation depth (Fig. 6) shows that thresholds improve with increasing  $\beta$ , which in turn increases the effective slope of the envelope at a filter's output without changing the number of envelope-peak samples. For the higher modulation rates, SAM thresholds are affected both by cochlear filtering and a possible central low-pass filter (Viemeister, 1979). As the modulation rate is increased, the SAM sidebands are increasingly removed from the carrier and are thereby attenuated by the skirts of the filter centered on the carrier, resulting in a lower envelope depth and, thus, poorer performance (Nuetzel and Hafters, 1981).

What factors limit performance at high and low modulation rates for SFM tones? If the index of modulation is kept constant as the modulation rate is decreased, there is by definition a linear decline in peak-frequency-deviation,  $\Delta f$ . The smaller  $\Delta f$  yields an AM response with a smaller envelope depth. Thus for an SFM tone, not only is performance at low rates degraded by reduced envelope slope (as was the case for SAM tones), but there is an additional reduction of envelope depth due to smaller  $\Delta f$ . Therefore, thresholds degrade more rapidly for SFM tones compared to SAM tones as the modulation rate is reduced. Note that even if  $\Delta f$  for an SFM tone is kept constant as the rate is lowered, there is no reason to believe that the lower end of the rate-by-threshold functions would be identical for SAM and SFM stimuli since the envelope shapes would not be identical (Edwards and Viemeister, 1996). At higher modulation rates of about 300–400 Hz, the SFM tone produces lower thresholds than the SAM tone. One reason for this is an increase in the peak-frequency-deviation with increasing rate and a corresponding increase in depth of the AM-like envelope. At even higher rates (700 Hz) neither stimulus can be lateralized, either because the components are not resolved or because of limitations imposed by a central low-pass filter.

What are the cues that promote off-frequency listening? The data of experiment III suggest that observers do not heavily rely on ITDs from filters on or near the carrier frequency of an SFM tone. If an AM-like cue is the primary

information from an SFM tone, then the rate associated with on-frequency listening would be twice the stimulus modulation rate and its depth would be considerably smaller than the response from off-frequency filters. Off-frequency filters produce an envelope rate that either has single peaks at the modulation rate or multi-modal peaks with energy at the fundamental frequency of modulation. The reliance on off-frequency listening is not at a major cost. One can estimate the available cues at different rates of frequency modulation at the outputs of a filterbank model from the Fourier spectrum of their envelopes. Figure 10 shows this output for both an SFM and SAM tone with a 3-kHz carrier and a 200-Hz modulation rate. Clearly, the predominant information is at the fundamental periodicity for both stimuli and the loss in information from neglecting the on-frequency filter is very small for the SFM tone since there is very little envelope-spectral-energy at the output of that filter to begin with. Consistent with these observations, others have provided evidence from monaurally presented stimuli that observers rely on off-frequency filters where the change in excitation pattern in response to frequency modulation is maximal (Moore and Glasberg, 1986; Moore and Sek, 1992; Hartmann and Hnath, 1982).

Finally, a concern throughout this study was how the phase response of auditory filters affect the processing of the FM stimulus. Equation (1) shows that the odd harmonics below the carrier of the SFM tone are inverted in phase relative to the remaining components. The phase characteristic of an auditory filter, of course, would affect the phase relations of this harmonic complex and thereby alter the time envelope at the output of that filter. Unless the filter has a linear phase (i.e., a delay) the envelope of any filtered waveform that contains more than two components within the filter passband will be affected. What then are the phase-by-frequency functions of auditory filters; how do they compare with the phase response of Gammatone filters; and, are changes in envelope from phase effects significant compared to external (stimulus) envelope modulation or that caused by a filter's magnitude response?

Kohlrausch and Sander (1995) suggest that the phase of an auditory filter is consistent with the phase pattern of a negative Schroeder-phase signal. A Schroeder-phase signal is a harmonic complex that has a quadratic phase-by-frequency function

$$\phi_n = \phi_1 - \pi n^2 / N, \quad (7)$$

where  $n$  is harmonic number,  $N$  is the total number of components in the complex, and the sign in front of  $\pi$  defines the sign of the Schroeder phase (Schroeder, 1970). The Gammatone filter used in our analysis has an antisymmetric phase that is increasing in slope below the point of inflection (at the filter's resonant frequency) and decreasing above it. The curvature of this filter's phase (its second derivative) changes signs at the resonant frequency, whereas the curvature of a Schroeder-phase signal is constant. Kohlrausch and Sander suggest that a Gammatone antisymmetric phase would result in similar masking period patterns for negative- and positive-Schroeder-phase maskers, which is not what they have observed in their data (i.e., they suggest that the

TABLE I. Effects of filter phase on modulation depth (dB).

Filter CF (kHz)	Phase: Antisymmetric	+ Schroeder	- Schroeder
2.75	-1.57	-2.22	-2.21
3.00	-19.5	-18.5	-18.4
3.25	-3.3	-4.3	-4.3

Gammatone phase is not the appropriate model for auditory-filter phases). It is not however clear to what extent the filter's amplitude response affects the predicted differences. The models used by Kohlrausch and Sander for comparison were the Gammatone and Strube's (1985, 1986) basilar-membrane (BM) model. The BM model has a phase function with a negative curvature that is constant throughout most of the filter's passband, but has an amplitude response that simulates only the passive properties of the basilar membrane and therefore shows poor frequency selectivity compared to psychophysical and physiological data. The Gammatone on the other hand shows frequency selectivity comparable to psychophysical data, but has the antisymmetric phase response.

Positive- and negative-Schroeder-phase harmonic complexes such as those used by Kohlrausch and Sander may be phase shifted by an allpass filter with an appropriate negative curvature such that one signal would have a high peak factor and the other signal a low peak factor. If these two signals are then filtered with the amplitude response of a Gammatone filter, one observes little difference in the envelopes of the resultant waveforms at the output of the filter (i.e., the Gammatone amplitude response is substantially more significant than its phase response in determining the modulation envelope). Even if the Gammatone filter had a negative quadratic phase, similar predictions would be observed for the positive- and negative-Schroeder-phase maskers.

To observe how these phase-by-frequency functions affect the envelope of an SFM tone, we made comparisons under three conditions: (1) Gammatone antisymmetric phase; (2) negative-Schroeder-phase; (3) positive-Schroeder-phase. All stimuli were defined by summation of 13 primary harmonics of the SFM tone, and then filtered with an allpass filter with the assigned phase condition by phase shifting each component. The stimulus was then filtered with the amplitude response of a Gammatone filter by attenuating each component before summation. Thus, the filtering properties of a hypothetical Gammatone filter with a phase response defined by one of the three conditions was modeled. An SFM tone with a modulation rate of 250 Hz and a carrier of 3 kHz was the stimulus. The filter had a center frequency of 2.75, 3.0, or 3.25 kHz. The Hilbert envelopes of the filtered signals were then used to calculate an amplitude modulation depth in units of  $20 \log_{10}[(\max - \min)/(\max + \min)]$ .

Results are shown in Table I. The negative- and positive-Schroeder-phase filters produced identical modulation depth, but with slightly asynchronous envelope minima. The difference between the antisymmetric phase and Schroeder phases are generally small, approximately 0.6 dB for filters centered at 2.75 kHz, 1 dB at 3 and 3.25 kHz. The

smallest envelope modulation, as expected, occurred at the filter centered on the carrier of the stimulus. Why are the results so similar for the various phase conditions? The primary components of an SFM tone are the carrier, the first and perhaps the second sidebands; further sidebands are not likely to contribute significantly to the envelope of the filtered waveform. As shown in Fig. 1, the first and second sidebands are approximately 5 and 16 dB lower in amplitude than the carrier. For a filter centered on the carrier, the first and second sidebands are further attenuated by 6 and 17 dB, respectively (assuming a Gammatone filter). The first and second sidebands for the modulation rate of 250 Hz are contained within the frequency region of 2.5–3.5 kHz. For components in this region, the filter produces a phase shift that is well modeled by a linear function of frequency for all three phase conditions ( $r^2 > 0.999$ ). A linear phase shift is of course only a waveform delay, leaving its envelope unaffected; this nearly linear phase also explains the shift in the envelope minima when using filters with negative versus positive Schroeder phases. If we include the effects of the second through fourth sidebands (2.0–4.0 kHz), the deviation from linearity increases, but not substantially (Gammatone:  $r^2 = 0.96$ ; Schroeder:  $r^2 = 0.98$ ). For an off-frequency filter centered at 2.75 kHz, the Gammatone-phase curvature becomes more pronounced and the fit to linearity is reduced ( $r^2 = 0.94$  for the 2- to 4-kHz region) while it remains the same for the Schroeder-phase filters ( $r^2 = 0.98$ ). Although it is difficult to generalize these results (since both the auditory filter phase and amplitude functions are level dependent: Allen, 1983; Ruggero *et al.*, 1992) it seems that the magnitude response of the Gammatone filter and not its phase response is the primary determinant of envelope depth for a filtered SFM tone, at least for the stimulus conditions examined here.

## VII. SUMMARY

(1) The binaural system shows optimum interaural-delay sensitivity when SFM tones are modulated at rates of 200–400 Hz, which is somewhat higher than those rates that are optimum for SAM tones (100–200 Hz). Lowest thresholds of about 100–200  $\mu$ s for SFM tones at  $\beta = 1$  are not as good as the 80–100  $\mu$ s lowest thresholds observed for SAM tones.

(2) An SFM tone presented to one ear interacts interaurally with an SAM tone presented to the other ear, but only for equivalent modulation rates. Interaction is maintained, however, for interaural carrier-frequency disparities of up to 1 kHz. Strongest interactions occur for modulation rates between 100 and 250 Hz. Using bandpass masking noise it was shown that observers can use information from auditory filters either above or below the stimulus carrier frequency, although when conflicting information is present, they attend to the high-frequency regions.

(3) SAM-tone data were used to set the parameters of a stochastic interaural cross-correlation model whose predictions were then obtained for SFM tones of various rates and depths. Several differences between the SAM data and SFM data were well predicted by the model whose main features were an FM-to-AM conversion by bandpass filtering, and an

internal-noise limitation on estimating the peak of the cross-correlation function, and thereby, discrimination of lateral position.

## ACKNOWLEDGMENTS

Work supported by NIH Grants Nos. R29 DC03648-01, R01 DC00177, and R01 DC00100. I thank Drs. David M. Green, Ervin R. Haftner, Louis D. Braida, Patrick M. Zurek, Brent W. Edwards, and Malvin C. Teich for many helpful comments. I also thank members of the Sensory Communication Group and the Research Laboratory of Electronics at MIT.

<sup>1</sup>A fractal renewal process (FRP) is generated from sample times between adjacent events that are independent and selected from the same fractal probability distribution (Lowen and Teich, 1995). The survival function for this renewal process decays as a power law

$$t_i \sim \text{power}\{c/t^D - c/\zeta^D\} \text{ for } t < \zeta \\ 0 \text{ otherwise,}$$

where  $t_i$  are independent samples from a power-law density,  $D$  is the fractal dimension set at 0.75 in current implementation (Teich and Lowen, 1994),  $\zeta = 0.4$  sec is a cutoff parameter for the waveform duration, and  $c$  is a normalization constant that ensures the area under the density integrates to unity. The FRP exhibits fractal behavior over time scales under the value of  $\zeta$ . The process consists of a set of points on the time axis, but it may be recast as a real-valued Bernoulli process that alternates between two states, for example zero and unity. This alternating function would then start at zero, switch to unity at the time corresponding to the first event, switch back to zero at the second event and so on. A number of identical and independent alternating FRPs may be summed to yield a binomial process with the same fractal dimension as the single alternating FRP (Lowen and Teich, 1995) which may then be scaled by the instantaneous level of the stimulus to be used as the driving function for the Poisson point process

$$\text{Poisson}\{\mu_i = X(t) \sum \text{FRP}(t)\} \text{ over } \{\tilde{N}(t, t + \delta) = 0, 1, 2, \dots\},$$

where  $\mu_i$  is the time-dependent parameter of the nonstationary process,  $X(t)$  is the instantaneous pressure waveform, and  $\tilde{N}$  is the number of events within a time specified by the discrete waveform sampling period ( $\delta$ ). The result is a fractal-binomial-noise-driven Poisson point process, a stimulus-level dependent doubly stochastic point process (DSPP) that models fractal spike generation by nerve fibers.

<sup>2</sup>The centrality weighting function,  $p(\tau)$  implemented here has the shape of a Butterworth bandpass filter. Previous models have used either a function with a Gaussian density shape (Shackleton *et al.*, 1992; Saberi, 1996), or a hybrid function (Stern *et al.*, 1988b; Saberi, 1995) consisting of a uniformly flat maximum between  $\tau = \pm 150 \mu\text{s}$  and an exponential rolloff for  $|\tau| \geq \pm 150 \mu\text{s}$ . The Butterworth function was selected because it has the advantage of a relatively flat region near  $\tau = 0$ , a steep rolloff, but without discontinuities in its derivative at the rolloff transition point (i.e., as does the exponential model at  $\pm 150 \mu\text{s}$ ).

<sup>3</sup>The headphone transfer function, measured with a probe-tube microphone (Etymotic ER-7) inside the ear canal was flat within 0.8 dB from 2750 to 3250 Hz (the frequency region where many of the experiments were focused on). In some experiments additional controls were used to ensure that envelope cues resulting from headphone properties were not significant. For example, in some control cases, inverse digital filtering of the SFM signal inside the observer's ear canal was used to flatten the headphone transfer function. It should also be noted that any cues from the filtering of the SFM tone by the headphone transfer function is not pure AM but rather an FM-AM cue (an envelope modulation of a carrier that is changing in frequency).

<sup>4</sup>It is also worth documenting that attempts were made to measure masking-level differences (MLD) for high-frequency SFM tones in continuous noise. However, as has also been reported for SAM tones, no MLDs were observed for SFM stimuli.

<sup>5</sup>Previous models have used one of two measures of frequency coding. One model uses the straightness (rms error) of peak trajectories (Stern *et al.*, 1988b; Trahiotis and Stern, 1994) and others use frequency integration (Stern and Colburn, 1978; Shackleton *et al.*, 1994; Saberi, 1996). Both measures produce identical results in the analysis described here.

<sup>6</sup>By using a similar masking method, Moore and Sek (1994) have shown that monaural discrimination of SFM from SAM tones, at least at low rates of modulation (10 Hz), is also based on listening to off-frequency filters and that the specific listening region (above or below the carrier) is not critical to the discrimination task.

Abramowitz, M., and Stegun, I. (1964). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* (Dover, New York).

Allen, J. B. (1983). "Magnitude and phase-frequency response to single tones in the auditory nerve," *J. Acoust. Soc. Am.* **73**, 2071–2092.

Bernstein, L. R., and Trahiotis, C. (1992). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," *J. Acoust. Soc. Am.* **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (1995). "Binaural spectral interference in detection and discrimination paradigms," in *Advances in Hearing Research: Proceedings of the 10th International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump, C. Koppl, H. Fastl, and H. Oekinghaus (World Scientific, Singapore), pp. 354–361.

Bielek, K. H. (1975). "Spektrale Dominanz bei der interauralen Signalanalyse," Diploma thesis, Ruhr-University Bochum.

Blauert, J. (1981). "Lateralization of jittered tones," *J. Acoust. Soc. Am.* **70**, 694–698.

Blauert, J., and Divenyi, P. L. (1988). "Spectral selectivity in binaural contralateral inhibition," *Acustica* **66**, 267–274.

Buell, T. N., and Trahiotis, C. (1993). "Interaural temporal discrimination using two sinusoidally amplitude-modulated, high-frequency tones: Conditions of summation and interference," *J. Acoust. Soc. Am.* **93**, 480–487.

Carlson, A. B. (1986). *Communication Systems: An Introduction to Signals and Noise in Electrical Communication*, 3rd ed. (McGraw-Hill, New York).

Carney, L. H. (1993). "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.* **93**, 401–417.

Colburn, H. S. (1973). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," *J. Acoust. Soc. Am.* **54**, 1458–1470.

Colburn, H. S., and Esquissaud, P. (1976). "An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli," *J. Acoust. Soc. Am.* **59**, S23(A).

Cooper, G. R., and McGillem, C. D. (1986). *Modern Communications and Spread Spectrum* (McGraw-Hill, New York).

Couch, L. W. II (1987). *Digital and Analog Communication Systems* (Macmillan, New York), 2nd ed.

David, E. E., Guttman, N., and van Bergeijk, W. A. (1959). "Binaural interaction of high-frequency stimuli," *J. Acoust. Soc. Am.* **31**, 774–782.

Edwards, B. W., and Viemeister, N. F. (1996). "Masking of a brief probe by sinusoidal FM," *J. Acoust. Soc. Am.* (submitted).

Green, G. G. R., and Kay, R. H. (1975). "On the localisation by humans of frequency-modulated tones," *J. Physiol. (London)* **252**, 59–60.

Green, G. G. R., Heffer, J. S., and Ross, D. A. (1976). "The detectability of apparent source movement effected by interaural phase modulation," *J. Physiol. (London)* **260**, 49.

Haftner, E. R., and De Maio, J. (1975). "Difference thresholds for interaural delay," *J. Acoust. Soc. Am.* **57**, 181–187.

Haftner, E. R., and Dye, R. H. (1983). "Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number," *J. Acoust. Soc. Am.* **73**, 1708–1713.

Haftner, E. R., Buell, T. N., and Richards, V. M. (1988). "Onset coding in lateralization: it's form, site, and function," in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 647–676.

Haftner, E. R., Dye, Jr., R. H., Wenzel, E. M., and Knecht, K. (1990). "The combination of interaural time and intensity in the lateralization of high-frequency complex signals," *J. Acoust. Soc. Am.* **87**, 1702–1708.

Harris, G. C. (1960). "Binaural interactions of impulsive stimuli and pure tones," *J. Acoust. Soc. Am.* **32**, 685–692.

Hartmann, W. M., and Hnath, G. M. (1982). "Detection of mixed modulation," *Acustica* **50**, 297–312.

- Henning, G. B. (1974a). "Detectability of interaural delay in high-frequency complex waveforms," *J. Acoust. Soc. Am.* **55**, 84–90.
- Henning, G. B. (1974b). "Lateralization and the binaural masking-level difference," *J. Acoust. Soc. Am.* **55**, 1259–1262.
- Henning, G. B. (1980). "Some observations on the lateralization of complex waveforms," *J. Acoust. Soc. Am.* **68**, 446–454.
- Henning, G. B., and Ashton, J. (1981). "The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay," *Hearing Res.* **4**, 185–194.
- Holdsworth, J., Nimmo-Smith, I., Patterson, R., and Rice, P. (1988). "Implementing a Gammatone filter bank," in *Annex C of the SVOS final report (Part A: The auditory filter bank)*, APU Report 2341.
- Hukin, R. W., and Darwin, C. J. (1994). "Lateralization of a frequency-modulated 500-Hz tone," *J. Acoust. Soc. Am.* **95**, 2895–2896.
- Inglis, A. F. (1988). *Electronic Communications Handbook* (McGraw-Hill, New York).
- Jeffress, L. A., Blodgett, H. C., Sandel, T. T., and Woods III, C. L. (1956). "Masking of tonal signals," *J. Acoust. Soc. Am.* **28**, 416–426.
- Joris, P. X., and Yin, T. C. T. (1992). "Responses to amplitude-modulated tones in the auditory nerve of the cat," *J. Acoust. Soc. Am.* **91**, 215–232.
- Kay, R. H. (1982). "Hearing modulation in sounds," *Physiol. Rev.* **62**, 894–975.
- Kohlerausch, 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.
- Kohlerausch, A., Par, S.v.d., and Houtsuma, A. J. M. (1995). "A new approach to study binaural interaction at high frequencies," in *Advances in Hearing Research: Proceedings of the 10th International Symposium on Hearing*, edited by G. A. Manley, G. M. Klump, C. Koppl, H. Fastl, and H. Oeckinghaus (World Scientific, Singapore), pp. 343–353.
- Leakey, D. M., Sayers, B. McA., and Cherry, C. (1958). "Binaural fusion of low- and high-frequency sounds," *J. Acoust. Soc. Am.* **30**, 222.
- Levitt, H. L. (1971). "Transformed up-down methods in psychophysics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lowen, S. B., and Teich, M. C. (1995). "Estimation and simulation of fractal stochastic point processes," *Fractals* **3**, 183–210.
- Lowen, S. B., and Teich, M. C. (1996). "The periodogram and Allan variance reveal fractal exponents greater than unity in auditory-nerve spike trains," *J. Acoust. Soc. Am.* **99**, 3585–3591.
- McFadden, D., and Pasanen, E. G. (1974). "High frequency masking level differences with narrow-band noise signals," *J. Acoust. Soc. Am.* **56**, 1226–1230.
- McFadden, D., and Pasanen, E. G. (1975). "Binaural beats at high frequencies," *Science* **190**, 394–396.
- McFadden, D., and Pasanen, E. G. (1976). "Lateralization at high frequencies based on interaural time differences," *J. Acoust. Soc. Am.* **59**, 634–639.
- McFadden, D., and Moffitt, C. M. (1977). "Acoustic integration for lateralization at high frequencies," *J. Acoust. Soc. Am.* **61**, 1604–1608.
- Moore, B. C. J., and Glasberg, B. R. (1986). "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
- Moore, B. C. J., and Sek, A. (1992). "Detection of combined frequency and amplitude modulation," *J. Acoust. Soc. Am.* **92**, 3119–3131.
- Moore, B. C. J., and Sek, A. (1994). "Discrimination of modulation type (amplitude modulation or frequency modulation) with and without background noise," *J. Acoust. Soc. Am.* **96**, 726–732.
- Nordmark, J. O. (1976). "Binaural time discrimination," *J. Acoust. Soc. Am.* **60**, 870–880.
- Nuetzel, J. M., and Hafter, E. R. (1976). "Lateralization of complex waveforms: Effects of fine structure, amplitude and duration," *J. Acoust. Soc. Am.* **60**, 1339–1346.
- Nuetzel, J. M., and Hafter, E. R. (1981). "Lateralization of complex waveforms: Spectral effects," *J. Acoust. Soc. Am.* **69**, 1112–1118.
- Ruggero, M. A., Rich, N. C., and Recio, A. (1992). "Basilar membrane responses to clicks," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 85–92.
- Saberi, K. (1995). "Lateralization of comodulated complex waveforms," *J. Acoust. Soc. Am.* **98**, 3146–3156.
- Saberi, K., and Hafter, E. R. (1995). "A common neural code for frequency- and amplitude-modulated sounds," *Nature (London)* **374**, 537–539.
- Saberi, K. (1996). "Observer weighting of interaural delays in filtered impulses," *Percept. Psychophys.* **58**, 1037–1046.
- Saberi, K., and Hafter, E. R. (1996). "Experiments on auditory motion discrimination," in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, 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.
- Shackleton, T. M., Meddis, R., and Hewitt, M. J. (1992). "Across-frequency integration in a model of lateralization," *J. Acoust. Soc. Am.* **91**, 2276–2279.
- Shear, G. D. (1987). "Modeling the dependence of auditory lateralization on frequency and bandwidth," master thesis, Carnegie Mellon University.
- Stern, R. M., and Colburn, H. S. (1978). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," *J. Acoust. Soc. Am.* **64**, 127–140.
- Stern, R. M., Shear, G. D., and Zeppenfeld, T. (1988a). "Lateralization predictions for high-frequency binaural stimuli," *J. Acoust. Soc. Am.* **84**, S80.
- Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (1988b). "Lateralization of complex binaural stimuli: A weighted-image model," *J. Acoust. Soc. Am.* **84**, 156–165.
- 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.
- Teich, M. C. (1989). "Fractal character of the auditory neural spike train," *IEEE Trans. Biomed. Eng.* **36**, 150–160.
- Teich, M. C., and Lowen, S. B. (1994). "Fractal patterns in auditory nerve-spike trains," *IEEE Eng. Med. Biol. Mag.* **13**, 197–202.
- Trahiotis, C., and Bernstein, L. R. (1990). "Detectability of interaural delays over select spectral regions: Effects of flanking noise," *J. Acoust. Soc. Am.* **87**, 810–813.
- Trahiotis, C., and Stern, R. M. (1994). "Across-frequency interaction in lateralization of complex binaural stimuli," *J. Acoust. Soc. Am.* **96**, 3804–3806.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Wetherill, G. B., and Levitt, H. (1965). "Sequential estimation of points on a psychometric function," *Br. J. Math. Stat. Psychol.* **18**, 1–10.
- Westerman, L. A., and Smith, R. L. (1985). "Rapid adaptation depends on the characteristic frequency of auditory nerve fibers," *Hearing Res.* **17**, 197–198.
- Yost, W. A., Wightman, F. L., and Green, D. M. (1971). "Lateralization of filtered clicks," *J. Acoust. Soc. Am.* **50**, 1526–1531.
- Zurek, P. M. (1985). "Spectral dominance in sensitivity to interaural delay for broadband stimuli," *J. Acoust. Soc. Am. Suppl. 1* **78**, S18.
- Zwicker, E. (1952). "Die Grenzen der Horbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones," *Acustica* **2**, 125–133.