Microsecond temporal resolution in monaural hearing without spectral cues?

Katrin Krumbholz[a] and Roy D. Patterson
Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

Andrea Nobbe and Hugo Fastl
Institute for Human-Machine Communication, Technical University Munich, Arcisstrasse 21, D-80333 München, Germany

(Received 26 April 2002; revised 5 October 2002; accepted 16 December 2002)

The auditory system encodes the timing of peaks in basilar-membrane motion with exquisite precision, and perceptual models of binaural processing indicate that the limit of temporal resolution in humans is as little as 10–20 *microseconds*. In these binaural studies, pairs of continuous sounds with microsecond differences are presented simultaneously, one sound to each ear. In this paper, a monaural masking experiment is described in which pairs of continuous sounds with microsecond time differences were combined and presented to both ears. The stimuli were matched in terms of the excitation patterns they produced, and a perceptual model of monaural processing indicates that the limit of temporal resolution in this case is similar to that in the binaural system. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1547438]

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

I. INTRODUCTION

Physiological studies have shown that primary auditory-nerve fibers phase lock to the temporal fine structure of sinusoids up to frequencies as high as 5 kHz in the cat (Johnson, 1980). Psychophysical studies indicate that the binaural system can detect interaural time differences of a few tens of microseconds. Perceptually, interaural time differences are interpreted in terms of where the source is relative to the head, or how spatially compact it is (Durlach and Colburn, 1978). In this paper, we report a monaural perception experiment which suggests that the auditory system can also detect monaural time differences of a few tens of microseconds in continuous sounds even when they have matched excitation patterns. In this case, the perception involves distinguishing whether the sound contains one or two sources, or in the limit, how perceptually coherent the sound is. These source characteristics would appear to be defined by the monaural temporal structure of the sound rather than its spatial or spectral attributes.

The experiments were motivated by a monaural model of temporal processing that was recently used to explain a monaural masking release that is presumed to be based on the temporal microstructure of the sounds (Krumbholz *et al.*, 2001). The model and the masking release are introduced briefly in subsections I A and I B to explain the motivation for the research, which is introduced in subsection I C.

A. Models of monaural temporal processing

Much of the temporal information in the neural firing patterns generated in the cochlea is preserved up to the level of the superior olivary complex (Oertel, 1997). Beyond the brainstem, however, the temporal precision of neural firing deteriorates as the volley of firing progresses to higher stages in the system (Rouilly *et al.*, 1979). Time-interval models of auditory processing (e.g., Meddis and Hewitt, 1991a; Patterson *et al.*, 1995) assume that there is a mechanism that measures the time intervals between peaks in the phase-locked activity of auditory-nerve fibers, and thereby converts the unstable temporal information in the phase-locked response into a more stable time-interval code. The obvious assumption, although it is not explicitly stated, is that the time-interval representation might emerge at the stage of the inferior colliculus in the midbrain. This is the level in the system where the upper limit of phase locking drops to about 500 Hz, and there is a mandatory synapse with strong convergence of ascending pathways (Irvine, 1992).

Studies of temporal processing often employ a stimulus referred to as iterated rippled noise (IRN) in an effort to neutralize processing in the tonotopic dimension and focus attention on temporal processing. IRN is constructed by delaying a copy of a random noise by d ms, adding it back to the original noise, and iterating the process n times. The stimulus has a noise-like waveform with a degree of temporal regularity that increases with increasing numbers of iterations, n. When the delay, d, is between about 1 and 30 ms, IRN elicits a two-component perception, with a noisy hiss and a buzzy tone whose pitch corresponds to $1/d$. IRN is interesting because it enables the generation of sounds with any degree of tonality from almost completely tonal to completely atonal while leaving the gross temporal structure and the excitation pattern of the stimulus essentially unchanged. Perceptual differences between IRNs with matched excitation patterns are used to demonstrate the importance of time-interval analysis in auditory processing; Yost *et al.* (1996) and Patterson *et al.* (1996) have shown that the pitch and pitch strength of IRN are more readily explained by time-

[a]Current address: Institute for Medicine (IME), Forschungszentrum Jülich, D-52425 Jülich, Germany.

 0001-4966/2003/113(5)/2790/11/$19.00 © 2003 Acoustical Society of America

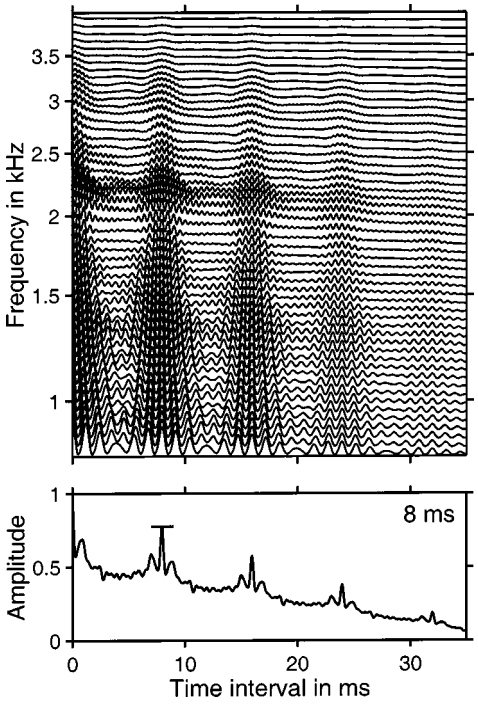

FIG. 1. Upper panel: Auditory image (AI) of an IRN with a delay of 8 ms and 16 iterations. The solid lines show time-interval histograms of the neural activity in different channels of the tonotopic array; the ordinate is the best frequency of the channel. Lower panel: Summary AI computed by averaging across channels in the AI; the short horizontal line shows the normalized height of the first peak in the summary image.

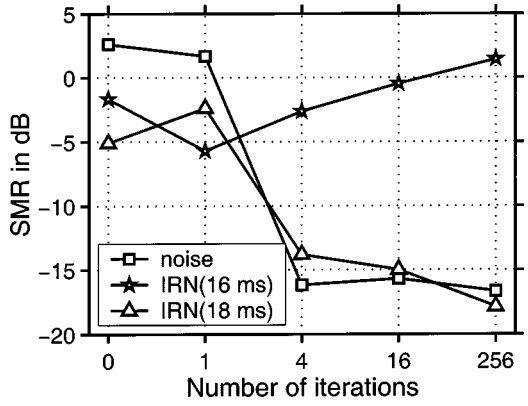

FIG. 2. Results from the IRN masking experiment of Patterson and Datta (1996). The abscissa is the number of iterations in the masker (0 iterations indicates that the masker was a noise). The signal was a noise (squares), or an IRN with 256 iterations and a delay of either 16 (stars) or 18 (triangles) milliseconds. The data show the average for three listeners and threshold is expressed in terms of the signal-to-masker ratio (SMR).

interval models of auditory processing than by models based on the excitation patterns of the stimuli.

In these time-interval models, a summary of the neural processing up to the level of the midbrain is simulated with a tonotopic array of autocorrelation functions (Meddis and Hewitt, 1991a; Slaney and Lyon, 1990) or time-interval histograms (Patterson *et al.*, 1992; Patterson, 1994). The upper panel in Fig. 1 shows the simulated neural pattern in response to an IRN with a delay of 8 ms and 16 iterations. The pattern was produced with the Auditory Image Model (AIM; Patterson *et al.*, 1995). The abscissa is time interval; the ordinate is frequency as it is represented along the frequency dimension of the cochlea. Each line in this representation shows the kind of time-interval histogram that might be calculated from the stream of spikes flowing from individual auditory-nerve fibers at a given frequency. The vertical ridges at 8 ms and integer multiples thereof demonstrate the presence of temporal regularity in the neural activity across the tonotopic array. The representation is referred to as an ''auditory image'' (Patterson, 1994). The overall form of the temporal regularity is similar in all channels, which illustrates how the stimulus neutralizes the tonotopic dimension of auditory processing. The lower panel of Fig. 1 shows the normalized sum of the activity across channels. It is referred to as a ''summary auditory image,'' and it is this representation that is typically used to summarize responses to stimuli in quantitative evaluations of time-interval models.

It is perhaps worth noting that in these representations, the autocorrelation-lag, or time-interval dimension is presented as if it were a spatial dimension similar to the frequency dimension, and that the auditory image might be thought of as the pattern of activity in a two-dimensional array of neurons at the level of the brainstem or the midbrain. This is a convenient way to conceive the processing of temporal information, but a spatial array of time intervals is not necessary for the functioning of these models; any method of representing concentrations of activity at different time intervals would suffice.

B. A monaural masking release based on temporal microstructure

Patterson and Datta (1996) reported that a noise masks an IRN much more effectively than an IRN masks a noise, even when the stimuli have the same energy and are filtered in the same spectral region. The effect is illustrated in Fig. 2, which shows the masked threshold for a noise (squares) and an IRN with a delay of 16 ms and 256 iterations (stars) in the presence of five different maskers—a random noise (0 iterations) and an IRN with 1, 4, 16, or 256 iterations and the same 16-ms delay. Threshold is expressed in terms of the signal-to-masker ratio (SMR), which is the difference between the overall level of the signal at threshold and the overall level of the masker. All of the stimuli were highpass filtered at 800 Hz, which is 12.6 times the reciprocal of the delay. The filtering ensures that the excitation patterns of the IRNs do not have sets of harmonically spaced peaks, and the long delay ensures that the stimuli contain frequency components in a frequency region with accurate phase locking to support time-interval processing. The figure shows that a noise signal is effectively masked by noise and by IRN with one iteration. But, as the number of iterations in the masker increases and its fine structure becomes more regular, there is a release from masking by as much as 18 dB. Similarly, the IRN signal with 256 iterations is effectively masked by IRN with 256 iterations. But, as the number of iterations in the masker decreases and its fine structure becomes less regular, threshold decreases. However, in this case the release from masking is only about 7 dB. Thus, there is an asymmetry of masking between noise and IRN, in that a noise (0 iterations)

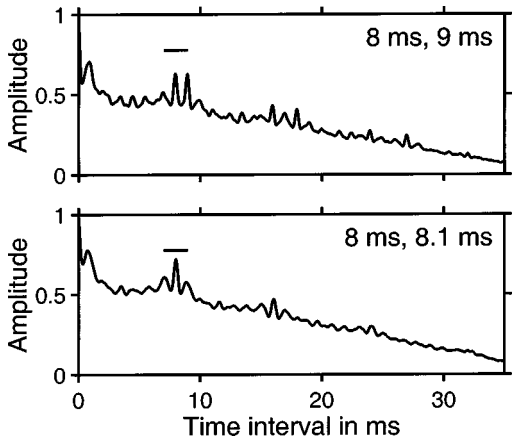

FIG. 3. Summary AIs of ''merged IRNs'' consisting of an 8-ms masker IRN and a signal IRN with a delay of either 9 (upper panel) or 8.1 (lower panel) milliseconds. The signal was added with a SMR of 0 dB. The IRNs were computed with 16 iterations of the delay-and-add process. The short horizontal lines show the height of the first peak in the summary AI of the masker alone, shown in the lower panel of Fig. 1.

masks an IRN (stars) much more effectively than an IRN (iterations$\geqslant$4) masks a noise (squares). Recently, Krumbholz *et al.* (2001) extended the range of the observations and showed that, for delays within the pitch region, the basic masking asymmetry between noise and IRN was essentially independent of the spectral region. This means that it is very unlikely that models based on the excitation patterns of the stimuli could explain the data. Krumbholz *et al.* (2001) explain the asymmetry of masking in terms of the fine-grain temporal information in the neural activity patterns of the stimuli.

There is a third condition in the study of Patterson and Datta (1996) in which the signal is an IRN with a delay of 18 ms and 256 iterations; it has a noticeably lower pitch than the IRN masker. The thresholds (triangles) show that the 18-ms IRN is masked by noise to about the same degree as the 16-ms IRN, as would be expected. However, when the number of iterations in the masker is four or more, the listener can use the difference in delay between signal and masker to detect the signal, and threshold falls to about the same levels as for the noise signal. This stands in contrast to the increase in threshold when the IRN signal has the same delay (16 ms) as the IRN masker. It is this condition where the IRNs have different pitches that provides the motivation for the current study.

The upper panel of Fig. 3 shows the summary auditory image when a 9-ms IRN signal is added to an 8-ms IRN masker with a signal-to-masker ratio (SMR) of 0 dB. The summary auditory image of the 8-ms IRN masker on its own was presented in the lower panel of Fig. 1. The summary auditory image of this ''merged'' IRN (Patterson *et al.*, 2000) exhibits peaks at both the masker delay (8 ms) and its integer multiples, and at the signal delay (9 ms) and its integer multiples. The heights of the peaks associated with the masker are, however, somewhat smaller than those for the masker alone; for reference, the short horizontal line shows the height of the 8-ms peak in the masker-alone condition. Handel and Patterson (2000) measured the pitch strength of merged IRNs with relatively large delay differences and showed that the pitch strength was greatly reduced except at octave delay ratios. Patterson *et al.* (2000) measured the pitch strength of merged IRNs with very small delay differences where the pitches of the two IRNs merge. They showed that there is still a pronounced reduction in pitch strength even when the delay difference was as little as 6% of the basic delay. So, the reduction in peak height remains audible to the listener after the pitches have merged. When the delay difference is less than about 2%, the time-interval peaks associated with the signal and masker merge in the summary auditory image (Patterson *et al.*, 2000); the lower panel of Fig. 3 shows the case where the masker is an 8-ms IRN and the signal is an 8.1-ms IRN. The form of the summary image is very similar to that for the 8-ms IRN masker on its own (Fig. 1), but the heights of the peaks are reduced, corresponding to the perceptual reduction in pitch strength. Note that the merging of the time-interval peaks for small delay differences is due to auditory processing that the stimuli undergo prior to the time-interval processing stage; Patterson *et al.* (2000) attribute it to a loss of phase-locking information in the midbrain. In any event, the degree to which the peaks merge provides information about the peripheral processing of the stimuli.

C. A masking paradigm for measuring monaural temporal resolution

In this paper, the reduction in pitch strength caused by the merging of IRNs with small delay differences is used to estimate the resolution with which time-interval information is represented in the auditory system. The listener's task is to detect a signal IRN in the presence of a masker IRN with a slightly different delay. The delay and level of the masker IRN were fixed and the level of the signal IRN was varied to determine the threshold as a function of the difference between the signal delay and the masker delay. The summary images in Figs. 1 and 3 suggest that threshold would drop as the delay of the signal begins to diverge from that of the masker. The function relating threshold to the difference in delay between the signal and masker will be referred to as a ''time-interval masking pattern.'' The traditional masking pattern reveals properties of frequency resolution in terms of the elevation of threshold for a sinusoidal signal in the frequency region around a narrow-band noise masker. Similarly, the time-interval (TI) masking pattern reveals properties of temporal resolution in terms of the elevation of threshold for an IRN signal with a delay in a time-interval region around the delay of an IRN masker.

II. METHOD

The time-interval masking pattern was measured for different masker delays and in different spectral regions to assess the effect of auditory filter width and phase locking on temporal resolution. The study of Krumbholz *et al.* (2001) showed that both variables have a pronounced effect on the asymmetry of masking between IRN and noise.

A. Stimuli

The masker and signal IRNs were generated with 16 iterations of the add-original algorithm (Yost *et al.*, 1996), using a gain, *g*, of unity. The stimuli were generated digitally with 16-bit resolution and a sampling rate of 25 kHz (TDT AP2). In order to produce the very small IRN delay differences required in this experiment, the IRN was generated in the spectral domain, by multiplying the Fourier spectrum of a Gaussian noise with the complex, comb-filter spectrum of an add-original IRN with delay, *d*, and *n* iterations

$$H(d,n)=\sum_{k=0}^{n} g^{k}[\cos(2\pi kdf)+i\cdot\sin(2\pi kdf)],$$

where *i* is the imaginary unit and *f* is frequency. The signal and masker were played out through separate channels of the TDT DD1 D/A converter and antialiasing filtered with a 10-kHz cutoff (TDT FT6-2). They were separately attenuated according to the signal and masker levels of the current trial using two programmable attenuators (TDT PA4). Finally, the signal and masker were added in an analog mixer (TDT SM3) and presented diotically to the listener, who was seated in a double-walled, sound-attenuating booth.

The stimuli were filtered into frequency bands with an equivalent rectangular bandwidth of 2.2 kHz; the lower cutoff frequency, F_c, was 0.4, 0.8, 1.6, 3.2, or 6.4 kHz. In order to avoid edge tones, the lower and upper edges of the frequency bands were rounded with a quarter cycle of a cosine function whose width was 0.2 kHz on the lower edge and 1 kHz on the upper edge (see Fig. 1 in Krumbholz *et al.*, 2001). A background of continuous pink noise was introduced to mask any distortion products that might be audible in the region below the passband of the stimuli. The pink noise was low-pass filtered half an octave below the lower cutoff frequency, F_c, of the current filter condition; the slope beyond the cutoff frequency was 96 dB/octave. The unfiltered pink noise had a level of 31.3 dB SPL in the 1/3-octave band around 1 kHz. In the 0.8- and 3.2-kHz filter conditions, thresholds were measured for several different masker delays, d_m, ranging from 4 to 64 ms for $F_c=0.8$ kHz, and from 4 to 32 ms for $F_c=3.2$ kHz. Initially, we intended to measure the same range of masker delays in both filter conditions. However, pilot work showed that the task was too difficult for $d_m=64$ ms in the 3.2-kHz filter condition. In the other filter conditions, only one masker delay, 16 ms, was used. Each masker delay was combined with six different signal delays. The delay difference, *dd*, between signal IRN and masker IRN ranged from 0 μs to a maximum of 800 μs. Pilot work revealed that signals with very small delay differences were more readily detectable in the higher filter conditions, so the maximum *dd* varied with filter cutoff, F_c, from 800 μs when F_c was 0.4 kHz, to 50 μs when F_c was 6.4 kHz. The signal and masker were gated on and off simultaneously with 25-ms cosine-squared ramps and a total duration of 800 ms (between the 0–V points).

B. Procedure

Masked threshold was measured using a two-interval, forced-choice adaptive procedure. The trials consisted of two 800-ms observation intervals that were separated by 500 ms. One interval contained the masker alone; the other interval contained masker plus signal. The listener's task was to indicate which interval contained the signal by pressing one of two buttons on a response box. Feedback was given after each response. The adaptive parameter was the level of the signal. At the beginning of each threshold run, signal level was set to a value well above the anticipated threshold. Signal level was varied according to a 3-down, 1-up rule which tracks the signal level that yields 79% correct responses (Levitt, 1971). The step size for the level changes was 5 dB up to the first reversal in signal level, 3 dB up to the second reversal, and 2 dB for the rest of the ten reversals that made up one threshold run. Each threshold estimate was the average of the last eight reversals of signal level. Threshold was measured three times in each condition; the data points and error bars in the figures show the means and standard errors of the three estimates. In some cases, the error bars are smaller than, and thus covered by, the symbols. The order in which conditions were measured was counterbalanced between the three threshold runs. The masker had a nominal overall level of 60 dB SPL. In order to reduce the salience of level cues and to encourage listeners to base their decisions on differences in sound quality, the level of the stimuli was randomized over a 10-dB range in every interval of every trial. The random level rove was expected to increase threshold in the conditions where signal and masker IRNs had the same delay, and thereby enhance the effect of sound quality differences associated with delay differences between signal and masker.

C. Listeners

A total of five listeners participated in the current experiment: authors KK and AN, as well as three students, who were paid for their services at an hourly rate. The listeners were between 20 and 32 years of age and had no history of hearing impairment or neurological disease.

III. RESULTS

The average threshold data for the detection of an IRN signal in an IRN masker are presented in Figs. 4 and 5 for filter cutoffs of 0.8 and 3.2 kHz, respectively. Threshold is expressed in terms of the signal-to-masker ratio, SMR. The parameter in both figures is the delay of the masker IRN; the number of iterations in the signal and masker IRNs was fixed at 16. In the upper panels of Figs. 4 and 5, threshold is plotted as a function of the absolute delay difference, *dd*, between the signal and masker IRNs; in the lower panels, threshold is plotted as a function of the relative delay difference, that is, delay difference normalized by the delay of the masker IRN, d_m. Each function in Figs. 4 and 5 is a ''time-interval (TI) masking pattern'' for a specific combination of masker delay and filter condition. When *dd* was 0 μs, the only detection cue was the loudness difference between the masker alone and masker plus signal. The levels of both stimuli were randomized over a 10-dB range, so threshold SMR in this condition was relatively high (about 2 dB). As soon as *dd* was increased to a few tens of *microseconds*,

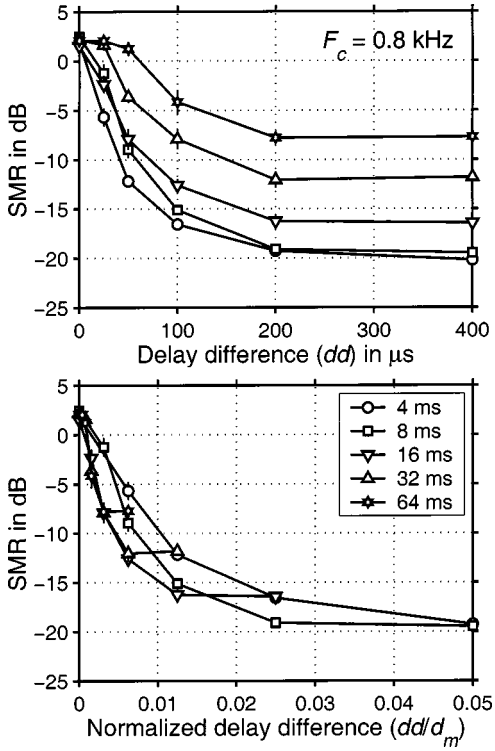

FIG. 4. Average masked threshold for detection of an IRN signal as a function of the delay difference, dd, between signal and masker IRNs; the parameter is the delay of the masker IRN (legend in lower panel). The filter cutoff was 0.8-kHz. The panels show the same thresholds but the abscissa in the lower panel is *normalized* delay difference, that is, dd divided by the masker delay, d_m. The error bars show the standard error of the threshold estimates.

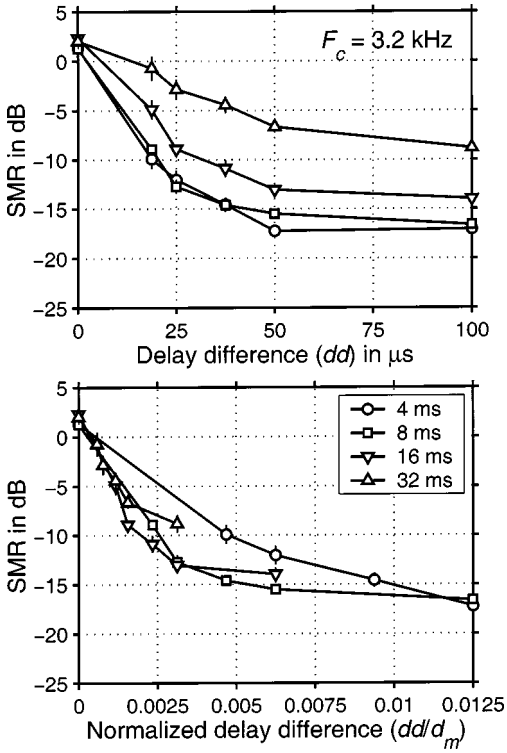

FIG. 5. Average masked threshold for detection of an IRN signal as a function of the delay difference, dd, between signal and masker IRNs; the parameter is the delay of the masker IRN (legend in lower panel). The filter cutoff was 3.2-kHz. The panels show the same thresholds but the abscissa in the lower panel is *normalized* delay difference, that is, dd divided by the masker delay, d_m. The error bars show the standard error of the threshold estimates.

however, there was a small difference in pitch strength between the two intervals, and threshold dropped by more than 10 dB in some conditions. Thereafter, threshold decreased at an ever slower rate, approaching an asymptote when dd was 200–400 μs for $F_c = 0.8$ kHz, and when dd was 50–100 μs for $F_c = 3.2$ kHz. The initial slope of the TI masking pattern was steeper, and the asymptotic level lower, for the smaller masker delays. The normalized making patterns in the lower panels of Figs. 4 and 5 show that, for d_m in the range 16–64 ms, the rate of the initial descent was inversely proportional to d_m. The masking patterns for $d_m = 4$ and 8 ms were similar in terms of their widths and asymptotic levels. The TI masking patterns were much narrower in the 3.2-kHz filter condition (Fig. 5) than in the 0.8-kHz filter condition (Fig. 4); note that in Fig. 5, the scale of the abscissas is stretched by a factor of 4 relative to Fig. 4.

Figure 6 shows the average TI masking patterns for a masker delay of 16 ms and filter cutoff frequencies ranging from 0.4 to 6.4 kHz. As before, threshold is expressed in terms of the SMR and plotted as a function of dd. Figure 6 shows that the width of the TI masking pattern *decreased* as the filter cutoff frequency increased from 0.4 to 6.4 kHz.

Overall, the TI masking patterns reveal remarkable temporal resolution. When the delay of the masker IRN is short (4–8 ms) or when the filter cutoff is high (3.2–6.4 kHz), a difference of a few tens of microseconds between the delays of the signal and masker is sufficient to produce an appreciable release from masking. Indeed, in the highest filter condition ($F_c = 6.4$ kHz), a delay difference of just 12.5 μs produced a masking release of about 7 dB relative to the condition where dd is 0 μs. The size of the delay difference producing a significant masking release, and the fact that the width of the TI masking pattern decreases with *increasing* filter cutoff means that it is unlikely that the present data could be explained by spectral differences between the stimuli when one considers the spectral resolution of the internal excitation pattern. In most of the experimental conditions, the lower filter cutoff, F_c, was greater than 12 times

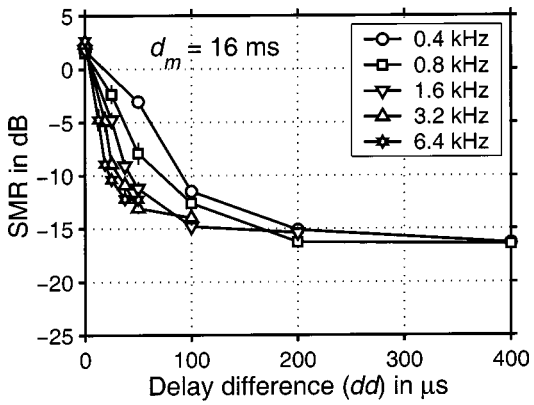

FIG. 6. Average threshold for detecting an IRN signal in the presence of a 16-ms IRN masker as a function of delay difference, dd; the parameter is the low-frequency cutoff, F_c, of the filter passband. The error bars show the standard error of the threshold estimates.

the reciprocal of the IRN delay, so the excitation patterns of the IRNs were like those for random noise filtered to the same frequency region (Krumbholz *et al.*, 2001). In some of the conditions, however, the excitation patterns of the IRNs would be expected to exhibit a few resolved harmonics by the relatively lenient criterion of Shackleton and Carlyon (1994). The fact that there are no discontinuities in the TI masking patterns suggests that resolved harmonics did not play a part in the processing, and that the masking release is associated with a single mechanism operating on the differences in the monaural *temporal* structure between the masker alone and the masker plus signal.

IV. SIMULATION OF TIME-INTERVAL MASKING PATTERNS WITH A TIME-INTERVAL MODEL OF AUDITORY PROCESSING

A computational version of Licklider's (1951) autocorrelation model of pitch was used to determine whether a simple time-interval model of auditory temporal processing could account for the basic effects in the present data. Specifically, we used a modified version of the auditory image model (AIM) presented in Patterson *et al.* (1995), implemented with the DSAM/AMS software package and MATLAB. The auditory model consisted of (i) a second-order bandpass filter with cutoffs at 0.45 and 8.5 kHz; (ii) a 60-channel gammatone filterbank (gtf) with center frequencies between 0.1 and 8 kHz, evenly distributed on the ERB scale; (iii) half-wave rectification, square-root compression, and fourth-order low-pass filtering with a 0.8-kHz cutoff (hcl) in each frequency channel; (iv) a channel-by-channel time-interval analysis performed by strobed temporal integration (sti), which is similar to autocorrelation. This version of AIM is referred to as the $gtf/hcl/sti$ model, and a detailed description of it can be found in Sec. IV of the paper by Krumbholz *et al.* (2001), who used the model to explain the masking asymmetry between noise and IRN.

The stages of the $gtf/hcl/sti$ model prior to the time-interval analysis simulate peripheral auditory processing up to the level of the auditory nerve: the initial bandpass filter represents the operation of the middle ear, the gammatone filterbank (gtf) simulates the spectral analysis performed by the cochlea, the half-wave rectification, compression, and low-pass filtering (hcl) simulate the transformation of the basilar-membrane response to the neural activity pattern (NAP) flowing up the auditory nerve. Strobed temporal integration (sti) produces a time-interval histogram of the neural activity in each tonotopic channel, which is similar to the all-order time-interval histograms produced by autocorrelation. However, sti has the advantage of preserving short-term temporal asymmetry that listeners hear (Patterson and Irino, 1998), and at the same time, it requires far less computation. The tonotopic array of time-interval histograms is referred to as an auditory image (AI), and the structure in the upper panel of Fig. 1 is the AI produced by the model in response to an IRN with an 8-ms delay.

The decision statistic was derived from the ''summary AI,'' produced by averaging the AI across frequency (as in the lower panel of Fig. 1). The summary image was normalized to the value at 0 ms which is a measure of the overall level of the stimulus. In the experiment, the roving level paradigm discouraged the use of overall level as a detection cue. The lower limit of pitch for complex sounds occurs when the period exceeds about 33 ms (Krumbholz *et al.*, 2000). Accordingly, the summary AI was limited to time intervals less than 35 ms, and the modeling was restricted to conditions where the masker delay was less than 35 ms. The stimuli in the simulations were generated using the same software as in the experiment. The maskers were set to a fixed level of 60 dB SPL. Signals were generated with SMRs ranging from -27 to 14 dB in 3-dB steps and then added to the maskers as in the experiment. The stimuli had a duration of 810 ms. For each stimulus condition, summary AIs for 15 different random samples of the stimuli were averaged to produce the summary AI used to calculate the decision statistic. Moreover, the summary AI for a given stimulus sample was itself an average of 20 summary AIs generated at 35-ms intervals throughout the stimulus duration; the first summary AI was calculated 105 ms after stimulus onset, at which point the model response had reached steady state.

A. Decision measure, D^2

Listeners reported using different detection cues in different experimental conditions. When the delay difference between the signal and masker IRNs was large, or the masker delay was small, the signal IRN produced a pitch that was distinct from that of the masker IRN, and listeners based their judgments on whether the stimulus contained one or two pitches. For small delay differences, or large masker delays, the pitches of the signal and masker merged, and listeners based their judgments on the strength, or salience, of the pitch. In the simulations, we used a Euclidean distance, D^2 (Meddis and Hewitt, 1991a) to measure the differences between the summary AI for the signal plus masker and that of the masker alone. D^2 is the integral of the squared difference between the summary AIs, so it includes differences at all time intervals within the summary images. For each experimental condition (i.e., each combination of F_c, d, and dd), D^2 was calculated as a function of SMR. Threshold was defined as the SMR at which D^2 reached a criterion level, C, and this criterion was the main parameter in the fitting process. All of the conditions of the experiment were fitted simultaneously, with a fixed value of C, which was then varied to find the value that minimized the root-mean-square (rms) deviation between the simulated and observed thresholds. The version of AIM in Krumbholz *et al.* (2001) produced a good fit to the data without modification. At the same time, it appeared that the fit could be improved by (i) modifying the *time-interval weighting function* associated with the lower limit of pitch and/or (ii) limiting the *order* of the peaks that contribute to D^2.

B. Time-interval weighting

There is a progressive reduction in pitch strength when approaching the lower limit of pitch. In the default version of AIM, the reduction in pitch strength and the lower limit of pitch are implemented with a linear weighting function that decreases from unity when the time interval is 0 ms, to zero when it is 40 ms. It is this which produces the progressive

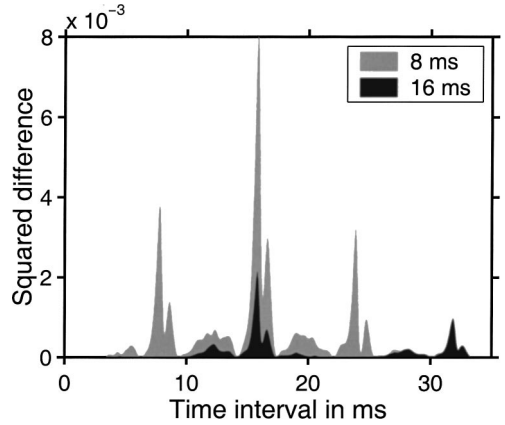

FIG. 7. Squared difference between the summary AIs for masker-plus-signal and masker alone, plotted as a function of time interval. In the case of the gray area, the masker and the signal had delays of either 8 and 8.1 ms (gray shading), or 16 and 16.1 ms (black shading).

reduction of activity level across time interval in the AI of the upper panel in Fig. 1, and thus, the slope in the summary AI of the lower panel in Fig. 1. This linear time-interval weighting function in turn produces an approximately linear decrease in the weight of the squared difference between the summary images for masker and masker plus signal across time interval. The black foreground in Fig. 7 shows the squared difference for a 16-ms masker IRN combined with a 16.1-ms signal IRN plotted as a function of time interval; the SMR is 0 dB and F_c is 0.8 kHz. The squared difference is concentrated at the signal delay, the masker delay, and their integer multiples. The gray background shows the squared difference for an 8-ms masker with an 8.1-ms signal. The first peak in the squared difference function for the 8-ms masker is about twice as large as the first peak in the difference function for the 16-ms masker because of the time-interval weighting function.

The relative weight of the time intervals was varied with a power function that reduces to the linear weighting function when the exponent, α, is unity. The weighting function is $w(ti)=1-(ti/40)^{\alpha}$, where ti is time interval and α is a parameter that determines the degree of nonlinearity. Figure 8 shows the time-interval weighting function for values of α from 0.25 to 8; they all decrease from unity at $ti=0$ ms to zero at $ti=40$ ms, but the relative weight of the short and long intervals varies considerably. When $\alpha<1$, the function is concave and longer time intervals contribute relatively less to the decision statistic; when $\alpha>1$, the function is convex and the longer time intervals contribute relatively more. The vertical, dotted line in Fig. 8 marks the time interval (35 ms) where the summary images were truncated.

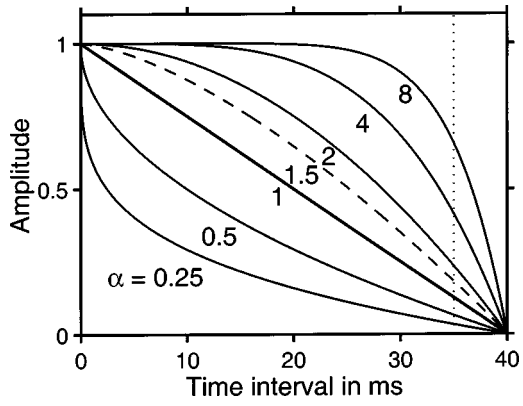

FIG. 8. Time-interval weighting functions for values of α from 0.25 to 8, plotted as a function of time interval; the vertical dotted line marks the time interval where the AI was truncated (35 ms). The function for $\alpha=1.5$, which produced the best fit to the data, is shown by the dashed line.

C. Peak-order limit

Peaks in the summary AI at multiples of a basic delay are referred to as ''higher-order'' peaks (Yost, 1996a,b). The presence of higher-order peaks is due to the fact that autocorrelation and sti include time intervals between nonadjacent peaks in the neural activity pattern, as well as first-order intervals. Kaernbach and colleagues have pointed out that the order of the time intervals that contribute to a given peak in the AI is usually greater, or equal, to the peak order (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001).

Figure 7 shows that the squared difference associated with the second-order peak (near 16 ms) can be greater than that associated with the first-order peak (near 8 ms), despite the linear time-interval weighting in the underlying summary images. The reason is that the separation between the peaks of the signal and masker increases in proportion to peak order. So, in the current example, the delay difference of 100 μs between the 8-ms masker and the 8.1-ms signal produces a peak separation of 100 μs for the first-order peaks, but a peak separation of 200 μs for the second-order peaks, and the larger peak separation leads to a larger squared difference for the second-order peaks. It is also the case that stimuli with shorter delays produce more higher-order peaks within the time-range of the summary images than stimuli with longer delays. In the version of the model proposed by Krumbholz *et al.* (2001) all peaks with delays less than 35 ms contribute to the D^2 measure, and the higher-order peaks enable the model to simulate the general reduction in threshold that occurs as delay *decreases* in the asymmetry of masking experiment.

There are other studies, however, where listeners appear to make little, if any, use of higher-order peaks, indicating that higher-order peaks may affect the masking properties of the stimulus more, and in a somewhat different way than they do its pitch and pitch strength. Yost (1996a) and Yost *et al.* (1996) found that listeners had difficulty discriminating IRN in which the delayed copy is added back to the ''original'' noise (IRNO) from IRN in which the delayed copy was added to the ''same'' noise (IRNS). The first-order peaks in these stimuli have the same height but the higher-order peaks in IRNO are larger, and so the IRNO might have been expected to have a stronger pitch. It appears that higher-order peaks have no influence on the pitch or pitch strength, as long as they are not *larger* than the first peak (Yost, 1996b, 1997). In IRN, peak height usually decreases with increasing peak order. However, when IRN is produced with a negative gain factor, g, the first-order peak is inverted and it is replaced by its adjacent smaller side peaks that dominate the perception when the number of iterations, n, is small. When

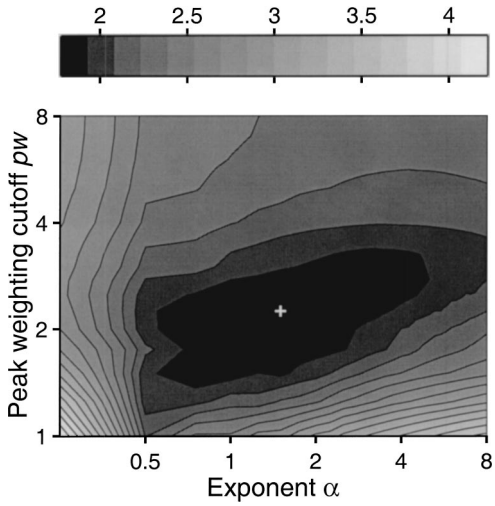

FIG. 9. Minimum root-mean-square (rms) deviation between simulated and observed thresholds, as a function of α (the parameter of the time-interval weighting function), and pw (the parameter of the peak-weighting function); the shading is black for smaller deviations and white for larger deviations (scale bar across the top of the figure). The small white cross marks the combination α of and pw for which the rms deviation is minimum (1.85 dB).

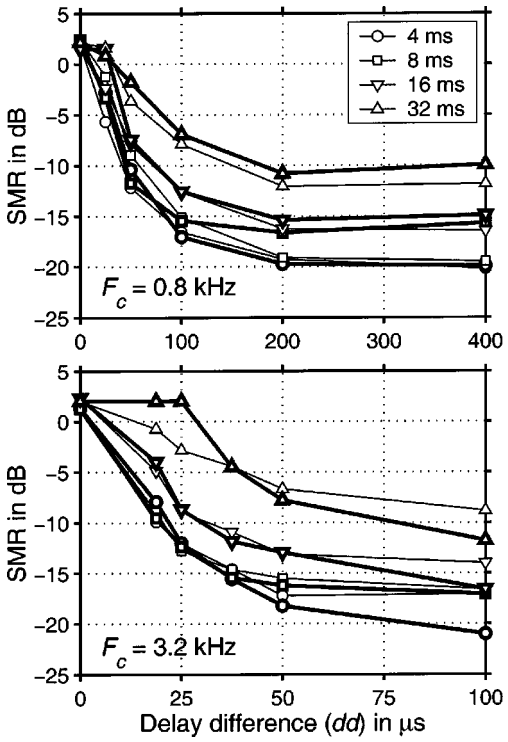

FIG. 10. Simulated thresholds (bold symbols and lines) and observed thresholds (thin symbols and lines) as a function of delay difference; the filter cutoff is 0.8 kHz (upper panel) or the 3.2-kHz (lower panel). The parameter is the delay of the masker IRN.

n increases, however, these side peaks can be smaller than the second-order peak at twice the delay, which is not inverted when g is negative. In this case, the pitch corresponds to the second-order peak (Yost, 1996b, 1997).

The data from the current study appear to be ambiguous with regard to the use of higher-order peaks. On the one hand, the width of the TI masking function decreased with delay for maskers with delays between 64 and 8 ms, which suggests that higher-order peaks are used. On the other hand, there was no further decrease for masker delays below 8 ms, which suggests that there is some relatively low limit on the order of the peaks that can be used. Accordingly, AIM was modified to include a "peak-weighting" function, $p(ti)$, that reduced the contribution of time intervals from unity to zero between one order of peak and the next. That is, for a given peak order, pw, the function decreased linearly from unity at $ti=pw\times d_m$ to zero at $ti=(pw+1)\times d_m$, where d_m is the masker delay. The upper limit on pw is effectively 8 since the shortest delay in the current experiment was 4 ms, and its eighth-order peak would occur at 32 ms.

D. Modeling results

The simulation of the TI masking patterns with the modified version of AIM was performed repeatedly using time-interval weighting functions with α ranging from 0.25 to 8 in steps of 0.25, and peak-weighting functions with pw ranging from 1 and 8 in steps of 0.25. For each combination of α and pw, the fit was evaluated in terms of the rms deviation between the simulated and observed thresholds. Figure 9 shows the rms deviation between simulated and observed threshold plotted with α on the abscissa and pw on the ordinate (white representing large, black representing small, deviations). The figure shows that there is a local minimum in the region where pw is 1.5 to 2.5 and α is 1 to 3. The minimum rms deviation is just 1.85 dB and it occurs at $pw=2.25$ and $\alpha=1.5$, marked by a white cross. At this point, the time-interval weighting function is slightly convex (see the dashed line in Fig. 8), and second-order peaks in the summary images contribute to the decision measure, but higher-order peaks do not. The TI masking patterns produced by this version of the model are presented by the bold symbols and lines in Figs. 10 and 11; the thin symbols and lines show the observed thresholds.

Overall, the model produces a surprisingly good fit to the data. The upper and lower panels of Fig. 10 show the data for the 0.8- and the 3.2-kHz filter conditions, respectively; the parameter is the masker delay (legend in upper panel). The initial descent of the simulated TI masking pattern is very similar to that of the data, except in the case of the 32-ms masker in the 3.2-kHz filter condition where the tip of the simulated masking pattern is too broad. The width of the TI masking pattern increases with the delay of the masker as in the data. The asymptotic level at the larger delay differences increases with increasing masker delay as in the data, although the absolute values at the largest delay differences are a little too low when the filter cutoff is 3.2 kHz.

Figure 11 shows the results for the 16-ms masker; in this case, the parameter is filter cutoff, F_c (legend in upper panel). The width of the TI masking pattern *decreases* with increasing filter cutoff (upper panel of Fig. 11), which suggests that temporal resolution (as measured by IRN masking) is affected by the duration of the impulse response of the auditory filter, which is roughly inversely proportional to filter center frequency. In the model, the duration of the auditory filter impulse response determines the width of the ver-

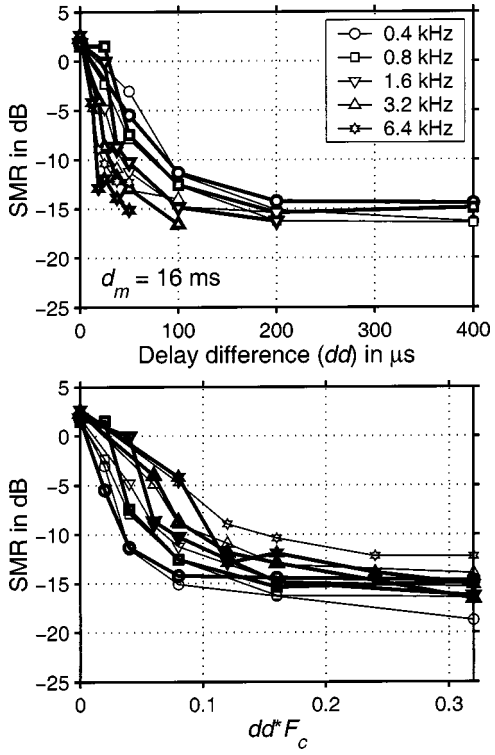

FIG. 11. Simulated thresholds (bold) and observed thresholds (thin) for the 16-ms masker as a function of delay difference; the parameter is filter cutoff, F_c. The abscissa is absolute dd in the upper panel and dd multiplied by F_c in the lower panel.

tical ridges in the AI of IRN, just as the frequency resolution of the auditory filters determines the width of the excitation pattern of a sinusoid. Consequently, the width of the ridges decreases with increasing filter frequency (see the upper panel in Fig. 1). This, in turn, explains the model's ability to account for the decrease in the width of the TI masking pattern with increasing filter cutoff frequency. In the lower panel of Fig. 11, threshold is plotted as a function of dd times F_c. This panel shows that the width of the simulated masking pattern, like that of the observed masking pattern, does not decrease in proportion to filter cutoff. The panel also shows that the model is unable to explain why the asymptotic threshold level increases with increasing filter cutoff (lower panel of Fig. 11).

V. DISCUSSION

The previous section showed that a simple time-interval model of auditory processing can produce a remarkably good fit to the present IRN masking data. The model explains the absolute width of the TI masking pattern and it accounts for the effects of masker delay and frequency region. The modeling results indicate that first- and second-order peaks in the time-interval histograms contribute to the masking release but that higher-order peaks do not, and that contribution of longer time intervals decreases approximately linearly with time interval. In the highest filter condition ($F_c = 6.4$ kHz), a delay difference of just 12.5 μs between the signal and masker IRNs produced a significant masking release (Fig. 6). The masker delay was 16 ms, and so the second-order peak at 32 ms was close to the lower limit of pitch. As a result, it would not be expected to contribute much to the detection of the signal (see the black area in Fig. 7). Thus, to the extent that the model is appropriate, the present data suggest that the monaural system processes time-interval information with a resolution of 10–20 microseconds, which is similar to the resolution of the binaural system.

The temporal resolution revealed by the TI masking pattern (tens of microseconds) is two orders of magnitude greater than the values obtained with traditional measures like gap detection or modulation detection (Forrest and Green, 1987). The discrepancy suggests that the thresholds in the TI masking pattern are based on a different perceptual cue and a different neural code. The cue in the gap and modulation experiments is rapid fluctuation in the temporal envelope of the sound, whereas the cue in the TI masking experiment is stationary—a reduction in pitch strength or precision of the perception. In models of envelope processing, it is assumed that the cue is a temporal fluctuation, and that it is represented by a *temporal* code up to the highest processing levels. In time-interval models of auditory processing, it is assumed that the cues are based on stable time-interval histograms with relatively long time constants that may involve a *spatial* representation of the time-interval information in the monaural stimulus (e.g., Patterson *et al.*, 1995).

Rate discrimination measurements yield estimates of temporal acuity that are closer to the microsecond resolution reported in the present study than the estimates from traditional measurements of temporal resolution. Rate discrimination threshold for sinusoidally amplitude-modulated (SAM) noise is about 6% for base rates up to about 300 Hz (Hanna, 1992). A 6% difference in rate at 300 Hz corresponds to a difference of 200 μs in the repetition period of the modulator. Kaernbach and Bering (2001) report that listeners could discriminate the rate of periodic click trains with an acuity of about 1.25% when the rate was 250 Hz and the click trains were high-pass filtered at 4.5 kHz to remove spectrally resolved harmonics. A 1.25% difference in rate at 250 Hz corresponds to a difference of 50 μs in the period of the sound, which is not much larger than the value suggested by the data from the current experiments. However, Kaernbach and Bering's results might be overly optimistic; using essentially the same stimuli, Cullen and Long (1986) reported rate discrimination thresholds that were considerably larger than those reported by Kaernbach and Bering. Discrimination of the frequency of a sinusoid can be interpreted to indicate temporal acuity that surpasses even the 12.5-μs resolution suggested by the current data. At 1 kHz, frequency discrimination threshold is about 0.2%. So, if changes in the frequency of sinusoids are mainly mediated by temporal cues in the phase-locking range, threshold at 1 kHz would correspond to temporal acuity on the order of 2 μs. However, changing the frequency of a sinusoid produces a change in the place of the excitation along the cochlear partition, and so this measure is probably confounded by spectral cues.

It is interesting to compare the temporal resolution derived from TI masking patterns with the frequency resolution derived from spectral masking patterns. The resolution with which temporal and spectral information is represented in the

auditory system might be expected to reflect their relative importance for the formation of perceptions. When IRN threshold is plotted as a function of the relative delay difference as in the lower panels of Figs. 4 and 5, the TI masking pattern is found to be surprisingly narrow. In the 0.8-kHz condition (Fig. 4), the width of the masking patterns is less than 1% of the masker delay, and the percentage is even smaller in the higher frequency region (Fig. 5). In percentage terms, spectral masking patterns are more than an order of magnitude wider than TI masking patterns. The relative width of spectral masking patterns is determined by the width of the excitation pattern produced by the narrowband masker, which is approximately 10%–15% of the masker frequency.

The paradigm used to generate and analyze the TI masking patterns is reminiscent of the binaural laterality paradigm. In that case, temporal resolution is measured in terms of the listener's ability to detect minute differences between the neural patterns produced by two similar sounds presented concurrently, but to separate ears. In the monaural case, the two stimuli are combined before entering the auditory system, and temporal resolution is measured in terms of the listener's ability to detect minute temporal differences in the neural pattern of the masker due to the presence of the signal. The two paradigms both yield resolution estimates on the order of 10 μs and time-interval models appear to be a reasonable basis for explaining the main effects in both domains.

In most of the current models of monaural temporal processing, the transformation from the time domain to the time-interval domain is modeled by an autocorrelation process; they effectively compute an all-order interval histogram for each channel of the tonotopic array produced in the cochlea. Although autocorrelation models can explain many perceptual phenomena in general terms (Meddis and Hewitt, 1991a,b, 1992; Meddis and O'Mard, 1997), they have been shown to be inadequate in several respects: (i) Time-interval histograms produced by autocorrelation are symmetric in the time-interval dimension. The autocorrelogram does not preserve temporal asymmetries in the stimulus which listeners may well perceive (Patterson and Irino, 1998). (ii) The autocorrelation process places no particular limit on the *length* of the time intervals that are represented in the histograms, whereas psychophysical evidence suggests that human listeners can accurately process time intervals up to 33 ms but no longer (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). (iii) In autocorrelation models, there is no limit on the *order* of the time intervals that appear in the histograms. Kaernbach and colleagues (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001) constructed click trains exhibiting temporal regularity in the higher-order interclick intervals but not in the first-order intervals, and concluded that ''the mechanism operates more easily on first-order...temporal regularities'' (Kaernbach and Bering, 2001, p. 1047). Moreover, pitch-strength studies suggest that higher-order peaks in the time-interval histograms, which are composed of higher-order time intervals, do not contribute to pitch strength (Yost, 1996a; Yost *et al.*, 1996; Patterson *et al.*, 2000). (iv) There are experiments involving discrimination of IRNs with positive and negative gain which show that the second-order peak contributes to the discrimination when it is similar in size to the first-order peak (Yost, 1996b, 1997). The current study corroborates the use of second-order peaks but indicates that peaks beyond second order are not used.

In the current simulation, the time-interval analysis was performed by strobed temporal integration (sti), instead of autocorrelation; sti measures time intervals from local peaks in the neural activity and, as a result, it is sensitive to temporal asymmetry in the stimulus. This is a fundamental difference between sti and autocorrelation. The contribution of higher-order peaks in the time-interval histograms was limited by multiplying the histograms with a peak-order weighting function, and the longer time intervals were reduced by a time-interval weighting function to explain the lower limit of melodic pitch. These modifications could be applied to the autocorrelogram and/or summary autocorrelogram in a similar way, but they would still represent aspects of autocorrelation that must be modified and/or limited if it is to explain pitch perception in detail.

VI. CONCLUSIONS

The current study indicates that listeners can discriminate a coherent sound, consisting of a single IRN, from a less coherent sound, consisting of two IRNs, when the delays of the two IRNs differ by only a few tens of microseconds. The magnitude of the difference and the fact that the difference decreases when the stimuli are restricted to higher frequency regions mean that it is unlikely that the discrimination is based on spectral cues. Rather, the discrimination appears to be based on small temporal differences in the monaural neural patterns produced by masker and masker plus signal.

The just-discriminable difference in IRN delay decreased with masker delay from 64 down to 8 ms, but it did not decrease further as delay decreased to 4 ms. This suggests that higher-order peaks in the time-interval histograms contributed to the resolution at the shorter masker delays, but that there is a limit on the order of the peaks that can be used. Signals with small delay differences were more readily detectable in the higher filter conditions, indicating that the resolution is affected by the duration of the impulse response of the auditory filters.

A computational model of time-interval processing (Patterson *et al.*, 2000; Krumbholz *et al.*, 2001) was modified to limit the contribution of longer time intervals and higher-order peaks, and in this modified form it could explain the main effects in the data accurately.

ACKNOWLEDGMENT

Research supported by the Medical Research Council (Grant No. G9901257).

Cullen, Jr., J. K., and Long, G. (**1986**). ''Rate discrimination of high-pass filtered pulse trains,'' J. Acoust. Soc. Am. **79**, 114–119.

Durlach, N. I., and Colburn, H. S. (**1978**). ''Binaural phenomena,'' in *Hearing, Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. IV.

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.

Hanna, T. E. (**1992**). ''Discrimination and identification of modulation rate,'' J. Acoust. Soc. Am. **91**, 2122–2128.
Handel, S., and Patterson, R. D. (**2000**). ''The perceptual tone/noise ratio of merged, iterated rippled noises with octave, harmonic and nonharmonic delay ratios,'' J. Acoust. Soc. Am. **108**, 692–695.
Irvine, D. R. F. (**1992**). ''Physiology of the auditory brain stem,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 153–231.
Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.
Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.
Kaernbach, C., and Bering, C. (**2001**). ''Exploring the temporal mechanism involved in the pitch of unresolved harmonics,'' J. Acoust. Soc. Am. **110**, 1039–1048.
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.
Krumbholz, K., Patterson, R. D., and Nobbe, A. (**2001**). ''Asymmetry of masking between noise and iterated rippled noise: Evidence for time-interval processing in the auditory system,'' J. Acoust. Soc. Am. **110**, 2096–2107.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.
Meddis, R., and Hewitt, M. J. (**1991a**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. J. (**1991b**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.
Meddis, R., and Hewitt, M. J. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Oertel, D. (**1997**). ''Encoding of timing in brain stem auditory nuclei of vertebrates,'' Neuron **19**, 959–962.
Patterson, R. D. (**1994**). ''The sound of a sinusoid: Time-interval models,'' J. Acoust. Soc. Am. **96**, 1419–1428.
Patterson, R. D., and Datta, A. J. (**1996**). ''The detection of iterated rippled noise (IRN) masked by IRN,'' Br. J. Audiol. **30**, 148.
Patterson, R. D., and Irino, T. (**1998**). ''Modeling temporal asymmetry in the auditory system,'' J. Acoust. Soc. Am. **104**, 2967–2979.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception, Proceedings of the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). ''The relative strength of tone and noise components of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3286–3294.
Patterson, R. D., Yost, W. A., Handel, S., and Datta, A. J. (**2000**). ''The perceptual tone/noise ratio of merged iterated rippled noises,'' J. Acoust. Soc. Am. **107**, 1578–1588.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). ''The lower limit of melodic pitch,'' J. Acoust. Soc. Am. **109**, 2074–2084.
Rouilly, E., deRibaupierre, Y., and deRibaupierre, F. (**1979**). ''Phase-locked responses to low frequency tones in the medial geniculate body,'' Hear. Res. **1**, 213–226.
Shackleton, T. M. and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529-3540.
Slaney, M., and Lyon, R. F. (**1990**). ''A perceptual pitch detector,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, New Mexico* (IEEE, New York), pp. 357–360.
Yost, W. A. (**1996a**). ''The pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3329–3335.
Yost, W. A. (**1996b**). ''The pitch of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 511–518.
Yost, W. A. (**1997**). ''Pitch strength of iterated rippled noise when the pitch is ambiguous,'' J. Acoust. Soc. Am. **101**, 1644–1648.
Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.