Melody recognition using three types of dichotic-pitch stimulus

Michael A. Akeroyd[a)]
Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030 and Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, United Kingdom

Brian C. J. Moore and Geoffrey A. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 5 January 2001; revised 21 May 2001; accepted 8 June 2001)

The recognition of 10 different 16-note melodies, constructed using either dichotic-pitch stimuli or diotic pure-tone stimuli, was measured. The dichotic pitches were created by placing a frequency-dependent transition in the interaural phase of a noise burst. Three different configurations for the transition were used in order to give Huggins pitch, binaural-edge pitch, and binaural-coherence-edge pitch. Forty-nine inexperienced listeners participated. The melodies evoked by the dichotic stimuli were consistently identified well in the first block of trials, indicating that the sensation of dichotic pitch was relatively immediate and did not require prolonged listening experience. There were only small improvements across blocks of trials. The mean scores were 97% (pure tones), 93% (Huggins pitch), 89% (binaural-edge pitch), and 77% (binaural-coherence-edge pitch). All pairwise differences were statistically significant, indicating that Huggins pitch was the most salient of the dichotic pitches and binaural-coherence-edge pitch was weakest. To account for these differences in salience, a simulation of lateral inhibition was applied to the recovered spectrum generated by the modified equalization cancellation model [J. F. Culling, A. Q. Summerfield, and D. H. Marshall, J. Acoust. Soc. Am. **103**, 3509–3526 (1998)]. The height of the peak in the resulting ''edge-enhanced'' recovered spectrum reflected the relative strength of the different dichotic pitches.
 [DOI: 10.1121/1.1390336]

PACS numbers: 43.66.Pn, 43.66.Hg, 43.66.Ba, 43.75.Cd [SPB]

I. INTRODUCTION

Since Cramer and Huggins' (1958) pioneering research, it has been known that pitch sensations can be created by the binaural interaction of noise stimuli. These ''dichotic pitches'' are analogous to the visual objects that can be seen in random-dot stereograms (e.g., Julesz, 1971); the stimulus at each ear gives no pitch sensation, but, when presented binaurally, disparities between the two ears lead to the perception of pitch. We report below a study of the recognition of melodies produced using three types of dichotic-pitch stimulus and also using pure tones.

The types of dichotic-pitch stimulus used were the ''Huggins pitch'' (Cramer and Huggins, 1958), the ''binaural-edge pitch'' (Klein and Hartmann, 1981), and the ''binaural-coherence-edge pitch'' (Hartmann, 1984; Hartmann and McMillon, 2001). For all of the stimuli, the presence of a single frequency-dependent transition in the interaural phase of a broadband noise gives rise to the perception of pitch. This percept is similar to that of a pure, although faint, tone. These transitions are illustrated schematically in Fig. 1. In a Huggins-pitch stimulus (left panel) the interaural phase changes progressively from 0 to 2π (equivalent to 0) radians across a narrow frequency region. The width of this region has sometimes been varied in previous studies; here, we use a fixed value of 16% of the center frequency of the transition.[1] In a binaural-edge-pitch stimulus (middle panel) the interaural phase changes abruptly from 0 to π radians. In a binaural-coherence-edge-pitch stimulus (right panel) the interaural phase changes abruptly from 0 radians to a random value. Although variations of these dichotic-pitch stimuli have been reported, differing in the interaural phase of the carrier noise in the case of Huggins pitch or the spectral direction of the transition in interaural phase in the case of the two edge pitches, we deal here only with the ''prototypical'' variations shown in Fig. 1.

The value of the perceived pitch has been measured by asking listeners to match the pitch using a pure tone of adjustable frequency. For Huggins pitch, the matching frequency is commonly found to be equal to the center frequency of the transition (e.g., Culling *et al.*, 1998). For binaural-edge pitch, Klein and Hartmann (1981) found that the distribution of matching frequencies was bimodal, with one peak slightly above and the other peak slightly below the transition frequency. Subsequent measurements conflict with this result, however, indicating a unimodal distribution with a peak centered on the transition frequency (Frijns *et al.*, 1986; Culling *et al.*, 1998); at present there is no agreement as to whether the distribution is bimodal or unimodal. For the particular variation of binaural-coherence-edge pitch used here, Hartmann and McMillon (2001) found a unimodal distribution, but with a peak placed approximately 5%–10% above the transition frequency.

Theories differ as to the nature of the binaural processing that creates the sensation of pitch from the transition in interaural phase (e.g., Licklider, 1959; Durlach, 1962; Klein

[a)]Author to whom correspondence should be addressed; electronic mail: maa@biols.susx.ac.uk

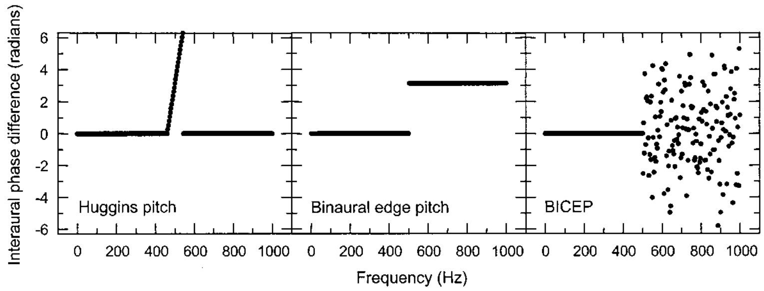

FIG. 1. Interaural phase difference plotted as a function of frequency for the three types of dichotic-pitch stimulus. The transition frequency was 500 Hz in each case. Each dot represents the interaural phase difference of one frequency component (spacing=3.3 Hz). Note that the interaural phase difference for the binaural-coherence-edge pitch (abbreviated to "BICEP") stimulus was random above the transition frequency and thus would have differed in each of the experimental stimuli from that plotted. The interaural amplitude difference of all the frequency components was zero, apart from the frequency components higher than the transition frequency in the BICEP stimulus, for which the interaural amplitude difference was random.

and Hartmann, 1981; Raatgever and Bilsen, 1986; Culling *et al.*, 1998; Akeroyd and Summerfield, 1999). The two most developed models of dichotic pitch are the central-spectrum model (e.g., Raatgever and Bilsen, 1986; Frijns *et al.*, 1986) and the modified-equalization-cancellation model (e.g., Culling *et al.*, 1998). In both of these models, the sensation of pitch is assumed to be produced by a peak in a spectrum calculated on the basis of binaural cues, but the method of determining this "binaural spectrum" differs across models. The sensation of a dichotic pitch is assumed to be due to the presence of a peak in the binaural spectrum, by analogy with the peak in a monaural excitation pattern created by a pure-tone stimulus. As the purpose of this study was not to test experimentally the different models of dichotic-pitch sensation, for simplicity we limit the present discussion to the modified-equalization-cancellation model.

Two operations are fundamental to this model. First, the signals at each ear are passed through an array of auditory filters. Second, processes of equalization (of overall differences in interaural time and intensity) and then cancellation are applied to each filter output, with the goal of minimizing the power of the remainder after cancellation (Durlach, 1972). The minimum value of the power is termed the "residual activation." The model allows *independent* time delays to be used in each filter channel when applying the time equalization. The result is a spectrum of residual activation versus frequency and is termed the "recovered spectrum." Figure 2 shows example recovered spectra for the three types of dichotic-pitch stimuli used here, each with a transition frequency of 500 Hz (the computational details of the calculation of these spectra are reported in Sec. IV below). For both the Huggins-pitch stimulus (solid line) and the binaural-edge-pitch stimulus (dotted line), there is a single peak at the transition frequency. For the binaural-coherence-edge-pitch stimulus (dashed line), there is instead a high-pass recovered spectrum with a sloping edge. It is presumed that a process of lateral inhibition is applied to the edge in the recovered spectrum, so creating a single peak placed slightly above the transition frequency.

It seems reasonable to assume that the strength of the pitch sensation is related to the height of the peak in the recovered spectrum. If so, binaural-edge pitch should be slightly weaker than the version of Huggins pitch used here. (Culling *et al.*, 1998, Fig. 4, showed that the height of the peak in the recovered spectrum of Huggins pitch is dependent upon the bandwidth of the transition in interaural phase. The present simulations were based on the same bandwidth, 16%, as used in our experiment.) A quantitative prediction of the strength of binaural-coherence-edge pitch cannot be made without a function representing the amount of lateral inhibition. It would seem reasonable, nevertheless, to expect that a peak introduced by this extra stage of processing would not be as large as a peak directly present in the recovered spectrum for the Huggins-pitch and binaural-edge-pitch stimuli. Hence, binaural-coherence-edge pitch should be weaker than the other two pitches.

One way of estimating the strength of a dichotic pitch is by adjusting the signal-to-noise ratio of a pure tone presented

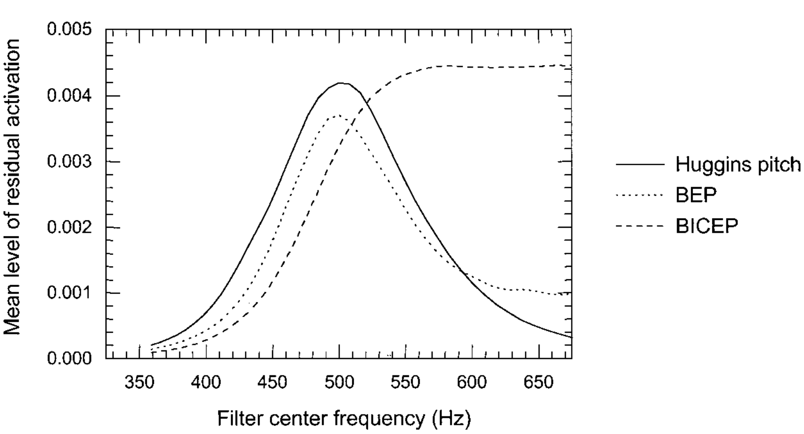

FIG. 2. Recovered spectra for a Huggins-pitch stimulus (solid line), binaural-edge-pitch stimulus (dotted line) and binaural-coherence-edge-pitch stimulus (dashed line). The calculations underlying the spectra are summarized in Sec. IV.

TABLE I. The frequencies of each of the 16 notes forming the 10 melodies used in the experiment.

Name of melody	Frequency of Note															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Au Clair de la Lune	C_5	C_5	C_5	D_5	E_5	E_5	D_5	D_5	C_5	E_5	D_5	D_5	C_5	C_5	C_5	C_5
Frere Jacques	C_5	D_5	E_5	C_5	C_5	D_5	E_5	C_5	E_5	F_5	G_5	G_5	E_5	F_5	G_5	G_5
Twinkle, Twinkle Little Star	C_5	C_5	G_5	G_5	A_5	A_5	G_5	G_5	F_5	F_5	E_5	E_5	D_5	D_5	C_5	C_5
This Old Man	F_5	D_5	F_5	F_5	F_5	D_5	F_5	F_5	G_5	F_5	$D_5^\#$	D_5	C_5	D_5	$D_5^\#$	$D_5^\#$
God Rest Ye Merry Gentlemen	D_5	D_5	A_5	A_5	G_5	F_5	E_5	D_5	C_5	D_5	E_5	F_5	G_5	A_5	A_5	A_5
Yankee Doodle	F_5	F_5	G_5	A_5	F_5	A_5	G_5	C_5	F_5	F_5	G_5	A_5	F_5	F_5	E_5	E_5
Good King Wenceslas	F_5	F_5	F_5	G_5	F_5	F_5	C_5	C_5	D_5	C_5	D_5	E_5	F_5	F_5	F_5	F_5
Chimes of Big Ben	A_5	F_5	G_5	C_5	C_5	G_5	A_5	F_5	A_5	F_5	G_5	C_5	C_5	G_5	A_5	F_5
Lead Us Heavenly Father Lead Us	C_5	E_5	G_5	G_5	A_5	G_5	F_5	E_5	E_5	F_5	G_5	C_5	E_5	D_5	C_5	C_5
Bobby Shaftoe	F_5	F_5	F_5	$A_5^\#$	A_5	C_6	A_5	F_5	C_5	C_5	C_5	F_5	E_5	G_5	E_5	C_5

in noise so as to match the pitch strength. Klein and Hartmann (1981) reported that the strength of binaural-edge pitch was slightly greater than that of Huggins pitch, although the differences between the two pitches were small and somewhat variable across frequency. We know of no equivalent measurements of the strength of binaural-coherence-edge pitch, although Hartmann and McMillon (2001) asked their listeners to make informal comparisons of the two pitches. They reported that a subset of their listeners found binaural-edge pitch to be the stronger, although none reported a ‘‘striking’’ (p. 303) difference in strength. An indirect estimate can be obtained from the variability in pitch matches. Hartmann and McMillon (2001) reported a standard deviation of approximately 3% of the transition frequency for binaural-coherence-edge pitch. This value is only slightly greater than the value of approximately 2% reported by Frijns *et al.* (1986) for binaural-edge pitch, suggesting that, within the margins of error, the two edge pitches may have about the same pitch strength. However, the possibility of individual differences in matching accuracy makes it difficult to draw firm conclusions from comparisons across studies.

The first question addressed by this study was: do the three types of dichotic pitch differ in salience, as might be expected on the basis of the above arguments? To answer this question, we used a melody recognition task. We were also interested in the salience of dichotic pitches relative to the clear pitches produced by pure-tone stimuli. Therefore we included a set of melodies produced by sequences of pure tones.

The second question of interest was: are dichotic pitches perceived without extensive training or does their perception require prolonged experience? To answer this question, we measured the degree to which melody recognition improved across blocks of trials. The experiment was conducted as part of a practical class, hence allowing a relatively large group (49) of inexperienced listeners to participate.

II. METHOD

A. Stimuli

The stimuli were created using MATLAB. Each stimulus consisted of a melody defined by a train of 16 notes of 300-ms duration, with each note separated by a silence of 100-ms duration. Ten melodies were used; they were the same as those used by Moore and Rosen (1979). They were based on well-known melodies, but were modified by slight distortions of the rhythm so as to consist of 16 equal-duration notes. These modifications required long-duration notes to be split into two or more equal-duration shorter notes. Moore and Rosen reported that their listeners did not find the modifications to be particularly disturbing. The frequencies corresponding to the notes in each melody are listed in Table I. The minimum, mean, and maximum frequencies were, respectively, 524 Hz ($=C_5$), 707 Hz ($\approx F_5$), and 1046 Hz ($=C_6$). Pilot experiments indicated that the chosen frequency range gave clear sensations of dichotic pitch, at least for experienced listeners.

In the pure-tone condition, each individual note in a melody was a tone burst with a frequency equal to that of the note. The pure-tone stimuli were presented diotically. In the three dichotic-pitch conditions, each individual note in a melody was a burst of dichotic bandpass noise which contained an interaural phase transition centered on the frequency of the note. Each dichotic-pitch stimulus was created in the spectral domain by rectangular filtering (0–4000 Hz passband) two matched 6000-point buffers representing the left and right channels of a diotic Gaussian noise sampled at 20 000 Hz. The phases of frequency components in the spectral buffer representing one channel were then modified. For each Huggins-pitch stimulus, a linear shift of 0 to 2π radians was added to the phases for frequency components from 8% below to 8% above the frequency of the note. For each binaural-edge-pitch stimulus, a shift of π radians was added to the phase of frequency components above the frequency of the note. For each binaural-coherence-edge-pitch stimulus, the amplitude and phase of frequency components above the frequency of the note in one spectral buffer were generated independently of those in the other spectral buffer. Subsequently the signal waveforms for the left and right channels were created by applying an inverse discrete Fourier transform to the two spectral buffers. A 30-ms raised-cosine ramp was applied to the onset and offset of each note. The melodies were recorded onto CD(R) for presentation to listeners using Sennheiser HD-414 earphones. All the earphones were driven in parallel from the same amplifier. The overall level of the stimuli was about 70 dB SPL at each ear.

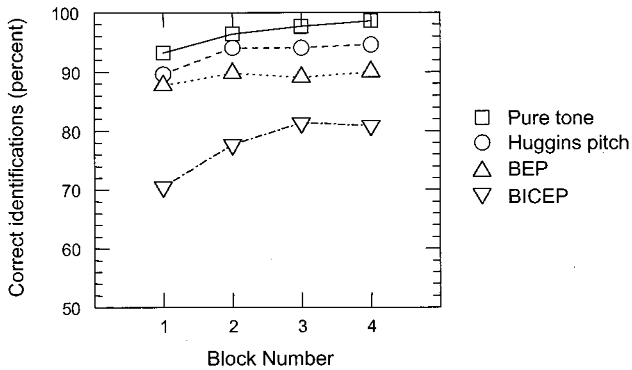

FIG. 3. Mean percentage of correct identifications in each of the four blocks of the test phase of the experiment for each type of pitch. The chance level is 10%. ''BEP'' is used as an abbreviation for binaural-edge pitch and ''BICEP'' is used as an abbreviation for binaural-coherence-edge pitch.

B. Procedure

The listeners were trained using a two-stage method. First, they received training to help them learn the name of each melody. Each of the 10 melodies, played with the pure-tone stimuli, was presented twice. Second, they received training to help them hear the pitches associated with each type of stimulus. They heard each of the 10 melodies, presented twice, first played with the Huggins-pitch stimuli, then with the binaural-edge-pitch stimuli, and finally with the binaural-coherence-edge-pitch stimuli. Throughout all of the training the name of a melody (also recorded on the CD) was announced before its presentation.

The test phase of the experiment was divided into four blocks. In each block, listeners heard each of the 40 possible combinations of melody and type of stimulus, presented in a random order (the ordering was constrained so that neither the same melody nor the same type of stimulus were presented in successive trials). They were required to identify each melody immediately after its presentation and to write their identifications on a score sheet. Feedback (the name of the tune) was provided after a delay of four seconds. The listeners scored their own responses as the experiment proceeded, and the response sheets were later checked for accuracy by the experimenters.[2]

Forty-nine undergraduate students enrolled at the University of Cambridge participated as listeners. The results reported below are based on the responses of 44 listeners, as responses were excluded for one listener who had a self-reported hearing loss and for four other listeners because they performed perfectly (the exclusion of these four sets of results does not affect the statistical analyses reported below).

III. RESULTS AND DISCUSSION

The mean scores across listeners are shown in Fig. 3. The symbols indicate the identification scores for the pure-tone stimuli (squares), Huggins-pitch stimuli (circles), binaural-edge pitch stimuli (upward-pointing triangles), and binaural-coherence-edge-pitch stimuli (downward-pointing triangles). There was no consistent pattern in the identification scores for the different melodies. Although some listeners showed patterns of errors indicating difficulty with specific melodies, these patterns were not consistent across listeners. Overall, each of the melodies was about equal in identifiability.

Three major outcomes can be identified. First, for all the types of stimuli, the identification scores improved from block 1 to block 2 and thereafter remained relatively stable. A Wilcoxon matched-pairs signed-ranks test (Siegel and Castellan, 1988), calculated on the mean responses across the four types of stimulus, showed that mean scores for blocks 1 and 2 differed significantly ($T=39$, $N=32$; $p<0.005$) but that scores for blocks 2 and 3 or for blocks 3 and 4 did not differ significantly ($T=184$, $N=34$, $p>0.05$; and $T=249$, $N=32$, $p>=0.05$, respectively). The initial improvement suggests a small practice effect. We believe that it represents an improvement in learning to attach names to the melodies, because the effect occurs for the pure-tone stimuli also, whose pitches were presumably clear and easily identified.

Second, the identification scores were consistently high in the first block for the three types of dichotic pitch stimulus. The high performance was especially noteworthy for the Huggins-pitch stimuli and binaural-edge-pitch stimuli, for which 41 out of 44 subjects scored better than 60% on the first block of trials (chance corresponds to 10%). This result supports the notion that the sensation of dichotic pitch is relatively immediate and does not require prolonged experience. Furthermore, the pitch appears to be ''musical,'' in that it readily supports melody recognition. It is of interest that the static interaural differences present in our stimuli were sufficient to support melody recognition. This contrasts with recent results of Culling (2000), obtained using stimuli similar to those of Kubovy *et al.* (1974). The latter presented eight continuous sinusoids to each ear, via earphones. The sinusoids had frequencies corresponding to the notes in a musical scale. Seven of the sinusoids were delayed by 1 ms at one ear. The remaining sinusoid was delayed by 1 ms at the other ear, with the result that this component had an ITD that was shifted by 2 ms from the ITDs of the other components. The shifted component was heard to stand out perceptually. A sequence of ITD shifts in different components was clearly heard as a melody. This melody was completely undetectable when listening to the input to one ear alone. Kubovy *et al.* interpreted their results as indicating that differences in relative phase at the two ears can allow an auditory ''object'' to be isolated in the absence of any other cues. However, Culling presented evidence that the transitions in interaural phase were the dominant cue allowing melody recognition. Static differences in interaural phase were found to be a weak cue ''that only a subset of listeners were able to exploit'' (Culling, 2000, p. 1768). Our results indicate that, for noise stimuli, static differences in the interaural phase can be used for melody recognition by the great majority of relatively untrained subjects.

The third major outcome was that identification scores differed across the four types of stimulus. The rank ordering (of highest scores to lowest scores) was: pure tones, Huggins pitch, binaural-edge pitch, and binaural-coherence-edge pitch. This ordering was observed in each of the four blocks as well in the across-block mean. Furthermore, the ordering

TABLE II. Results from a group of Wilcoxon matched-pair signed-rank tests applied to each pairwise comparison of the types of stimuli. In each test the value of T was less than the critical value of T required for a significance level of 0.0083 (used as it is equal to the conventional significance level of 0.05 divided by 6, which was the number of tests); thus each comparison shows a statistically significant effect. "BICEP" is used as an abbreviation for "binaural coherence edge pitch."

	Type of stimulus			
Type of stimulus	Pure tone	Huggins pitch	Binaural-edge pitch	BICEP
Pure tone	—			
Huggins pitch	$T=38$, $N=32$	—		
Binaural-edge pitch	$T=0$, $N=32$	$T=72$, $N=32$	—	
BICEP	$T=7$, $N=42$	$T=14$, $N=42$	$T=63$, $N=44$	—

was the same or similar in the individual scores for each listener; almost every listener scored worst for the binaural-coherence-edge-pitch stimuli. Table II reports the results of a set of Wilcoxon matched-pairs signed-ranks tests, which show that each of the six pairwise differences between the types of stimuli was statistically significant.

This outcome suggests that all of the pitch sensations produced by the dichotic-pitch stimuli are less salient than the pitch sensation produced by pure-tone stimuli. Furthermore, of the three types of dichotic pitch, Huggins pitch is the most salient, binaural-edge pitch is intermediate, and binaural-coherence-edge pitch is the weakest. This conclusion is further supported by the result that some of the listeners made errors in identification of the binaural-coherence-edge-pitch stimuli but made no errors for the other stimuli. These results are broadly consistent with the analysis described in the Introduction based upon the modified equalization-cancellation model and shown earlier in Fig. 2. However, to make the analysis more rigorous, a model of edge enhancement must be developed in order to account for the existence of the binaural-coherence-edge pitch. This is done in the next section.

It should be noted that there is nothing in the following model of edge enhancement that requires the use of the *recovered* spectrum generated by the modified equalization-cancellation model. It could also be applied to the spectrum of binaural activity, at a fixed internal time delay, generated by the central-spectrum model (e.g., Raatgever and Bilsen, 1986; Frijns *et al.*, 1986), or it could also act an implementation of the (nonmodified) equalization-cancellation explanation of binaural-coherence-edge pitch (Hartmann and McMillon, 2001). We expect that these models will give similar results, but may require different values of the free parameter that determines the relative magnitudes of the spectrum and the edge enhancement. For simplicity, however, and also because the purpose of the present study was not to compare experimentally the various models, we limit the analysis to the recovered spectrum.

IV. A MODEL OF EDGE ENHANCEMENT

Figure 2 shows the recovered spectra for the three types of dichotic pitches used in the experiment. The recovered spectra were calculated as follows. First, bursts of each dichotic-pitch stimulus were synthesized, each of 300-ms duration and 40-dB spectrum level. The parameters of the transition in interaural phase defining the dichotic pitches were the same as those used in the experiment, with the modification that the transition frequency was fixed at 500 Hz. The left and right waveforms of the stimuli were passed through matched 41-channel gammatone filterbanks coded in MATLAB (Patterson *et al.*, 1995; Slaney, 1998), with center frequencies ranging from two equivalent rectangular bandwidths ("ERBs," Glasberg and Moore, 1990) below 500 Hz to 2 ERBs above 500 Hz and with filters spaced at 0.1-ERB intervals. The output of each filter was then passed through a model of a high-spontaneous-rate fiber, with a threshold of 45 dB, that computed the probability of firing (Meddis *et al.*, 1990, Table II). The equalization and cancellation processes were combined by subtracting the probability of fiber output for a left channel from the probability of fiber output for the corresponding right channel, as a function of a time delay applied to one channel. Of the set formed by the remainder-after-cancellation at each of the time delays, the remainder with the smallest mean value was measured. This process was repeated for each frequency channel, so giving the recovered spectrum of the residual activation as a function of frequency. The recovered spectra shown in Fig. 2 are means across 25 independent bursts of each dichotic-pitch stimulus.

As may be seen, the mean recovered spectrum of the binaural-coherence-edge-pitch stimuli (dashed line) shows a sloping edge rather than a peak. Next, we describe a model for generating a peak in this spectrum. This model is effectively a computational implementation of a process of lateral inhibition applied to the recovered spectrum. Illustrations of the operation of lateral inhibition, applied to a sloping edge like that seen in the recovered spectrum for binaural-coherence-edge pitch, can be found in von Békésy (1959, Fig. 8) and Small and Daniloff (1967, Fig. 5). The model is based on extracting the second derivative (i.e., the curvature) of the recovered spectrum and inverting its sign. In the transition from a low-level flat portion of the recovered spectrum to a rising edge, the curvature is positive, so the inverse is negative. In the transition from a rising edge to a high-level plateau, the curvature is negative, so the inverse is positive. Thus, the inverse of the curvature has the appropriate properties for simulating the effects of lateral inhibition.

In detail, the first-order derivatives of the recovered spectrum were calculated using

$$\frac{dr_z}{dz} = r_z - r_{z-1}, \tag{1}$$

and then the second-order derivatives were calculated using

$$\frac{d^2 r_z}{dz^2} = \frac{dr_{z+1}}{dz} - \frac{dr_z}{dz} = r_{z+1} - 2r_z + r_{z-1}, \tag{2}$$

where z is the frequency coordinate (i.e., the channel number), and r_z is the level of the recovered spectrum (i.e., the residual activation) in the zth frequency channel. It should be remembered that the separation of the frequency channels is 0.1 ERBs and so the denominator in these two equations is also 0.1 ERBs. Next, minor fluctuations in the second-order differences were smoothed by averaging the values across ± 4 adjacent channels, corresponding to ± 0.4 ERBs:

$$\begin{aligned} \left(\overline{\frac{d^2 r_z}{dz^2}}\right) &= \frac{1}{4+4+1} \sum_{z-4}^{z+4} \frac{d^2 r_z}{dz^2} \\ &= \tfrac{1}{9}(r_{z+5} - r_{z+4} - r_{z-4} + r_{z-5}). \end{aligned} \tag{3}$$

We refer to the function defined by Eq. (3) as the "edge-enhancement function." This function was then inverted in sign, scaled, and added to the original recovered spectrum so as to simulate the process of lateral inhibition, giving an "edge-enhanced recovered spectrum." The value of the scaling factor is a free parameter in the model; here a value of 50 was chosen so as to give a peak of reasonable height in the edge-enhanced recovered spectrum for the binaural-coherence-edge-pitch stimuli. The edge-enhanced recovered spectrum, E_z, is thus given by

$$E_z = r_z - 50\left(\overline{\frac{d^2 r_z}{dz^2}}\right). \tag{4}$$

One consequence of choosing a factor of 50 was that, in the final edge-enhanced recovered spectrum, both the original recovered spectrum r_z and the edge-enhancement function [Eq. (3)] each contribute 50% of the value of the peak of Huggins pitch, although the results are insensitive to the exact value of the constant.

Figure 4 shows the results of applying each step of this "edge-enhancement" model to the recovered spectra of the three dichotic-pitch stimuli considered here. In each panel, the solid line represents Huggins pitch, the dotted line represents binaural-edge-pitch, and the dashed line represents binaural-coherence-edge pitch. Panel A shows the same recovered spectra that were illustrated in Fig. 2. Note that there are clear peaks in the functions for Huggins pitch and binaural-edge pitch but not for binaural-coherence-edge pitch. Panel B shows the first-order differences [see Eq. (1)] and panel C shows the unsmoothed second-order differences [see Eq. (2)]. Note that, for both Huggins pitch and binaural-edge pitch, the peak in the recovered spectrum near 600 Hz corresponds to a value for the first difference that is approximately zero and to a value for the second difference that is a local minimum. Panel D shows the smoothed second-order differences; i.e., the edge-enhancement function [see Eq. (3)]. Note that the smoothing process has removed the minor fluctuations that result from the randomness inherent to all noise stimuli. Last, panel E shows the edge-enhanced recovered spectrum [see Eq. (4)]. Note that, for Huggins-pitch and binaural-edge pitch, the frequency location (and the relative height) of the peak in the recovered spectrum is preserved,

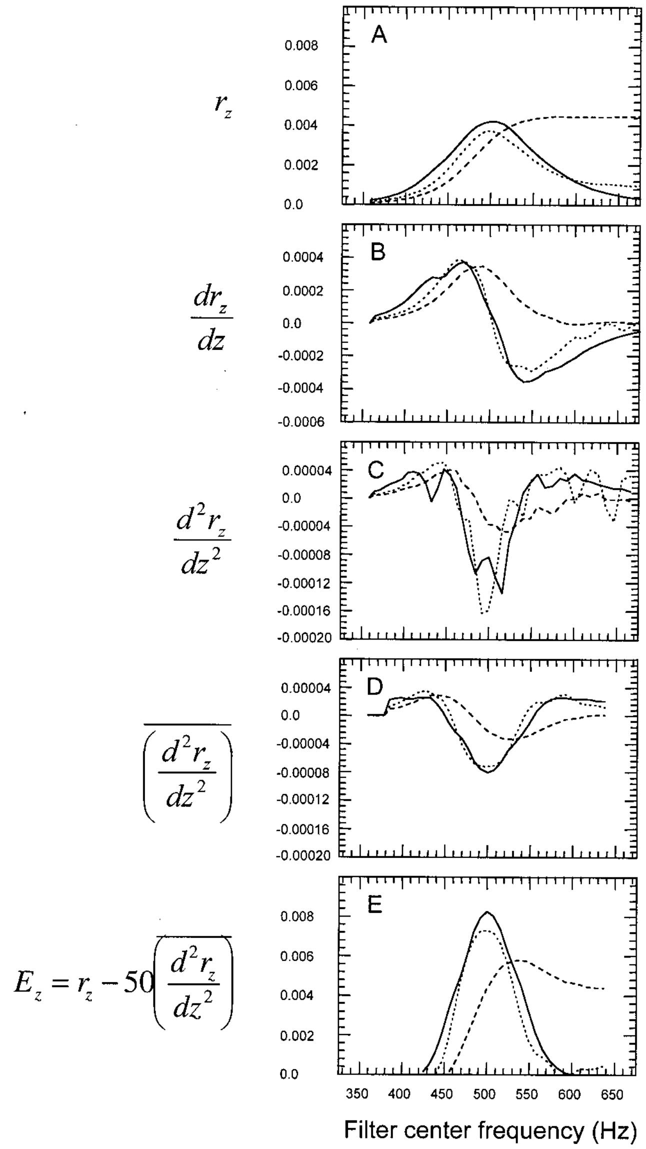

FIG. 4. The results from each step of the "edge-enhancement" model. Panel A shows the recovered spectra for a Huggins pitch stimulus (solid line), binaural-edge-pitch stimulus (dotted line), and binaural-coherence-edge-pitch stimulus (dashed line); these are the same as those illustrated in Fig. 2. Panel B shows the first-order differences [see Eq. (1)]. Panel C shows the unsmoothed second-order differences [see Eq. (2)]. Panel D shows the smoothed second-order differences; i.e., the edge-enhancement function [see Eq. (3)]. Panel E illustrates the edge-enhanced recovered spectrum [see Eq. (4)].

but for binaural-coherence-edge pitch, a peak is present in the edge-enhanced recovered spectrum that was not present in the original recovered spectrum.

There are two effects in panel E of Fig. 4 that are of interest. First, the frequency location of the peak in the edge-enhanced recovered spectrum for binaural-coherence-edge pitch is at approximately 540 Hz. This frequency corresponds to a shift of about 8% above the transition frequency of 500 Hz. Such a magnitude of shift is compatible with Hartmann and McMillon's (2001) measurements of the value of the perceived pitch for binaural-coherence-edge pitch.

Second, the magnitude of the peak in the edge-enhanced recovered spectra is greatest for Huggins pitch, marginally smaller for binaural-edge pitch, and distinctly smaller for binaural-coherence-edge pitch. This ordering of the peak heights resembles the pattern of results in our experiment and is consistent with the assumption that the strength of these dichotic pitches is related to the height of the peak in the enhanced recovered spectrum. It should be noted, however, that this effect is partially determined by the value of the free parameter used in Eq. (4); our choice of 50 was chosen, in part, to obtain this ordering of the three dichotic-pitch stimuli.

V. SUMMARY

(1) We measured identifiability of melodies created using three types of dichotic-pitch stimuli (Huggins pitch, binaural-edge pitch, binaural-coherence-edge pitch) as well as using pure-tone stimuli. The identification scores were consistently high for all of the dichotic-pitch stimuli, demonstrating that the pitch sensations evoked by dichotic-pitch stimuli are "musical."

(2) The identification scores were high for the dichotic-pitch stimuli even in the first block of testing, indicating that the sensation of dichotic pitch is relatively immediate and does not require prolonged experience.

(3) Differences in identification scores were observed for the three types of dichotic pitch, Huggins pitch giving the highest scores and binaural-coherence-edge pitch the lowest scores. These differences are consistent with analyses based upon the modified equalization-cancellation model of dichotic pitch, extended to include a model for enhancement of edges in the recovered spectrum. The data are consistent with the idea that the strength of a dichotic pitch is related to the height of a peak in the enhanced recovered spectrum.

ACKNOWLEDGMENTS

We thank Brian Glasberg for his help, and an anonymous reviewer, Frans Bilsen and Sid Bacon (Associate Editor) for their comments in the review process. M.A.A. is supported by a MRC (U.K.) Career Development Award. G.A.M. is supported by a scholarship from Defeating Deafness (U.K.).

[1]Cramer and Huggins (1958) used a bandwidth of 10%, meaning that the width of the interaural phase shift from 90° to 270° was 10% of the center frequency. This definition of bandwidth is usual if, like Cramer and Huggins, the interaural phase shift is constructed using an analog all-pass filter. If, like in the present experiment, the stimuli are constructed digitally in the frequency domain, it is more convenient to define the value of the bandwidth as the width of the interaural phase shift from 0° to 360°. Our bandwidth of 16% thus corresponds to 8% in Cramer and Huggins' terminology.

[2]Listeners were asked to write down their response to each tune *before* the correct answer was given, and then to mark the response as correct or incorrect with a tick or a cross. The experimenter checked that they were doing this as the experiment proceeded. The response sheets were subsequently checked to ensure that the ticks and crosses correctly reflected the tune names that were written down. This was always the case.

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A fully temporal account of the perception of dichotic pitches," Br. J. Audiol. **33**, 106–107.

Cramer, E. M., and Huggins, W. H. (**1958**). "Creation of pitch through binaural interaction," J. Acoust. Soc. Am. **30**, 413–417.

Culling, J. F. (**2000**). "Auditory motion segregation: A limited analogy with vision," J. Exp. Psychol. **26**, 1760–1769.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998**). "Dichotic pitches as illusions of binaural unmasking. I. Huggins pitch and the binaural edge pitch," J. Acoust. Soc. Am. **103**, 3509–3526.

Durlach, N. I. (**1962**). "Note on the creation of pitch through binaural interaction," J. Acoust. Soc. Am. **34**, 1096–1099.

Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic, New York).

Frijns, J. H. M., Raatgever, J., and Bilsen, F. A. (**1986**). "A central spectrum theory of binaural processing. The binaural edge pitch revisited," J. Acoust. Soc. Am. **80**, 442–451.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Hartmann, W. M. (**1984**). "Binaural coherence edge pitch," J. Acoust. Soc. Am. **75**, S22.

Hartmann, W. M., and McMillon, C. D. (**2001**). "Binaural coherence edge pitch," J. Acoust. Soc. Am. **109**, 294–305.

Julesz, B. (**1971**). *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, IL).

Klein, M. A., and Hartmann, W. M. (**1981**). "Binaural edge pitch," J. Acoust. Soc. Am. **70**, 51–61.

Kubovy, M., Cutting, J. E., and McGuire, R. M. (**1974**). "Hearing with the third ear: dichotic perception of a melody without monaural familiarity cues," Science **186**, 272–274.

Licklider, J. C. R. (**1959**). "Three auditory theories," in *Psychology, a Study of a Science*, edited by S. Koch (McGraw-Hill, New York).

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse," J. Acoust. Soc. Am. **87**, 1813–1816.

Moore, B. C. J., and Rosen, S. M. (**1979**). "Melody recognition with reduced pitch and interval information," Q. J. Exp. Psychol. **31**, 229–249.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A model architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.

Raatgever, J., and Bilsen, F. A. (**1986**). "A central spectrum theory of binaural processing. Evidence from dichotic pitch," J. Acoust. Soc. Am. **80**, 429–441.

Siegel, S., and Castellan, N. J., Jr. (**1988**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).

Slaney, M. (**1998**). "Auditory Toolbox: A MATLAB toolbox for auditory modeling work, Version 2," Technical Report No. 1998-010 (Interval Research Corporation).

Small, A. M., and Daniloff, R. G. (**1967**). "Pitch of noise bands," J. Acoust. Soc. Am. **41**, 506–512.

von Békésy, G. (**1959**). "Neural funneling along the skin and between the inner and outer hair cells of the cochlea," J. Acoust. Soc. Am. **31**, 1236–1249.