Dichotic pitches as illusions of binaural unmasking. I. Huggins' pitch and the "binaural edge pitch"

John F. Culling,[a] A. Quentin Summerfield, and David H. Marshall
MRC Institute of Hearing Research, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

(Received 21 February 1996; revised 2 March 1998; accepted 4 March 1998)

The two most salient dichotic pitches, the Huggins pitch (HP) and the binaural edge pitch (BEP), are produced by applying interaural phase transitions of 360 and 180 degrees, respectively, to a broadband noise. This paper examines accounts of these pitches, concentrating on a ''central activity pattern'' (CAP) model and a ''modified equalization-cancellation'' (mE-C) model. The CAP model proposes that a dichotic pitch is heard at frequency f when an individual across-frequency scan in an interaural cross-correlation matrix contains a sharp peak at f. The mE-C model proposes that a dichotic pitch is heard when a plot of interaural decorrelation against frequency contains a peak at f. The predictions of the models diverge for the BEP at very narrow transition bandwidths: the mE-C model predicts that salience is sustained, while the CAP model predicts that salience declines and that the dominant percept is of the in-phase segment of the noise. Experiment 1 showed that the salience of the BEP was sustained at the narrowest bandwidths that could be generated (0.5% of the transition frequency). Experiment 2 confirmed that the pitch of a BEP produced by a 0.5% transition bandwidth was close to the frequency of the transition band. Experiment 3 showed that pairs of simultaneous narrow 180-degree transitions, whose frequencies corresponded to vowel formants, were perceived as the intended vowels. Moreover, the same vowels were perceived whether the in-phase portion of the noise lay between the two transition frequencies or on either side of them. In contrast, different patterns of identification responses were made to *diotic* band-pass and band-stop noises whose cutoff frequencies corresponded to the same formants. Thus, the vowel-identification responses made to the dichotic stimuli were not based on hearing the in-phase portions of the noise as formants. These results are not predicted by the CAP model but are consistent with the mE-C model. It is argued that the mE-C model provides a more coherent and parsimonious account of many aspects of the HP and the BEP than do alternative models. [S0001-4966(98)05906-2]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Pn [RHD]

INTRODUCTION

Dichotic pitches arise through processes of binaural interaction when broadband noises are presented to the two ears (i.e., the pitches cannot be heard monaurally). The present investigations of these phenomena were motivated by the observation that the dominant account of these pitches, the CAP model (Bilsen, 1977), invokes a process similar to ''across-frequency grouping by common interaural time delay (ITD).'' This is a putative process in which energy in different frequency regions originating from the same source would be grouped together by virtue of possessing the same ITD. However, several recent experiments (Culling and Summerfield, 1995; Hukin and Darwin, 1995; Darwin and Hukin, 1997) have cast doubt on the idea that auditory analysis includes the capacity to group energy in this way. Culling and Summerfield proposed instead that each frequency channel in the binaural system operates independently when recovering signals from noise. They embodied this concept in a multi-channel model of binaural unmasking (the mE-C model). This model makes acceptably accurate predictions of the clarity and frequency of dichotic pitches. In the course of the present investigations, it also emerged that, despite its intuitive appeal, the CAP model has some previously unreported shortcomings. Consequently, this article and its companion (Culling *et al.*, 1998) have two objectives. The first is to demonstrate that the CAP model, employing an across-frequency process, does not predict dichotic pitches correctly. The second is to show that the mE-C model, without such a process, can explain them well.

Four dichotic pitches have been described:[1] the Huggins pitch (Cramer and Huggins, 1958), the binaural edge pitch (Klein and Hartmann, 1981), the Fourcin pitch (Fourcin, 1970), and the dichotic repetition pitch (Bilsen and Goldstein, 1974). These four phenomena fall into two classes, which differ in their method of generation. The Fourcin pitch (FP) and the dichotic repetition pitch (DRP) are generated by applying large interaural delays (>1 ms) to broadband noise. To generate the FP, two independent noises are presented simultaneously and binaurally. The two noises have different interaural delays. If an interaural phase shift of 180 degrees is given to one of the noises, the period of the perceived pitch frequency is equal to the difference between the two interaural delays (Fourcin, 1970; Bilsen and Goldstein, 1974; Bilsen, 1977). Without the interaural phase shift, the

[a]New address: University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, United Kingdom.

overall pitch is ambiguous (Bilsen, 1977). To generate the DRP, a single noise is presented binaurally with a large interaural delay (2–20 ms). The period of the perceived pitch frequency corresponds to the interaural delay (Bilsen and Goldstein, 1974). Although there are similarities between the stimuli which give rise to these two pitches, the FP is much more salient perceptually than the DRP. A companion paper (Culling *et al.*, 1998) compares explanations for the FP and the DRP. It argues that the FP is produced by the same mechanism that underpins the Huggins pitch and the binaural edge pitch, while the DRP is produced by a different mechanism.

The Huggins pitch (HP) and the binaural edge pitch (BEP) are generated by producing an interaural phase transition at a particular frequency within a broadband noise. For the HP, the noise is identical at the two ears below the transition frequency. At the transition frequency, the interaural phase relationship changes sharply with increasing frequency, shifting through 360 degrees over a narrow bandwidth (e.g., 6% of the transition frequency). Above the transition band, the noise is identical at the two ears. For the BEP, a similar transition occurs, but over a range of only 180 degrees, so that the noise is out-of-phase between the ears above the transition frequency and in-phase below, or *vice versa*. HP and BEP stimuli both evoke the perception of a pure-tone-like pitch, corresponding approximately to the transition frequency, which can be heard against the background of the noise (Guttman, 1962; Klein and Hartmann, 1981; Frijns *et al.*, 1986). The more salient percept is provided by the HP, but the BEP is also clearly audible by naive listeners. Most listeners hear the HP lateralized to one side or the other. Some listeners also hear the BEP lateralized away from the mid-line.

I. MODELING HP AND BEP

A. The central activity pattern (CAP) model

The central activity pattern (CAP) (Bilsen, 1977; Raatgever and Bilsen, 1986; Frijns *et al.*, 1986) is similar to an interaural cross-correlation matrix[2] of the kind proposed by Jeffress (1948, 1972) to account for the lateralization of sound sources. The CAP is a matrix which displays a map of interaural correlation by frequency and interaural delay. To explain the HP, Raatgever and Bilsen (1986) suggested that the matrix might be scanned across frequency at a chosen internal delay to produce a ''central spectrum,'' whose structure would determine the pitch and timbre of the perceived sound. The CAP model does not include an explicit mechanism for choosing which of the many possible across-frequency scans is selected. Rather, Raatgever and Bilsen (1986) suggested that the mechanism ''recognises and selects the frequency spectrum information by making use of cues like harmonicity and depth of modulation or *a priori* knowledge of spectral features'' (p. 431). Despite the lack of an explicit selection mechanism, there is intuitive appeal in the idea that attention can be directed selectively to an individual across-frequency scan, since in real listening situations one would expect such scans to display the spectra of sound sources which lie on different azimuths. The CAP model adheres to this principle rigorously: dichotic pitches are explained by features found in the spectrum of a single scan; evidence from several scans is not combined.

Raatgever and Bilsen illustrated the workings of the CAP model with a computational procedure in which the interaural phase relationship at each frequency was used to generate a pattern of activity which was a sinusoidal function of internal delay [Raatgever and Bilsen, 1986, Eq. (6)]. We refer to their formulation as the ''original'' version of the CAP model. The sinusoidal functions were accepted as an approximation to the corresponding cross-correlation functions. In fact, they are equivalent to cross-correlation functions produced with infinitely long time windows and infinitely narrow frequency channels. The following paragraphs demonstrate that these assumptions materially affect the predictions of the CAP model.

Figure 1(a) and (c) shows CAPs generated by the original version of the model.[3] Figure 1(a) illustrates the CAP of a HP stimulus with the transition frequency at 600 Hz and a transition bandwidth of 6%.[4] The model produces peaks at the transition frequency in some across-frequency scans. For example, the inset illustrates such a peak in the scan taken at an interaural delay of 0.83 ms. Similar peaks can also occur in across-frequency scans of BEP stimuli. Figure 1(c) shows the CAP of a BEP stimulus with the transition frequency at 600 Hz and a transition bandwidth of 6%. The inset contains the across-frequency scan taken at an interaural delay of 1.25 ms. In discussing the BEP, Frijns *et al.* (1986) drew attention to the sharp peak at 600 Hz which is found in this scan.

A difficulty in relating the predictions of the original version of the CAP model to the performance of listeners is that the model assumes unrealistically high values for the frequency resolution of the auditory system. Figure 1(b) and (d) shows the results of convolving the CAPs of Fig. 1(a) and (c) across frequency with a rounded-exponential-shaped moving-average filter which increases in bandwidth with frequency in accordance with estimates of the bandwidth of auditory filters [Moore and Glasberg, 1983, Eq. (3)]. Such an integration across frequency in order to obtain more realistic patterns was suggested, but not performed, by Raatgever and Bilsen (1986). We refer to the formulation of the CAP model that incorporates this smoothing filter as the ''smoothed'' version. After smoothing, the peaks in across-frequency scans are very much less apparent, but, in this example, remain visible.

An important prediction of the CAP model is that the strength of the BEP, like that of the HP, should be reduced or eliminated when stimuli have very narrow transition bandwidths. This prediction is tested in experiment 1. To illustrate the prediction, Fig. 2 shows scans generated by the original (dotted lines) and smoothed (solid lines) versions of the model for HP and BEP stimuli with transitions at 600 Hz. The scans were taken at a delay of 0.83 ms for the HP and 1.25 ms for the BEP (similar to the insets of Fig. 1). Scans have been plotted for transition bandwidths, w, which are 0.5%, 1%, 8%, and 64% of the transition frequency. In the limit, as w is reduced, the transition band of the HP becomes infinitesimally small and so the stimulus becomes a diotic noise. In practice, the limit is reached when w is less than or

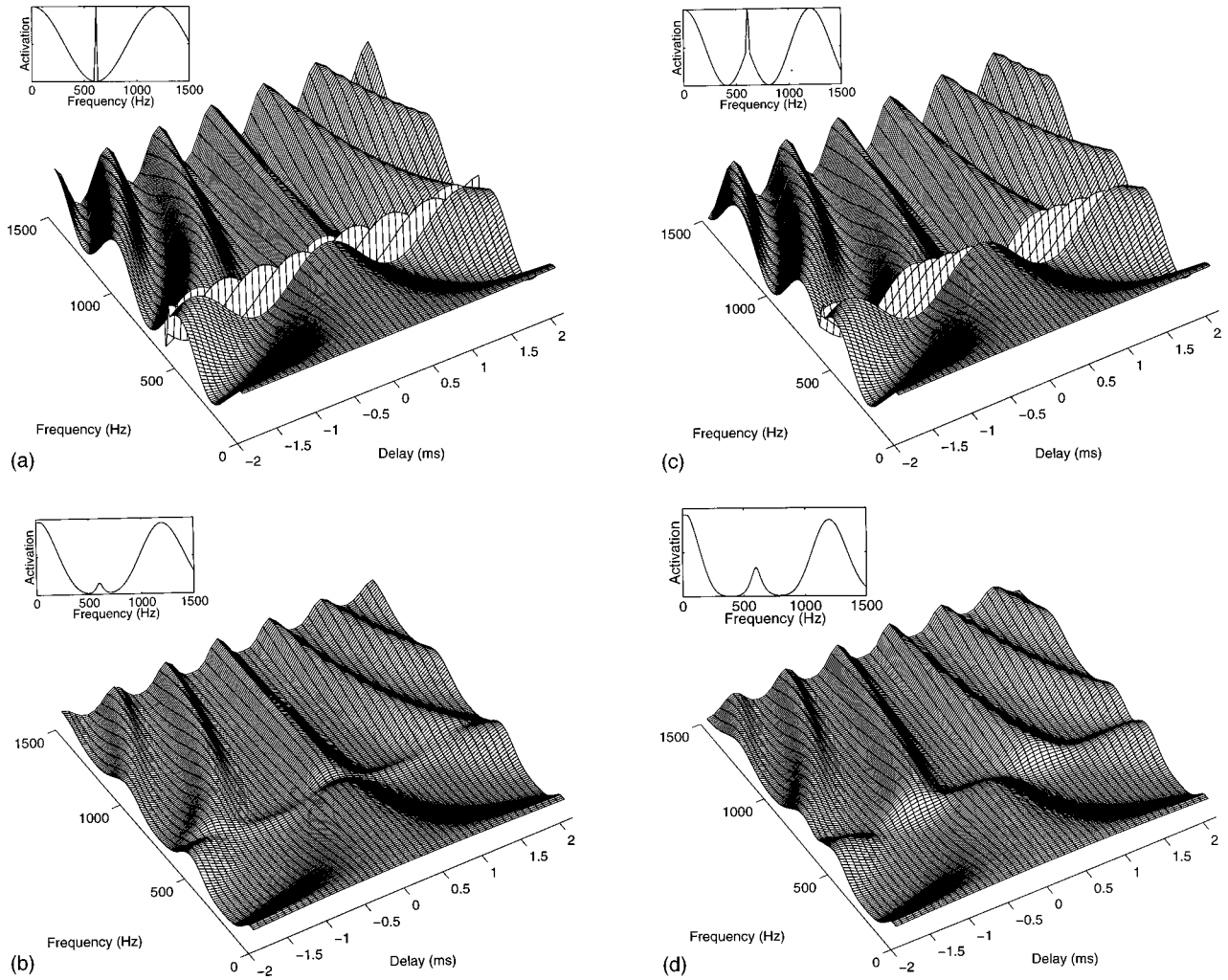

FIG. 1. The effect of the assumptions made by Raatgever and Bilsen [1986, Eq. (6)] in predicting the HP and BEP: (a) ''original'' CAP for a HP, produced from Eq. (6) of Raatgever and Bilsen (1986); (b) ''smoothed'' CAP produced by convolving the original CAP in panel (a) with a rounded-exponential-shaped, moving-average filter whose bandwidth varied with frequency according to Moore and Glasberg [1983, Eq. (3)]; (c) original CAP for a BEP; (d) smoothed CAP for the original CAP in panel (c). The insets show across-frequency scans taken at optimal internal delays for detecting peaks at the transition frequencies (0.83 ms for HP and 1.25 ms for BEP).

equal to the frequency spacing between adjacent bins in the Fourier spectrum of the stimulus. For the stimuli used to create Fig. 2, the limit is reached when w is 0.5%. For this value of w the transition bandwidth for the HP stimulus is effectively 0%. Such a limit does not exist for the BEP.

For $w=1\%$ and $w=8\%$, the scans produced by the original version of the model (dotted lines) contain spectral peaks at the 600-Hz transition frequency which are characterized by high amplitude and narrow bandwidths. According to the CAP model, these ''sharp peaks'' are responsible for the perception of the dichotic pitch. The sharp peaks are produced by the changing phase within the transition band. This phase shift compensates for the internal delay at just one frequency within the band, to produce maximal cross correlation at that frequency. Since the phase changes so abruptly with frequency, the same effect does not occur at closely adjacent frequencies, where the idealized cross correlation is consequently much lower. It is a necessary requirement of the model that peaks with broader bandwidths, such as those which occur at 1200 Hz in the panels of Fig. 2, are *not* treated as candidate pitches. Consistent with this view, when $w=64\%$ (bottom panel in each column), where the stimuli no longer give rise to tonal pitches (Cramer and Huggins, 1958), the peaks at the 600-Hz transition frequency are broader than when $w=8\%$.

In contrast to the scans produced by the original version of the model, the scans produced by the smoothed version include peaks at the transition frequency which reduce in height and width as the transition bandwidth is reduced from 8%, through 1%, to 0.5% for both the HP and BEP. Since a higher, narrower peak occurs for $w=8\%$ than for $w=0.5\%$, $w=1\%$, or $w=64\%$, the smoothed version of the model predicts that both the HP *and* the BEP should be heard most clearly at intermediate values of w. In particular, it should be harder to hear the BEP when w equals 0.5% or 1% than when w equals 8%.

In discussing the BEP, Frijns *et al.* (1986) distinguished the tonal quality of the dichotic pitch from the percept of a low-pass or high-pass noise which may also be heard, depending on whether the in-phase portion of the noise is below or above the transition frequency. The CAP account predicts that these percepts will determine the pattern of

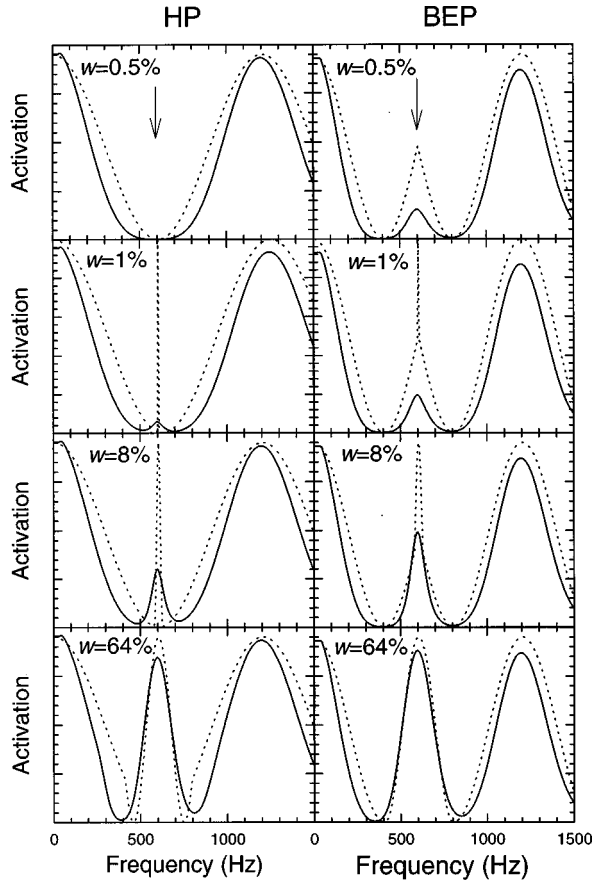

FIG. 2. Across-frequency scans from central activity patterns. The dotted lines are scans derived from the original CAP model of Raatgever and Bilsen [1986, Eq. (6)]. The solid lines are the same scans spectrally smoothed by a rounded-exponential-shaped, moving-average filter with bandwidth varying according to Moore and Glasberg [1983, Eq. (3)]. In different panels the patterns for the HP and BEP have nominal transition bandwidths, w, of 0.5%, 1%, 8%, and 64%. The vertical arrows indicate the frequency of the perceived pitch of 600 Hz. The scans are taken at internal delays which are optimal for detecting peaks at the appropriate frequencies (0.83 ms for HP and 1.25 ms for BEP). (Note that the effective bandwidth is 0% when w is set to 0.5% for the HP stimulus.)

listeners' responses when dichotic pitches cannot be heard. This prediction is tested in experiments 2 and 3.

B. The augmented equalization-cancellation (aE-C) model

Durlach (1962) pointed out that the HP can be explained by the equalization cancellation (E-C) model of binaural masking release (Durlach, 1960, 1972). In this model, the signals from each ear are equalized by any or all of a specified set of transformations (including changes in interaural delay, phase, and level), and are then cancelled by addition or subtraction. The E-C model was originally developed to account for the detection of tones in noise, and so did not include separate operations within different frequency channels. In line with this formulation of E-C, Klein and Hartmann (1981) also applied transformations to the whole signal when accounting for HP. In this case, the E-C process is particularly simple. Since the waveforms at the two ears are largely identical, they can be cancelled over most of the frequency range by subtraction without any prior equalization process, leaving a narrow band of noise centered on the transition frequency which would be heard as a warbling tone.

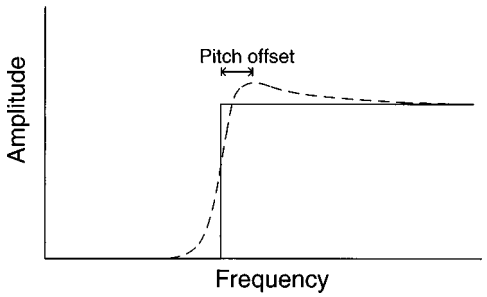

FIG. 3. Schematic illustration of the effect of lateral inhibition on the internal representation of a high-pass noise. The stimulus spectrum (after broadband interaural cancellation for a BEP) is represented by the solid line. The perceived spectrum, following lateral inhibition, is represented by the dashed line.

In the case of the BEP, broadband E-C gives rise to a low-pass or high-pass noise which would not directly be heard as a tone. When Klein and Hartmann (1981) discovered the BEP, they suggested that listeners do indeed hear a high-pass or low-pass noise, but that a process of central lateral inhibition enhances the edge of the noise, giving rise to the perceived tone (Fig. 3). Another process of lateral inhibition may be responsible for a different pitch sensation, called the ''edge pitch'', which arises monaurally when a high-pass or low-pass noise is presented (Small and Daniloff, 1967; Fastl, 1971; Fastl and Stoll, 1979; Klein and Hartmann, 1981). Consequently, Klein and Hartmann termed their phenomenon the ''binaural edge pitch'' or BEP.

Klein and Hartmann collected pitch-matching data to support their analogy between the BEP and the monaural edge pitch. Listeners were required to match monaural and binaural edge pitches to pure tones. The resulting matches were offset from the frequency of the cutoff or phase transition by a frequency difference of 3%–8%. Figure 3 shows schematically how lateral inhibition would give rise to such an offset. In the case of the monaural edge-pitch, Klein and Hartmann found that pitch matches were consistently offset into the noise, whereas for BEP stimuli, matches were offset to both higher and lower frequencies, giving rise to a bimodal distribution of matches. It was possible to account for the bimodal distribution by invoking a feature of the E-C model which states that cancellation can occur either by adding or by subtracting the waveforms at the two ears. If addition is performed, a BEP stimulus which is in-phase below the transition frequency would give rise to a low-pass residue, while if subtraction is performed, the same stimulus would produce a high-pass residue. So, depending on which cancellation operation is applied, the binaural edge pitch can be offset in either direction, compatible with the observed bimodal distribution. Thus, Klein and Hartmann concluded that a process of broadband E-C could explain the HP, and that broadband E-C augmented by lateral inhibition on the residue could explain the BEP. We refer to the latter account as the ''augmented equalization-cancellation model'' (aE-C).

Other results have not supported these conclusions consistently. Frijns *et al.* (1986) repeated Klein and Hartmann's pitch-matching experiments for the BEP and found that lis-

teners matched the dichotic pitch close to the transition frequency. The distribution of matches had a standard deviation of 1.5% with no evidence of a bimodal distribution of matches, thereby undermining one of the predictions of the aE-C model. Hartmann (1984a) sought to measure the lateral inhibition directly using the pulsation threshold method (Houtgast, 1972), but found no evidence for it. These two results imply that lateral inhibition does not underlie the BEP. However, Hartmann (1984b) has described a further dichotic pitch whose existence is difficult to explain without invoking central lateral inhibition. The binaural coherence edge pitch (BICEP) occurs when noise below a specified transition frequency is uncorrelated (i.e., statistically independent) at the two ears, while above that frequency it is correlated. A pitch is heard which corresponds to the transition frequency. The aE-C model can easily account for the BICEP. Interaural cancellation by subtraction leaves a low-pass residue. Lateral inhibition at the high-frequency edge of the residue generates the pitch. In summary, therefore, the balance of current evidence favors the existence of central lateral inhibition, but is equivocal about its role in the BEP.

This uncertainty motivates the exploration of alternative bases for the BEP. The aE-C model invokes central lateral inhibition to explain the BEP because it incorporates broadband E-C which does not itself generate a dichotic pitch. In the next section we consider a modified form of the E-C which can account for the BEP directly, without requiring central lateral inhibition as an additional process.

C. The modified equalization-cancellation (mE-C) model

A third account of the HP and BEP is provided by a modified version of the E-C model (mE-C) which was designed to account for the binaural masking release of broadband sounds, such as speech (Culling and Summerfield, 1995). Since no tone is physically present in the stimuli, the perception of the HP and BEP is, according to the mE-C model, an illusion of binaural unmasking. In this model, auditory frequency analysis is simulated by analyzing the waveforms presented to each ear with a gamma-tone filterbank (Patterson *et al.*, 1987, 1988). Mechanical to neural transduction in each of the resulting frequency channels is simulated with a model of hair cell transduction (Meddis, 1986, 1988). The time-varying excitation in corresponding frequency channels (f_L and f_R) from each ear is equalized in two steps. First, the rms levels are equated. Second, the internal delay is sought at which the residue, R, after subtraction (i.e., cancellation) of the excitation from each ear is minimized. Thus, the residue within a channel is a function of internal delay, τ, which is evaluated for $-5\ \text{ms} < \tau < 5\ \text{ms}$. The residue is weighted according to an exponentially tapering temporal window with a time constant, T, of 50 ms [Eq. (1)]:

$$R(\tau) = \int_0^{3T} (f_L(t) - f_R(t+\tau)) e^{-t/T}\, dt. \qquad (1)$$

The process is repeated in each frequency channel independently (i.e., permitting different adjustments of rms level and τ). The minimum residue in each channel is taken as a measure of the strength of the signal in that channel. Effectively, the model detects the degree of interaural decorrelation present in each channel, since excitation which correlates at internal delays within the range ± 5 ms cancels, while uncorrelated noise does not. Signals which have a different interaural phase from the masking noise are detected because they disrupt the interaural correlation of the noise. In this respect the model is similar to Colburn's model of binaural unmasking (Colburn, 1973, 1977). A key feature of the mE-C model is that it performs equalization and cancellation in each frequency channel *independently*.[5]

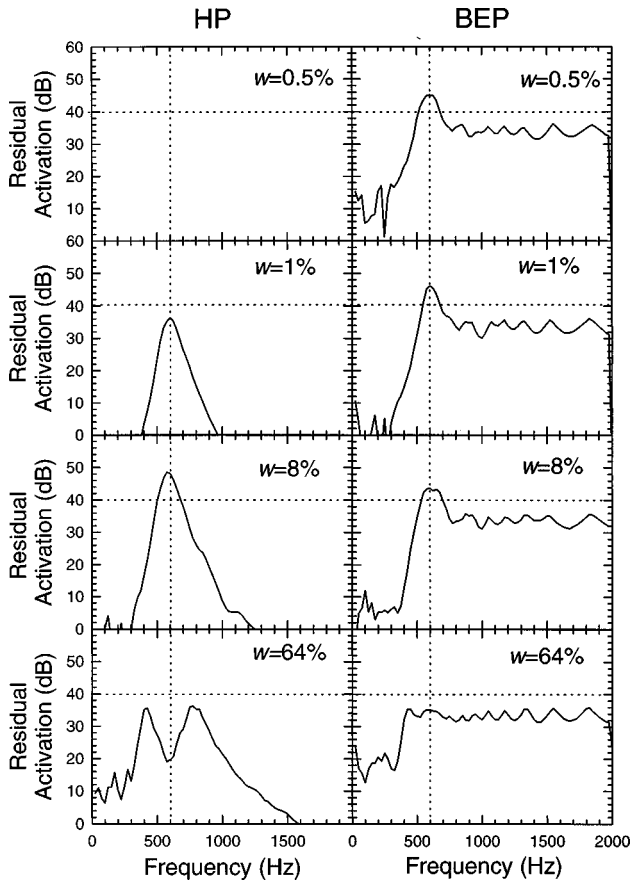

FIG. 4. The spectra recovered by the mE-C model of Culling and Summerfield (1995) for the HP and BEP for transition bandwidths, w, of 0.5%, 1%, 8%, and 64%. The vertical arrows indicate the frequency of the perceived pitch of 600 Hz. The dotted horizontal line defines an arbitrary threshold, which could predict whether or not the peaks were audible. (Note that the effective bandwidth is 0% when w is set to 0.5% for the HP stimulus.)

Figure 4 contains residual-activation spectra generated by the mE-C model for the same HP and BEP stimuli as were analysed by the CAP model in Fig. 2. In each panel, the stimulus contains an interaural phase transition at 600 Hz. For the HP, the prediction is straightforward and similar to that described in Sec. I B. The noise is in-phase both above and below the transition frequency and consequently cancels almost completely at zero internal delay. Close to the transition, however, noise with different interaural phases enters the same frequency channel and consequently cannot be cancelled completely at any internal delay, resulting in the peak in the residual-activation spectrum recovered by the model. For the BEP, the prediction is less obvious. On one side of

the transition (below 600 Hz in Fig. 4), the noise is in-phase, but on the other side it is 180 degrees out-of-phase. The in-phase noise is completely cancelled, leaving no residual activation below the transition frequency. The noise which is 180 degrees out-of-phase is partly cancelled because each channel admits only a narrow band of frequencies; a phase difference of 180 degrees at each of these frequencies is approximately equivalent to an internal delay of half the period of the center frequency of the channel. Channels close to the transition frequency itself, however, admit energy with a range of widely differing interaural delays which cannot be canceled, leaving a peak in the residual-activation spectrum. When $w \leqslant 8\%$, the peak is about 10 dB higher than the residual activation in the higher frequency channels which receive out-of-phase noise. According to the mE-C model, it is this peak which gives rise to the perception of the BEP.

In order to show how the model might predict the detection of the HP and BEP for different transition bandwidths, a dotted horizontal line has been drawn (arbitrarily) at 40 dB. If this level of residual activation were required for listeners to detect spectral features in residual-activation spectra relative to internal noise (whose effects are not simulated in the mE-C model), then the model would predict that listeners should hear the HP for a transition bandwidth of 8%, but not for 0.5%, 1%, or 64%, and that they should hear the BEP for 0.5%, 1%, and 8%, but not for a 64% transition bandwidth.[6] Thus, the mE-C model diverges from the CAP model in predicting that BEP should be heard strongly in stimuli containing very narrow transition bandwidths.

D. Empirical questions

The following experiments investigated the perceptual properties of the BEP and HP. The results are used to compare the predictions of the CAP, aE-C, and mE-C models. Experiment 1 shows that BEPs produced by narrow interaural phase transitions ($\leqslant 0.5\%$ of the transition frequency) are as perceptually prominent as BEPs produced by wider transitions. Experiment 2 confirms that the pitch evoked by BEP stimuli containing narrow transitions is matched to that of a pure tone at the transition frequency. Experiment 3 shows that two BEPs generated by narrow phase transitions at the formant frequencies of a vowel can be used in combination to evoke the perception of that vowel. These results are difficult for the aE-C model and the CAP model to explain, but receive a straightforward explanation from the mE-C model.

II. EXPERIMENT 1

Cramer and Huggins (1958) measured the ability of listeners to detect the HP as a function of the bandwidth of the interaural phase transition. Because of technical limitations, the narrowest bandwidth which they could explore was 3% of the transition frequency. They found that listeners could detect the HP reliably for transitions of 3% and 6%, but that the pitch was harder to detect when broader transitions were employed. Van Tilburg (1974) reported that the ease of detection of the HP also declines for transitions narrower than 3%. The present experiment extends these investigations by including both the HP and the BEP, each generated with a wide range of transition bandwidths.

A. Stimuli

The stimuli were triplets of 409.6-ms noises containing interaural phase transitions at three different frequencies. They were generated digitally with 16-bit amplitude quantization and a 10-kHz sampling rate and filtered in the frequency domain. To make each stimulus, three white noises of 409.6-ms duration were synthesized using the method recommended by Klatt (1980) in which 16 consecutive numbers from a pseudorandom number generator are summed to form each output sample. These noises were low-pass filtered at 2 kHz. A copy of each noise was further filtered in order to produce a linear phase transition between specified frequencies. The copy was combined with the original to form a stereo file containing an interaural phase transition. The three noises making each triplet contained transitions centered on 500, 600, and 700 Hz. In the case of the BEP, only stimuli which were in-phase below the transition frequency were generated. Each resulting sound file was shaped with 10-ms raised-cosine onset and offset ramps. Finally, the stereo sound files were concatenated to form an ascending sequence of dichotic pitches. This method was repeated to create ten stimuli based on different samples of noise in each of the following conditions. Each of the two dichotic pitches (HP and BEP) was created with eight transition bandwidths, which were 0.5%, 1%, 2%, 4%, 8%, 16%, 32%, and 64% of the transition frequency, giving 16 conditions and 160 stimuli in all. The smallest transition bandwidth, nominally 0.5%, was a single bin in the Fourier transform (2.4 Hz) for all transition frequencies. In the case of the HP, a transition from 0 to 360 degrees in a single bin is equivalent to no transition at all, so the ''0.5%'' members of the HP stimulus set were simply diotic noises. Results from these stimuli are therefore plotted at 0% in Fig. 5 but shall be refered to in the text as ''0.5%'' to preserve the symmetry of the experimental design. In addition to these stimuli, ten further diotic noises were created to form a control condition.

The digital stimuli were converted to analog using a Loughborough Sound Images delta-sigma digital-to-analog converter and presented to listeners via Sennheiser HD414 headphones in a double-walled sound attenuating chamber at 66 dB(A). Stimulus levels were measured with a B&K artificial ear type 4153, with a flat-plate adapter type DB0843, a half-inch microphone type 4134, and a sound-level meter type 2235 on its ''fast'' setting.

B. Procedure

Four listeners, including the first author, who had normal hearing at audiometric frequencies from 0.25–8 kHz inclusive and who had participated in previous psychoacoustic experiments, attended one 40-min session. The session was broken into four 10-min runs in which the 160 experimental stimuli were each presented once and the ten control stimuli were each presented eight times. The random stimulus ordering was changed for each run. Listeners were required to

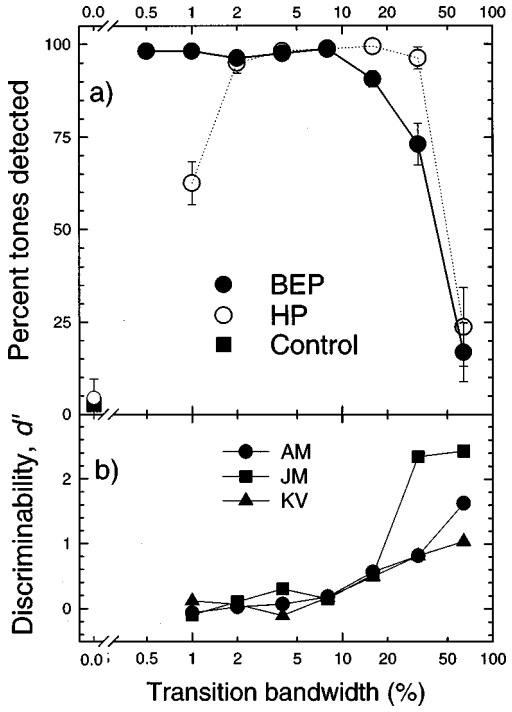

FIG. 5. Upper panel: The percentage of BEPs and HPs detected by four listeners in experiment 1 as a function of the transition bandwidth, which is expressed as a percentage of the transition frequency. Error bars mark ± 1 standard error of the mean. Note that the effective bandwidth was 0% when w was set to 0.5% for the HP stimuli, and so is plotted at 0%. Lower panel: Values of the discrimination index, d', for three listeners who attempted to discriminate the tonal prominence of BEPs with transition bandwidths of 0.5% from BEPs with the larger transition bandwidths plotted on the abscissa.

report whether or not they heard a sequence of ascending tones in the noise via a single key-press on the keyboard of a VDU. No feedback was given.

C. Results

Figure 5(a) shows the percentage of trials on which the tone sequence was detected as a function of the transition bandwidth, with data averaged over listeners. The HP was detected on nearly 100% of trials when the transition bandwidth ranged from 4% to 32% of the transition frequency. The corresponding range for the BEP was 0.5% to 8%. In accordance with van Tilberg's (1974) observations, the proportion of trials on which the HP was detected declined as the transition bandwidth of the HP was reduced. For the widest transition bandwidths, detection declined for both pitches in the way originally observed by Cramer and Huggins (1958) for the HP, although the BEP was detected less frequently than the HP for $w=32\%$.

The control stimuli (diotic noises) yielded an overall false-alarm rate of 3%. Compared to this rate, all other conditions (save the HP with $w=0.5\%$, which was also diotic noise) showed significantly higher detection rates [binomial probability, 160 trials and $p(\text{hit})=0.03$, $p<10^{-6}$ for all conditions with data pooled across listeners].

A two-way analysis of variance covering type of dichotic pitch (BEP versus HP) and transition bandwidth (0.5%, 1%, 2%, 4%, 8%, 16%, 32%, and 64%) showed significant main effects of pitch type [$F(3,1)=76.49$, $p<0.005$] and transition bandwidth [$F(3,7)=70.93$, $p<0.0001$] and a significant interaction between the two [$F(7,21)=125.66$, $p<0.0001$]. Tukey pairwise comparisons indicated that the interaction was produced by differences between detection rates for the BEP and HP at transition bandwidths of 0.5%, 1%, and 32% ($q=35.94$, 13.66, and 8.87, respectively, $p<0.01$). Comparisons among the different transition-bandwidth subconditions of the BEP condition showed that there were no significant differences among bandwidths of 0.5%, 1%, 2%, 4%, 8%, and 16%. For the HP condition, those among the 2%, 4%, 8%, 16%, and 32% subconditions did not differ significantly from one another. All other comparisons (bar 0.5% versus 64% in the HP condition) differed significantly ($p<0.05$).

D. Discussion

Experiment 1 shows that listeners can reliably detect BEP's which are produced by interaural transitions with very narrow bandwidths (0.5%, 1%, and 2% of the transition frequency). This result is consistent with the predictions of the aE-C and mE-C models. The aE-C model requires only that there should be a transition in interaural phase from in-phase to out-of-phase for a dichotic pitch to be heard. Such a transition will form an edge after broadband interaural cancellation; the perceived pitch is predicted by assuming lateral inhibition (Fig. 3). The mE-C model requires only that a frequency channel centered on the transition frequency should contain widely differing interaural phases for a peak to appear at the transition frequency in the residual activation spectrum (Fig. 4). The result is problematic, however, for the account of the BEP offered by the CAP model. As illustrated in Fig. 1(d) and in Fig. 2, the CAP model requires a progressive change in interaural phase across frequency to produce a sharp peak within the transition band. As the transition band narrows, so must the sharp peak. In the limit, the sharp peak disappears (Fig. 2), yet experiment 1 shows that, although the HP declines in this way, the salience of the BEP remains high even at the smallest transition bandwidths.

In order to underline this point, a supplementary experiment was conducted on a separate group of three listeners. These listeners were naive to the purposes of the experiment. In a two-interval forced-choice procedure, they were required to discriminate the loudness/salience of the pitches evoked by pairs of BEP stimuli with different transition bandwidths. On each trial, one stimulus had a transition bandwidth of 0.5%, while the other stimulus had one of the wider transition bandwidths (1%–64%). Listeners indicated which interval contained the stimulus with the clearer or louder pitch. The resulting values of the discriminability index, d', are plotted in Fig. 5(b). None of the listeners were able to discriminate stimuli with bandwidths in the range 1%–8% from stimuli with a bandwidth of 0.5% ($-0.2<d'<0.2$). Thus, BEP stimuli with transition bandwidths ranging from 0.5% to 8% cannot be distinguished from one another on the basis of the strength of the pitch percept. Bandwidths of 16%, 32%, and 64% gave progressively more positive values of d' which, according to the marking scheme employed, indicates that the resulting BEPs were

less salient than those produced by smaller transition bandwidths. Negative d' values, indicating that the 0.5% bandwidth was less salient than the comparison stimulus, were not observed consistently in any condition, showing that there is no perceptible loss of salience for the BEP as transition bandwidth is reduced.

A legitimate criticism of the main part of experiment 1 is that it provides no guarantee that listeners actually heard BEPs, only that they could distinguish stimuli containing a 180-degree interaural phase transition from a diotic noise. Listeners might have made the distinction by noting that the lateralization of BEP stimuli is relatively diffuse, because the noise is out-of-phase above the transition frequency, whereas the lateralization of diotic noise is compact. There are, nonetheless, two reasons for believing that listeners' detection responses were based on hearing dichotic pitches: (1) without feedback, listeners other than the first author had no guidance in their use of alternative cues; (2) for the broader transition bandwidths, which produced equally diffuse localization but a less salient dichotic pitch, detection rates dropped substantially. Nonetheless, it is important to demonstrate that listeners do hear tones of an appropriate frequency when they listen to BEP stimuli synthesized with very narrow transition bandwidths. This was the primary aim of experiment 2.

III. EXPERIMENT 2

In experiment 2, listeners compared the pitches evoked by a range of BEP and HP stimuli with different transition frequencies against that of a single pure tone. If the pitch of a BEP or HP stimulus with transition frequency f is equivalent to that of a pure tone of frequency f, then listeners should judge stimuli with higher transition frequencies to have a higher pitch than the tone, and those with lower transition frequencies to have a lower pitch. If listeners respond in this way to BEP stimuli with very narrow interaural transition bandwidths, and without feedback, it indicates that, contrary to the predictions of the CAP model, a gradual phase transition is not necessary to evoke a pitch which corresponds closely to the transition frequency. Rather, it suggests that listeners detect a narrow band of binaural excitation at the transition frequency, as predicted by the mE-C model. Thus the primary focus of experiment 2 was on the perception of BEP stimuli. The HP stimuli were included for purposes of comparison.

A. Stimuli

The stimuli were generated, presented, and calibrated in a similar manner to those of experiment 1. The HP stimuli had a fixed transition bandwidth of 6% of the transition frequency. The BEP stimuli had transition bandwidths of $\approx$0.4% (a single bin). In separate conditions, the BEP stimuli were prepared with in-phase noise either above (BEPa) or below (BEPb) the transition frequency. Figure 4 shows recovered spectra only for BEPb stimuli. The positions of spectral peaks produced by the mE-C model are determined by the CFs of interaural phase transitions, since the CFs determine the maximally decorrelated parts of the spectrum. Hence, recovered spectra for BEPa stimuli are similar to those plotted in Fig. 4, but contain residual activation below the peak, rather than above it. (Spectra for both orientations of the transition are illustrated in Fig. 7 for the stimuli of Experiment 3.) The transition frequencies of the different stimuli were distributed around 600 Hz, the frequency which Bilsen (1977) and Raatgever and Bilsen (1986) reported to produce the most potent dichotic pitches. The transition frequencies were mistuned from 600 Hz by -64, -48, -32, -16, -8, -4, -2, -1, 0, 1, 2, 4, 8, 16, 32, 48, and 64 Hz. Thus, there were 17 mistunings$\times$3 dichotic pitches$=$51 stimuli in all. Each stimulus consisted of three 409.6-ms dichotic pitches with the same transition frequency (but synthesized from separate samples of noise) alternating with three 409.6-ms, 600-Hz pure tones, beginning with the dichotic pitch. The level of the noise was 66 dB and that of the tone was 29 dB. These values were chosen so that the pure tone and dichotic pitch had approximately equal loudness when the transition bandwidth of the HP was 3%.

B. Procedure

The four listeners who participated in experiment 1 attended ten 30-min sessions. In each session, they listened to each stimulus ten times in a random sequence and classified the dichotic pitch as either higher or lower than the pitch of the pure tone via the keyboard of a VDU.

C. Results

The results for each listener are shown separately in Fig. 6. The fitted curves were derived using a logistic regression based on Eq. (2), where y is the percentage of ''lower pitch'' judgments, x is the mistuning in Hz, and b and k are free parameters, which jointly control the location and steepness of the slope of the curve. Three of the four subjects showed more accurate discrimination of the pitch of the HP stimuli than of the pitch of the BEP stimuli, as reflected in the steeper slopes in the fitted functions. Estimates of the dichotic pitch transition frequencies whose pitches correspond to that of the 600 Hz tone can be determined by evaluating the offset between 600 Hz and the fitted curves as they cross the 50% point on the y axis [Eq. (3)]:

$$y = \frac{100}{1 + ke^{bx}}, \tag{2}$$

$$\text{offset} = \frac{-\log_e k}{b}. \tag{3}$$

Table I lists the transition frequencies (with 95% confidence intervals) at the 50% point of the fitted curves for each of the three binaural conditions. For the three listeners whose fitted functions showed steeper slopes for the HP than the BEP, the confidence intervals of the offset are also smaller. For the HP, the offset in the fitted curve is within 2% of 600 Hz for each of the four listeners, and the mean offset is 0.11% of 600 Hz. For the BEP, the offsets are within 3% of 600 Hz for each listener and the mean offset is within 1%, for each condition.

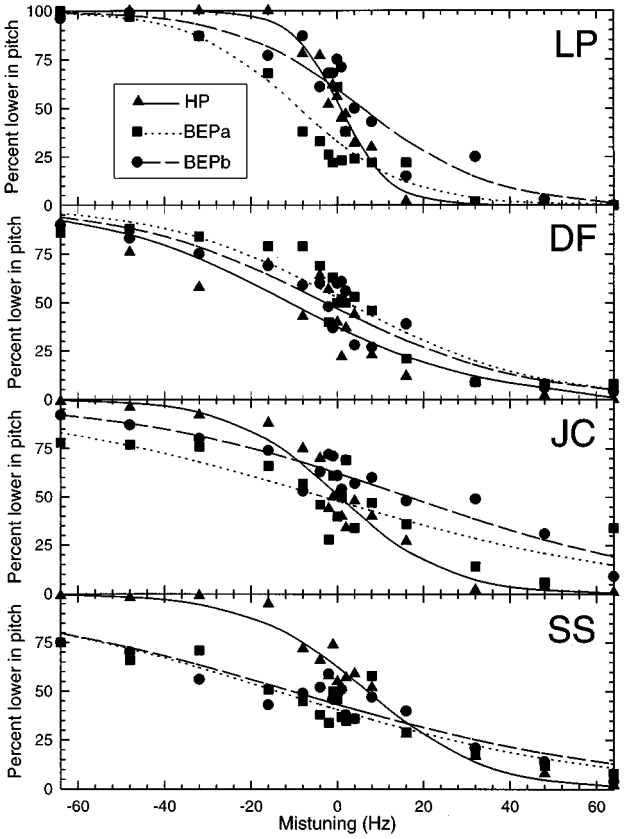

FIG. 6. The percentage of stimuli from each condition in experiment 2 (HP, triangles; BEPa, squares; BEPb, circles) which were judged lower in pitch than a 600 Hz comparison tone, as a function of the mistuning of the transition frequency from 600 Hz, for each subject. Logistic curves, fitted using two free parameters [see Eq. (1)], are also shown for each condition (HP, solid; BEPa, dotted; BEPb, dashed).

D. Discussion

Frijns *et al.* (1986) drew a distinction between the tonal percept which BEP stimuli produce and a low- or high-pass noise percept, which may also be heard. According to the CAP model, if the interaural transition is very narrow, the tonal percept should disappear, leaving only the high- or low-pass percept. The primary objective of experiment 2 was to show that the narrow transitions in the BEP condition of experiment 1 give rise to the perception of a pitch which is tonal in nature and is equivalent to the transition frequency, rather than giving rise only to the perception of a low- or high-pass noise. The important outcome of experiment 2, therefore, is that listeners perceived a tonal pitch in the BEP stimuli which they matched closely to the reference frequency. Moreover, there was only a negligible offset between the fitted logistic curve at its 50% point and 600 Hz. The tolerance of the offset was similar in size to the 1.5% tolerance of matches observed by Frijns *et al.* for BEP stimuli with wider transition bandwidths. Thus, contrary to the predictions of the CAP model, very narrow transition bandwidths do not prevent the perception of a tonal pitch in BEP stimuli.

Three of the four listeners discriminated HP stimuli more accurately than BEP stimuli. In fact, one of these three subjects (SS) showed no discrimination of the BEP during his first seven runs, while performing well with the HP. Data collection was restarted from scratch once he began to perform above chance with BEP stimuli. More accurate discrimination with the HP may reflect the lesser prominence of the peak in interaural decorrelation which is recovered from a BEP compared to that from a HP (e.g., Fig. 4, $w=8\%$).

TABLE II. Vowel percepts formed by different combinations of formant frequencies.

	*F*2	
*F*1	975 Hz	1925 Hz
225 Hz	OO	EE
625 Hz	AR	ER

IV. EXPERIMENT 3

A. Rationale

Experiment 3 provides a further test of the prediction of the mE-C model that BEP's defined by narrow interaural phase transitions give rise to dichotic pitches. The experiment exploited characteristics of a stimulus developed by Culling and Summerfield (1995). They combined two first- and two second-formant frequencies to produce sounds akin to the vowels in British-English pronunciations of the words ''hard'' (AR), ''heed'' (EE), ''haired'' (ER), and ''who'd'' (OO) (Table II). Dichotic stimuli were synthesized containing two BEPs with narrow ($<1.2\%$) interaural phase transitions at pairs of these formant frequencies in order to evoke the perception of vowels in noise. Two sets of dichotic (BEP) stimuli were generated. In one set, the noise was in-

TABLE I. Results of experiment 2. Transition frequencies (in Hz) which were matched by each listener to the 600-Hz comparison tone for each dichotic pitch, with 95% confidence intervals (conf. intvl.) and the % age mistuning which these matches represent.

	HP			BEPa			BEPb		
Listener	Frequency (Hz)	95% conf. intvl.	% age offset	Frequency (Hz)	95% conf. intvl.	% age offset	Frequency (Hz)	95% conf. intvl.	% age offset
LP	600.8	598.8–603.1	+0.13%	591.2	586.4–595.0	−1.46%	605.9	601.2–611.4	+0.98%
JC	600.5	597.9–603.3	+0.08%	596.8	583.0–609.3	−0.53%	616.5	610.6–623.8	+2.75%
DF	589.0	581.1–595.3	−1.82%	602.3	596.4–608.5	+0.38%	597.3	591.6–602.7	−0.44%
SS	607.0	604.3–610.0	+1.12%	586.0	577.6–592.9	−2.32%	589.5	581.3–596.4	−1.75%
$\bar{x}$	599.3		−0.11%	594.1		−0.98%	602.3		+0.38%

phase between the two transition frequencies (BEP1). In the other set, the noise was 180 degrees out-of-phase between the transition frequencies (BEP2).

The formants were centered on 225, 625, 975, and 1925 Hz. Table II shows that no formant on its own uniquely specifies a particular vowel; only in combination do the formants define the different vowels. A limitation in this design is that dichotic pitches are not audible for the majority of listeners above 1500 Hz (Cramer and Huggins, 1958; Yost, 1991). Thus, a formant defined by a BEP at 1925 Hz should not be detectable. However, listeners can obviate this problem. They can identify OO and AR from two audible formants below 1000 Hz (225 and 975 Hz for OO; 625 and 975 Hz for AR). They can identify EE and ER by detecting the lower formant of those vowels (225 Hz for EE and 625 Hz for ER) and determining that a second formant at 975 Hz is absent. Nonetheless, because of this complication, the predictions which follow are made more strongly for OO and AR than for EE and ER.

The CAP model predicts that stimuli containing BEPs with very narrow transition bandwidths are not heard as containing a tonal dichotic pitch. Instead, they are heard as either high- or low-pass noise. Thus, if two transitions are combined in a single stimulus, band-pass or band-stop noise percepts should result. For this reason, the experiment also included two sets of diotic control stimuli which were band-pass and band-stop noises with the cutoffs placed at the formant frequencies. The CAP model makes two predictions for the pattern of identification responses that should be given to these diotic (control) stimuli and to the dichotic (BEP) stimuli. First, the BEP1 and the diotic band-pass stimuli should receive the same pattern of responses; likewise, the BEP2 and diotic band-stop stimuli. Second, neither set of responses should correspond systematically to the intended vowels.

To the extent that the aE-C model draws a strong analogy between the detection of edge pitches by the monaural and binaural systems, it predicts that listeners will hear the intended vowels both in the dichotic (BEP) conditions and the diotic (control) conditions. In the dichotic (BEP) conditions, formants will be defined by binaural edge pitches close in frequency to the interaural phase transitions. In the diotic (control) conditions, formants will be defined by edge pitches at roughly the same frequencies. Thus, the aE-C model predicts that responses in all four conditions should correspond to the intended vowels.

The mE-C model predicts that the dichotic stimuli will elicit percepts of the intended vowels. Excitation will be largely canceled in channels with center frequencies away from the interaural phase transitions, but not in channels close to the frequencies of the transitions themselves, leaving two peaks in the residual activation spectrum which can be interpreted as formants. The model itself does not predict how the diotic (control) stimuli will be identified. However, in our experience the first-order percept of a low-, band-, or high-pass diotic noise is determined by the spectrum of the noise, rather than its edge pitch(es). Therefore, we expect the pattern of identification responses to the diotic (control) stimuli to differ from that to the dichotic (BEP) stimuli, and to be based on listeners interpreting the spectral profiles of the noise bands as if they were the spectral envelopes of vowels. In the case of some stimuli, this strategy should yield percepts that differ systematically from the intended vowels. Thus, the mE-C model predicts that listeners will identify the stimuli as the intended vowels in the dichotic conditions, and will give different, and generally less accurate, patterns of responses in the diotic conditions. Moreover, these latter patterns should differ between the band-pass and band-stop conditions.

TABLE III. Predictions of three models for the accuracy of vowel-identification responses in experiment 3. The entries ''Incorrect (pattern A)'' and ''Incorrect (pattern B)'' imply that different patterns of misidentifications would occur.

	Condition			
	Diotic		Dichotic	
Model	Band-pass	Band-stop	BEP1	BEP2
aE-C	Correct	Correct	Correct	Correct
CAP	Incorrect (pattern A)	Incorrect (pattern B)	Incorrect (pattern A)	Incorrect (pattern B)
mE-C	Incorrect (pattern A)	Incorrect (pattern B)	Correct	Correct

In summary, the CAP model predicts that the dichotic (BEP) stimuli will elicit the same pattern of identification responses as the diotic (control) stimuli (at least for AR and OO), but that neither set of responses will correspond to the intended vowels. The aE-C model also predicts that both the dichotic (BEP) and the diotic (control) stimuli will elicit the same pattern of identification responses, but that the response patterns will correspond to the intended vowels. Finally, the mE-C model predicts that the dichotic (BEP) stimuli will elicit percepts of the intended vowels, while the diotic (control) stimuli will elicit different patterns of identification responses in which the band-stop and band-pass noises are themselves interpreted as defining the spectral envelopes of vowels. These predictions are summarized in Table III.

B. Stimuli

The stimuli were generated, presented, and calibrated in a similar manner to experiments 1 and 2. White noise was synthesized and was filtered in the time domain by a 512-point linear FIR filter in order to produce a 3-dB/oct spectral roll-off. The resulting pink noise was filtered in the frequency domain in order to produce both BEP and diotic control stimuli using the formant frequencies listed in Table II. The dichotic stimuli were, in separate conditions, in-phase between the two transition frequencies (BEP1) or out-of-phase between the two transition frequencies (BEP2). The corresponding diotic control stimuli were band-pass and band-stop noises using the same transition frequencies as cut-off frequencies. Each corresponding pair of stimuli (e.g., BEP1 and band-pass) was generated using the same noise sample. The 180-degree interaural phase transitions and the diotic low-pass and high-pass cutoffs were one analysis bin wide (2.4 Hz), giving transition bandwidths of 0.13%–1.1% of the transition frequency, depending on the center fre-

quency of the transition. Eight stimuli based on different samples of noise were created for each of the four vowels in each of the four conditions, giving 128 experimental stimuli. In addition, four practice stimuli (one for each vowel) were created which incorporated HPs at the formant frequencies with 6% transition bandwidths. The stimuli were presented at 64 dB.

C. Procedure

Five listeners, including the first two authors, all of whom had previously participated in vowel identification experiments, attended a single 40-min session. Each session consisted of four runs. Each run began with 20 practice trials in which the listeners identified the HP practice stimuli five times each, in a random order without feedback. The rest of the run consisted of two blocks. In two runs, the first block contained the diotic control stimuli, while in the other two runs the first block contained the dichotic stimuli. The ordering of the blocks was counterbalanced across runs. The stimuli were presented in a random order, which changed for each run.

Pilot experiments showed that some of the control stimuli compellingly evoked percepts of the vowel in the word ''hoard'' (OR), which lay outside the nominal stimulus set. In order to accommodate this effect, listeners were permitted five response categories, including OR, but were advised that they should not necessarily expect to use all of them equally often. Listeners identified the vowel presented on each trial with a single key-press on a VDU. No feedback about the accuracy of responses was given.

D. Results

Table IV contains the confusion matrices for each of the four conditions pooled across the five listeners. These data were analyzed in three ways: first, to assess overall accuracy; second, to compare the pattern of responses between the four conditions; and third, to compare the patterns of responses with the predictions of a model of vowel identification.

In the dichotic conditions, with the data pooled over subjects, each of the four stimuli in each condition was identified as the intended vowel more often than chance (from binomial probability, $p \ll 0.01$, for each vowel). The response category corresponding to the intended vowel received the greatest proportion of the responses made to each of the eight stimuli. Individually, four of the five listeners identified the stimuli as the intended vowel significantly more often than chance according to binomial probability ($p \ll 0.01$ for 512 trials and 5 choices). The fifth listener's responses were at chance (21% correctly identified). In the diotic control conditions, listeners identified some stimuli consistently as the intended vowel, but made consistently different responses to other stimuli. In the pooled data, the response category corresponding to the intended vowel received the highest proportion of responses made to a stimulus in only five out of the eight possible cases.

The second analysis tested the prediction of the mE-C model that the same pattern of responses would be made in the BEP1 and BEP2 conditions, while two different patterns of responses would be made in the band-pass and band-stop conditions. The patterns of results in the pooled data in each pair of conditions were compared using a chi-squared (χ^2) test. The χ^2 value provides a metric for assessing the degree to which the patterns of responses differ between pairs of conditions. Table V lists the χ^2 values for all pairwise comparisons among the four conditions. The patterns of responses made in the dichotic (BEP1 and BEP2) conditions differ significantly from each other, but only at the $p<0.05$ level, while the patterns in every other pair of conditions differ significantly at the $p<0.001$ level ($\chi^2_{(19)}>43.82$). The values in parentheses are the corresponding values of $\chi^2_{(9)}$ computed from the submatrices containing responses to OO and AR (the vowels for which both formants would have been clearly audible). They show the same pattern as the overall analysis.

The third analysis tested the prediction that listeners would classify the stimuli in the dichotic conditions by locating formants at the frequencies of the interaural phase transitions, but would classify the stimuli in the diotic con-

TABLE IV. Confusion matrices for the four conditions of experiment 3, pooled across five listeners. Figures in parentheses are the numbers of responses predicted by the vowel classifier described in the Appendix.

		Target			
Condition	Response	AR	EE	ER	OO
Band-pass	AR	188 (217)	21 (93)	160 (130)	9 (58)
	EE	1 (6)	24 (18)	2 (15)	0 (11)
	ER	10 (43)	221 (171)	154 (142)	29 (27)
	OO	15 (18)	48 (21)	0 (17)	177 (57)
	OR	106 (37)	6 (17)	4 (16)	105 (168)
Band-stop	AR	3 (35)	1 (7)	1 (24)	35 (10)
	EE	4 (77)	310 (210)	21 (126)	39 (149)
	ER	262 (112)	0 (23)	9 (110)	39 (23)
	OO	16 (73)	9 (68)	9 (47)	199 (116)
	OR	35 (23)	0 (12)	280 (13)	8 (22)
BEP1	AR	178 (219)	14 (2)	36 (73)	13 (17)
	EE	36 (2)	185 (285)	43 (8)	45 (8)
	ER	66 (40)	32 (14)	194 (167)	44 (18)
	OO	15 (21)	80 (4)	25 (15)	201 (200)
	OR	25 (36)	9 (14)	22 (57)	17 (77)
BEP2	AR	171 (219)	13 (2)	33 (73)	20 (17)
	EE	21 (2)	216 (285)	62 (8)	69 (8)
	ER	84 (40)	22 (14)	182 (167)	33 (18)
	OO	27 (21)	63 (4)	25 (15)	187 (200)
	OR	17 (36)	6 (14)	18 (57)	11 (77)

TABLE V. A pairwise comparison of the patterns of data produced by the four conditions of experiment 3 (Table IV). The values given are χ^2 values which are larger the greater the differences. Values in parentheses were computed from the submatrices for the AR and OO stimuli.

	BEP1	BEP2	Band-stop
Band-pass	671.6 (238.5)	791.5 (294.9)	1686.3 (590.6)
Band-stop	905.4 (327.7)	828.7 (288.2)	
BEP2	32.4 (21.1)		

trol conditions by interpreting the spectral profiles of the band-pass and band-stop noises as the spectral envelopes of vowels. In order to evaluate both parts of this prediction on an equal footing, it was necessary to assume that listeners extracted formants not only from the dichotic stimuli but also from the diotic control stimuli. The vowel classifier described in the Appendix includes rules for extracting formants from both types of stimulus. In the dichotic (BEP) stimuli, the classifier assumes that formants are detected at the frequencies of both interaural phase transitions in OO and AR, but only at the lower-frequency transition in EE and ER. In the diotic (control) condition, the classifier assumes that formants are detected at locations within the bands of noise. The classifier includes rules for estimating the similarity of formant frequencies extracted from stimuli to stored templates, and for converting the similarities to predicted numbers of identification responses. These predictions are listed in Table IV as the numbers in parentheses.

The accuracy of predictions was assessed by computing product-moment correlation coefficients between the predicted and observed numbers of identification responses. Analyses were conducted on the pooled data from the two diotic control conditions and on the pooled data from the two dichotic conditions, with 40 identification scores to be predicted in each case. The coefficient of correlation was 0.89 for the dichotic conditions; the vowel classifier correctly predicted that the majority of identification responses made to each of the eight stimuli would correspond to the intended vowel. The coefficient of correlation was 0.61 for the diotic conditions. One way of judging the overall accuracy of this prediction is to note that 11 cells in the matrices for the diotic conditions in Table IV received 100 or more responses. The classifier correctly predicted that 8 of these 11 would receive more than 100 responses. A more critical test is to establish whether the accuracy of prediction declines when the rules for locating formants in the diotic and dichotic conditions are reversed. If it is assumed that formants are located at the band edges in the diotic conditions, the correlation between observed and predicted numbers of responses falls from 0.61 to 0.39. If it assumed that formants are located within the noise bands in the dichotic conditions, the correlation falls from 0.89 to 0.47. Overall, therefore, although the classifier does not make completely accurate predictions particularly for the diotic conditions, its performance is consistent with the prediction of the mE-C model that listeners locate formants at the frequencies of the interaural phase transitions in the dichotic conditions, but within the noise bands in the diotic conditions.

E. Discussion

1. Predictions of the CAP model

The CAP and mE-C models make divergent predictions with respect to experiment 3. The CAP model predicts that listeners should not be able to identify the intended vowels in the dichotic stimuli, primarily because the bandwidth of the interaural transitions is too narrow to generate peaks at the formant frequencies in across-frequency scans.

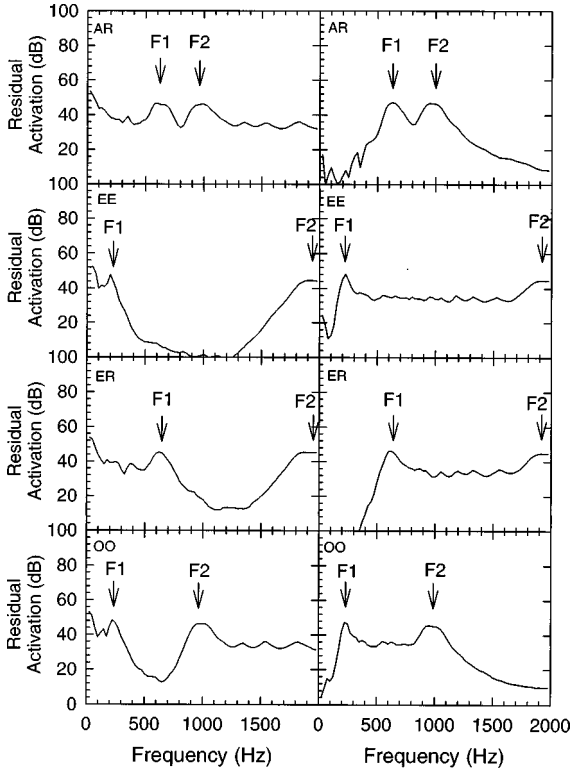

FIG. 7. Residual-activation spectra from the mE-C model for the BEP1 stimuli (left column) and BEP2 stimuli (right column) used in experiment 3, showing peaks at the two transition frequencies (marked with arrows) in each stimulus. Descending each column, stimuli are identified as exemplars of ''AR,'' ''EE,'' ''ER,'' and ''OO.''

If dichotic pitches are not heard, the CAP account predicts that listeners should hear the dichotic stimuli as band-pass and band-stop noises, in which case the dichotic stimuli should have received the same pattern of identification responses as the diotic stimuli. In fact, the patterns differed significantly from one another (Table V).

2. Predictions of the mE-C model

The results of experiment 3 are broadly consistent with the predictions of the mE-C model. Listeners gave identification responses to the dichotic stimuli that were consistent with detecting formants at the frequencies of the interaural phase transitions in OO and AR, and with detecting a low-frequency formant in EE and ER, while noting the absence of a high-frequency formant. Figure 7 shows the recovered spectra generated by the model in response to each of the stimuli. Clearly, the model performs ''too well'' at high frequencies: the recovered spectra contain two peaks at the frequencies of the interaural phase transitions of all eight stimuli. Thus, these spectra are incompatible with demonstrations that listeners are unable to detect interaural phase transitions at frequencies as high as 2000 Hz (Yost, 1991). Bernstein and Trahiotis (1996) have shown that the deterioration of binaural masking release with frequency can be modeled by incorporating the change from phase locking to the waveform to phase locking to the envelope with increasing frequency. The mE-C model uses the Meddis (1986, 1988) hair-cell model for this purpose. That model substantially overestimates the degree of phase locking at high fre-

quencies. A peripheral model which is more accurate in this respect should yield recovered spectra that reflect the information available to listeners more accurately.

V. GENERAL DISCUSSION

The Huggins pitch (HP) is produced by introducing an interaural phase transition of 360 degrees over a narrow bandwidth to a diotic noise. A closely related phenomenon, the binaural edge pitch (BEP), is generated with an interaural phase transition of 180 degrees. In each case, listeners hear a tone with a pitch which matches the center frequency of the transition, suggesting that a binaural process produces an output spectrum with a single spectral peak at the transition frequency. The experiments described in this paper demonstrate that most aspects of these phenomena do not require a special explanatory framework beyond that provided by the general principles of binaural unmasking which are embodied in the mE-C model. The model applies equalization of level and internal delay in each frequency channel independently and then cancels by interaural subtraction. This process detects the degree of interaural decorrelation in each frequency channel. It yields an output spectrum which is closely related to a plot of interaural decorrelation by frequency. That spectrum contains a single peak at the transition frequency of HP and BEP stimuli. An attraction of accounting for the HP and BEP with the mE-C model is that a satisfactory explanation can thus be achieved without invoking additional perceptual processes. This simplicity is not shared by the aE-C and CAP models.

The aE-C model incorporates a conventional model of binaural unmasking (broad-band E-C). However, in the case of the BEP the output of this process is a high- or low-pass noise, so the aE-C model needs to invoke central lateral inhibition as an additional process in order to produce a spectrum with a single peak near the transition frequency. The CAP model incorporates a conventional multi-channel cross-correlation process, typical of models of binaural sound localization, but in order to find spectra with peaks at the transition frequencies of HP and BEP stimuli the model requires an additional across-frequency scanning process.

The following paragraphs review the strengths and weakness of the three alternative models of dichotic-pitch phenomena in the light of the results reported in this paper.

A. The aE-C model

The aE-C model invokes two processes to account for the BEP: (1) broadband E-C and (2) central lateral inhibition on the residue after cancellation. Broadband E-C is equivalent to applying the same equalizing and cancelling operations in each frequency channel. This process differs from processes performed by the mE-C model where different delays can be applied in different channels. Strong evidence in favor of using different delays is that the mE-C model provides the only complete account of the Fourcin pitch (Fourcin, 1970; Culling *et al.*, 1998). Supporting evidence comes from demonstrations that the model provides a straightforward account of the HP and BEP, as shown in the present paper, and of a range of other cases of binaural masking release involving broadband signals (Culling and Summerfield, 1995). Thus the balance of evidence favors the conclusion that binaural analysis involves independent E-C in different frequency channels, rather than broadband E-C.

Released from the constraints imposed by broadband E-C, it is not necessary to invoke central lateral inhibition to account for the BEP. The involvement of central lateral inhibition was originally supported by evidence that the pitch frequency of the BEP is offset from transition frequency by 3%–8%, as are monaural edge pitches from the cutoff frequencies of low- and high-pass noises. Subsequent experiments by Hartmann (1984a) and by Frijns *et al.* (1986) failed to corroborate the involvement of central lateral inhibition and failed to replicate the offset in pitch matches. Nonetheless, the existence of the binaural coherence edge pitch (Hartmann, 1984b) provides independent evidence of the existence of central lateral inhibition. These results would be reconciled if the mE-C process generated the primary auditory representation of the BEP, and central lateral inhibition enhanced edges in that representation. In this case, the largest contribution to the BEP would derive from the primary representation with a small additional contribution from lateral inhibition which results from less effective cancellation of the N_π noise.

Further evidence against the aE-C model was provided by experiment 3. This model predicts that listeners hear low- or high-pass noise when presented with a BEP stimulus. In experiment 3, two transitions at frequencies which corresponded to vowel formant frequencies were used to produce stimuli which sounded like vowels in noise. According to the aE-C model, binaural analysis converts these stimuli into band-pass or band-stop noises, depending on whether the noise was in-phase or out-of-phase between the two transition frequencies (and on whether the binaural system adds or subtracts the signals at the two ears). Then, lateral inhibition enhances the spectrum near the transition boundary exactly as for a monaural edge-pitch. Thus, the model predicts that listeners should give similar responses to these stimuli as to diotic band-pass and band-stop noises, and that listeners should hear the appropriate vowels in both cases. In fact, listeners heard the appropriate vowels from the dichotic stimuli, but produced a systematically different pattern of responses for the band-pass and band-stop noises. The aE-C model can only be reconciled with these results if the lateral inhibition process which it uses is assumed to be much stronger than the lateral inhibition process(es) which generate the monaural edge pitch. In conclusion, we suggest that the aE-C model should be rejected as an account of the HP and BEP, but that the conditions giving rise to central lateral inhibition should be explored further.

B. The CAP model

The CAP model is based on four interrelated ideas. First, an array of interaural cross-correlation functions is computed forming a two-dimensional central activity pattern in frequency and interaural delay. Second, attention can be focused on individual across-frequency scans taken at particular interaural delays. Third, in certain respects, these across-frequency scans, or ''central spectra,'' can be interpreted like monaural spectra; in particular, sharp peaks in scans, but not

broader peaks, give rise to the perception of tones. Fourth, the interaural delay of a scan determines the lateralization of the features detected within it.

Bilsen and his colleagues have argued that the CAP model provides a basis for explaining three aspects of the HP and BEP: (1) their frequencies; (2) their masking patterns; and (3) their lateralization. In the paragraphs which follow, we argue that there are plausible alternative accounts of the first two of these aspects, though not yet of the third.

1. Detection and pitch matching of the HP and BEP

A weakness of the CAP model is that it includes no explicit mechanism for choosing which across-frequency scan receives a listener's attention and determines the sound pattern that is heard. Instead, the model relies on some scans containing features of sufficient prominence that they command attention. In this respect, the explanatory power of the model is materially reduced when it is implemented with realistic frequency and temporal resolution in place of the idealized parameters used by Raatgever and Bilsen (1986). With this revision, features that appeared as prominent peaks in scans generated by the original version of the model may be diffused to the point where they should no longer command attention.

A critical test is provided by BEP stimuli created with very narrow interaural phase transitions. The original version of the model predicts that such dichotic pitches should be harder to detect than pitches created from broader transitions. The ''smoothed'' version of the model incorporating realistic frequency resolution predicts that dichotic pitches should not be heard in such stimuli. However, experiment 1 showed that the pitch sensation produced by such stimuli is as strong as that produced by stimuli with wider transitions. Experiment 2 showed that this pitch closely matches that of a tone whose frequency is equal to the transition frequency. Experiment 3 showed that stimuli containing two such transitions at frequencies which correspond to vowel formant frequencies can be identified as vowels in noise. Thus, the results of experiments 1, 2, and 3 contradict the predictions of the CAP model.

Further evidence against the predictions of the CAP model was provided by Culling and Summerfield (1995), Hukin and Darwin (1995), and Darwin and Hukin (1997). They demonstrated that listeners are very poor at grouping energy in separate frequency regions which share the same interaural delay. An across-frequency scanning mechanism of the kind used in the CAP model would have been able to perform such across-frequency grouping.

2. Measurements of the CAP

Raatgever (1980), Raatgever and Bilsen (1986), and Frijns *et al.* (1986) used measures of binaural masking release to support the idea that listeners can attend to individual across-frequency scans in the CAP. They measured the threshold for detection of a diotic tone in the presence of a dichotic pitch stimulus, which acted as a masker. It was argued that listeners could restrict attention to a single spatial location (the center) while performing the task. The CAP was sampled at specific frequency/internal-delay coordinates

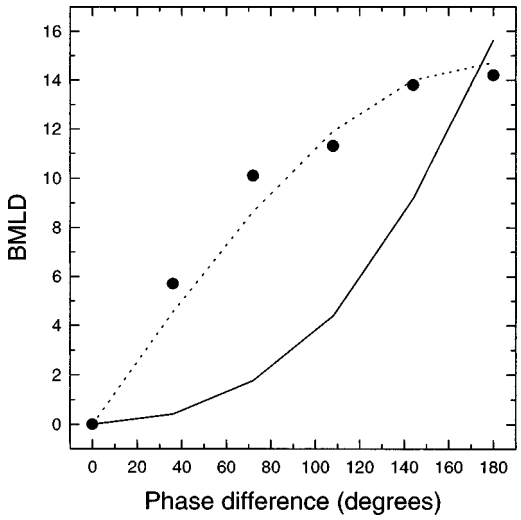

FIG. 8. The relationship between BMLDs and the relative interaural phase of masker and tone. The filled symbols are empirical data from Jeffress *et al.* (1952), the dashed line was fitted using Eq. (4), and the solid line was fitted using Eq. (7). Data from Jeffress *et al.* (1952) were digitized from a scanned image of their Fig. 1(f).

by using target tones of different frequencies and by applying interaural delays to the masker in order to place different parts of the CAP in the center of auditory space. For each interaural delay, thresholds were measured at a range of different frequencies. The BMLDs (re:NoSo) were plotted as a function of frequency and were compared with across-frequency scans generated by the CAP model. The resulting scans were qualitatively similar to thresholds predicted from the CAP [using Eq. (7) in Raatgever and Bilsen, 1986].

Potentially, these results provide powerful evidence listeners can attend to individual across-frequency scans, and thus that they are in a position to detect the features in such scans that support the perception of dichotic pitches. However, the following analysis demonstrates that the pattern of masking release across frequency reported by Raatgever (1980), Raatgever and Bilsen (1986), and Frijns *et al.* (1986) can be predicted without invoking the assumption that the listener has attended to a particular spatial location. There is an established relationship between the BMLD and the phase difference between the interaural phases of the target tone (ϕ_s) and the masking noise (ϕ_n). Durlach and Colburn (1978) pointed out that ''to a rough approximation [the BMLD] depends only on the phase difference ($\phi_n - \phi_s$) rather than on the individual values of ϕ_n and ϕ_s.'' This relationship is illustrated in Fig. 8 (filled symbols) with data from Jeffress *et al.* (1952). The interaural delays used in that study have been converted to the corresponding phases in the figure. A satisfactory fit to the data can be produced using Eq. (4) and is illustrated by the dotted line in Fig. 8:

$$\mathrm{BMLD} = \mathrm{BMLD}_{\max} |\sin(\phi/2)|. \tag{4}$$

In this equation, ϕ is the difference in interaural phase between signal and noise ($\phi_n - \phi_s$) and $\mathrm{BMLD}_{\max}$ is the BMLD in dB which is obtained with NoSπ versus NoSo. For a dichotic pitch stimulus, the interaural phase of the masker is frequency dependent. The threshold for detecting a tone of

frequency f is therefore dependent on the interaural phase of the masker at that frequency, ϕ_f. If, as in the experiments in question, an interaural delay τ_e has also been applied, Eq. (5) is required:

$$\phi_n = 2\pi f \tau_e + \phi_f. \tag{5}$$

Since $\phi_s = 0$, this is also the expression for $\phi_n - \phi_s$. Substituting Eq. (5) into Eq. (4), we obtain

$$\text{BMLD} = \text{BMLD}_{\max} |\sin([2\pi f \tau_e + \phi_f]/2)|. \tag{6}$$

For comparison, it is necessary to substitute Eq. (7) from Raatgever and Bilsen (1986) into their Eq. (6) to obtain

$$\text{BMLD} = 10 \log_{10}\left[\frac{10^{\text{BMLD}_{\max}/10}}{(10^{\text{BMLD}_{\max}/10} - 1)[\frac{1}{2} + \frac{1}{2}\cos(2\pi f \tau_i + \phi_f)] + 1}\right], \tag{7}$$

where τ_i is the internal delay. However, since this internal delay is interrogated using an interaural delay applied to the noise, $\tau_i = \tau_e$.

Although Eqs. (6) and (7) look different, each expression relates the BMLD monotonically to the difference in interaural phase ($0 < \pi < \phi$) between the tone and the masker at the frequency of the tone. The equations differ in that Eq. (6) defines a function whose slope decreases with increases in the phase difference, while Eq. (7) defines a function of increasing slope. In Fig. 8, the solid line was derived from Eq. (7). The accuracy of its fit to the data of Jeffress *et al.* (1952) can be compared with the empirical fit of Eq. (4) which is plotted as the dotted line.

Figure 9 compares the accuracy with which Eqs. (6) and (7) predict the BMLD data obtained by Frijns *et al.* (1986) using a BEP stimulus as the masker. The transition frequency was 600 Hz with a transition bandwidth of 36 Hz (6%). The open circles plot the empirical BMLDs; the predictions from Eq. (6) are shown as dotted lines, those from Eq. (7) as continuous lines. The data in the top panel were obtained with no interaural delay applied to the masker, those in the middle panel with a delay of 0.83 ms, and those in the bottom panel with a delay of 1.25 ms. This figure may be compared with Fig. 9 of Frijns *et al.* (1986, p. 449). The predictions derived in the present paper, based empirically on the relationship between BMLD and $\phi_n - \phi_s$, are as accurate as those based on the CAP model.

The foregoing analysis shows that it is not necessary to assume that listeners attended to a particular spatial location during the BMLD experiments of Raatgever (1980), Raatgever and Bilsen (1986), and Frijns *et al.* (1986). Therefore, the results of those experiments do not prove that listeners can attend to particular across-frequency scans in the CAP. Rather, they confirm that the size of the BMLD is largely determined by the relationship between the interaural phases of signal and masker within the frequency channel occupied by the signal.

3. The lateralization of the HP and BEP

It is an important feature of the original version of the CAP model that it can account for the lateralization of the HP and BEP, because the sharp peaks in across-frequency scans that are judged to be responsible for these dichotic pitches occur at the particular internal delays which correspond to the perceived lateralization(s) of the pitches. Raatgever and Bilsen (1986) demonstrated that the reported lateralizations for HP stimuli correspond closely with the predictions of the CAP model. They also demonstrated that the lateralization can be shifted in a consistent fashion by applying additional interaural delays to stimuli. Frijns *et al.* (1986) reported difficulty in collecting similar data for the BEP. Using a rather complex procedure, they confirmed that two of three listeners frequently perceived lateralizations consistent with the peak at +1.25-ms internal delay in Fig. 2(a). However, these listeners also frequently heard centrally located images. The third listener reported centrally located images almost exclusively.

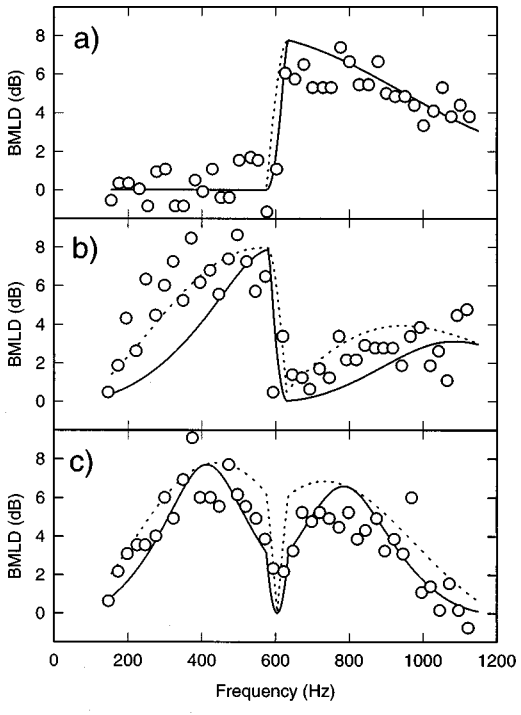

FIG. 9. The empirical and predicted BMLDs from Frijns *et al.* (1986) for a BEP masker (600-Hz transition frequency) and a diotic tone. In panels (a), (b), and (c) interaural delays of 0, 0.83, and 1.25 ms have been applied to the masker. The open symbols are the empirical data from Fig. 9 of Frijns *et al.* The solid lines are predictions based on their model [Eq. (7)]. The dotted lines are predictions based on the phase difference between tone and masker [Eq. (6)]. In each case, $\text{BMLD}_{\max}$ is a function of frequency which was derived by fitting a fourth-order polynomial function to the NπSo data of listener FB in Raatgever and Bilsen (1986, Fig. 9). The empirical data were digitized from scanned images of the published figures.

The mE-C model is not intended to predict the lateralization of sounds and we have not found a satisfactory way to extend it to predict the lateralization of the HP and BEP.[7] Thus the original version of the CAP model provides the only account of the lateralization of the HP. Its predictions with respect to the lateralization of the BEP are also supported by the data, but to a more limited extent.

VI. CONCLUSIONS

The arguments presented in this paper reinforce the conclusion that the aE-C model is inadequate to account for the

characteristics of dichotic pitches because of the limited explanatory power of broadband E-C. However, we believe that one component of that model, central lateral inhibition, warrants further study. The relative merits of the CAP and mE-C models can be summarized as follows. The mE-C model has the weakness that it does not predict the lateralization of dichotic pitches. The CAP model, on the other hand, accurately predicts the lateralization of the HP and to a lesser extent that of the BEP. In every other respect that we have investigated, the mE-C model provides a more satisfactory account. First, it automatically produces an output spectrum which is qualitatively consistent with the spectral features perceived in dichotic pitches. In comparison, the CAP model is under-specified in that it contains no explicit mechanism for choosing which of the available array of central spectra determines what is perceived, nor for distinguishing peaks in scans which are perceived from those which are not perceived. Second, the mE-C model is more parsimonious in that it does not assume any perceptual processes beyond those which are strictly necessary to account for binaural masking release, while the CAP model incorporates an across-frequency scanning process which is not required to account for any independent findings. Indeed, Culling and Summerfield (1995) and Hukin and Darwin (1995) sought independent evidence of such a process but found no evidence for its existence. Third, the mE-C model was designed with, and operates successfully using, realistic parameters of frequency and temporal resolution, while the CAP model is not robust when implemented with realistic parameters (Figs. 1 and 2). Fourth, the mE-C model correctly predicts the range of transition bandwidths over which the HP and BEP are audible, while the CAP model fails to predict that the BEP is clearly audible with very narrow transition bandwidths (experiments 1–3; Figs. 2 and 4).

These arguments favor the explanation for the Huggins pitch and the binaural edge pitch offered by the mE-C model. However, before its account can be fully accepted, the reasons for the perceived lateralization of dichotic pitches will need to be understood.

ACKNOWLEDGMENTS

The authors thank Johann Raatgever and Steve Colburn for helpful discussions and for suggesting improvements to Experiment 3, Alain de Cheveigné for useful comments on an earlier draft of this paper, and Frans Bilsen and Bill Hartmann for their thorough and insightful reviews.

APPENDIX: VOWEL CLASSIFIER

The pooled identification responses of listeners in experiment 3 were predicted by simple template-matching models of vowel classification derived from the PEAK procedure described by Assmann and Summerfield (1989). The templates were the center frequencies of the first two formants ($F1$ and $F2$) of each of the five vowels that were available as response categories. These frequencies were averages of measures from the speech of four adult male talkers of British English, each of whom produced five tokens of each vowel in an /h/–vowel–/d/ context. These values were used directly in modeling listeners' responses to the diotic band-pass and band-stop stimuli. They are listed in Table AI as the ''diotic'' templates. When considering the dichotic (BEP) stimuli, it was necessary to accommodate the strategy which listeners were presumed to have adopted; that is, of identifying OO and AR from both interaural transitions, while identifying EE and ER from the lower transition only, while noting the absence of a higher transition (Sec. IV A). Here $F1$ and $F2$ had the same values in the ''dichotic'' templates for OO, AR, and OR as in the diotic templates for those three vowels. For EE and ER, however, the frequency of $F1$ was assigned to *both* $F1$ and $F2$. These values are listed in Table AI as the ''dichotic'' templates. Frequencies for two formants were estimated in each stimulus in the diotic (control) conditions and in the dichotic (BEP) conditions in the following ways. In the diotic (control) conditions, it was assumed that the perceived locations of the formants must be well within the noise bands. Various different schemes for placing the formants were explored, but the results of the modeling were found to be relatively insensitive to this variable. In the *band-pass* condition, $F1$ was taken as the frequency one-quarter of the way through the noise band above its low-frequency edge, while $F2$ was taken as the frequency three-quarters of the way through the noise band. In the *band-stop* condition, $F1$ was taken as the mid-frequency of the lower-noise band. Where a band started at 0 Hz, its lower edge was taken as 150 Hz for the purpose of this calculation in order to avoid estimates of $F1$ outside the range found in adult male speech. $F2$ was taken as the mid-frequency of the upper noise band. Where a band extended to the Nyquist frequency, its upper edge was taken as 2250 Hz for the purpose of this calculation in order to avoid estimates of $F2$ outside the range found in adult male speech. In the BEP1 and BEP2 conditions, the frequencies of $F1$ and $F2$ were taken as the center frequencies of the two 180-degree interaural phase transitions for OO and AR; for EE and ER, the center frequency of the lower-frequency transition was assigned to both $F1$ and $F2$. The spectral distance d_i between stimulus s and each of the templates i was calculated using Eq. (A1):

TABLE AI. Frequencies in Hz of the templates corresponding to each response category and of the formant frequencies estimated in the four classes of stimuli used in experiment 3.

	Vowels									
	AR		EE		ER		OO		OR	
	$F1$	$F2$	$F1$	$F2$	$F1$	$F2$	$F1$	$F2$	$F1$	$F2$
	Diotic templates (Hz)									
	658	1001	269	2115	540	1640	281	1140	362	695
	Dichotic templates (Hz)									
	658	1001	269	269	540	540	281	1140	362	695
Condition	Formant estimates (Hz)									
Band-pass	713	888	650	1500	950	1600	413	788	⋯	⋯
Band-stop	388	1613	188	2088	388	2088	188	1613	⋯	⋯
BEP1	625	975	225	1925	625	1925	225	975	⋯	⋯
BEP2	625	975	225	1925	625	1925	225	975	⋯	⋯

$$d_i = \sqrt{\frac{(\log_{10}(f_{1s}) - \log_{10}(f_{1i}))^2 + (\log_{10}(f_{2s}) - \log_{10}(f_{2i}))^2}{2}}, \quad \text{(A1)}$$

where f_{1s} and f_{2s}, are the frequencies in Hz of the first and second formants estimated in the stimulus, and f_{1i} and f_{2i} are the frequencies in Hz of the first and second formants for template i. The number of responses, n_i, out of a total of 320 given to response category i was calculated using Eq. (A2) which incorporates the idea that the proportion of responses given to response category i is a function not only of the similarity of the stimulus to template i but also of its similarity to the other four templates:

$$n_i = 320\left(\frac{e^{-d_i/K}}{\Sigma_{i=1}^{5} e^{-d_i/K}}\right). \quad \text{(A2)}$$

The constant K was adjusted to minimize the squared error between predicted and observed numbers of responses. The resulting value of K was 0.1. The resulting values of n_i are tabulated in parentheses in Table IV.

[1]Several taxonomies of dichotic pitches can be defined. The one which is set out in the Introduction regards the ''MPS Pitch'' described by Bilsen (1977) and Raatgever and Bilsen (1986) as a multiple Huggins pitch which inherits its properties from the component pitches; since the Huggins pitch is the most salient of the pitches in our classification, a stimulus with many Huggins pitches is proportionately more salient. On the other hand, we have preserved the distinction between the Huggins pitch and the binaural edge pitch, although the distinction may be artificial. The binaural coherence edge pitch, described in an abstract by Hartmann (1984b) and in Sec. I A, of this paper, may be more deserving of separate classification.

[2]In generating the CAP, cross correlation was employed as a simple mathematical surrogate for a process of neural coincidence detection using a range of different delays. Jeffress (1948) suggested that localization on the basis of interaural time delays was based on an array of units in the medial superior olive which are connected to the two ears by axons with varying transmission times and which respond to simultaneous action potentials arriving from the two ears. In such a network, a given neuron will selectively respond to sound of a given frequency, determined by the place in the two cochleae from which it receives innervation, and a given interaural delay, determined by the relative delays imposed by the converging axons.

[3]Raatgever (1980) and Raatgever and Bilsen (1986) applied a two-dimensional envelope to the CAP which emphasized frequencies around 600 Hz and delays around 0 ms (Raatgever, 1980, eqs. IV.1 and IV.2). However, the delay dimension of this envelope has no effect upon the peak-to-valley ratio within an individual scan, and it is this ratio which defines a scan as ''well-modulated'' in the CAP model, causing that scan to command attention. Consequently we did not include the weighing in our illustrations of the CAP model.

[4]Those previous studies, which employed analog methods of stimulus generation (Cramer and Huggins, 1958; Bilsen, 1977; Raatgever and Bilsen, 1986), reported transition bandwidths which reflected the bandwidth over which half of the phase transitions took place. This specification was used because the transitions were not linear with frequency. Rather their steepness declined away from the center of the transition, gradually asymptoting to zero. Consequently no bandwidth for the complete transition could be specified. In the present report, bandwidths for those studies will be reported in the same way, but bandwidths for the digitally generated linear phase transitions employed here will be specified as for the complete transition (i.e., 6% in the present study $\approx$3% in Cramer and Huggins).

[5]Such independence is necessary, given that the model does not perform phase equalization, to account for two related observations: (1) a large release from masking occurs when speech is presented in the NπSo condition relative to NoSo (Licklider, 1948); and (2) this amount of masking release is greater than that observed when the noise is given a fixed interaural delay with respect to the speech (i.e., NπSo relative to NoSo) (Carhart *et al.*, 1968; Levitt and Rabiner, 1967a, b). If the model were constrained to apply the same cancelling delay in all channels, it would erroneously predict that more masking release would be measured in case 2 than case 1.

[6]The split peak which occurs for the HP when $w = 64\%$ occurs whenever the transition band is approximately one octave. When the transition band is narrower, the phase transition produces interaural decorrelation, which is detected by the mE-C model as a spectral peak. However, when the transition extends over an octave the phase transition is approximately equivalent to an interaural delay within the transition region. The duration of the delay is half the period of the center frequency of the transition band. So, for a 600-Hz transition, the delay is 3.33 ms. The model can apply this delay within the transition region and largely cancel the noise. At the edges of the transition region, there are sharp changes in interaural delay (between 0 and 3.33 ms), which prevent the model from performing cancellation and so cause lateral peaks in the recovered spectrum on either side of the transition. Informal listening indicates that these peaks can sometimes be heard.

[7]One way in which the mE-C model might be extended to predict the lateralization of the HP and BEP is as follows. The binaural detection of a tone could trigger perceptual segregation mechanisms that label the frequency channels close to the transition as containing a separate auditory object from the background noise. The auditory system might lateralize the perceived tone by pooling the cross-correlation functions in the labeled channels and identifying the largest peak in the pooled function. This approach is consistent with demonstrations (1) that segregation of concurrent sources of sound occurs prior to lateralization (e.g., Hill and Darwin, 1996), and (2) that the lateralization of broad-band sounds can be predicted from the interaural delay of the largest peak in pooled cross-correlation functions (Shackleton *et al.*, 1992). However, this strategy does not predict the lateralization of the HP and BEP correctly, because the cross-correlation function in the frequency channel centered on the transition of a HP stimulus contains a peak at an interaural delay of 0 ms, whereas the tone is clearly lateralized to one side or the other.

Assmann, P. F., and Summerfield, A. Q. (**1989**). ''Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency,'' J. Acoust. Soc. Am. **85**, 327–338.

Bernstein, L. R., and Trahiotis, C. (**1996**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Bilsen, F. A. (**1977**). ''Pitch of noise signals: Evidence for a 'central spectrum,' '' J. Acoust. Soc. Am. **61**, 150–161.

Bilsen, F. A., and Goldstein, J. L. (**1974**). ''Pitch of dichotically delayed noise and its possible spectral basis,'' J. Acoust. Soc. Am. **55**, 292–296.

Carhart, R., Tillman, T. W., and Johnson, K. R. (**1968**). ''Effects of interaural delays on masking by two competing signals,'' J. Acoust. Soc. Am. **43**, 1223–1230.

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

Colburn, H. S. (**1977**). ''Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,'' J. Acoust. Soc. Am. **61**, 525–533.

Cramer, E. M., and Huggins, W. H. (**1958**). ''Creation of pitch through binaural interaction,'' J. Acoust. Soc. Am. **30**, 858–866.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Culling, J. F., Marshall, D. H., and Summerfield, Q. (**1998**). ''Dichotic pitches as illusions of binaural unmasking. II. the Fourcin pitch and the Dichotic Repetition Pitch,'' J. Acoust. Soc. Am. **103**, 3527–3539.

Darwin, C. J., and Hukin, R. W. (**1997**). ''Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity,'' J. Acoust. Soc. Am. **102**, 2316–2324.

Durlach, N. I. (**1960**). ''Note on the equalization-cancellation theory of binaural masking level differences,'' J. Acoust. Soc. Am. **32**, 1075–1076.

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).

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

Fastl, H. (**1971**). ''Über Tonhöhenempfindungen bei Rauschen,'' Acustica **25**, 350–354.
Fastl, H., and Stoll, G. (**1979**). ''Scaling of pitch strength,'' Hearing Res. **1**, 293–301.
Fourcin, A. J. (**1970**). ''Central pitch and auditory lateralization,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, The Netherlands).
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.
Guttman, N. (**1962**). ''Pitch and loudness matches of a binaural subjective tone,'' J. Acoust. Soc. Am. **34**, 1996(A).
Hartmann, W. M. (**1984a**). ''A search for central lateral inhibition,'' J. Acoust. Soc. Am. **75**, 528–535.
Hartmann, W. M. (**1984b**). ''Binaural coherence edge pitch,'' J. Acoust. Soc. Am. **75**, K10.
Hill, N. I., and Darwin, C. J. (**1996**). ''Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning,'' J. Acoust. Soc. Am. **100**, 2352–2364.
Houtgast, T. (**1972**). ''The psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.
Hukin, R. W., and Darwin, C. J. (**1995**). ''Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel,'' J. Acoust. Soc. Am. **98**, 1380–1387.
Jeffress, L. A. (**1948**). ''A place theory of sound localization,'' J. Comp. Physiol. Psychol. **61**, 468–486.
Jeffress, L. A. (**1972**). ''Binaural signal detection: vector theory,'' in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic, New York).
Jeffress, L. A., Blodgett, H. C., and Deatherage, B. H. (**1952**). ''The masking of tones by white noise as a functions of the interaural phases of both components. I. 500 cycles,'' J. Acoust. Soc. Am. **24**, 523–527.
Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.
Klein, M. A., and Hartmann, W. M. (**1981**). ''Binaural edge pitch,'' J. Acoust. Soc. Am. **70**, 51–61.
Levitt, H., and Rabiner, L. R. (**1967a**). ''Binaural release from masking for speech and gain in intelligibility,'' J. Acoust. Soc. Am. **42**, 601–608.
Levitt, H., and Rabiner, L. R. (**1967b**). ''Predicting binaural gain in intelligibility and release from masking for speech,'' J. Acoust. Soc. Am. **42**, 620–629.
Licklider, J. C. R. (**1948**). ''The influence of interaural phase relations upon the masking of speech by white noise,'' J. Acoust. Soc. Am. **20**, 150–159.
Meddis, R. (**1986**). ''Simulation of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.
Meddis, R. (**1988**). ''Simulation of auditory-neural transduction: further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). ''An efficient auditory filterbank based on the gammatone function,'' paper presented to the I.O.C. speech group on auditory modelling at R.S.R.E, December 14–15.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). ''Spiral VOS final report, Part A: The auditory filter bank,'' Cambridge Electronic Design, Contract Report (A.P.U. 2341).
Raatgever, J. (**1980**). ''On the Binaural Processing of Stimuli with Different Interaural Phase Relations,'' doctoral dissertation, Delft University of Technology.
Raatgever, J., and Bilsen, F. A. (**1986**). ''A central theory of binaural processing. Evidence from dichotic pitch,'' J. Acoust. Soc. Am. **80**, 429–441.
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1992**). ''Across-frequency integration in a model of lateralization,'' J. Acoust. Soc. Am. **91**, 2276–2279.
Small, A. M., and Daniloff, R. G. (**1967**). ''Pitch of noise bands,'' J. Acoust. Soc. Am. **41**, 506–512.
van Tilburg, J. J. (**1974**). ''Central Residu Effekt Met Behulp van de Dichotische Toonhoogtegewaarwording volgens Huggins,'' Masters thesis, Delft University of Technology (unpublished).
Yost, W. A. (**1991**). ''Thresholds for segregating a narrow-band from a broadband noise based on interaural phase and level differences,'' J. Acoust. Soc. Am. **89**, 838–845.