Binaural models and the strength of dichotic pitches

William M. Hartmann[a] and Peter Xinya Zhang
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

Modern physiologically based models of the binaural system incorporate internal delay lines in the pathways from left and right peripheries to central processing nuclei. Different binaural models for the formation of dichotic pitch employ these delay lines in different ways. Consequently, the different models make different predictions for the relative strengths of dichotic pitches made with particular phase conditions. The differences are magnified for dichotic pitches at low frequencies where especially long delay lines may be required. Data from four low-frequency pitch strength experiments on pure-tone-like dichotic pitches (two on Huggins pitch and two on binaural coherence edge pitch) are consistent with models of the equalization-cancellation type and not consistent with the central activity pattern model. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1624072]

I. INTRODUCTION

Among the diverse phenomena arising in the study of human hearing, dichotic pitch is one of the most fascinating (Cramer and Huggins, 1958; Bilsen and Raatgever, 2000, 2002). A dichotic pitch stimulus consists of a white noise in the left ear and a white noise in the right ear, with nothing in the amplitude or phase spectrum of the noise in either ear alone that is frequency specific. Nevertheless, a listener has the impression of a clear pitch located somewhere within the head. The listener can match that pitch consistently to within a few percent (Hartmann, 1993). The origin of dichotic pitches is not to be found in the individual noises sent to left and right ears. Instead, it is in the relationship between the two noises, particularly their relative (interaural) phase relationship.

Dichotic pitches may be classified into two types, pure-tone-like and complex-tone-like. The former are made with an interaural phase relationship that encodes a single frequency leading to a binaural sensation rather like a sine tone in noise. These include the Huggins pitch (Cramer and Huggins, 1958; Guttman, 1962), the binaural edge pitch (Klein and Hartmann, 1980; Frijns *et al.*, 1986), and the binaural coherence edge pitch (Hartmann and McMillon, 2001). These pitches have been compared for a large number of listeners by Akeroyd *et al.* (2001). The complex-tone-like pitches are made with interaural phase relationships that somehow encode many harmonically related frequencies, and they sound like complex tones in noise. These effects include the Fourcin pitch (Fourcin, 1962, 1970), the dichotic repetition pitch (Bilsen, 1972; Bilsen and Goldstein, 1974), and the multiple phase shift pitch (Bilsen, 1976).

The present article focuses on two pure-tone-like dichotic pitches, the Huggins pitch (HP) and the binaural coherence edge pitch (BICEP). These two effects were studied to compare two well-known binaural models, the equalization-cancellation (EC) model (Durlach, 1960, 1972) and the central activity pattern (CAP) model (Raatgever and

Bilsen, 1986). Specifically, we asked which model best accounts for the relative strength of the different phase configurations for these two kinds of dichotic pitch.

According to the EC model, the binaural system *subtracts* the left- and right-ear signals to form a central spectrum.[1] The subtraction process is preceded by an equalization stage in which the amplitudes are adjusted and the phases are shifted so as to optimize the central spectrum for performing the task. For instance, it maximizes the signal-to-noise ratio in a signal detection task. In modern versions of the EC model, such as the modified EC model (e.g., Culling *et al.*, 1998a,b), the phase shifting operation is performed by binaural delay lines, tuned in frequency and in interaural delay. Therefore, different frequency bands may be equalized by different phase shifts. In the EC model, a dichotic pitch is detected as an anomalous point (peak or edge) along the internal frequency axis in the central spectrum, as modified by EC operations (Durlach, 1962).

According to the CAP model, the binaural system *adds* the left- and right-ear signals in channels that are tuned both in frequency and in interaural time delay (ITD) to create a central activity pattern in the frequency-ITD plane. In this model a dichotic pitch is detected as an excitation peak in that plane, with the pitch itself determined by the frequency axis and the lateralization of the pitch image determined by the ITD axis. Although addition is not a popular model for binaural processing, the Appendix shows that it can be regarded as equivalent to binaural cross-correlation in the tonotopic-ITD plane, as introduced by Jeffress (1948).

Both the EC and the CAP models depend upon interaural delay lines. The EC model uses the delay lines to perform the phase equalization operation. The CAP model uses the delay lines as a fundamental part of the binaural display. The models differ in their rules for binaural combination; the EC model employs subtraction but the CAP model uses addition.

The goal of the present article is to distinguish between the EC and CAP models. The key to our approach lies in the different binaural combination rules (subtraction versus addition) and in the character of the population of interaural

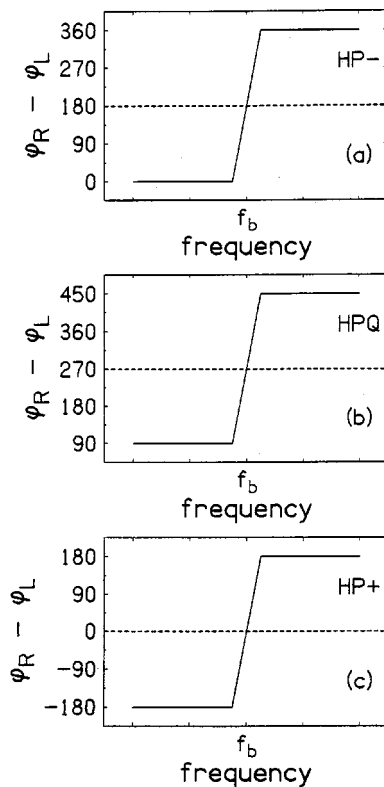[a]Electronic mail: hartmann@pa.msu.edu

FIG. 1. Although the phases of noise components for Huggins pitch (HP) are random, there is a strict relationship between the phases for left and right ears as a function of component frequency. Parts (a)–(c) show the interaural phase functions for HP−, HPQ, and HP+ respectively. In each case the interaural phases change by 360° in the boundary region. The center of the boundary region is the boundary frequency, f_b. Pitch matching experiments show that the Huggins pitch corresponds to f_b.

delay lines. In contemporary models of the binaural system, delay lines become less numerous, and therefore noisier on the average, as the delay increases to large values (Stern and Trahiotis, 1995). The decrease in fidelity with increasing interaural delay τ is commonly modeled by the central weighting function $p(\tau)$. In the EC model $p(\tau)$ is used to explain the larger masking level difference (MLD) in the $N0S\pi$ configuration compared to the $N\pi S0$ configuration, especially for low signal frequencies (Colburn, 1977). In the CAP model, the central weighting accounts for the unique lateralization of dichotic pitches (Raatgever, 1980; Raatgever and Bilsen, 1986; Bilsen and Raatgever, 2000).[2] The focus of this article is therefore on dichotic pitches at low frequencies where long delay lines may be required. The selective degradation of the binaural processing for long delay lines gives the experiments of this article their special leverage.

II. MODELS FOR THE HUGGINS PITCH

The Huggins pitch is made with a broadband noise that is identical in the two ears except for the interaural phases of the spectral components. Figure 1(a) shows the stimulus known as HP−, where the interaural phase difference is zero (or 360°, equivalent to zero) for all frequencies except for the boundary region where the interaural phase increases monotonically from 0 to 360°. The boundary region is centered on the boundary frequency f_b where the interaural phase is

180°. Figure 1(c) shows the stimulus called HP+, which is the same as HP− except for a global phase shift of 180°, equivalent to the simple inversion of one of the two channels. It is useful to introduce the concept of a *background* phase, 0° for HP− and 180° for HP+.

According to the EC model of Huggins pitch, the binaural system cancels the background components of HP− [Fig. 1(a)] by subtraction, leaving a residual consisting only of the components in the boundary region. The peak of the residual occurs where the interaural phase is 180°, and that peak corresponds to the perceived pitch. By contrast, detecting HP in the HP+ configuration [Fig. 1(c)] requires internal delay lines to adjust the phases of the low- and high-frequency components. These components must be shifted by about 180° (i.e., half the period corresponding to f_b) to make the effective interaural background phase close to zero. Only then can the low and high frequencies be cancelled in the subtraction process.

Because detection of HP+ requires interaural delay, and detection of HP− does not, the EC model predicts that HP− should be stronger. The difference might occur for all boundary frequencies, even for favorable frequencies like 500 Hz, because in any tuned channel it is not possible to completely cancel a constant π phase shift with a single internal delay. As the boundary frequency is reduced, for example, towards 100 Hz, the required delays increase towards 5 ms. Delays this long stretch the limits of any binaural model and the EC model clearly predicts that detection performance will be worse for HP+ compared to HP− for boundary frequencies in this range.

The operation of the CAP model is shown in Fig. 2 where the loci of the phase boundary frequencies are shown in parts (a) and (c) for HP− and HP+, respectively, and filled and open circles correspond to the lowest and highest boundary frequencies used in our experiments. Thus, according to the CAP model, HP+ and HP− are detected at different ITDs in the frequency-ITD plane. Because the interaural phase at the boundary frequency is zero for HP+, the image for HP+ appears directly in the center (ITD=0) so that it is not necessary for any delay lines to be involved. Because the interaural phase at the boundary frequency is π for HP−, the image for HP− appears at the left or right. For HP− with a boundary frequency of 100 Hz, the image appears at an ITD value of plus or minus 5 ms, where delay lines are sparse and the image is weak. In their 1986 paper on the CAP model, Raatgever and Bilsen used the $p(\tau)$ function developed by Colburn (1977). For an ITD of 5 ms, $p(\tau)$ drops to 1% of its small-ITD value. Therefore, the CAP model clearly predicts that detection performance will be worse for HP− compared to HP+ for boundary frequencies in this range.

One more model should be added to the pair of models under consideration, namely Green's 1966 variation on the EC model. In Green's variation, the cancellation process can be done either by addition or subtraction. As it turned out, Green's variation provided a uniquely convenient model for binaural edge pitch (Klein and Hartmann, 1980), and it may find increased support from recent physiology emphasizing the joint role of excitation and inhibition in the medial superior olive (Grothe, 2000; Brand *et al.*, 2002). According to
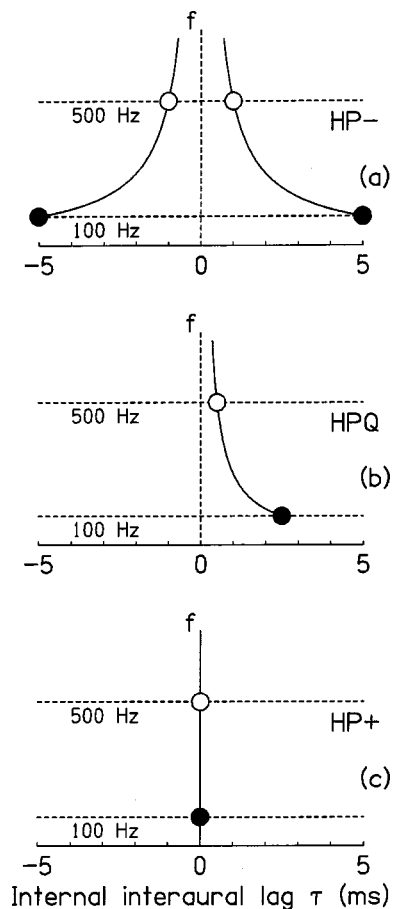
FIG. 2. Peaks of the central spectra according to the CAP model for HP−, HPQ, and HP+. The location of the peaks of the excitation are plotted in the frequency-ITD plane. (a) For boundary frequency f_b, the HP− pitch image appears at an ITD of $1/(2f_b)$, (b) the HPQ image appears at an ITD of $1/(4f_b)$, and (c) the HP+ image appears at an ITD of zero. If f_b is low then HP− ought to be weak because of its extreme laterality. The operation is indicated by the highest and lowest frequencies used in the experiments, indicated by open filled symbols.

Green's variation, both HP+ and HP− can be formed without internal delay lines. For HP+, cancellation except for the boundary region is achieved by addition. For HP−, cancellation except for the boundary region is achieved by subtraction. However, for interaural background phases other than 0 and 180° the model achieves only imperfect cancellation of the background. The most disadvantageous case occurs when the background phases become ±90°, and the component at the boundary frequency is subjected to the same cancellation as the background. Therefore, Green's variation would predict that the worst performance would occur for ±90°, a condition called HPQ (Q for "quadrature") as shown in Fig. 1(b).

III. EXPERIMENT 1—HUGGINS PITCH DETECTION

The goal of experiment 1 was to test the predictions of the models for HP for low boundary frequencies. According to the EC model, detection performance should be best for HP− and worst for HP+. According to the CAP model performance should be best for HP+ and worst for HP−. According to Green's version of the EC model, performance should be best for HP+ and HP− and worse for HPQ.

A. Method

The listener heard two noises in succession. One of them had a Huggins-type phase boundary region, the other did not. The ability of a listener to correctly distinguish between these noises was taken to be a measure of the strength of the HP.

1. Stimuli

Each noise stimulus had two channels. The left-ear channel was broadband white noise with 16 384 spectral components of equal amplitude and random phase. To create a HP stimulus, the right-ear channel was generated from the left-ear channel by changing the phases of components. Below the phase boundary region, the interaural phase was 0 for HP−, 90° for HPQ, and −180° for HP+. In the boundary region, the interaural phase increased linearly with frequency by 360° so that the interaural phases for high-frequency components were the same as for the low. Therefore, at the boundary frequency (center frequency of the boundary region) the interaural phases were 180° for HP− (that's why it's called HP−), 270° for HPQ, and 0° for HP+. These phase variations are shown in Fig. 1. The width of the boundary region was 6% of the boundary frequency.

To create the noise without the HP, the interaural phase was the same as for the low-or high-frequency regions of the HP noise. Therefore, the listener could not use a change in the background interaural phases (e.g., $N0$ vs. $N\pi$) to make decisions. The listener had to use information within the boundary region.

Noise stimuli were calculated by an array processor (Tucker Davis AP2) and converted to audio by 16-bit DACs (Tucker Davis DD1). The sample rate per channel was 20 ksps. With a continuously cycled memory buffer of 32 768 words, the period was 1.6384 s and the frequency spacing between adjacent components was 0.61 Hz. The maximum frequency of the broadband noise as converted was 10 kHz, and the output was low-pass filtered at 8 kHz at a rate of −115 dB/oct. The phase integrity of the two channels was tested by adding or subtracting them electronically at a point in the signal chain immediately prior to the headphone power amplifiers. The cancellation was good to at least 40 dB.

2. Listeners

There were three listeners, M (female, age 42), W (male, age 61), and X (male, age 26). All listeners had normal hearing except that W had a bilateral hearing loss above 8 kHz typical of males of this age. Listeners W and X were the authors and had considerable experience in dichotic listening. Listener M had no previous experience.

3. Procedure

On each trial, the listener heard two 500-ms noise intervals with a silent gap of 500 ms between them. One of the intervals included the HP. The listener's task was to decide which interval had the pitch and to press the corresponding button on a response box. There was no feedback.

The phases of the components were randomized on each interval. Each experimental run (block) included 80 trials,
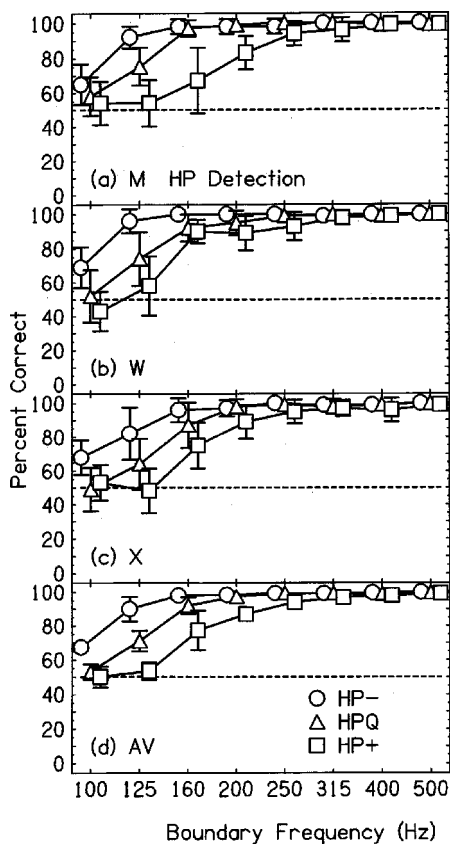
FIG. 3. Results of experiment 1—Huggins pitch detection for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency. Error bars are two standard deviations in overall length.

ten trials with each of eight nominal boundary frequencies: 100, 125, 160, 200, 250, 315, 400, and 500 Hz. On each trial the boundary frequency was varied by a random amount within a rectangular distribution having an overall width of 6% of the nominal boundary frequency.[3] The order of appearance of the boundary frequencies within a run was random. Every run presented a fixed type of Huggins stimulus—HP−, HPQ, or HP+. Each listener did ten runs of each type, in a haphazard order based on alternation and randomness. The level of the noises was 65 dB, making the spectrum level 26 dB re 10^{-12} W per square meter per Hertz. The listener heard the stimuli via Sennheiser HD 480-II headphones while seated in a double-walled sound-treated room.

B. Results

The results of experiment 1 are given in Fig. 3, which shows the percentage of correct identifications of the interval with the HP stimulus as a function of the boundary frequency for the three different phase configurations, or stimulus types.

1. Observations

Figure 3 shows two trends, common among the listeners. First, for each stimulus type (HP−, HPQ, or HP+) performance decreased for decreasing boundary frequency. Second, for boundary frequencies where performance was better than guessing but worse than perfect, there was a clear or-

dering among the stimulus types. The ordering was the same for all three listeners, and, consequently, for the average shown in the bottom panel. Performance was best on HP−, worst on HP+, and intermediate on HPQ. This is the ordering predicted by the EC model.

2. Statistics

Statistical tests of the results were complicated by both floor (near 50%) and ceiling (near 100%) effects in performance. To make the critical test between HP− and HP+ we included paired comparisons where at least one of the two configurations led to a correct response rate between 55% and 95%. For the three listeners and eight frequencies there were 15 such pairs. The difference in percentages of correct responses had the same sign (sign test) for all 15 pairs, in favor of HP−. A one-tailed t-test at the 0.05 level found a significant advantage for HP− on 14 of the 15 pairs.

IV. EXPERIMENT 2—HUGGINS PITCH DISCRIMINATION

Whereas experiment 1 studied HP detection, experiment 2 studied HP discrimination—specifically discrimination between two different boundary frequencies, which lead to two different pitches. An advantage of the discrimination experiment in studying the Huggins pitch effect is that it tests for the musical sense of highness or lowness that is the hallmark of pitch. Other investigators have used a sine-tone matching task (e.g., Guttman, 1962). A discrimination task was actually used in the first reports of dichotic pitch effects (Cramer and Huggins, 1958). Experiment 2 was a boundary-frequency discrimination experiment to compare HP−, HPQ, and HP+.

A. Method

The stimuli, listeners, and protocol were the same as in experiment 1. The difference was that in experiment 2 both intervals included a phase boundary region. The boundary frequency in the second interval was 9% (1.5 semitones) higher, or 9% lower, or the same as in the first interval. In a three-alternative forced-choice task the listener had to compare the pitches on the two intervals and indicate a choice (higher, lower, or the same) by means of push buttons. The advantage of the three-alternative task is that it reduced the guessing limit to 33% correct, affording a greater range for comparing the three HP configurations.

B. Results

The results of the HP discrimination experiment are shown in Fig. 4. For listeners M and W there was a clear ordering. Performance was best on HP−, worst on HP+, and intermediate for HPQ. For listener X, the critical data were ordered in the same way with one exception (200 Hz). Because the data from the three listeners were similar, the average shown in the bottom panel (d) of Fig. 4 exhibits a strong ordering suggesting low-frequency pitch strength in the order: HP−>HPQ>HP+. A statistical test parallel to the test used for experiment 1, in the range 38% (33+5) to 95%, led to 18 comparisons. All differences had the same sign, and
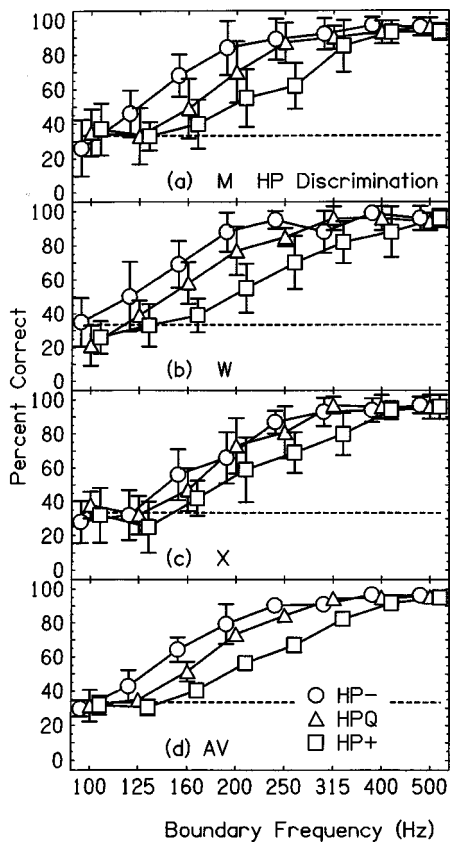
FIG. 4. Results of experiment 2—Huggins pitch discrimination for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency.

FIG. 5. Parts (a)–(c) show the interaural phases for the binaural coherence edge pitch (BICEP). Above the edge frequency, f_e, the interaural phase is fixed at 0°, 90°, or 180°. Below the edge frequency the interaural phase is completely random because the noises in the two ears are independent.

12 of them showed a significant advantage for HP− over HP+ at the 0.05 level. The ordering and statistical advantage are the same as found in experiment 1 and serve as further evidence in favor of the EC model.

V. EXPERIMENT 3—BICEP DETECTION

The binaural coherence edge pitch (BICEP) is created by a discontinuity in coherence where the binaural cross-correlation function is equal to zero (incoherent) on one side of the edge frequency and finite (coherent) on the other. The edge (or boundary) can take on several different forms. The noise components above the edge frequency can be incoherent while the components below the edge are coherent. Alternatively, the noise above the edge can be coherent while the noise below is incoherent. Hartmann and McMillon (2001) found that for low edge frequencies the BICEP was much stronger when the noise was coherent above the edge. (The interpretation of this result was that the critical element in BICEP perception is the ability to cancel the coherent components to create an edge in the central spectrum. When the noise is coherent only below a low-frequency edge there is not much to cancel and the BICEP is weak. To perceive an edge one must be aware of two different regions.) Therefore, experiment 3 used only the BICEP configuration with coherent noise above the edge, as shown in Fig. 5.

Analogous to HP−, HPQ, and HP+, the interaural phase of the coherent components was either 0°, 90°, or
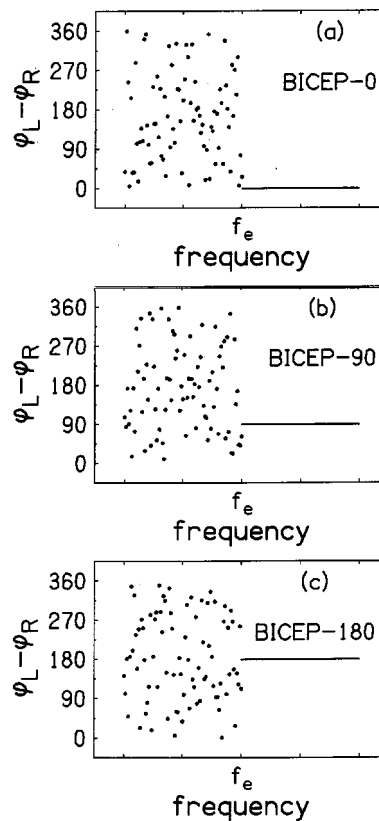
180°, called BICEP-0, BICEP-90, and BICEP-180, respectively. The EC model predicts that for low edge frequencies, BICEP-0 should be the easiest to hear and BICEP-180 should be the hardest. This prediction is based on the idea that the coherent components of BICEP-0 can immediately be cancelled by a subtraction operation whereas the coherent components of BICEP-180 require a phase shift (presumably mediated by internal delay lines) before cancellation. Again, in parallel with the predictions for HP, the CAP model predicts the reverse order of difficulty, and Green's variation of the EC model predicts that BICEP-90 should be the hardest.

The BICEP matching experiments described in Fig. 2 of the article by Hartmann and McMillon (2001) suggested that it was easiest to hear BICEP-180 (there called "pi-phase BICEP") because two of the five listeners were much more successful in matching pitches to that phase condition, at least for edge frequencies between 550 and 750 Hz. Brief experiments at low edge frequencies, however, did not lead to the same conclusion. Low edge frequency experiments were pursued further in the present work.

A. Method

The experiments reported here focused on the critical low-frequency region and use BICEP detection rather than the BICEP matching procedure of Hartmann and McMillon. The protocols and listeners were the same as in experiment
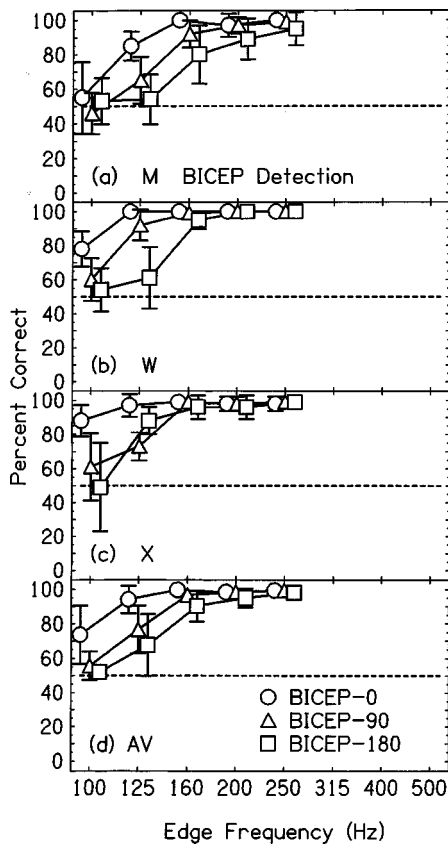
FIG. 6. Results of experiment 3—BICEP detection for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency.
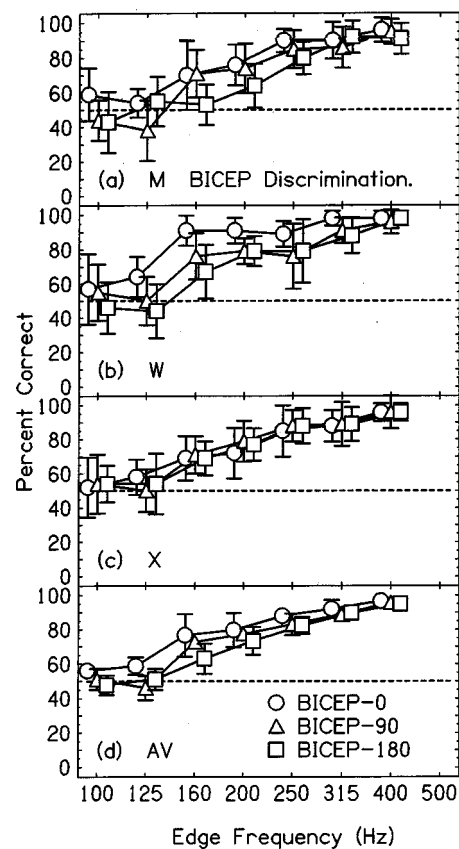

FIG. 7. Results of experiment 4—BICEP discrimination for three listeners (a,b,c) followed by the average (d) as a function of the boundary frequency.

1. The stimuli were also the same except that BICEP noises replaced HP noises. Edge frequencies were distributed randomly over a range of ±5% of the nominal edge frequency. Because preliminary experiments showed that all the listeners could easily detect the BICEP for edge frequencies above 250 Hz, the range of nominal edge frequencies was reduced to the five frequencies less than or equal to 250 Hz. Therefore, a run included only 50 trials.

B. Results

The results of the BICEP detection experiment are shown in Fig. 6. The plots there show nearly perfect performance for an edge frequency of 250 Hz and somewhat better than chance performance at 100 Hz. Below 250 Hz, performance was best for BICEP-0, worst for BICEP-180, and intermediate for BICEP-90. There was only one important exception to this rule, listener X at 125 Hz. The ordering of performance, with this one exception, supports the EC model for BICEP detection. A statistical test at the 0.05 level between BICEP-0 and BICEP-180, parallel to the test for experiment 1, led to seven paired comparisons with BICEP-0 having a significant advantage on all seven, in further support of the EC model.

VI. EXPERIMENT 4—BICEP DISCRIMINATION

Experiment 4 on BICEP discrimination resembled experiment 2 on HP discrimination in the same way that ex-

periment 3 on BICEP detection resembled experiment 1 on HP detection.

A. Method

The method and listeners for experiment 4 were the same as for experiment 2 except that the BICEP stimulus replaced the HP stimulus. Because the BICEP was harder to hear than the HP [as noted also by Akeroyd *et al.* (2001)] only a two-alternative task was used. The edge frequency on the second interval was either 9% higher or 9% lower than on the first.

B. Results

The results of the BICEP discrimination experiment are shown in Fig. 7. The figure shows that there was not much to be learned from the responses at 400, 100, or 125 Hz because of ceiling or floor effects. For the four frequencies in between, the data for M and W suggest that BICEP-0 and BICEP-90 are stronger than BICEP-180. The data for X showed similar pitch strengths for all three phase conditions. Averaging the data for the three listeners led to a plot that tends to agree with the ordering of strength for experiments 1–3. This ordering favors the EC model. The 0.05-level statistical test parallel to experiment 1 showed an advantage for BICEP-0 over BICEP-180 for 8 out of 12 comparisons for listeners M and W and for 0 out of 5 for listener X, overall 8 out of 17. No test was possible for the reverse hypothesis, which failed the sign test on 13 out of 17 pairs.

VII. DISCUSSION

A. Summary of the experiments

Four experiments on the detection and discrimination of Huggins pitch (HP) and binaural coherence edge pitch (BICEP) were performed to test models of binaural pitch formation. The experiments focused on the region of low boundary frequencies. The results of the four experiments all pointed to the same conclusion: the data agree with the predictions of the equalization cancellation (EC) model and do not agree with the predictions of the central activity pattern (CAP) model. The data also do not agree with the predictions of Green's variation on the EC model. These results might be interpreted as peculiar to the low-frequency region, but the low-frequency data in Figs. 3, 4, 6, and 7 fit smoothly onto the intermediate-frequency data, and there would seem to be no reason to postulate that a different mechanism of binaural pitch formation somehow begins to apply when the frequency becomes lower than about 300 Hz. Instead, the results of these experiments, and also the lateralization experiments of Zhang and Hartmann (2002), suggest a single mechanism that becomes progressively less efficient as the frequency decreases. Therefore, the experiments suggest that the EC model correctly describes binaural pitch formation and the other models do not. It is worth investigating the assumptions that led to this conclusion.

B. Assumptions

The central assumption is that the performance differences observed for different binaural stimuli can be attributed to limitations in internal delay lines as the delays become long. The logic of our experiments is based on the idea that different binaural models use those delay lines in different ways and thus make different predictions for the experimental stimuli. Therefore, it is important to the validity of the conclusions that the observed low-frequency performance differences be caused by the delay lines and not by some other aspect of binaural hearing or pitch formation.

In fact, the data as a whole cannot be understood by assuming that the only cause of poor low-frequency performance is the failure of long delay lines because for each model there are some stimuli that do not require delay lines at all, and yet performance is poor for these stimuli at low frequencies, specifically HP$-$ and BICEP-0 in the EC model and HP$+$ and BICEP-180 in the CAP model. To understand the low-frequency behavior seen for these cases requires a binaural model with a failure characteristic that depends on frequency. It is not uncommon for binaural models to require delay line weighting on both delay and frequency such that the $p(\tau)$ function becomes $p(\tau,f)$, falling off at both high and low frequencies (Raatgever, 1980; Stern et $al.$, 1988; Shackleton et $al.$, 1992). There is experimental evidence for added internal noise at low frequencies from MLD experiments, particularly the strong level dependence of the MLD (Wilbanks and Whitmore, 1967), which decreases binaural coherence and could disrupt the formation of dichotic pitch at low boundary frequencies.

Apart from the general decrease in performance with decreasing boundary frequency, there remain large observed differences between the different phase conditions, and current binaural models would attribute these differences to the use of internal delays. The observed differences seem consistent with that interpretation in that the differences between HP$+$ and HP$-$ as well as the differences between BICEP-0 and BICEP-180 appear for boundary (or edge) frequencies between 200 and 300 Hz, corresponding to delay lines between 2.5 and 1.7 ms long. Experiments on other binaural phenomena, e.g., the MLD, have been interpreted assuming delay lines that start to fail for comparable delays (van der Heijden and Trahiotis, 1999).

C. The spatial masking interpretation

An alternative account of the experiments reported here appeals to a form of masking or interference and is primarily based on subjective impressions. It can be illustrated for the Huggins pitches HP$+$ and HP$-$ in the context of the CAP model. According to the CAP model, HP$+$ is detected without delay lines, and therefore this phase condition should be the most easily detectable at low boundary frequency. This prediction is contradicted by experiments 1 and 2 which show that it is HP$-$ that is the most detectable phase condition. Bilsen (2000) has observed that in comparison with HP$-$ the pitch image in HP$+$ sounds as though it is masked more, meaning that the listener has the impression that there is greater interference from the background noise. Our listeners agree with that observation.

An interpretation that is consistent with that observation is as follows. Stimulus HP$+$ includes a background noise that is $N\pi$, a noise that is lateralized to both left and right sides of the head and tends to spread towards the center. Thus it encroaches on the spatial region occupied by the boundary region, which is S0. The background noise is broadband including a great deal of power below the boundary frequency. The encroachment then interferes with the detection of the pitch at the boundary frequency. By contrast, stimulus HP$-$ includes a background noise that is $N0$, which forms a compact image near the center of the head. The background noise image is so compact that it does not interfere with the boundary region, which is $S\pi$ and appears far to the left or right. As a consequence HP$-$ may be more detectable than HP$+$, consistent with experiment.

Despite its visceral appeal, there are a number of problems with this spatial interpretation: (1) There are no models of dichotic pitch based on contrasts along the spatial dimension; all models of dichotic pitch are based on contrasts along the tonotopic dimension. (2) A model of dichotic pitch based on spatial contrasts would be contrary to evidence indicating the dominance of tonotopic organization over spatial organization (Culling and Summerfield, 1995). (3) It is not enough for spatially based masking merely to exist. In order to preserve the CAP model, the spatial advantage of the more compact background-noise image for HP$-$, must outweigh the disadvantage of longer delay lines in the formation of the pitch image itself. (4) The subjective impression of greater interfering noise for the HP$+$ phase condition is not inconsistent with the EC model. Cancelling a noise stimulus having long interaural delays is expected to lead to a large residual noise because of jitter in the cancellation process.

Further, it is impossible for the EC process to cancel all the noise in a tuned channel using a single value of internal delay for HP+, where the background noise has π interaural phase, but it is trivially easy in the case of HP−, where the background noise phase is zero. (5) One would expect the spatial masking effect to make HP− more detectable than HP+ for all boundary frequencies, not just for low frequencies, contrary to our data.

D. Cross-correlation model

The cross-correlation model advanced by Colburn (1977) to account for the results of diverse experiments including those on the MLD makes predictions that are hard to distinguish from those of the EC model. In fact, Domnitz and Colburn (1976) have shown that in the limit of small signal-to-noise ratio, a normalized cross-correlation model is equivalent to an EC model in which the interaural phase difference of the noise is compensated by the equalization process. Like the CAP model, Colburn's model uses a cross-correlation function, but unlike the CAP model it attributes detection to dips, and not peaks, in the central excitation pattern. The model operation can be illustrated with HP−, for which the background noise is $N0$. The cross-correlation function, $\gamma(\tau)$, is maximum $[\gamma(0)=1]$ at zero interaural lag except in the boundary region where there is a dip in the maximum. This dip is very important perceptually because the binaural system is most sensitive to deviations from perfect ($\gamma=1$) cross-correlation (Gabriel and Colburn, 1981). Because the dip for HP− occurs for no internal delay, this model agrees with the EC model in predicting that low-frequency dichotic pitch should be strongest for HP−.

The cross-correlation model is naturally applied to an MLD experiment where a signal with an interaural phase that is different from the masker causes a just-detectable deviation from $\gamma=1$. The model immediately predicts that the greatest MLD should occur for the $N0S\pi$ condition, and with the introduction of the delay weighting, $p(\tau)$, the model predicts the smaller MLD for the $N\pi S0$ condition. In contrast with MLD experiments, the decorrelations presented by dichotic pitch stimuli are larger. The application of the cross-correlation model to dichotic pitches would require that the same principles continue to apply. It is a matter of parsimony to assume that they do.

E. The exponential CAP model

The Appendix introduces a modification to the CAP model including the assumption that the left and right inputs to the central processor are exponentially rectified versions of the input signals. This modification has the effect of passing the mathematical predictions of the CAP model through a nonlinear transformation, essentially the modified Bessel function I_0. As shown in the Appendix, the predicted CAP for tonotopic place ω and interaural delay τ becomes

$$\gamma(\omega,\tau)=I_0\{\{g_L^2+g_R^2+2g_Lg_R$$
$$\times\cos[\phi_R(\omega)-\phi_L(\omega)-\omega\tau]\}^{0.5}\}.$$

The synchrony coefficients g_L and g_R, for left- and right-ear pathways respectively, can be expected to depend on the interaural delay line.

The argument of the Bessel function, itself a function of tonotopic coordinate ω and ITD τ, is essentially the cross-correlation function from the CAP model. The CAP model predicts dichotic pitch images at peaks in the ω-τ plane. The sharper the peaks, the stronger the dichotic pitch ought to be. The purpose of the calculations described in this section was to discover whether this simple and plausible nonlinear transformation on the model could result in predictions in agreement with our experiments by sharpening the peaks.

In the context of the CAP model, g_L and g_R should be approximately the same in the perception of HP+, which is located at $\tau=0$. But they could be quite different in the perception of HP−, which is located off to the left or right because a longer delay line to one side might well introduce increased jitter which would decrease the synchrony coefficient on that side. *A priori*, it seemed possible that the nonlinear transformation could lead to a peak that was actually sharper for g_L not equal to g_R and hence predict better performance for HP− than for HP+.

Colburn (1973) and Stern and Colburn (1978) took values of g to be less than or equal to $\sqrt{20}$. Therefore, we performed calculations for many pairs of g_L and g_R less than 5. Plots of the calculated cross-correlation functions showed that by any reasonable measure of sharpness, peaks for $g_L \neq g_R$ are never sharper than the peaks for $g_L=g_R$. Therefore, even after the exponential modification and asymmetry, the CAP model continues to predict that HP+ should be stronger than HP−, contrary to experiment.

VIII. CONCLUSIONS

Four experiments on the detection and discrimination of Huggins pitch (HP) and binaural coherence edge pitch (BICEP) were performed to test models of binaural pitch formation. By studying these dichotic pitches at low frequencies, the experiments were able to distinguish among alternative models of pitch formation. The results favored the equalization-cancellation (EC) model or close relatives such as Colburn's cross-correlational model (1977) or the modified EC model of Culling *et al.* (1998a), elaborated by Akeroyd and Summerfield (2000).

The results did not support Green's variation on the EC model. Whereas Green's variation predicts that HPQ and BICEP-90 should be the most difficult of the HP and BICEP stimuli, the observed order of difficulty put these 90° conditions in between the 0° and 180° conditions. This conclusion casts some doubt on the validity of Green's variation as it has been applied to binaural edge pitch.

The results of the four experiments disagreed with the predictions of the central activity pattern (CAP) model. From an esthetic point of view, this conclusion is disappointing because the CAP model is the only model that integrates both the pitch and the lateralization of the pitch image in an economical way. Further, the predictions of the CAP model for the lateralization of HP+ and HP− agree qualitatively with observation. In the end, however, the four experiments

on pitch strength presented in this article all lead to the same conclusion favoring the EC model.

ACKNOWLEDGMENTS

APPENDIX: CROSS-CORRELATION AND THE CAP MODEL

The CAP model has been remarkably successful in accounting for a variety of binaural effects. The model is based on the concept of a central spectrum, derived by Raatgever and Bilsen (1986) by assuming that the signals to the left and right ears are simply added. The central spectrum is then the power spectrum of that sum for a given delay between the summed waveforms. With unit-amplitude noise signals and infinitely sharp auditory filters, Bilsen and Raatgever found a central spectrum at best frequency ω and internal delay τ given by

$$P(\omega, \tau) = 1 + \cos[\phi_R(\omega) - \phi_L(\omega) - \omega\tau], \tag{A1}$$

where the difference between signal phases $\phi_R(\omega)$ and $\phi_L(\omega)$ is the interaural phase for frequency ω.

The purpose of this appendix is to show how an equivalent result can be obtained by making the more usual assumption that the central representation is the result of a cross-correlation operation. Cross-correlation, like an AND gate in electronics, depends upon a product and not a sum.

1. Weak synchrony limit

In the limit of weak synchrony the ensemble-average $(\langle\!\langle\cdots\rangle\!\rangle)$ firing rates in left and right channels can be expanded as a unit spontaneous rate plus a modulation term that is linear in the signal,

$$\langle r_L(\omega, t) \rangle = 1 + c\cos[\omega t + \phi_L(\omega)], \tag{A2}$$

for the left-ear channel and similarly for the right. This simple form maintains the assumption that the auditory filter with best frequency ω is infinitely sharp. Modulation fraction c must be less than 1 to keep the firing rate positive.

The ensemble average cross-correlation function is

$$\langle \gamma(\omega, \tau) \rangle = \frac{1}{T_D} \int_0^{T_D} dt \langle r_L(\omega, t + \tau) r_R(\omega, t) \rangle, \tag{A3}$$

where T_D is the internal integration time of the neural cross-correlator. With the assumption that ωT_D is considerably larger than one, and that ensemble averages in left and right ears are independent, the cross-correlation function becomes

$$\langle \gamma(\omega, \tau) \rangle = 1 + \frac{c^2}{2}\cos[\phi_R(\omega) - \phi_L(\omega) - \omega\tau]. \tag{A4}$$

Equation (A4) for the cross-correlation function is the same as Eq. (A1) for the central spectrum except that the peaks and valleys in the ω-τ plane are not as large in the cross-correlation function. They are reduced by a factor of $c^2/2$. However, nothing in the application of the CAP model depends critically on the absolute sizes of peaks and valleys.

2. Strong synchrony—general

If the synchrony is strong, it is helpful to represent the neural firing rate as an exponential rectification of the input signal $x(t)$ (e.g., Stern and Colburn, 1978),

$$r_L(t) = \exp[g_L x_L(t)], \tag{A5}$$

where g_L is a measure of synchrony for the left ear input to a binaural cross correlator, and similarly for the right ear input. Again assuming infinitely sharp auditory filtering,

$$\langle \gamma(\omega, \tau) \rangle = \frac{1}{T_D} \int_0^{T_D} dt \exp\{g_L \cos[\omega(t + \tau) + \phi_L(\omega)]$$
$$+ g_R \cos[\omega t + \phi_R(\omega)]\}. \tag{A6}$$

The integral is a modified Bessel function I_0 for which a series formula is given by Abramowitz and Stegun (1964) (p. 375):

$$\langle \gamma(\omega, \tau) \rangle = I_0(\{g_L^2 + g_R^2 + 2g_L g_R$$
$$\times \cos[\phi_R(\omega) - \phi_L(\omega) - \omega\tau]\}^{0.5}). \tag{A7}$$

The argument of the square root is essentially the frequency and delay dependence of the CAP model. Unlike its more famous cousin J_0, which oscillates, the modified Bessel function I_0 is monotonic and smooth. For small argument x, $I_0(x) = 1 + x^2/4$. Thus, like the low-synchrony limit, the exponential rectifier provides a way to recover the CAP model, or a monotonic transformation of it, from the mathematical operation of cross-correlation. The cross-correlation operation represents the operation of the binaural system better than channel addition.

[1]Durlach's original articles developing the EC model paid little attention to the form of the final displayer. The idea of the central spectrum as a display seems to have originated in a report by Bilsen (1972).

[2]Were it not for the central weighting function a dichotic pitch predicted to be centered could instead be lateralized where the ITD is $\pm 1/f_b$.

[3]The boundary frequency was roved because the nominal boundary frequencies form simple ratios, and these have implications for pitch. By roving the boundary frequencies over the range of a semitone, the experimenters discouraged the listeners from making unwanted musical associations across different trials, especially in the pitch-sensitive discrimination experiments 2 and 4.

Abramowitz, M., and Stegun, I. (**1964**). *Handbook of Mathematical Functions*, National Bureau of Standards Applied Math Series no. 55 (USGPO, WAshington, DC).

Akeroyd, M. A., and Summerfield, A. Q. (**2000**). "The lateralization of simple dichotic pitches," J. Acoust. Soc. Am. **108**, 316–334.

Akeroyd, M. A., Moore, B. C. J., and Moore, G. A. (**2001**). "Melody recognition using three-types of dichotic pitch stimulus," J. Acoust. Soc. Am. **110**, 1498–1504.

Bilsen, F. A. (**1972**). "Pitch of dichotically delayed noise," in *Hearing Theory—1972*, edited by E. de Boer, B. L. Cardozo, and R. Plomp (IPO, Eindhoven, Holland), pp. 5–8.

Bilsen, F. A. (**1976**). "Pronounced binaural pitch phenomenon," J. Acoust. Soc. Am. **59**, 467–468.

Bilsen, F. A. (**2000**). Personal communication.

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.

Bilsen, F. A., and Raatgever, J. (**2000**). "On the dichotic pitch of simultaneously presented interaurally delayed white noises. Implications for binaural theory," J. Acoust. Soc. Am. **108**, 272–284.

Bilsen, F. A., and Raatgever, J. (**2002**). *Demonstrations of Dichotic Pitch*, compact disc, available from the authors at Perceptual Acoustics Laboratory, Applied Physics Department, Delft University of Technology, Delft, The Netherlands.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

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**, 413–417.

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

Culling, J. F., Marshall, D. H., and Summerfield, A. Q. (**1998b**). "Dichotic pitches as illusions of binaural unmasking II. The Fourcin pitch and the dichotic repetition pitch," J. Acoust. Soc. Am. **103**, 3527–3540.

Domnitz, R. H., and Colburn, H. S. (**1976**). "Analysis of binaural detection models for dependence on interaural target parameters," J. Acoust. Soc. Am. **59**, 598–601.

Durlach, N. I. (**1960**). "Note on the equalization and 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, Volume 2*, edited by J. Tobias (Academic, New York), pp. 369–462.

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.

Fourcin, A. J. (**1962**). "An aspect of the perception of pitch," Proc. 4th Int. Congress on Phonetic Sciences, Helsinki (Mouton, The Hague), pp. 355–399.

Fourcin, A. J. (**1970**). "Central pitch and auditory localization," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sitjhoff, Leiden), pp. 319–328.

Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.

Green, D. M. (**1966**). "Signal-detection analysis of equalization and cancellation model," J. Acoust. Soc. Am. **40**, 833–838.

Grothe, B. (**2000**). "The evolution of temporal processing in the medial superior olive, an auditory brainstem structure," Prog. Neurobiol. **61**, 581–610.

Guttman, N. (**1962**). "Pitch and loudness of a binaural subjective tone," J. Acoust. Soc. Am. **34**, 1996(A); Bell Telephone Laboratories memorandum MH-1232-NG-KN, 30 November, 1962.

Hartmann, W. M. (**1993**). "On the origin of the enlarged melodic octave," J. Acoust. Soc. Am. **93**, 3400–3409.

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

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Klein, M. A., and Hartmann, W. M. (**1980**). "Binaural edge pitch," J. Acoust. Soc. Am. **70**, 51–61.

Raatgever, J. (**1980**). "On the binaural processing of stimuli with different interaural phase relations," thesis, Delft (unpublished).

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

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.

Stern, R. M., and Colburn, H. S. (**1978**). "Theory of binaural interaction based on auditory-nerve data IV. A model for subjective lateral position," J. Acoust. Soc. Am. **64**, 127–140.

Stern, R. M., and Trahiotis, C. (**1995**). "Models of Binaural Interaction," in *Handbook of Perception and Cognition—Hearing*, edited by B. Moore (Academic, San Diego), pp. 347–386.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.

van der Heijden, M., and Trahiotis, C. (**1999**). "Masking with interaurally delayed stimuli: The use of internal delays in binaural detection," J. Acoust. Soc. Am. **105**, 388–399.

Wilbanks, W. A., and Whitmore, J. K. (**1967**). "Detection of monaural signals as a function of interaural noise correlation and signal frequency," J. Acoust. Soc. Am. **43**, 785–797.

Zhang, P. X., and Hartmann, W. M. (**2002**). "On the lateralization of the Huggins pitch," J. Acoust. Soc. Am. **112**, 2271(A).