Lateralization of Huggins pitch

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

(Received 25 June 2006; revised 18 July 2008; accepted 29 July 2008)

The Huggins pitch is a sensation of pitch generated from a broadband noise having a narrowband boundary region where the interaural phase difference varies rapidly as a function of frequency. Models of binaural hearing predict that the pitch image should be well lateralized. A direct psychophysical experimental method was used to estimate the lateral positions of Huggins pitch images with two different forms of phase boundaries, linear phase and stepped phase. A third experiment measured the lateral positions of sine tones with controlled interaural phase differences. The results showed that the lateralization of Huggins pitch stimuli was similar to that of the corresponding sine tones and that the lateralizations of the two forms of Huggins pitch phase boundaries were even more similar to one another. Both Huggins pitches and sine tones revealed strong laterality compression (exponent≈0.5). Ambiguous stimuli, with an interaural phase difference of 180°, were consistently lateralized on one side or the other according to individual asymmetries—an effect called "earedness." An appendix to this article develops a new first-order lateralization model, the salient phase density model, which combines attributes of previous models of dichotic pitch lateralization. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2977683]

PACS number(s): 43.66.Qp, 43.66.Pn [ARK] Pages: 3873–3887

I. INTRODUCTION

Dichotic pitches are pitch sensations arising from noises, presented by headphones, with no evident frequency specificity in either left- or right-ear channels. The frequency-specific information that leads to a tonal pitch sensation is encoded entirely in the phase relationship between the two channels. Several binaural models have been advanced to explain the creation of dichotic pitches, notably the equalization-cancellation (EC) model (Durlach, 1962, 1972), the modified EC model (Culling *et al.*, 1998a, 1998b), and the central activity pattern (CAP) model (Raatgever, 1980; Raatgever and Bilsen, 1986). In recent years, arguments for and against the various models have appeared in literature (Culling, 2000; Bilsen and Raatgever, 2000, 2001; Akeroyd and Summerfield, 2000; Hartmann and Zhang, 2003).

The best known dichotic pitch effect is the Huggins pitch (HP) (Cramer and Huggins, 1958), which is made with a broadband noise that is identical in the two ears except for a single phase boundary region. Figure 1 (light line) shows the stimulus known as HP−, where the interaural phase difference (IPD) is 0° (or 360°, equivalent to 0°) for all frequencies except for the boundary region. In the boundary region, the IPD increases monotonically from 0° to 360°. Figure 1 (heavy line) also shows the stimulus called HP+, which is the same as HP− except for an all-band phase shift of 180°. The width of the phase boundary is much smaller than a critical band, though its exact value is not critical. The width is typically chosen to be about 6% of the center frequency. It is useful to introduce the concept of a *background* IPD, 0° for HP− and 180° for HP+. When presented with HP stimuli through headphones, listeners hear, on top of the noise background, a pitch at the center frequency of the boundary region, which is defined as the *boundary frequency*.

A. Lateralization

In addition to its pitch, the tonal image created by the HP stimulus has a laterality. It appears to be centered in the head or off to the left or right. In thinking about the lateralization of HP, it is helpful to plot the interaural cross correlation in the frequency-lag plane, as shown in Fig. 2—a signal-based representation of the classic binaural model by Jeffress (1948). High cross correlation, illustrated by a bright region in the figure, occurs for frequencies and lags where the two channels are in phase. For instance, in Fig. 2(a) for HP− there is a bright stripe running from top to bottom at a lag of zero because of the in-phase background components. Bright stripes (hyperbolas) to the left and right of midline are centered on regions of strong correlation for frequencies and lags where the product of frequency and lag is an integral number of complete cycles. The pattern in Fig. 2(b) for HP+ is essentially the negative of the pattern in Fig. 2(a) because HP+ can be derived from HP− by a 180° phase shift of all components in one channel. Apart from the background, the cross-correlation functions in Fig. 2 represent the boundary regions as thin horizontal bands centered on the boundary frequency, f_c, here chosen to be 200 Hz. Consistent with the visual impression given by Fig. 2, HPs at the centers of the boundary regions are observed to be lateralized to the left or right for HP− and to be lateralized near the center for HP+.

The purpose of the experiments of the present article is to measure the lateral positions of HPs using a direct psychophysical method (Marks, 1974). In this method listeners assign numerical values, according to a fixed scale, to their estimates of the lateral positions of the pitch images. Previous experiments have also measured the lateralization of HP

[a)]Electronic mail: zhangx11@msu.edu
[b)]Electronic mail: hartmann@pa.msu.edu

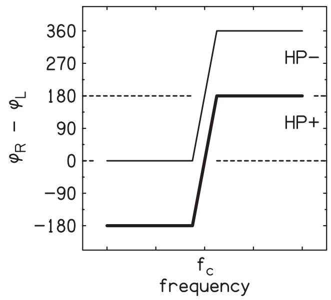

FIG. 1. IPD for HP stimuli of the linear-phase type. Although the phases of noise components for HP stimuli are completely random, there is a definite relationship between the phases for left and right ears as a function of component frequency. Light and heavy lines show the IPD for HP− and HP+, respectively. In each case the interaural phase changes by 360° in the boundary region as a linear function of the component frequency. The frequency at the center of the boundary region is defined as the boundary frequency, f_c. Pitch matching experiments show that the HP corresponds to f_c.

(Raatgever and Bilsen, 1977, 1986; Grange and Trahiotis, 1996; Akeroyd and Summerfield, 2000), but those experiments were indirect. They matched the lateral position of HPs by varying an interaural parameter of some other form of stimulus, used as an acoustic pointer. However, only if it is known that there is a linear relationship between the parameter of the pointer and the perception of location can this method lead to an absolute scale of localization. By contrast, the scale of interaural parameters used in matching methods has been found to be warped with respect to scales outside the domain of binaural stimuli (Yost, 1981). The advantage of the direct psychophysical method is that it removes the response from the domain of binaural stimuli and provides a more veridical measure of perceived lateralization. In addition to its direct psychophysical method, the present work employs a large number of stimulus conditions, including a wide range of boundary frequencies.

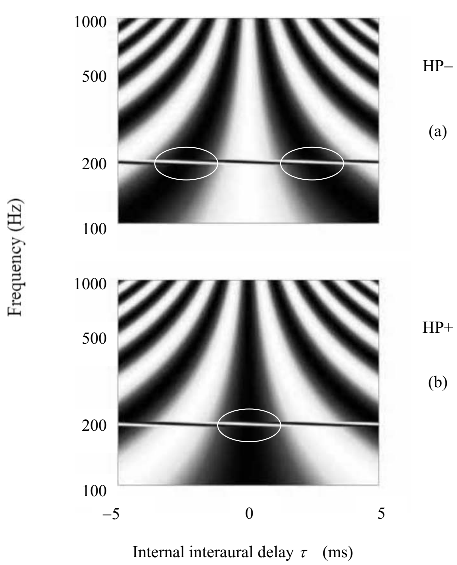

FIG. 2. Cross-correlation for HP− and HP+ plotted in the frequency-lag plane. The axes are linear in lag τ and logarithmic in frequency. Bright areas are centered on regions of high cross correlation. According to simple models, the HP is signaled by the bright narrow horizontal band of intense correlation in the middle of an otherwise dark band where the cross correlation is negative. Therefore, for boundary frequency f_c, HP− appears at a lag of $1/(2f_c)$ (hyperbolic rule) and HP+ appears at a lag of zero. Because the figure is signal based, the vertical axis is frequency, not tonotopic coordinate.

B. Lateralization models

It is necessary to have first-order predictions for the lateralization of HP images to guide the experiments. First-order predictions are mainly based on stimulus parameters and make minimal assumptions about auditory processing.

The CAP model of Raatgever and Bilsen (1986) simultaneously predicts both the existence of the HP and its lateralization. The model is "stimulus based" in that it assumes linear peripheral processing and perfect frequency resolution along the tonotopic axis of the receiver. It is based on a CAP that is very similar to the cross-correlation pattern shown in Fig. 2. Therefore, the lateralization of the image depends mainly on phases within the boundary region.

By contrast, the half-period rule introduced by Akeroyd and Summerfield (2000) (also entirely stimulus based) says that the location of the pitch image depends only on the interaural phase of the noise background. The image location is displaced from the noise background by an interaural time difference (ITD) corresponding to half a period of the phase boundary frequency. The advantage of the half-period rule is that it correctly predicts the lateralization of the pitch image when the interaural phases of the boundary region are random variables. However, Akeroyd and Summerfield (2000) showed that the rule fails for a boundary region having constant delay. Ultimately, they preferred a model that they called the reconstruction-comparison (RC) model, which includes assumptions about auditory filter bandwidth and nonlinear peripheral processing. The assumptions make the model more realistic physiologically at the cost of complexity. An important feature of the RC model is that the lateralization of dichotic pitches is jointly predicted by the interaural properties in the boundary region and in the background region.

The Appendix to this article presents a third model, the salient phase density (SPD) model that is a compromise between the CAP and RC models. It is stimulus based, but its predictions depend jointly on the interaural phases in the boundary region and the background region. The SPD model successfully predicts the approximate lateralizations for all the forms of HP that have been discussed in literature. It is a mathematically straightforward algorithm that makes predictions for other forms of dichotic pitch as well (Zhang and Hartmann, 2007).

The above brief background on models of dichotic pitch lateralization is adequate for the purposes of the present article. For the stimuli of this article, the CAP model, half-period rule, RC model, and SPD model all make essentially the same predictions. The models agree with Grange and Trahiotis (1996) in predicting that HPs with boundaries centered on a given ITD will be lateralized similar to narrow noise bands or sine tones with equivalent ITD. Therefore, it is not the purpose of the present article to compare the different models and discover which is best. Instead, the purpose is to compare all of them with direct psychophysical experiments.

C. First-order predictions

To make first-order predictions for comparison with experimental data requires that the models (CAP, RC, and SPD) be supplemented with assumptions about listener responses. For definiteness, this article begins by assuming that listeners lateralize stimuli exactly as predicted by the SPD model. As noted in the Appendix, the SPD model predicts that lateralization corresponds to an ITD of τ_p, called the "best-fit lag." When applied to the stimuli of this article (subsection of HP+ and HP− with added ITD in the Appendix), the best-fit lag corresponds to the peak of the cross-correlation function at the boundary frequency. Thus, for these particular stimuli, the SPD model agrees with the half-period rule because τ_p always turns out to be half a period removed from the ITD of the background.

In addition, it will be assumed that listeners scale their numerical responses linearly with τ_p according to an unchanging scale wherein the maximum allowed response is assigned to the maximum value of τ_p present in the stimuli. It is assumed that listeners learn about the maximum value of τ_p from their experience with the stimulus set. Together, the SPD model and the linear scaling assumptions constitute the first-order predictions to guide the experiments below.

II. EXPERIMENT 1: LINEAR-PHASE BOUNDARY

Experiment 1 measured the perceived lateralization of HPs by asking listeners to give numerical estimates of the locations of the pitch images. The basic experimental method was simple. We presented listeners with HP+ and HP− stimuli having various boundary frequencies and asked the listeners to estimate the lateral position of the pitch image on a fixed scale. According to the first-order predictions, we expected to find HP+ located at the center of the head independent of boundary frequency, as shown in Fig. 2(b). We expected to find HP− along a curved path similar to the hyperbola in Fig. 2(a).

A lateralization experiment of this kind instantly encounters problems. In a HP− experiment there is a close correspondence between pitch and expected lateral position. Listeners could easily learn to expect low pitches to appear to the extreme left or right and higher pitches to appear toward the center. In a HP+ experiment listeners would presumably experience no variation at all in lateral positions, making the experiment appear pointless. To solve these problems, our stimuli included an added frequency-independent interaural delay that varied randomly from trial to trial. This added frequency-independent ITD is called the "ITD" below. It should not be confused with the best-fit-lag, τ_p, which is equal to the ITD at the boundary frequency due to a combination of both the ITD (added ITD) and the boundary-region phase shifts introduced by the HP stimulus itself.

A. 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. The right-ear channel was generated from the left-ear channel by incrementing the phases of components in the boundary region as a linear function of component frequency, as shown in Fig. 1. The overall width of the boundary region was 6% of the boundary frequency.

There were five different nominal boundary frequencies: 200, 315, 500, 700, and 1000 Hz. The factor of 5 from bottom to top gave a good spread of expected lateral positions and spanned most of the existence region for reliable HP [see, however, Culling (1999)]. On each trial, the boundary frequency varied randomly over a range of $\pm 5\%$ of the nominal boundary frequency to try to reinforce the perceived independence of different trials. There were five values of ITD: −1000, −500, 0, 500, and 1000 μs. A positive delay refers to a lead in the right ear, created by incrementing the phase in the right-ear channel by a shift proportional to increasing frequency.

Noise stimuli were calculated by an array processor (Tucker Davis AP2) and converted to audio by 16-bit digital-to-audio converters (Tucker Davis DD1). The sample rate per channel was 20 ksps. With a continuously cycled memory buffer of 32 768 words, the entire period was 1.6384 s, and the frequency spacing between adjacent components was 0.61 Hz. The duration of the stimuli, as presented to the listener, was one entire period. The maximum frequency of the broadband noise as converted was 10 kHz, and the output was lowpass filtered at 5 kHz by Stanford SR640 filters with a slope of −115 dB/octave. 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. The level of the noises was 65 dB, making the spectrum level 28 dB. The listener heard the stimuli via Sennheiser HD 520-II headphones while seated in a double-walled sound-treated room.

B. Listeners

There were five listeners, C (female, age 61), L (female, 29), W (male, 62), X (male, 27), and Z (male, 28). All listeners were right handed and had normal hearing except that W had a mild 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 as did listener Z. Listeners C and L had little or no previous experience.

C. Procedure

Each experimental run consisted of 25 trials, using every combination of the five boundary frequencies and the five values of ITD, presented in random order. The stimulus was computed anew for each trial using a new set of random phases. During a trial, the listener could call up a single burst of HP stimulus as many times as desired. The listener was required to enter a number between −40 (extreme left) and

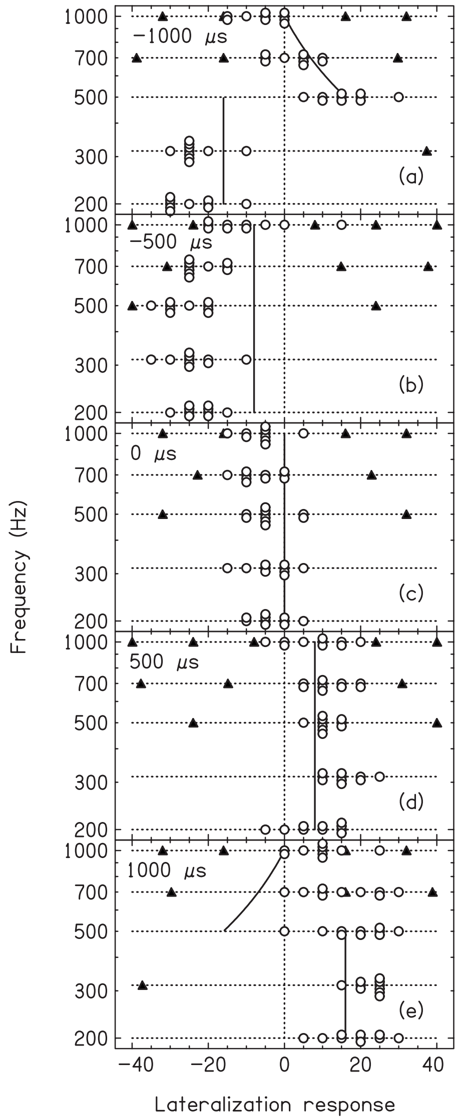

FIG. 3. Scatter plot showing lateral responses for HP+ for listener Z. The solid line gives the first-order prediction. The line bends as needed to keep the interaural phase less than 180°. The triangles indicate alias predicted points, separated from the solid line by 360°. Each panel shows a different ITD. In each panel the 50 circles show all the lateral responses made for five boundary frequencies in the course of ten runs.

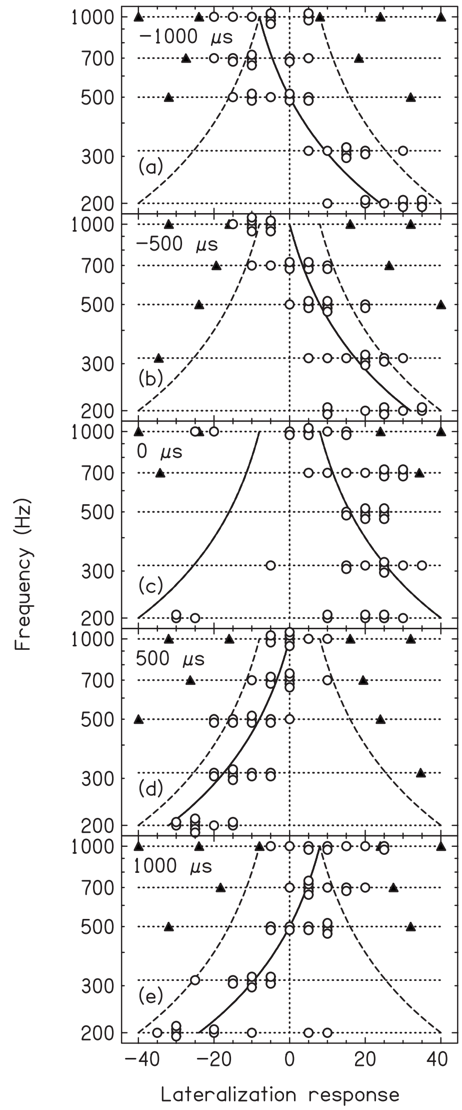

FIG. 4. Scatter plot showing lateral responses for HP− for listener Z, analogous to Fig. 3. The dashed lines delimit the central region defined by interaural phases of ±180°.

+40 (extreme right) as an estimate of the location of the pitch image. There was no feedback. Every experimental run used a fixed type of stimulus, HP− or HP+. This procedural element kept the background interaural phase constant except for the added ITD. Successive runs alternated between these two types. Overall, a listener did ten runs of each type.

D. Results

The analysis of the results required that the data from the runs be unpacked and displayed separately by stimulus type (HP+ or HP−), by boundary frequency, and by ITD. For each listener there were ten plots in Experiment 1, five values of the ITD for each stimulus type. The most revealing data format is a scatter plot consisting of one point for each individual response. As an example, Figs. 3 and 4 show the scatter plots for listener Z for HP+ and HP−, respectively. The data from listener Z are typical: the variability, indicated by scatter in the data, is neither the smallest for Z (the smallest is for listener X) nor is it the largest (the largest is for listener L).[1]

On the scatter plots, the stimulus variable (independent variable), namely, the boundary frequency, is on the vertical axis and the listener's lateralization responses (dependent variable) are on the horizontal axis. This unusual presentation was chosen because the first-order predictions are that responses will reflect the best-fit lag, τ_p, which appears on the horizontal axis in Fig. 2.

The solid lines in Figs. 3 and 4 give the expected values of lateral position from the first-order predictions. The lines are normally vertical in Fig. 3 because for HP+ the best-fit lag is simply the applied ITD. However, as the boundary frequency ω_c increases, the best-fit interaural phase ϕ_p ($\phi_p = \omega_c \tau_p$) surpasses 180°. According to the principle of centrality (Stern *et al.*, 1988) listeners lateralize narrow-band stimuli within the range of −180 to +180°. Consequently, for some boundary frequencies and ITDs, the first-order predic-

TABLE I. RMSE and correlation coefficient (r) between experimental lateral responses and the first-order predictions. The last column on the right indicates an average over listeners. RMSE values can be compared with the scale −40 to +40, corresponding to locations on the extreme left and extreme right.

Listener		C	L	W	X	Z	Average
Linear-phase	RMSE	13.05	16.17	9.58	8.78	7.59	11.03
HP	r	0.90	0.88	0.87	0.86	0.89	0.88
Stepped-phase	RMSE	13.58	⋯	9.39	8.89	8.93	10.19
HP	r	0.91	⋯	0.89	0.88	0.92	0.90
Sine-tones	RMSE	16.08	⋯	9.73	10.06	11.29	11.79
HP	r	0.91	⋯	0.88	0.81	0.89	0.87

tions bend toward the center, shifted laterally by the reciprocal of the boundary frequency, to satisfy the centrality principle.

The solid line predictions on the graph include the assumption that over the course of experimental runs listeners learn about the leftmost and rightmost stimuli and use the range from −40 to +40 based on this experience. The largest τ_p in the experiments was 2500 μs, and it occurred for HP− with an ITD of zero and a boundary frequency of 200 Hz. Thus, a linear scale was established for the predictions whereby a τ_p of zero was expected to equate to a response of 0 and a τ_p of ± 2500 μs was expected to equate to a response of ± 40. Therefore, the predicted lateral response for HP− in Fig. 4 is given by

$$L = \frac{40}{2500 \times 10^{-6}} \tau_p = \frac{40}{2500 \times 10^{-6}} \left(\mathrm{ITD} \pm \frac{1}{2 f_c} \right), \qquad (1)$$

where terms in the parentheses are expressed in seconds and the $\pm$ sign is chosen to minimize the absolute value of the response.

The dashed curves in Fig. 4 correspond to phase shifts of $\pm 180°$. If centrality holds then all the data points ought to fall between these dashed curves. Alias points, where the interaural phase differs by 360° from the first-order predictions, are shown by triangles. The triangles were included on the graphs to test the idea that lateral position estimates outside the central region might preferentially cluster at alias points.

In general, visual inspection of Figs. 3 and 4 shows that the lateral responses of listener Z tended to follow the predicted solid curves qualitatively, but not perfectly. For instance, for HP+ (Fig. 3), the data points did not cluster about the solid lines but were shifted toward the outside, i.e., the absolute values of the position estimates were larger than expected. For HP− (Fig. 4), the data with 0-ITD did not follow the hyperbolic functions but showed a weaker dependence on frequency.

An interesting effect appears for ambiguous stimuli where the ϕ_p is 180°, and for which the first-order predictions suggest lateralization on left and right sides equally. Ambiguous stimuli occur in Fig. 3 at 500 Hz with ITDs of ± 1000 μs and 1000 Hz with ITDs of ± 500 μs. They occur in Fig. 4 for all frequencies with ITD of 0 μs and for 1000 Hz with ITDs of ± 1000 μs. For these ambiguous stimuli, listener Z tended to respond on the rightmost of the time (73%), indicating a rightward bias.

The rightward bias is also evident in Fig. 3(e) where listener Z usually responded on the right (85%) for frequencies above 500 Hz when centrality predicts responses on the left or centered. Similar findings also occurred for other listeners in that lateral responses for large values of τ_p showed left/right biases for stimuli that were not actually ambiguous. These biases were made evident by lateral position estimates that were systematically contrary to the predictions of centrality. These biases will appear in summary data below.

It is evident from Figs. 3 and 4 that the lateral position estimates by listener Z agreed with the first-order predictions only approximately, and this conclusion can be reached for the other listeners as well by examining figures of the form of Figs. 3 and 4. The average results for each listener in Experiment 1 are shown by open circles in summary figures, Figs. 6 and 7. Figures 6 and 7 use the same coordinate system as Figs. 3 and 4. These summary figures will be discussed later in this article.

E. Quantitative comparison with the first-order predictions

The lateral responses made by the listeners were compared with the first-order predictions by computing two statistics. The first was the rms error (RMSE) between the responses and the predictions. For each listener the calculation of RMSE was based on the average of the ten repeated measures for each of 50 conditions (5 boundary frequencies $\times$ 5 ITDs $\times$ 2 boundary types). Absolute values were taken for the ambiguous stimuli before averaging. For nonambiguous stimuli, there were a few cases (never more than 2 out of the 50 conditions) when listeners responded on the opposite side to the predictions because of the individual lateralization biases noted above. To prevent these outliers from dominating the calculation of RMSE, only the data points on the same side as predictions were included. Thus, some RMSE values were calculated for 48 or 49 points instead of 50.

The RMSE is shown in the top block of Table I. The scale for the RMSE is the scale of the responses, −40 to +40. Thus typical values from the table (about 10) correspond to about one-quarter of the distance from the center to the extreme—a rather large discrepancy.

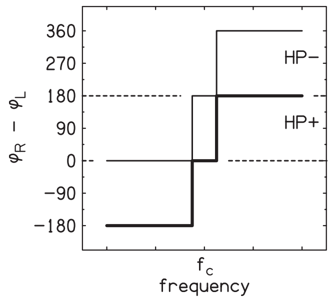

FIG. 5. Interaural phases for HP− and HP+ for stepped-phase stimuli.

The second statistic is the correlation coefficient between the data and first-order predictions. It too is shown in the top block of Table I. The r values are all greater than 0.85, indicating that the experimental data approximately followed the prediction in spite of the rather large rms discrepancies.

III. EXPERIMENT 2: STEPPED-PHASE BOUNDARY

In a stepped-phase version of the HP, the boundary region is made with a constant interaural phase, ϕ_o, different from the background phase (Yost, 1991; Grange and Trahiotis, 1996). If ϕ_o differs from the background phase by 180°, then the first-order predictions for the lateral position of the stepped-phase image are identical to the predictions for the linear-phase image in Experiment 1 (see Appendix).

This prediction is not an obvious one because there is a big difference between the stepped-phase stimulus and the linear-phase stimulus. When heard by itself, in the absence of the background, the stepped-phase boundary can easily be lateralized, and this lateralization does not seem to be different from the lateralization with the background present. By contrast, the linear-phase boundary alone, shown by the bright horizontal segment at 200 Hz in Fig. 2(a) or Fig. 2(b), cannot be consistently lateralized (Zhang, 2006). This latter result is hardly surprising. The linear-phase boundary region contains all possible interaural phases with a uniform distribution over a frequency range that is much narrower than a critical band. Unlike the tendency to hear the pitch at the average frequency (f_c), there appears to be no tendency to lateralize based on some form of average interaural phase. Instead, the lateralization of the linear-phase boundary depends on a contrast between the boundary region and the background.

Experiment 2 used a stepped-phase stimulus with constant phase ϕ_o differing by 180° from the background. Because the stepped-phase boundary region can be lateralized on its own, it was conjectured that this form of boundary region might lead to better agreement with the first-order predictions than did the linear-phase stimulus. On the other side was the first-order prediction that the two types of phase boundaries should be identically lateralized.

A. Method

Experiment 2 was identical to Experiment 1 except that the phases in the boundary region did not increase smoothly with frequency but were constant. As shown in Fig. 5, the boundary phases were fixed at 0° (HP+) or at 180° (HP−), causing the interaural phase to exhibit a step with respect to the background phases. The four most reliable (least variance about their individual means) listeners, C, W, X, and Z from Experiment 1, participated in the experiment.

B. Results

The averaged lateral responses for HP+ and HP− in Experiment 2 are shown by filled circles in summary figures, Figs. 6 and 7. Although we had conjectured that a stepped-phase boundary might lead to different lateral responses, the results actually provided a dramatic exhibition of reproducibility. The Pearson correlation coefficients between the lateral responses in Experiments 1 and 2 were very high with an average of 0.97, as shown in the top row of Table II. This result shows that listeners lateralized the HP stimuli similarly for the linear-phase and stepped-phase boundaries, consistent with the first-order predictions.

Parallel to Experiment 1, the rms error and correlation were calculated to compare the results of Experiment 2 with the first-order predictions. The calculated values are shown in the middle block of Table I. The resulting values were similar to those for the same listeners in Experiment 1.

IV. EXPERIMENT 3: SINE TONES

The HP sounds like a sine tone in noise. The HP stimulus boundary is tonotopically local like a tone. By treating the Huggins effect as a tone in noise, the EC model makes the connection to the masking level difference (MLD) (Durlach, 1962). The CAP model and the RC model predict that the HP with a boundary centered on a given ITD is lateralized the same as a tone with comparable ITD. The SPD model uses a sine-tone cross-correlation template to determine HP lateralization. For these reasons, we performed Experiment 3, which studied the lateralization of sine tones for comparison with the lateralization of HPs. A sine tone with 0-IPD (sin−0) is analogous to HP+. A sine tone with 180°-IPD (sin−π) is analogous to HP−.

A. Method

Experiment 3 for tones was identical to Experiments 1 and 2 for HP except that a sine tone replaced the HP stimulus. Effectively, the stimulus for Experiment 3 was equivalent to the stimuli of Experiment 1 or 2 with all spectral components removed except for the single component at the boundary frequency. Thus, there was no noise background. The sine-tone level was 45 dB to approximate the loudness of the HP in Experiments 1 and 2 according to informal estimates. Listeners C, X, W, and Z did ten runs for sin−0 interleaved with ten runs for sin−π.

B. Results

Averaged results of Experiment 3 appear as "+" symbols in Figs. 6 and 7 for sin−0 and sin−π, respectively. Visually comparing the + symbols with the open and filled circles in these figures, one can observe that the + symbols and the circles tend to follow the same trend, although the

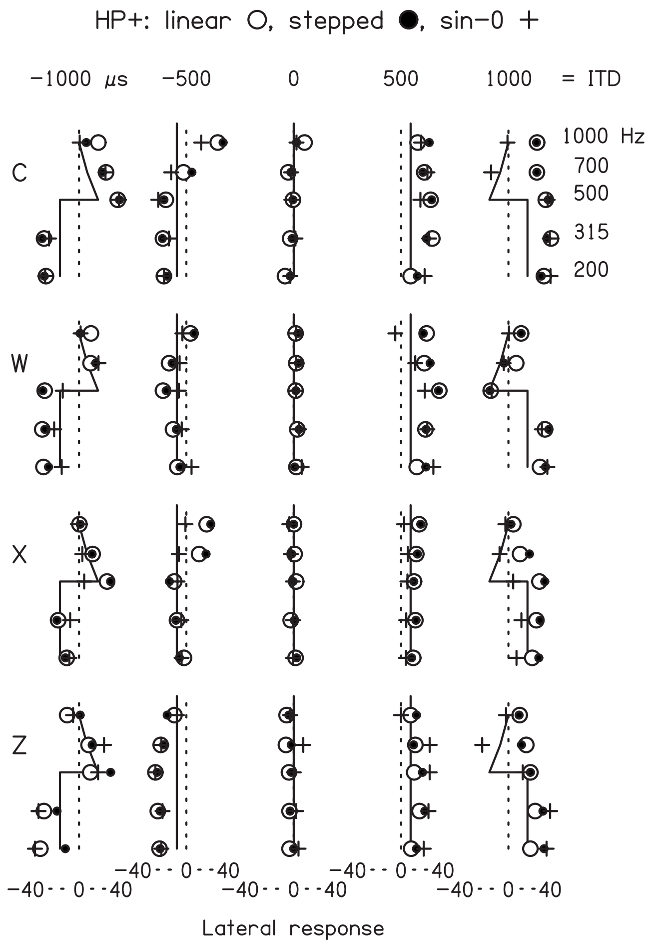

FIG. 6. Results of Experiments 1–3 for HP+ and sin−0 for four listeners. As for the scatter plot illustrated in Fig. 3, the vertical axis shows the boundary frequency (or sine-tone frequency) and the horizontal axis shows the listener's lateral responses averaged over ten trials. The vertical dashed line indicates the zero of lateral position. The solid line shows the first-order prediction.

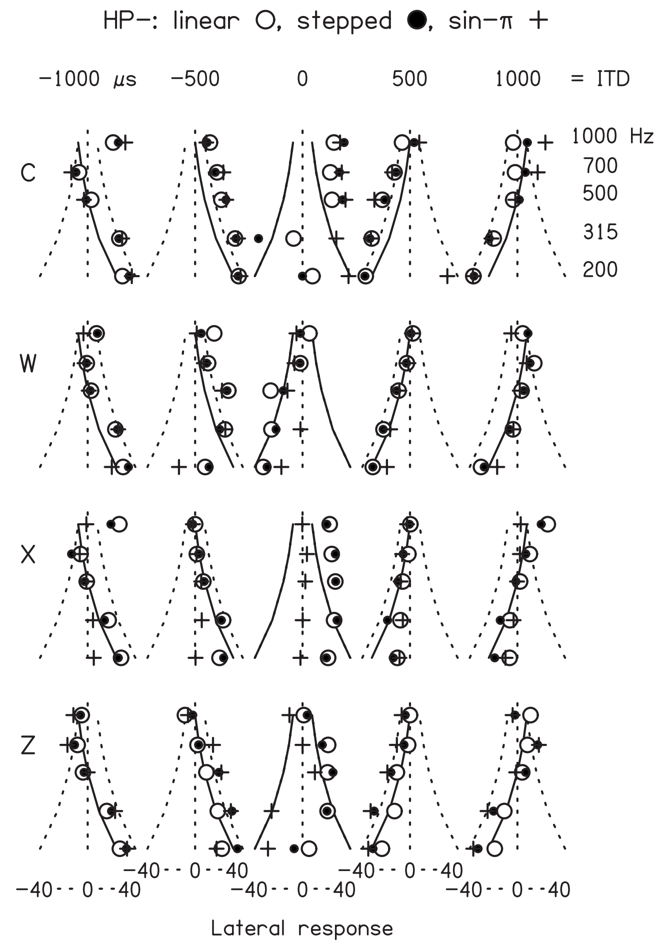

FIG. 7. Results of Experiments 1–3 for HP− and sin−π for four listeners. The form is the same as for the scatter plot illustrated in Fig. 4. The vertical dashed line indicates the zero of lateral position. The solid line shows the first-order prediction.

open and filled circles are more similar to each other. For the four listeners who did all three experiments, the correlation coefficients comparing the sine-tone results with the HP results in Experiments 1 and 2 are high (the bottom two rows in Table II), and average to 0.86 and 0.88, respectively. These results suggest that there is a common mechanism for lateralizing HPs and sine tones. Similar to Experiments 1 and 2, the RMSEs and correlation coefficients were calculated to compare the experimental data with the first-order predictions. They are shown in the bottom block of Table I. The results are comparable to the results of Experiments 1 and 2 for HP.

V. DISCUSSION

A. Summary of the experiments

The best comparison between the three experiments is obtained by examining the individual scatter plots for each listener, similar to Figs. 3 and 4.[1] Figures 6 and 7 for the mean values of lateral responses serve as summaries of the individual scatter plots. Of course, averaging lateral responses leads to possible cancellation between positive (right) and negative (left) responses, especially for the ambiguous stimuli. Such cancellation would distort the interpretation by falsely representing responses as too close to the midline. Fortunately, because listeners usually responded with a bias to one side or the other for ambiguous and near-ambiguous stimuli, such distortions were infrequent. That fact can be seen in Figs. 6 and 7 where the symbols indicating lateral responses for ambiguous stimuli are located well away from the midline.

In Fig. 6 (HP+) there are 25 data points for each listener, corresponding to 25 conditions (5 boundary frequencies $\times$ 5 ITDs). Each data point is an average of ten repeated measures from ten runs. The first-order predictions (solid curves) are that listeners' lateral responses should form straight lines except for accommodations to the centrality principle. In Fig. 7 (HP−), there are similarly 25 data points

TABLE II. Correlation coefficients of lateral position responses across three different stimuli. Linear and stepped refer to two types of HP boundary region.

Listener	C	W	X	Z	Average
Linear vs stepped	0.98	0.97	0.97	0.96	0.97
Linear vs sine	0.82	0.88	0.84	0.89	0.86
Stepped vs sine	0.84	0.90	0.83	0.93	0.88

for each listener, and the predictions are that the lateral responses should form a hyperbolic function per the solid curves.

Visual observation of Figs. 6 and 7 confirms that judgments tended to follow the predictions qualitatively. For instance, for frequencies below 700 Hz, lateral responses for nonambiguous conditions were 100% on the same side as the predictions for listeners X and Z, 96% for listener C, and 92% for W. The exceptions for C and W occurred for sine tones. Moreover, for HP+, the lateral responses (especially below 700 Hz) had little frequency dependence as predicted. By contrast, the corresponding responses for HP− demonstrated considerable frequency dependence, as expected, though not as much as predicted. Furthermore, when comparing across the three types of stimuli (i.e., linear-phase boundary, stepped-phase boundary, and sine tones), each listener's lateral responses were very similar, and the three symbols in Figs. 6 and 7 overlap fairly well. The sine-tone data for listener X are systematic exceptions.

B. Correlation coefficients

Pearson correlation coefficients comparing the lateral responses in the three different experiments were calculated for each listener. These correlation coefficients were based on all ten trials for each individual stimulus, with absolute values used for ambiguous stimuli. Therefore, the data sets used to calculate the correlation coefficients were slightly more complete than the data used in Table I.

The correlation coefficients are shown in Table II. All the values in the table are above 0.82, indicating similarity among the lateral responses for the three types of stimuli. The large correlations in the top row (two forms of HP) suggest that the detailed information within the boundary region does not influence the lateralization of HP. For each listener, the correlation coefficient in the top row is always greater than those in the bottom two rows. One-tailed paired t-tests for the four listeners show that the difference is significant at the 0.02 level. These results show more similarity in lateralization between HPs created with different boundary types than between HPs and sine tones.

C. Fitted function

To represent the lateralization data in a systematic way, we plotted lateralization responses as a function of the best-fit lag, τ_p. These plots are given in Figs. 8–10 for linear-phase HP, stepped-phase HP, and sine tones, respectively. The plots show responses averaged over the ten trials for each of the 25 stimuli in an experiment, with absolute values taken for ambiguous stimuli. The circles represent the responses for HP+ ($\sin - 0$), and the triangles represent the responses for HP− ($\sin - \pi$). Ideally, all the data points should fall in the first and third quadrants, indicating lateral responses on the expected side. This was true of almost all the data. Occasionally, data points fell into the wrong quadrants, and those points (very few and never more than two out of the 50 points for any given listener on a single figure) are not included in the figures. This omission made it possible to show the plots for all listeners in one figure. The y-axis labels for positive τ_p are displayed on the right axis, and the labels for negative τ_p are displayed on the left axis. The dashed straight lines show the first-order predictions, which assume a fixed scale factor that assigns lateral responses of ± 40 to τ_p of ± 2500 μs.

The lateralization data for each listener were fitted by a function of the form

$$L = a \cdot \tau_p^q + b, \tag{2}$$

where τ_p is in μs.

The parameters allow for an individual offset bias (b), a scale factor (a), and a laterality compressive exponent (q). As shown in the lateralization estimation experiments by Yost (1981) and by Zhang and Hartmann (2006), the perceived lateral positions of sine tones are not linear functions of the IPD when IPDs beyond 90° are included in the stimulus set. Instead, the lateral judgments are compressed at large IPDs.

The fitting process computed the difference between the data points and the functional form and minimized the RMSE for the data points shown in the plots. The three-parameter function, including compression, always agreed with the data better than did the first-order predictions. The RMSEs are always smaller and the correlation r values are always larger in Table III compared to Table I. A paired t-test shows that the improvement in RMSEs is significant at the 0.01 level. The best fits to the experimental data are shown as solid curves in Figs. 8–10. The optimal parameters (a, b, and q) are also shown in Table III.

In Table III, the parameters for different listeners were quite different. By contrast, for a given listener, the q-parameters, representing the amount of compression, and the b-parameters, representing the left/right bias, were very similar across the three types of stimuli. The only exception was listener X, for whom sine tones led to much more compression ($q=0.17$) than did HP stimuli ($q=0.41$), and for whom left/right bias (b) was different for different stimuli.

Several variations on the fitting procedure were done in order to determine the importance of the parameters. Forcing the fit to be linear ($q=1$) but allowing for individual differences in offset and slope led to smaller values of RMSE compared to the original first-order estimates with $b=0$ and $a=1$. The difference was significant at the 0.0001 level in a one-tailed paired t-test. However, this best linear fit resulted in a RMSE that was larger by a factor of 1.5 compared to the RMSE for the compressive three-parameter fit, shown in Table III. The difference was significant with a p-value of 2×10^{-5}. This statistical result confirms the importance of laterality compression.

In a second variation, the τ_p in Eq. (2) was replaced by the ϕ_p and the RMSE was again minimized. This substitution increased the RMSE significantly, $p < 0.05$. As was found in the sine-tone lateralization experiments by Zhang and Hartmann (2006), the perceived lateral positions for HPs are more consistent with τ_p than with ϕ_p. In other words, lateral position corresponds better to ITD than to IPD.

There is evidence that listeners adjust their response scales to the range of best-fit lags arising in a given experiment run. The HP+ runs, summarized in Fig. 6, had a maxi-

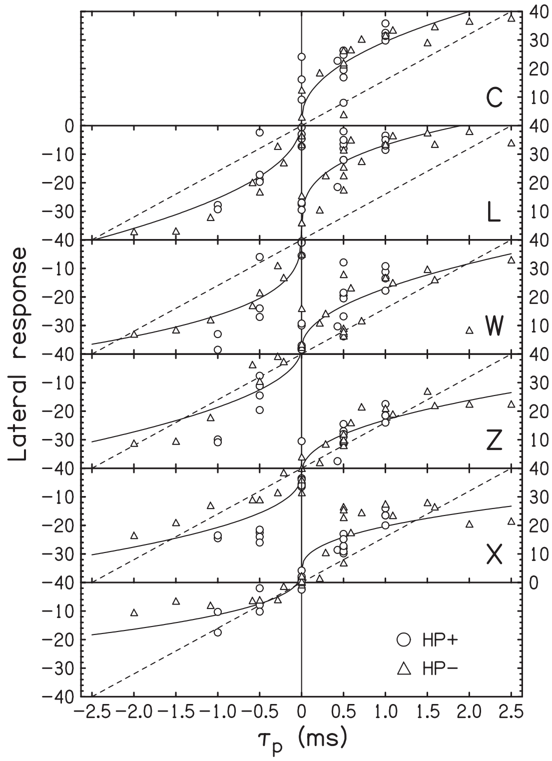

FIG. 8. Mean lateral responses for HP stimuli with linear-phase boundary, plotted against τ_p. Responses are limited to the range −40 to +40. Only responses with the same sign as τ_p are plotted. The dashed lines correspond to the first-order predictions. The solid curves are the best-fitting three-parameter compressive functions. All points with $\tau_p=0$ correspond to lateral responses near zero, never near +40 or −40.

mum τ_p of 1000 μs. The HP− runs, summarized in Fig. 7, had a maximum τ_p of 2500 μs. The first-order predictions were based on a scaling that assumes a maximum of 2500 μs. Consequently, if listeners readjust their responses, one expects that responses in Fig. 6 should tend to be more

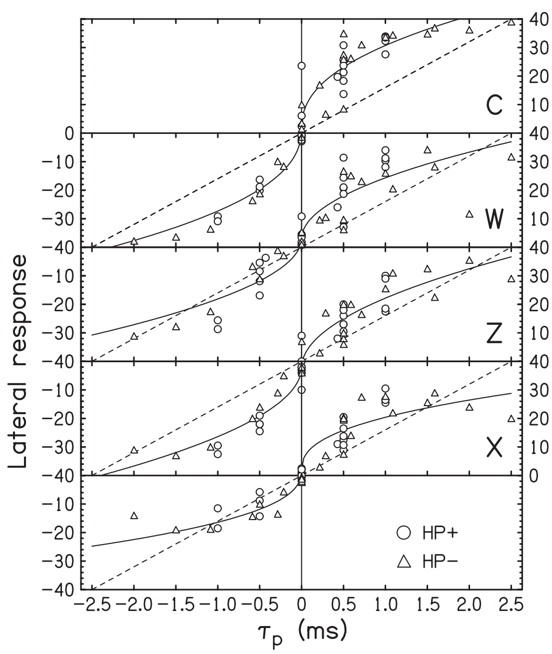

FIG. 9. Same as Fig. 8, except for stepped-phase boundary.

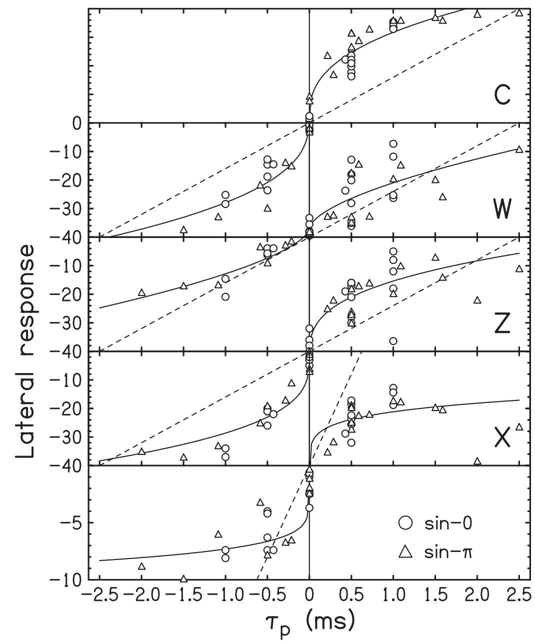

FIG. 10. Same as Figs. 8 and 9, except for sine tones. The vertical range for listener X is from −10 to +10, smaller than for other listeners.

lateral (larger magnitudes) than first-order predictions. Furthermore, because of the compression at large τ_p, one expects that compression should be especially evident in Fig. 7, where the large τ_p values should cause the observed dependence on τ_p to be weaker than first-order predictions. Both of these expected effects are evident in Figs. 6 and 7.

To test the idea that blocking trials on stimulus type affects the estimate of compression, the optimum parameters were recalculated fitting the compressive function to lateral response data for each type of phase boundary (i.e., HP+ or HP−) separately—similarly for the corresponding sine tones. For a given listener and a given stimulus, the q-parameter

TABLE III. Best-fitting parameters for the three-parameter compressive function. Units for a are μs^{-1}. By comparison, for the first-order model, $a=0.016\ \mu s^{-1}$, $b=0$, and $q=1$. RMSE and correlation coefficient (r) for the best fit are also included. RMSE values can be directly compared with those in Table I.

Listener		C	L	W	X	Z	Average
Linear phase	a	1.07	3.27	0.66	0.91	0.84	
	b	2.0	3.4	2.2	4.2	−1.9	
	q	0.47	0.32	0.50	0.41	0.45	
	RMSE	6.98	6.22	7.20	4.99	4.90	6.06
	r	0.95	0.97	0.91	0.92	0.95	0.94
Stepped phase	a	1.39	⋯	0.63	1.08	0.72	
	b	1.7	⋯	3.1	2.0	−2.1	
	q	0.44	⋯	0.51	0.41	0.51	
	RMSE	5.72	⋯	6.43	4.77	4.98	5.47
	r	0.97	⋯	0.93	0.95	0.97	0.95
Sine tones	a	2.02	⋯	0.24	1.86	2.54	
	b	1.9	⋯	3.1	−1.3	−2.0	
	q	0.39	⋯	0.61	0.17	0.34	
	RMSE	4.29	⋯	5.75	1.72	5.07	4.21
	r	0.98	⋯	0.90	0.94	0.97	0.95

TABLE IV. Number of left (L) or right (R) judgments for 110 ambiguous stimuli. The last row shows the percentage of all responses on each side for each listener. Greatly different percentages for left and right ears are an indication of earedness.

	C		L		W		X		Z	
Stimulus	L	R	L	R	L	R	L	R	L	R
Linear	22	88	6	103	61	49	0	109	23	80
Stepped	18	91	⋯	⋯	63	47	0	109	32	65
Sine	9	101	⋯	⋯	83	20	36	61	62	38
Total	15%	85%	5%	94%	63%	35%	11%	85%	35%	55%

was smaller for HP− than for HP+ in 12 cases out of 13, indicating larger compressive effect for HP−, confirming the expectation. The one exception was listener Z for linear-phase boundary. A one-tailed paired *t*-test showed that this effect is significant at the 0.0005 level.

D. Earedness

The stepped-phase HP stimulus centered on an IPD of 180° is binaurally symmetric. A linear-phase HP stimulus centered on 180° is almost symmetric. There is no reason to expect the pitch image to be on the left or on the right. However, Experiments 1 and 2 showed that listeners lateralized ambiguous HP stimuli to one side or the other rather reliably. Experiment 3 showed that most listeners also have biases for ambiguous sine tones that are similar to their biases for HP stimuli.

These observations of inexplicable bias can be made quantitative for the 11 ambiguous conditions. There were four ambiguous conditions for HP+: 500 Hz with ITDs of ± 1000 μs and 1000 Hz with ITDs of ± 500 μs. There were seven ambiguous conditions for HP−: 1000 Hz with ITDs of ± 1000 μs, and all the five stimuli with ITDs of zero. There were ten repeated measures for each of the 11 ambiguous conditions. Thus, there are 110 points for ambiguous stimuli on each of the original scatter plots such as Figs. 3 and 4. Table IV shows the number of responses on left and right for each listener and each stimulus. Sometimes the numbers on left and right do not add up to 110 because the listener responded "zero" for a few points, and these were counted neither for the left nor for the right. The bottom row of Table IV shows the overall percentage on each side for each listener. It is clear that listeners had preference for one side or the other. In Table IV, among the 13 (left-right) pairs of numbers, there are eight cases in which the bias is strong (75% or more for a preferred side). In four other cases the bias is moderate and toward the same preferred side.

This tendency to hear an ambiguous stimulus on one side or the other conjures the concept of left-eared or right-eared listeners, analogous to left-handed or right-handed persons (Hartmann *et al.*, 2004). Table IV shows that listeners C, L, and X performed as right-eared listeners and that listener W appeared to be a left-eared listener. Listeners demonstrated a somewhat different level of earedness for HP stimuli and sine tones. Listener Z even became a left-eared listener for the sine tones, whereas he was right eared for the HP [but see the interaural level difference (ILD) discussion below].

Besides the earedness effect for the ambiguous stimuli, listeners C, L, X, and Z were so right eared that, for the nonambiguous stimuli of HP+ with a boundary frequency of 700 Hz and an ITD of +1000 μs, they preferentially chose the alias points on the right. There was no tendency to do the reverse when the ITD was −1000 μs (see Fig. 6). For these 700 Hz stimuli (± 1000 μs and both linear-phase and stepped-phase boundaries), the percentages of total responses on the right side were very high: 95% for C, 90% for L, 98% for X, and 88% for Z. By the symmetry of this pair of stimuli, the expected value would be 50%. Left-eared listener W did not show a strongly consistent bias for these stimuli. For all listeners, this bias with nonambiguous stimuli did not appear for sine tones.

The earedness effect for HP can also be dramatically demonstrated in an informal two-step experiment. First, a listener puts on headphones and hears the HP− stimulus (without added ITD). A compact disk, with this and other dichotic stimuli, is available (Bilsen and Raatgever, 2002). This HP− stimulus is ambiguous. The listener is asked to say whether the pitch image is on the left or right. Then the headphones are reversed. Usually, the listener discovers, to his or her surprise, that the pitch image is still lateralized on the same side of the head. Culling (Hartmann *et al.*, 2004) performed this brief experiment using 36 listeners. He found that 12 listeners heard the pitch on the right side and 14 listeners heard the pitch on the left side, independent of headphone orientation. Ten listeners were unsure about the lateralization for one or both orientations, and two listeners actually said that the pitch image moved from one side to the other when the headphones were reversed. There was no correlation between earedness and handedness.

Apparently there is a strong tendency for a given listener to lateralize HP− to a particular side, at least for a constant boundary frequency. In Culling's experiment, more than 75% of the listeners were either left eared or right eared. In our experiments, the earedness found for individual listeners in HP was largely maintained for sine tones with an IPD of 180°. However, earedness seemed to have a reduced influence for tones while centrality had increased influence.

The ear preference found in our experiments and in Culling's contrasts with the results of Hafter *et al.* (1969) who found that listeners lateralize the signal in an NoSπ stimulus about equally often to the left and right. The difference might be attributed to method. Hafter *et al.* (1969) presented only IPD of 180° throughout the entire experiment and varied signal-to-noise ratio as a parameter. As suggested by the reduced earedness that we found for sine tones, part of the difference might arise from the difference between a HP stimulus and a MLD stimulus. In the end, the sidedness tendency seen in HP experiments seems strong enough to justify the concept of left-eared or right-eared listeners, necessarily leaving some room for the binaural analog to ambidextrous.

A trivial explanation for the strong left/right bias observed for ambiguous stimuli would attribute the effect to an audiometrically better ear and consequently an effective ILD. This explanation is plausible for ambiguous sine tones, for which lateralization depends on ILD. It is much less plausible for HPs. Although Grange and Trahiotis (1996) found that the lateral positions of HPs could be shifted somewhat by ILDs, informal observations by our listeners agreed with Raatgever and Bilsen (1986) that lateralization is very insensitive to ILDs, i.e., the HP seems to be a "time image." This distinction between sine tones and HP may be responsible for the anomalous reversal of earedness seen for listener Z when sine stimuli replaced HP stimuli. For HPs, it is more likely that the observed earedness reflects an asymmetry in midbrain processing.

VI. CONCLUSION

The lateral position of the pitch image was measured for two forms of HP, HP+ and HP−, using a direct psychophysical method. Experiments with five different boundary frequencies and five different frequency-independent ITDs showed that, for each listener, the experimental lateralizations of HPs correlated strongly with the lateralizations of corresponding sine tones. This result agrees with the predictions from the CAP model (Raatgever, 1980; Raatgever and Bilsen, 1986) and the RC model (Akeroyd and Summerfield, 2000). It suggests that a common localization mechanism applies to quite-different stimuli.

The experiments found even higher correlations between the lateral responses for HP made with the linear-phase and stepped-phase boundaries. The average correlation coefficient was 0.97. This result leads to the conclusion that the detailed information within the boundary region does not affect the lateral position of HP. The result is somewhat surprising because the stepped-phase boundary region can be lateralized on its own while the linear-phase boundary region cannot. The SPD model, presented in the Appendix, shows a similar insensitivity to the details of the phase boundary region and predicts that both forms of HP will be lateralized the same.

The experiments were motivated by a first-order hypothesis consisting of the SPD model and scaling assumptions. The SPD model predicts that lateralization of dichotic pitches is determined by the best-fit lag, τ_p, which corresponds to the peak of the cross-correlation function for these stimuli. The scaling assumptions are as follows: (1) lateralization judgments are linearly related to the best-fit lag values, τ_p, and (2) listeners apply the full range of lateralization responses to the full range of τ_p under all circumstances. The experimental results were approximately in agreement with that hypothesis in that the average lateral responses tended to be monotonically related to the predictions. However, detailed agreement could only be obtained by incorporating several modifications. Least important of the modifications was an offset, left or right, for each listener. Signs of the offsets were unrelated to a listener's earedness, left or right.

Much more important were individual scale factors and nonlinear compression in the relationship between τ_p and the estimates of lateral position. Incorporating those modifications reduced the rms discrepancy between predictions and data by a factor of 2. Compressive exponents for HP− tended to be about 0.5. Consequently, the hyperbolic dependence of lateral position on boundary frequency expected from the first-order predictions for HP− does not normally hold. The square root of that hyperbolic dependence holds much better. The observed importance of laterality compression indicates the value of the direct psychophysical method used in our experiments. Compression cannot be observed using an acoustical pointer if the pointer itself is laterally compressed. Other direct experiments show that ILDs, ITDs, and IPDs are all compressed perceptually (Yost, 1981; Zhang and Hartmann, 2006).

There was good evidence that the individual scale factors and compressive exponents depend on the experimental context—most likely on the range of τ_p occurring in an experimental block. Compressive exponents were smaller (greater compression) for HP− than for HP+ because of the greater range in τ_p for HP−.

For ambiguous stimuli, such as HP− and sin$-\pi$ with no added ITD, the signal has a best-fit phase, ϕ_p, of 180° ($S\pi$). Listeners usually lateralized these binaurally symmetric stimuli consistently on one side. This observation led to the concept of earedness, i.e., left-eared and right-eared listeners.[2] Surveys by Hartmann *et al.* (2004) found no correlation between earedness and handedness. The phenomenon of earedness indicates an asymmetry in the binaural system. It may someday have clinical significance in midbrain neurology.

ACKNOWLEDGMENTS

This work was supported by the NIDCD under Grant No. DC 00181. Many of the experiments were performed at the Hearing Research Center in the Department of Biomedical Engineering at Boston University, partially supported by NIDCD Grant No. DC 00100. We are grateful to Dr. H. S. Colburn, Dr. F. A. Bilsen, and Dr. M. A. Akeroyd for useful discussions and to Dr. J. F. Culling for permission to report his otherwise unpublished data that helped us (and him) conclude that there are right-eared and left-eared listeners. Ms. Yongfang Zhu helped with statistical matters. Dr. Armin Kohlrausch and two anonymous reviewers provided useful comments on earlier versions of this report.

APPENDIX: THE SALIENT PHASE DENSITY MODEL

The SPD model is an intuitively plausible stimulus-based model for a first-order calculation of the lateralization of a dichotic pitch. Like all models of dichotic pitch, the SPD model begins by isolating a narrow band of frequencies as distinct from the background because of an unusual interaural phase property within the narrow band. This step is explicit in the RC model of Akeroyd and Summerfield (2000). In the SPD model, the bandwidth of the isolated region is not critical; the model assumes that it is determined by an auditory filter approximately centered on the phase boundary region. With N spectral components in the isolated region, the left-ear signal is

$$x_L(t) = \sum_{n=1}^{N} \cos(\omega_n t + \phi_{L,n}), \tag{A1}$$

where ω_n and $\phi_{L,n}$ are the angular frequency and phase of the nth component. The component amplitudes have all been set equal to 1.0 for simplicity because the results of the calculations below do not depend in any important way on the amplitudes. The right-ear signal $x_R(t)$ is similarly defined. With no loss of generality, it can be assumed that frequencies ω_n are separated by a constant frequency spacing, $\delta\omega = 2\pi/T_D$, where T_D is the stimulus duration.

The SPD model begins with the cross-correlation function between the two ears,

$$\gamma(\tau) = \int_0^{T_D} dt\, x_L(t)\, x_R(t+\tau), \tag{A2}$$

and for the signal as described,

$$\gamma(\tau) = \frac{1}{N}\sum_{n=1}^{N} \cos(\omega_n \tau + \Delta\phi_n), \tag{A3}$$

where τ is the lag and $\Delta\phi_n$ is the IPD for component n, $\Delta\phi_n = \phi_{R,n} - \phi_{L,n}$. This interaural phase function is the essential physical ingredient in a dichotic pitch stimulus.

It is convenient to write the component frequencies in terms of the boundary frequency, ω_c, so that $\omega_n = \omega_c + \delta\omega_n$. The boundary frequency is important because the cross-correlation function for a narrow noise band oscillates with frequency ω_c. Also, ω_c is the perceived pitch for a HP stimulus (Hartmann, 1993).

Because a dichotic pitch like the HP resembles a sine tone in noise, the SPD model assumes that the binaural system compares $\gamma(\tau)$ with the cross-correlation function for a sine tone, namely, with the template function $\cos(\omega_c\tau + \phi)$, where ϕ is the interaural phase of the template sine tone. The value of ϕ that leads to the best match with $\gamma(\tau)$, namely, ϕ_p, will correspond to the model prediction for the listener's lateralization of the pitch image. Specifically, the pitch image will be lateralized according to an interaural delay $\tau_p = \phi_p/\omega_c$.

Mathematically, the best value of ϕ in the overall least-squares errors sense is that value which maximizes the overlap integral,

$$F(\phi) = \frac{1}{T}\int_{-T}^{T} d\tau\, \gamma(\tau) \cos(\omega_c \tau + \phi), \tag{A4}$$

where T is the range over which the binaural system can register ITDs. Typically T is assumed to be between 1 and 4 ms.

The product of $\gamma(\tau)$ and $\cos(\omega_c\tau + \phi)$ includes frequency summation terms and frequency difference terms. For the summation terms the argument of the cosine function in Eq. (A4) has a range of approximately $2\omega_c \times 2T$. Because this range typically corresponds to at least several cycles of oscillation, the integral of the frequency summation terms includes cancellations that makes it smaller than the integral of the difference terms. Ignoring the summation terms leads to

$$F(\phi) = \frac{1}{2T}\int_{-T}^{T} d\tau \frac{1}{N}\sum_n \cos(\delta\omega_n \tau + \Delta\phi_n - \phi), \tag{A5}$$

or

$$F(\phi) = \frac{1}{N}\sum_n \frac{\sin(\delta\omega_n T)}{\delta\omega_n T} \cos(\Delta\phi_n - \phi). \tag{A6}$$

The sinc function above is a slowly varying prefactor that is nearly independent of component frequency n. For the smallest $\delta\omega_n$, namely, $\delta\omega_n = 0$, the sinc function is 1.0. For $\delta\omega_n$ equal to its largest value the function decreases only by a little. This decrease can be calculated assuming that the isolated region is determined by an auditory filter (critical band). Auditory filter bandwidths are about 16% of the center frequency ω_c and the product $\omega_c T$ is normally not greater than π—corresponding to the dashed lines in Figs. 4 and 7. Thus $\delta\omega_n T$ is never greater than 0.08π and the sinc function is never smaller than 0.99. If the bandwidth or the maximum ITD is doubled the sinc function only drops to 0.96. Therefore, to a reasonable approximation the sinc function is independent of frequency, and

$$F(\phi) = \frac{1}{N}\sum_n \cos(\Delta\phi_n - \phi). \tag{A7}$$

In this approximation, the frequency dependence has disappeared. What remains is only the IPD values themselves with no reference to their ordering as functions of frequency. Therefore, it is possible to replace the sum with an integral over the interaural phase density $\rho(\phi)$,

$$F(\phi) = \int_{-\infty}^{\infty} d\phi' \rho(\phi') \cos(\phi' - \phi). \tag{A8}$$

The infinite limits on the integral allow for a completely general representation of phase density as long as it is normalized to unity. The cosine function in the integral is periodic; consequently a 2π translation of any portion of the phase density has no effect on F.

Because function F involves only the distribution of interaural phases, the SPD model works only along this phase axis, where the interaural phases are considered to be lateralization influences at the level of the midbrain. Frequency resolution within the isolated band plays no role.

The above development, starting from Eq. (A4) and up to this point, has been purely mathematical, using quantitatively reasonable approximations. The next step in the model makes an assumption. It replaces the density of physical interaural phases by a density of *salient* interaural phases. Salient phases are lateralization influences that are different from the background. In fact, the HP exists only because these phases are different from the background. Therefore, the concept of salient phases is logically unavoidable. The value of ϕ which maximizes the overlap F in Eq. (A8) is the

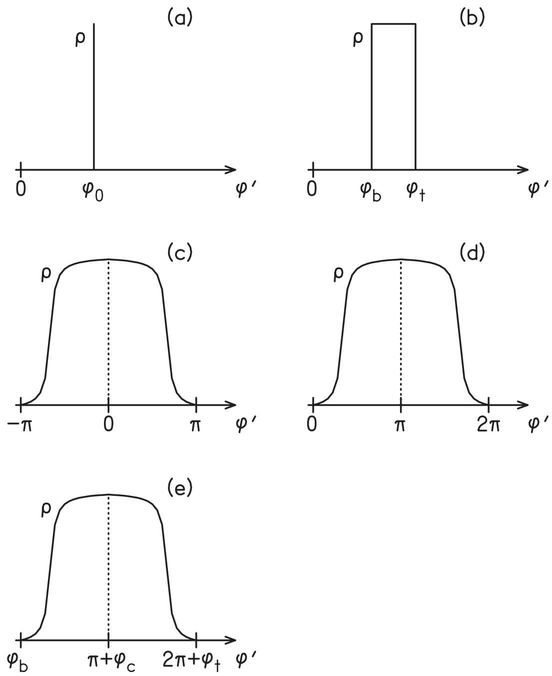

FIG. 11. Salient phase densities for (a) stepped-phase HP, (b) stepped-delay HP, (c) linear-phase HP+, (d) linear-phase HP−, and (e) linear-phase HP− with added ITD. Panels (a) and (b) are simply stimulus phase densities. Panels (c)–(e) make use of the salience principle. In panel (e) the phases ϕ_b, ϕ_c, and ϕ_t are equal to the added ITD multiplied by the angular frequencies of the boundary region bottom, center, and top.

best-fit phase lag, ϕ_p, and the best-fit lag, which predicts lateralization, is $\tau_p=\phi_p/\omega_c$. Several interesting results follow immediately.

Stepped phase. For a stepped-phase HP, SPD $\rho(\phi')$ equals $\delta(\phi'-\phi_o)$, where ϕ_o is the single, specific value of interaural phase within the boundary. This density function, from which the background phase has been eliminated, is shown in Fig. 11(a). Then

$$F(\phi)=\cos(\phi_o-\phi), \tag{A9}$$

which is maximized when $\phi=\phi_o$, leading to the prediction that the perceived location of the pitch image, namely, ϕ, is actually ϕ_o. This prediction for the lateralization of the pitch image is valid for any ϕ_o. However, if the value of ϕ_o is close to the interaural phase of the background then the HP may be weak or inaudible. Thus, a prediction for the lateralization does not guarantee the existence of the pitch effect. As applied in Experiment 2 of the body of this article, ϕ_o is 180° removed from the background interaural phase. Thus the pitch effect is strong, and the first-order prediction for the lateralization of the pitch image is 180° removed from the lateralization of the background—equivalent to the half-period rule.

Stepped delay. In the case of stepped delay, the boundary region is created by delaying the components in the boundary by a fixed delay time, t_d. Then $\rho(\phi)$ is a rectangular density,

$$\rho(\phi)=\frac{1}{\phi_t-\phi_b} \quad \text{for } \phi_b<\phi<\phi_t,$$

$$\rho(\phi)=0 \quad \text{otherwise}. \tag{A10}$$

Here $\phi_b=\omega_b t_d$ and $\phi_t=\omega_t t_d$ represent the bottom and the top of the phase boundary region, respectively. This density function is shown in Fig. 11(b). Then,

$$F(\phi)=\frac{\sin(\phi_t-\phi_b)}{\phi_t-\phi_b}\cos(\phi-\phi_c), \tag{A11}$$

where ϕ_c is the interaural phase at the center of the boundary region. Function F is maximized if $\phi=\phi_c$, predicting that the pitch image is lateralized the same as a sine tone with interaural phase of $\phi_c=\omega_c t_d$, corresponding to an interaural delay of t_d. Akeroyd and Summerfield (2000) noted that the stepped-delay stimulus is an important test case because the RC model predicts that the lateralization is determined by the boundary region, in agreement with experiment, but the half-period rule does not make this prediction. The SPD model also correctly predicts that the lateralization is determined by the boundary region.

Linear phase. The linear-phase case is the first for which the salience principle of the SPD model becomes crucial. For linear-phase HP+ and HP−, the interaural phases in the boundary region extend over the entire 360° range, and the physical IPD density functions are rectangular over that range. An integral of the form of Eq. (A8) over such phase densities would lead to a $F(\phi)$ that is identically zero with no prediction for the lateralization.

When the density of IPDs is replaced by the SPD the *effective* range of interaural phase does not extend over 360° because those interaural phases that are close to the interaural phase of the background are associated with the background and not with the pitch image. Only those interaural phases associated with the pitch image are salient. Consequently the SPDs have forms sketched in Figs. 11(c) and 11(d). Each SPD is approximately rectangular except that it goes to zero near the phase corresponding to the background. Interaural phases near the background phase are not salient, and they do not contribute to the lateralization of the pitch image. The exact shape of the SPD depends on the listener. It is likely that different listeners have different tendencies to assign phases to the tonal image and the background. As will be seen below, important predictions from the model do not depend on the fine details of the densities or on the way that not-salient interaural phases are eliminated from the effective phase density.

The densities shown in Figs. 11(c) and 11(d) correspond to HP+ and HP− in the special case where there is no added ITD. The densities are centered on $\phi'=0$ and $\phi'=\pi$, respectively. For these special symmetries, Eq. (A8) becomes

$$F(\phi)=\cos(\phi)\int_{-\infty}^{\infty} d\phi'\,\rho(\phi')\cos(\phi'). \tag{A12}$$

For HP+, the integral is positive because of the concentration of density around $\phi'=0$. Therefore, F is maximized by choosing $\cos(\phi)=1$, so that $\phi_p=0$, indicating that HP+ is lateralized in the center of the head. For HP−, the integral is negative because of the concentration of density around $\phi'=\pi$. Therefore F is maximized by choosing $\cos(\phi)=-1$, i.e., $\phi_p=\pm\pi$, indicating that HP− is lateralized to the extreme left or right. These predictions, made by maximizing F, do not depend on the details of the SPD, and they agree qualitatively with perceived lateralizations.

Random phase. Akeroyd and Summerfield (2000) introduced a form of HP in which the interaural phase is a random variable in the boundary region. From the point of view of the CAP model (Bilsen and Raatgever, 2000), which makes the normally harmless assumption of infinitely sharp frequency resolution, the random-phase stimulus looks peculiar because the CAP is infinitely ragged. Experimentally, however, the perceived lateralization for random phases agrees with the lateralization of HP+ and HP− for the linear-phase boundaries shown in Fig. 1. Informal experiments in our laboratory found that the lateralizations of random phases, linear phases, and stepped phases were all identical, depending only on the background phase. The idea that random phases should lead to identical lateralizations as the well-ordered linear or stepped phases is counterintuitive, and it led Akeroyd and Summerfield (2000) to propose the half-period rule which says that lateralization of any Huggins-type pitch, no matter how made, is half a period removed from the background.

For any particular choice of random-phase boundary regions, the SPD has no simple structure except that it becomes zero near the background phase. However, the ensemble average (Hartmann, 1998) SPD for random phases is identical to the SPD for HP+ or HP− as shown in Figs. 11(c) and 11(d), depending on the background phases. Therefore, in the average sense, the SPD model can successfully predict the unintuitive result that the lateralization for random phase is not different from lateralization for linear-phase HP. The reason is that, with the narrow-band approximation of Eq. (A8), the lateralization prediction does not depend on the size of the integral in Eq. (A12). It depends only on the *sign* of the integral. Possibly the expected variability of the lateralization depends on the size of the integral, but variability is beyond the scope of the present investigation. Furthermore, for any particular choice of random-phase boundary regions, the salience principle is enough to ensure that the pitch image will be lateralized away from the background, though how far away is not predictable.

HP+ and HP− with added ITD. When HP+ and HP− are created using linear phase or stepped phase and a constant ITD is added to the entire stimulus, as in the stimuli studied in this article, the symmetries that led to Eq. (A12) no longer apply. It is necessary to revert to Eq. (A8).

The added ITD has several effects. One effect is that there is no longer a unique background IPD. The background IPD depends on frequency, but it is not a rapidly varying function of frequency. The SPD model deals with this problem by the additional assumption that the background IPDs that are most important are those that come from frequencies in the neighborhood of the boundary region. This assumption is consistent with the notion that dichotic pitches are contrast effects; the existence of the binaural coherence edge pitch (Hartmann and McMillon, 2001) supports this notion. Because the added ITD is small, the choice of the frequency neighborhood of the boundary is not very critical. Even in

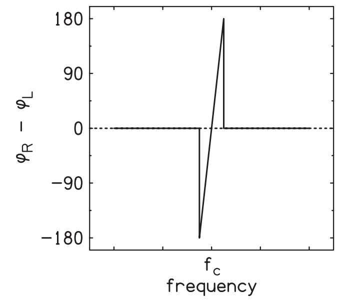

FIG. 12. The HP− stimulus with a boundary region that is shifted by 180°. This stimulus is an interesting test for the SPD model.

the worst case, where the added ITD is 1 ms and the boundary frequency is 1 kHz, the phase changes by less than 60° over a critical band.

The other effect of the added ITD is to shift the densities shown in Figs. 11(c) and 11(d) along the ϕ' axis. An example is shown in Fig. 11(e) for HP−. With the assumption that the relevant approximate background phase is established by neighboring frequency regions, the effect of the shift in SPD is to cause an equal shift in the value of ϕ that maximizes $F(\phi)$. Therefore, for these particular stimuli, the SPD model predicts that the pitch image will be lateralized approximately at τ_p, equal to the lag corresponding to the peak in the cross-correlation function. The approximation is better for small added ITDs because the variation in background IPD is smaller. These predictions, together with the scaling assumptions, lead to the first-order estimates for pitch lateralization in Figs. 3, 4, 6, and 7 of the body of this article. The impossibility of determining a single precise background phase might have measurable consequences on the variability of localization responses—beyond the scope of the present work.

Discontinuous phase boundaries. The discontinuous interaural phase function shown in Fig. 12 is made by applying a large phase shift, such as 180°, to the boundary region for HP− from Fig. 1. This transformation changes the phase at the center of the boundary region by 180°. Although the shifted boundary looks very different, this transformation leaves the SPD function unchanged from Fig. 11(d). Therefore, this stimulus makes an interesting test case for the SPD model because the prediction of the model is unusual. The prediction is that the lateralization does not correspond to the center of the boundary region but instead it remains the same as for HP−. Unpublished experiments in our laboratory with the shifted phase boundary show that the lateralization is indeed unchanged, as predicted by the SPD model.

Final notes. As applied in the present article, the SPD model was not seriously tested. Other models, the CAP model and the RC model, behave similarly given the stimuli used, and they might equally well have served as part of the first-order predictions. Furthermore, the lateralization prediction by the SPD model, specifically τ_p, happens to agree with the center of the phase boundary region and with the peak of the interaural cross-correlation function for the phase boundary region for these stimuli.

To avoid the confusion caused by the above coincidences, it is worth remembering that according to the model,

τ_p corresponds to the interaural phase of a sine tone having a cross correlation that best agrees with the cross-correlation function for a spectral region that includes the phase boundary. The value of τ_p is determined by the entire cross-correlation function for that region, not just its peak. The distinction is likely to be important for more complicated dichotic stimuli.

Although the development of the SPD model begins with the cross-correlation function, the unique dependence on the interaural phase density alone is a property that emerges from the mathematics in this Appendix. The introduction of the salience principle assumes that this phase density is accessible to the binaural system so that it can be modified by downweighting interaural phases similar to the interaural phases of the background. One of the benefits of the salience principle is that the spectral region passed by the auditory filter does not need to be precisely defined. Contributions to the phase density that come from that part of the background which is included in the region are eliminated in the contrast enhancement operation described by the principle.

[1]Scatter plots like Figs. 3 and 4 for all listeners and for all experiments in this article are available in an auxiliary report (Zhang and Hartmann, 2008) or in the Ph.D. thesis by Zhang (2006).

[2]This article is not the first report of an ear preference for Huggins pitch location. Dr. M. A. Akeroyd, Dr. F. A. Bilsen, and Dr. J. F. Culling have all remarked on it informally.

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

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. (**2001**). "On the compatibility of binaural models with dichotic pitch data," Proceedings of the 17th ICA, 4D.13.02, Rivista Italiana di Acustica, Rome, Italy (unpublished), Vol. **25**, pp. 168–171.

Bilsen, F. A., and Raatgever, J. (**2002**). "Demonstrations of dichotic pitch compact disc," Delft University of Technology, Delft, The Netherlands. This CD is available at fabilsen@xs4all.nl.

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

Culling, J. F. (**1999**). "The existence region of Huggins pitch," Hear. Res. **127**, 143–148.

Culling, J. F. (**2000**). "Dichotic pitches as illusions of binaural unmasking III—The existence region of the Fourcin pitch," J. Acoust. Soc. Am. **107**, 2201–2208.

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

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

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

Durlach, N. I. (**1972**). "Binaural signal detection—equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Vol. **2**, pp. 369–462.

Grange, A. N., and Trahiotis, C. (**1996**). "Lateral position of dichotic pitches can be substantially affected by interaural intensive differences," J. Acoust. Soc. Am. **100**, 1901–1904.

Hafter, E. R., Bourbon, W. T., Blocker, A. S., and Tucker, A. (**1969**). "A direct comparison between lateralization and detection under conditions of antiphasic masking," J. Acoust. Soc. Am. **46**, 1452–1457.

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

Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer, New York), pp. 44–45.

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

Hartmann, W. M., and Zhang, P. X. (**2003**). "Binaural models and the strength of dichotic pitches," J. Acoust. Soc. Am. **114**, 3317–3326.

Hartmann, W. M., Zhang, P. X., and Culling, J. F. (**2004**). "Earedness: left-eared and right-eared listeners," J. Acoust. Soc. Am. **115**, 2534.

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

Marks, L. (**1974**). *The New Psychophysics* (Academic, New York).

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

Raatgever, J., and Bilsen, F. A. (**1977**). "Lateralization and dichotic pitch as a result of spectral pattern recognition," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 443–453.

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.

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.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.

Yost, W. A. (**1991**). "Thresholds for segregating narrow-band noise from broadband noise based on interaural phase and level differences," J. Acoust. Soc. Am. **89**, 838–844.

Zhang, P. X. (**2006**). "Detection and localization of virtual tones and virtual images," Ph.D. thesis, Michigan State University, East Lansing, MI.

Zhang, P. X., and Hartmann, W. M. (**2006**). "Lateralization of sine tones—Interaural time vs phase," J. Acoust. Soc. Am. **120**, 3471–3474.

Zhang, P. X., and Hartmann, W. M. (**2007**). "Salient phase density model for the lateralization of binaural pitches," J. Acoust. Soc. Am. **122**, 2968.

See EPAPS Document No. E-JASMAN-124–008811 for Figures 1 through 26 on scatter plots showing lateralization of Huggins pitch—supplement. For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.