Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models

Matthew J. Goupell[a] and William M. Hartmann
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 17 November 2005; revised 26 March 2007; accepted 30 March 2007)

In the first two articles of this series, reproducible noises with a fixed value of interaural coherence (0.992) were used to study the human ability to detect interaural incoherence. It was found that incoherence detection is strongly correlated with fluctuations in interaural differences, especially for narrow noise bandwidths, but it remained unclear what function of the fluctuations best agrees with detection data. In the present article, ten different binaural models were tested against detection data for 14- and 108-Hz bandwidths. These models included different types of binaural processing: independent-interaural-phase-difference/interaural-level-difference, lateral-position, and short-term cross-correlation. Several preprocessing transformations of the interaural differences were incorporated: compression of binaural cues, temporal averaging, and envelope weighting. For the 14-Hz bandwidth data, the most successful model postulated that incoherence is detected via fluctuations of interaural phase and interaural level processed by independent centers. That model correlated with detectability at $r=0.87$. That model proved to be more successful than short-term cross-correlation models incorporating standard physiologically-based model features ($r=0.78$). For the 108-Hz bandwidth data, detection performance varied much less among different waveforms, and the data were less able to distinguish between models. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2734489]

PACS number(s): 43.66.Ba, 43.66.Pn, 43.66.Qp [AK] Pages: 1029–1045

I. INTRODUCTION

In a previous article, to be called "Article I," Goupell and Hartmann (2006) studied the ability of listeners to detect a small amount of interaural incoherence. The experiments employed selected noises in which the interaural coherence was fixed at a value of 0.992, where the interaural coherence was defined as the maximum of the cross-correlation function, as computed over the entire duration (500 ms) of the stimulus. Physical analysis of the noises studied the fluctuations in interaural phase difference (IPD) and interaural level difference (ILD). It was found that these fluctuations were increasingly variable across different noises for decreasing bandwidth.

In the psychoacoustical experiments of Article I, the listener's task was to distinguish between the incoherent noises (coherence=0.992) and diotic noises with a coherence of 1.0. In spite of the fact that the incoherent noises all had the same coherence, the experiments showed that for narrow bandwidths the incoherence was much more readily detectable in some noises than in others. Listeners found it significantly easier to detect incoherence when the fluctuations in IPD or ILD were larger. As the bandwidth increased, the incoherence became equally detectable in all the different noises, consistent with a model in which detection is predictable from interaural coherence alone.

The stimuli for the experiments of Article I were selected based on large or small fluctuations in interaural phase or level. For any given noise, the fluctuations were measured by the standard deviations over time in IPD or ILD. The corresponding variations in detectability, especially when the bandwidth was as narrow as 14 Hz, indicated that these fluctuation measures have considerable perceptual validity. However, it is possible, even likely, that some other measure of stimulus fluctuation would correlate better with human perception of incoherence.

It is also not clear that IPD and ILD fluctuations, as used in Article I, should be considered as comparably important. The experiments showed that variability in IPD and variability in ILD led to similar variability in detectabilty, but phase and level fluctuations are so strongly correlated within an ensemble of noises that comparable data do not clearly demonstrate comparable importance. For instance, it is possible that listeners only responded to interaural phase fluctuations. Experiments using controlled level fluctuations would then lead to significant effects only because the level fluctuations are so strongly correlated with phase fluctuations.

The purpose of the present article is to address the uncertainties left by Article I by testing a variety of different binaural detection models against incoherence detection data. The binaural detection models were derived from models previously used to study the masking-level difference (MLD). This was a sensible approach because the MLD is closely related to incoherence detection (Durlach *et al.*, 1986; Bernstein and Trahiotis, 1992). Although the set of models tested affords a notable variety, it must be acknowledged in advance that the set is not exhaustive.

[a]Electronic mail: goupell@kfs.oeaw.ac.at

Domnitz and Colburn (1976) summarized the two major types of binaural models historically presented to explain the MLD phenomenon. The first type uses interaural parameter differences, IPD and ILD. For example, the vector model (Jeffress *et al.*, 1956) predicts the largest release from masking for a signal phase difference of 180° (NoSπ). Another example is the lateralization model (Hafter, 1971), in which a signal is detected by a shifted lateral image that is formed by combining IPD and ILD. The second type of binaural model includes energy and cross-correlation models. When an out-of-phase tone is added to homophasic noise, the interaural correlation of the entire stimulus is reduced. Models such as the equalization-cancellation (EC) model by Durlach (1963) and the correlation model of Osman (1971) fall into this category.

There is still debate as to which type of model best describes binaural detection. Gilkey *et al.* (1985) found that wideband reproducible-noise masking data were incompatible with several interaural parameter models. On the other hand, Colburn *et al.* (1997) showed that the EC model was incompatible with reproducible noise data from Isabelle and Colburn (1991). Several recent articles have favored EC-like models to describe binaural detection data (Breebaart *et al.*, 1999; Breebaart and Kohlrausch, 2001; Breebaart *et al.*, 2001a, b, c). However, not all of the data can be described by EC-like models. Breebaart and Kohlrausch (2001) found that correlation and energy models cannot entirely describe thresholds for stimuli with a fixed-interaural correlation. Using narrowband multiplied noise, Breebaart *et al.* (1999) found that neither interaural difference parameters nor the EC model could account for the results of experiments that included static level differences.

Article I showed that incoherence detection cannot be understood in terms of coherence derived from the cross-correlation of the stimulus as computed over a long duration. A second article, Article II (Goupell and Hartmann, 2007), found the same negative result for coherence computed over short-duration stimuli. Therefore, the present article focuses on interaural parameter models and on cross-correlation models that include a physiologically-based preprocessor.

II. EXPERIMENT 1: 14-Hz BANDWIDTH

The purpose of Experiment 1 was to obtain incoherence detection data from a large set of narrowband noises that were randomly generated and unselected so as to be a fair representation of all noises with a given bandwidth, duration, and interaural coherence. The detection data were collected in order to test the models presented in this article.

A. Stimuli

A collection of 100 dual-channel noises with 14-Hz bandwidth was created for Experiment 1. It was the same collection from which particular noises were selected in Article I. In the present experiment all 100 noises were used to avoid any bias.

Each noise was constructed from equal-amplitude random-phase components that spanned a frequency range of 490–510 Hz with a frequency spacing of 2 Hz. Components between 495 and 505 Hz had equal amplitudes of unity. Frequencies below 495 and above 505 Hz were attenuated with a raised-cosine spectral window. The 3-dB bandwidth was 14 Hz. An orthogonalization procedure guaranteed that the interaural coherence of each noise was precisely 0.992.

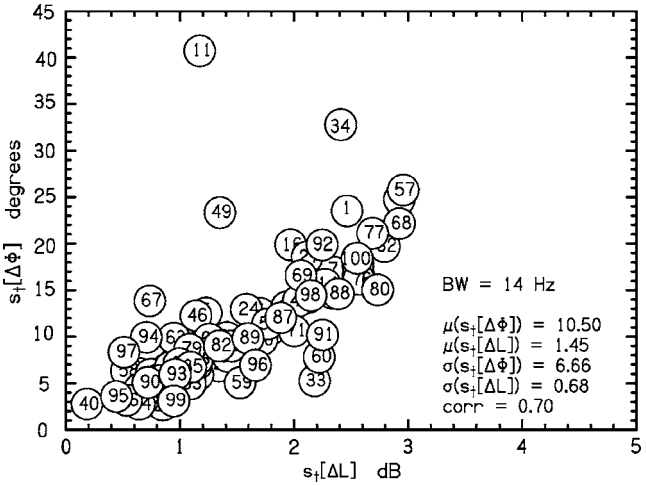

FIG. 1. Fluctuations of IPD vs fluctuations of ILD for the collection of 100 reproducible noises having a 14-Hz bandwidth used in Experiment 1. Each noise is labeled by a serial number indicating only the order of creation. The means, standard deviations, and IPD-ILD correlation of the distributions are reported.

As in Articles I and II, our attention focused on the interaural phase difference,

$$\Delta\Phi(t) = \phi_R(t) - \phi_L(t), \tag{1}$$

and the interaural level difference,

$$\Delta L(t) = 20 \log_{10}\left[\frac{E_R(t)}{E_L(t)}\right], \tag{2}$$

where ϕ is the phase and E is the envelope calculated from the analytic signal. Fluctuations in these interaural differences were initially defined in terms of their standard deviations over time, computed over the duration of the stimulus T, and indicated by the functions

$$s_t[\Delta\Phi] = \sqrt{\frac{1}{T}\int_0^T [\Delta\Phi(t) - \overline{\Delta\Phi}]^2 dt} \tag{3}$$

and

$$s_t[\Delta L] = \sqrt{\frac{1}{T}\int_0^T [\Delta L(t) - \overline{\Delta L}]^2 dt}. \tag{4}$$

The fluctuations $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ were calculated for each noise, and average quantities, indicated by an overbar, refer to a time-averaged interaural difference for the noise—normally very close to zero. These fluctuations are shown in Fig. 1 for the 100 noises, labeled by serial number (order of creation). Figure 1 also indicates the mean, standard deviation, and correlation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, computed over the ensemble of 100 noises.

As in Article I, noise stimuli were presented to the listeners in three observation intervals, each with a total duration of 500 ms and with 30-ms Hanning windows for attack

and decay. Noises were computed by a Tucker-Davis AP2 array processor (System II) and converted to analog form by 16-bit DACs (DD1). The buffer size was 4000 samples per channel and the sample rate was 8000 samples per second (sps). The noise was low-pass filtered with a corner frequency of 4 kHz and a −115-dB/octave rolloff. The noises were presented at 70±3 dB SPL with levels determined by programmable attenuators (PA4) operating in parallel on the two channels prior to the low-pass filtering. The level was randomly chosen in 1-dB steps for each of the three intervals within a trial to discourage the listener from trying to use level cues to perform the task.

B. Procedure

Listeners were seated in a double-wall sound-attenuating room and used Sennheiser HD414 headphones. The 100 noises were presented in sets of ten as ordered by serial number. Thus, the first set had noises 1–10, the second set had noises 11–20, and so on. Six runs were devoted to listening to a set of ten reproducible noises. Listeners completed each set before moving on to the next set.

The structure of runs, trials within a run, and the data collection procedure were the same as in Articles I and II. It is briefly described as follows: A noise could be presented either incoherently (the dichotic presentation of x_L and x_R) or it could be presented coherently (the diotic presentation of x_L). A run consisted of 60 trials, where each of the ten reproducible noises in a set was presented incoherently a total of six times. Thus, a listener heard an individual noise incoherently a total of 36 times (six runs times six presentations per run).

On each trial the listener heard a three-interval sequence. The first interval was the standard interval, which was always a coherent noise. The second interval was randomly chosen to be either incoherent or coherent. The third interval was the opposite of the second (e.g., if the second interval was coherent, the third interval was incoherent). The two coherent presentations were randomly selected from the remaining nine reproducible noises in the set except that they were required to be different from the x_L and x_R in the incoherent "odd" interval and to be different from one another. The interinterval duration was 150 ms. The listener was required to decide which of the two latter intervals was the incoherent interval. As described in Articles I and II, listeners were allowed to indicate that they were confident about a response, leading to a confidence adjusted score (CAS) on a scale of 0 to 72 for a run of 36 trials. Although the data collection procedure also kept track of the percentage of correct responses (P_c), it was found that the CAS provided greater "dynamic range" by preventing most of the ceiling effects for the most successful listeners. Confidence ratings and multipoint decision scales have been shown by signal detection theory to be valid psychophysical techniques (Egan *et al.*, 1959; Schulman and Mitchell, 1966).

C. Listeners

Experiments in this article employed three male listeners from Article I—D, M, and W. Listeners D and M were between the ages of 20 and 30 and had normal hearing according to standard audiometric tests and histories. Listener W was 65 and had a mild bilateral hearing loss, but only at frequencies four octaves above those used in the experiment. Listeners M and W were the authors.

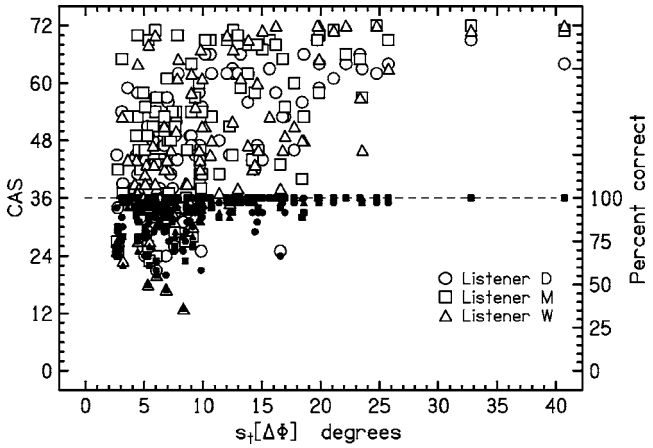

FIG. 2. All the detection data for the 100 noises from Experiment 1 for three listeners, D, M, and W, are plotted twice, once as the number correct—on a scale from 0 to 36 (0 to 100%), and once as CAS—on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase in an attempt to give some order to the plot.

D. Results

The results from experiments using all 100 noises can be seen in Fig. 2. The open symbols show the CAS while the closed symbols show the number of correct responses, essentially equivalent to the P_c. These values are plotted as a function of $s_t[\Delta\Phi]$ only to give some order to the plot, not because $s_t[\Delta\Phi]$ is thought to be the best model for detection. Figure 2 illustrates the advantage of using the CAS over P_c because the number of correct responses reaches a ceiling, especially for listener M. The CAS increases the dynamic range of the experiment, though it has not completely removed ceiling effects.

Agreement between the listeners for individual noises is difficult to see in Fig. 2, but agreement is actually good. The interlistener Pearson correlation was 0.73 for D and M, 0.71 for D and W, and 0.80 for M and W. These interlistener correlations are smaller than those reported for the ten noises in Article I—approximately 0.9 on average. The reason for the difference is probably that members of the entire collection of 100 noises are less distinctive than are the five largest and five smallest fluctuation noises used in Article I.

III. MODELS FOR INCOHERENCE DETECTION

In order to discover the stimulus features that best predict human perception of incoherence, models of perception were constructed using transformed interaural parameters, as described below, and the models were tested against the large set of perceptual data from Experiment 1.

A. Model preprocessing assumptions

Several assumptions, common to all models, were made to reflect auditory preprocessing of the complex incoherent

stimuli. Two free parameters are introduced in the following as well as a scale of lateralization for the $\Delta\Phi(t)$ and $\Delta L(t)$.

1. Temporal averaging

The fluctuation measures used to construct stimulus sets for Article I were based on instantaneous values of interaural differences as they appeared with our 8 ksps sample rate. But it is not evident that, for example, a large interaural difference with a duration of only 0.125 ms would receive much respect from the binaural system. Therefore, the present models include a parametric temporal averaging operation, following other models, e.g., Viemeister (1979), in using an exponential averaging window to represent temporal modulation transfer functions of the form

$$\Delta\Phi'(t) = \hat{e}[\Delta\Phi(t)] = \frac{\int_0^{T_D} \Delta\Phi(t-t')e^{-t'/\tau}dt'}{\int_0^{T_D} e^{-t''/\tau}dt''} \quad (0 < t < T), \tag{5}$$

and

$$\Delta L'(t) = \hat{e}[\Delta L(t)] = \frac{\int_0^{T_D} \Delta L(t-t')e^{-t'/\tau}dt'}{\int_0^{T_D} e^{-t''/\tau}dt''} \quad (0 < t < T). \tag{6}$$

Parameter T is the duration of the stimulus, and the time constant τ was a free parameter. The averaging window, with running variable t', was terminated when t' became greater than t or when the weight of the exponential function dropped to 0.1, which determined the upper limit of the integration T_D.

2. Compression of binaural cues

A small static interaural difference leads to a small displacement in the lateral position of the auditory image from a centered position. A greater interaural difference leads to a greater displacement, but increasing interaural differences produce diminishing returns because the laterality is a compressive function of interaural differences. A perceptual model for fluctuations can easily adopt this effect from static experiments. The compression functions, to be called "laterality compression," used in the present analysis were exponential fits to the data from Yost's 1981 experiments. They are of the form

$$\Psi'_{\Delta\Phi}(t) = 10\ \mathrm{sgn}[\Delta\Phi'(t)](1 - e^{-|\Delta\Phi'(t)|/40}), \tag{7}$$

and

$$\Psi'_{\Delta L}(t) = 10\ \mathrm{sgn}[\Delta L'(t)](1 - e^{-|\Delta L'(t)|/8}), \tag{8}$$

where $\Psi'_{\Delta\Phi}(t)$ and $\Psi'_{\Delta L}(t)$ are on a scale of lateral position that ranges from −10 to 10. In Eq. (7), the weighting constant of the exponential is 40°. In Eq. (8), the weighting constant of the exponential is 8 dB. These functions correspond to the experimentally determined lateral position of a sine tone at a frequency of 500 Hz, the center frequency of our noise bands. A further benefit of the compressive laterality transformation is that IPD and ILD are put on the same scale so that they can be easily combined in mathematical models.

3. Critical envelope value weighting

Maxima can occur in the IPD, $\Delta\Phi(t)$, at times when the envelope in one ear is very small. But if the envelope is near zero, the listener may not be able to detect this fluctuation in $\Delta\Phi(t)$. Therefore, it would be wrong for a model to give much weight to this phase fluctuation. We sought to reduce the problem by discounting phase fluctuations that coincided with very small envelope values[1] by employing a weighting function,

$$w_g(t) = \begin{cases} 1 & \text{if } E_L(t) \text{ and } E_R(t) \geqslant gE_{\mathrm{rms}} \\ 0 & \text{if } E_L(t) \text{ or } E_R(t) < gE_{\mathrm{rms}}, \end{cases} \tag{9}$$

where E_L and E_R are Hilbert envelopes for left and right channels, and E_{rms} actually indicates a comparison with corresponding left- or right-channel overall rms values. Parameter g is the critical envelope fraction, a free parameter. If the envelope in either channel is less than g times the rms envelope, then the weight is set to zero. Otherwise, the weight is set to one.

After all the modeling assumptions, the *transformed* IPD and ILD are described by the notation

$$\Psi_{\Delta\Phi}(t) = \Psi'_{\Delta\Phi}(t)w_g(t) \tag{10}$$

and

$$\Psi_{\Delta L}(t) = \Psi'_{\Delta L}(t). \tag{11}$$

Because the allowed values of the preprocessing parameters τ (exponential averaging) and g (envelope weighting) include the entire physical range, the transformed interaural differences admit the possibility of no transformation. The exception is in the laterality compression, which was always applied to models 1–7.

B. Models for binaural combination

Ten different binaural combination models with adjustable parameters were studied. Each model produced a decision statistic intended to predict the detectability of incoherence. The models and their parameters were then independent variables in regressions comparing predictions with listener detection performance.

The models tested three different hypotheses concerning binaural combination: (1) the independent-interaural-difference or independent-centers model, (2) the lateral-position or lateralization model, and (3) the short-term cross-correlation model. In models of the independent-difference type, averaged fluctuations in IPD and in ILD are combined with a relative weighting parameter a. In models of the lateral-position type, an image location is calculated based on IPD and ILD values that are combined with a time/intensity trading parameter b. The decision statistic is then based on fluctuations in that location. In the short-term-cross-correlation models, only the IPD is used, as will be shown later in this section.

The models are based on transformed (i.e., preprocessed) values of IPD and ILD combined in different ways. It should be noticed that there is no important distinction between the IPD and the interaural time difference (ITD) in this work. With bandwidths as narrow as ours, the ITD can

be determined from the IPD by dividing by the band center frequency of 500 Hz. Consequently, although the models are expressed in terms of IPD, they could equally well be expressed in terms of ITD with no important changes.

Model 1: Sum of interaural differences. A simple model of the independent-interaural-difference type hypothesizes that incoherence is detected on the basis of a linear combination of the standard deviation in transformed IPD and the standard deviation in transformed ILD. The standard deviation of a transformed interaural difference is

$$s_t[\Psi] = \sqrt{\frac{1}{T}\int_0^T [\Psi(t) - \overline{\Psi}]^2 dt}, \tag{12}$$

where $\Psi(t)$ is either the transformed IPD or ILD, and the integral spans the entire stimulus of duration T. Therefore, the sum of transformed standard deviations of IPD and ILD is

$$d_1 = a s_t[\Psi_{\Delta\Phi}] + (1-a) s_t[\Psi_{\Delta L}]. \tag{13}$$

This model has three free parameters: a, τ, and g. The *nontransformed* fluctuations in IPD and ILD were, in fact, the basis for choosing stimuli in Articles I and II. There, it was found that larger values of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ correlated with a greater detectability of incoherence in noises for a given value of coherence.

Model 2: Sum of mean square variations. As a close relative to the decision statistic d_1, an independent-differences model could use the square of the fluctuation, as introduced by Isabelle and Colburn (1987) in connection with a masking level difference experiment with reproducible stimuli,

$$d_2 = a s_t^2[\Psi_{\Delta\Phi}] + (1-a) s_t^2[\Psi_{\Delta L}]. \tag{14}$$

This model has the same free parameters as model 1. This model "…intended to capture the subjective increase in image width caused by the addition of a target tone to the narrowband masker…" (Isabelle and Colburn, 2004).

Model 3: Sum of integrations. An alternative decision statistic is based on an integration of the absolute value of the IPD and ILD over the duration of the stimulus. In this model the contributions of the IPD and ILD are computed separately,

$$d_3 = a\frac{1}{T}\int_0^T |\Psi_{\Delta\Phi}(t)|dt + (1-a)\frac{1}{T}\int_0^T |\Psi_{\Delta L}(t)|dt. \tag{15}$$

This model has the same free parameters as model 1. We do not know of any precedent for such an independent-integration model in the literature.

Model 4: Sum of threshold deviations. A fourth kind of decision statistic measures the fraction of the time that interaural differences are far from zero (the center position). This thresholded statistic is defined as

$$d_4 = a\frac{1}{T}\int_0^T W[h,\Psi_{\Delta\Phi}(t)]dt + (1-a)\frac{1}{T}\int_0^T W[h,\Psi_{\Delta L}(t)]dt, \tag{16}$$

where

$$W[h,\Psi(t)] = \begin{cases} 1 & \text{if } \Psi(t) \geqslant h \\ 0 & \text{if } \Psi(t) < h. \end{cases} \tag{17}$$

In addition to the same three free parameters of other models, model 4 has a fourth free parameter, h, to set the level of threshold. Since both transformed interaural parameters are on the same scale of lateral position, it was assumed that the threshold is the same for both interaural differences.

Webster (1951) proposed a similar model that used only deviations of IPD to determine the influence of interaural phase on masking thresholds. Our model permits large deviations in either IPD or ILD to be the basis for incoherence detection. Model 4 reduces to Webster's model for $a=1$.

Model 5: Standard deviation of the lateral position. Model 5 comes from a suggestion by Hafter (1971) that a signal might be detected by a shift in the lateral position of an image formed by combining IPD and ILD with a time-intensity trading ratio. Model 5 is the first model of three in this article based on a time-varying lateral position, and it hypothesizes that the standard deviation of fluctuations in the lateral position describes incoherence detection. The key distinction is that in a lateral-position model, a fluctuation in phase can cancel a fluctuation in level, but such cancellation is not possible in an independent-centers model such as models 1–4. The lateral position itself can be defined as

$$\Psi_z(t) = b\Psi_{\Delta\Phi}(t) + (1-b)\Psi_{\Delta L}(t), \tag{18}$$

where b is a dimensionless time-intensity trading parameter for transformed interaural differences. The overall time-intensity trading ratio is a combination of b and the laterality-compression factors in Eqs. (7) and (8). Then the standard deviation of the lateral position becomes

$$d_5 = s_t[\Psi_z(t)] = s_t[b\Psi_{\Delta\Phi}(t) + (1-b)\Psi_{\Delta L}(t)], \tag{19}$$

where there are three free parameters: b, τ, and g.

Model 6: Integration of the lateral position. The particular model that Hafter proposed in 1971 was actually a model based on the integrated absolute value of lateral-position incorporating time-intensity trading. Converted to use transformed variables, the model gives

$$d_6 = \frac{1}{T}\int_0^T |\Psi_z(t)|dt = \frac{1}{T}\int_0^T |b\Psi_{\Delta\Phi}(t) + (1-b)\Psi_{\Delta L}(t)|dt. \tag{20}$$

Here, the instantaneous lateral position corresponds to the fluctuation because it is assumed that the undisplaced position corresponds to $z=0$. This model has the same free parameters as model 5.

Model 7: Threshold deviation of the lateral position. Deviations that exceed a threshold value constitute events, and the durations of these events are summed in a decision statistic given by

$$d_7 = \frac{1}{T}\int_0^T W[h, \Psi_z(t)]dt. \tag{21}$$

As for model 4, W has the value 1 if Ψ_z is greater than h and is zero otherwise. In addition to the three free parameters of the other lateral-position models (models 5 and 6), model 7 has a fourth free parameter, h, to set the level of threshold.

Model 8: rms deviation of the short-term cross-correlation function. In connection with the MLD, Osman (1971) proposed a model based on the interaural cross-correlation computed over the entire observation interval. An alternative computes the cross-correlation as a function of running time t,

$$\gamma(t) = \frac{\int_{t-\Delta t}^{t} x_L(t')x_R(t')dt'}{\sqrt{\int_{t-\Delta t}^{t} x_L^2(t_1)dt_1 \int_{t-\Delta t}^{t} x_R^2(t_2)dt_2}}, \tag{22}$$

where $x_L(t')$ is the left-channel waveform and $x_R(t')$ is the right-channel waveform. This cross-correlation function is evaluated at zero lag because an incoherence detection experiment includes no offset ITD. The integration window Δt is brief. For instance, Isabelle and Colburn (2004) take it to be the inverse of the center frequency of the noise band.

Rewritten in terms of the Hilbert envelope and phase, the running cross-correlation is

$$\gamma(t) = \frac{\int_{t-\Delta t}^{t} E_L(t')E_R(t')\cos[\omega t' + \Phi_L(t')]\cos[\omega t' + \Phi_R(t')]dt'}{\sqrt{\int_{t-\Delta t}^{t} |E_L(t_1)|^2\cos^2[\omega t_1 + \Phi_L(t_1)]dt_1 \int_{t-\Delta t}^{t} |E_R(t_2)|^2\cos^2[\omega t_2 + \Phi_R(t_2)]dt_2}}. \tag{23}$$

Isabelle and Colburn (2004) showed that $\gamma(t)$ is approximately given by the cosine of the instantaneous interaural phase difference when the bandwidth is small, as for Experiment 1. For the bandwidth of 14 Hz, the Hilbert envelope and phase should vary on the time scale of $1/14 = 74$ ms. For the center frequency of $f_c = 500$ Hz, this time scale is much slower than the period, $\Delta t = 2$ ms. It then can be assumed that E_L, E_R, Φ_L, and Φ_R are approximately constant over the integration intervals in Eq. (23). Over one period of the stimulus the denominator reduces to the product $E_L E_R$, which then cancels the envelope factors in the numerator. Therefore the short-term cross-correlation (STCC) function can be approximated as

$$\gamma(t) \approx \cos \Delta\Phi(t). \tag{24}$$

The deviation from the diotic value is $1 - \gamma(t)$ and the transformed deviation is

$$\Psi_{CC}(t) = \hat{e}\{1 - \cos[\Delta\Phi(t)]\}w_g(t). \tag{25}$$

The transformed deviation in Eq. (25) includes exponential temporal averaging, which potentially reduces the effectiveness of brief lateral excursions, and it incorporates critical envelope weighting wherein a deviation from perfect correlation is not noticed if an envelope becomes too small.

The root-mean square of the transformed deviation then forms a decision statistic

$$d_8 = \left[\frac{1}{T}\int_0^T \Psi_{CC}^2(t)\, dt\right]^{1/2}. \tag{26}$$

Like the other models, the short-term cross-correlation incorporates temporal averaging and envelope weighting. Unlike the other models, it does not include laterality compression so that the interaural phase remains in units of radians. Because d_8 does not include any form of ILD, it does not include IPD-ILD weighting, and it has only two free parameters, τ and g. Unlike models 1 and 5, which compute a standard deviation, the decision statistic d_8 was computed as a deviation of Ψ_{CC} from zero to represent the deviation from a diotic noise.

Model 9: Integration of the short-term cross-correlation function. Just as models 3 and 6 integrated the absolute value of model percepts, model 9 integrates the absolute deviation,

$$d_9 = \frac{1}{T}\int_0^T \Psi_{CC}(t)\, dt. \tag{27}$$

By definition, $\Psi_{CC}(t) \geqslant 0$. Model 9 has the same free parameters as model 8, τ and g. Also, like model 8, laterality compression was not included in the transformed variable so that $\Delta\Phi(t)$ is in radians.

Model 10: Threshold of the short-term cross-correlation function. Just as models 4 and 7 were based on thresholded values of model percepts, model 10 integrates a thresholded short-term cross-correlation,

$$d_{10} = \frac{1}{T}\int_0^T W[h, \Psi_{CC}(t)]dt. \tag{28}$$

Model 10 has three parameters, τ, g, and h, where h sets the *magnitude* of the threshold deviation. Per Eq. (25) the magnitude of the deviation, $\Psi_{CC}(t)$, can be as large as 2—the difference between $\cos(0)$ and $\cos(\pi)$. As in the other STCC models, the laterality compression was omitted.

C. Models compared with Experiment 1

The ten models presented above were tested against the data from Experiment 1. A linear regression of the form $y = mx + b$ was used to evaluate the effectiveness of a model to describe incoherence detection. The y variable was the CAS for the individual listeners or for an average over listeners. The x variable was d_n from one of the ten models. Figure 3 shows example regressions for model 1. The solid line is the line of best fit. The dotted lines have the same slope as the

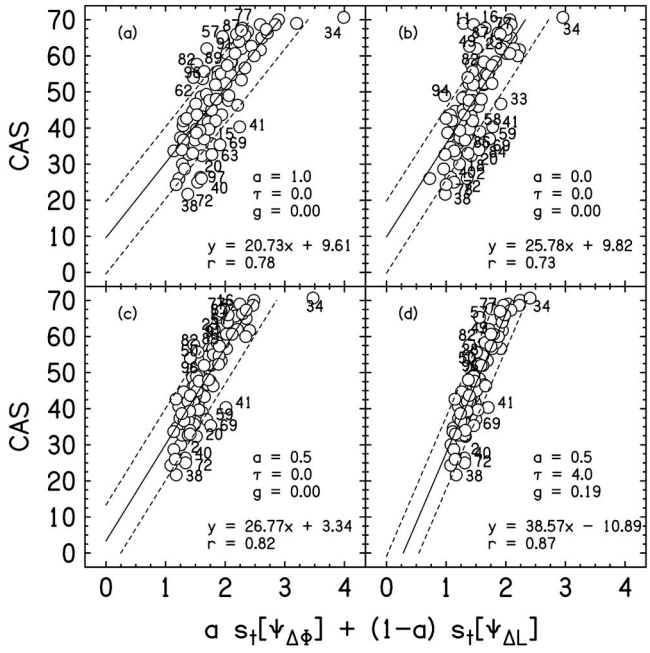

FIG. 3. Example linear regressions for model 1. Laterality compression is applied to all the noises. The solid line is the line of best fit; the equation and value of τ are reported. The dotted lines have the same slope as the solid line, but have intercepts that differ by ±10 CAS. Noises that fall outside the dotted lines are numbered by the serial number in Fig. 1. The plots show the advantages of (c) using both IPD and ILD, and (d) envelope weighting and temporal averaging.

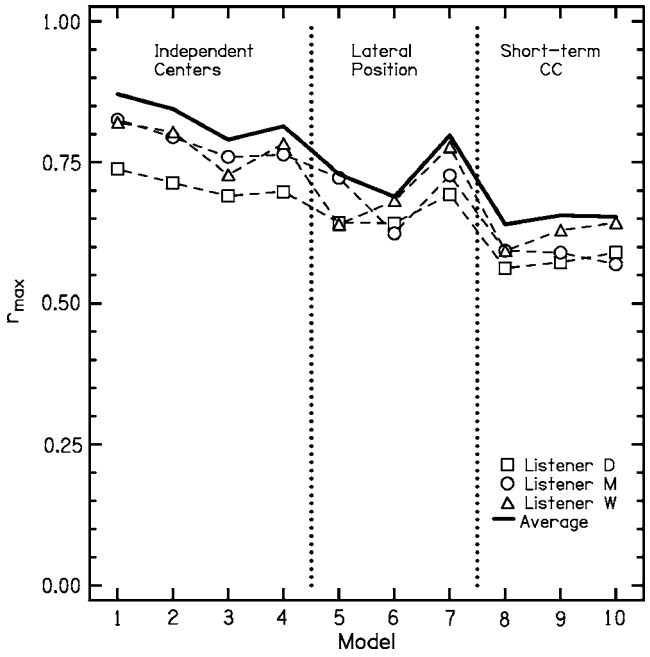

FIG. 4. The comparison of CAS scores for the 14-Hz noises of Experiment 1 with ten models. The value of r_{max} shows the correlation between the experimental CAS scores for the 100 noises and predictions by each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. It is not the average of r_{max} averaged over the listeners.

solid line but are displaced vertically by plus or minus 10 CAS units. Noises falling outside of the dotted lines region are numbered on the plot. Figure 3(a) shows model 1 using only the laterality-compressed IPD ($a=1$) without exponential averaging and threshold weighting. Figure 3(b) shows model 1 using only the laterality-compressed ILD ($a=0$). Figure 1 shows that noise 57 has large fluctuations in IPD and ILD but experiments, including those of Article I, showed that listeners found it relatively difficult to detect the incoherence in this noise. By contrast, Fig. 3(a) shows stimulus 57 to the left of the line of best fit. This shows that when the laterality compression is included in the model, the fluctuations are actually comparatively small, which is more in line with the detection data. Figures 3(c) and 3(d) show equal weighting of IPD and ILD ($a=0.5$), respectively, without and with temporal averaging and envelope weighting.

The linear correlation coefficient, r, was used to compare the results of the regressions. The maximum r, r_{max}, was found by independently varying all the free parameters over a reasonable space. For example, model 1 has three free parameters a, τ, and g. The range of a was 0 to 1 with a 0.01 increment; the range of τ was 0 to 10 ms with a 0.5-ms increment (tests with larger values of τ will be described later); the range of g was 0 to 0.5 with an increment of 0.01. Therefore, for model 1, 400 000 linear regressions were performed ($100\times20\times50\times4$ listeners). For the threshold models 4 and 7, the range of h was 0 to 10 with a 0.25 increment. (Recall that the laterality-compressed IPD and ILD are on a scale of −10 to 10.) For threshold model 10, the range of h was 0 to 2 with a 0.01 increment. A power law regression equation was also used to fit the data, but it did not improve the correlation between the experimental data and the models.

1. Comparison of model types

The results of the regressions are shown in Fig. 4 for the best combination (largest r) of all the parameters for each model. Figure 4 presents results for individual listeners and for the average listener. Therefore, the r_{max} of the averaged data is not the average r_{max} of the listeners. It is interesting that the most successful models agree better with the average listener than they do with any single listener. A similar result was found by Isabelle and Colburn (2004) in modeling MLDs for reproducible noise. Given that the models are simple signal processing algorithms whereas human listeners have complicated individual tendencies, that is the sort of result that one would expect from a model that correctly represents the general population.

Figure 4 shows that the models of the independent-interaural-difference type (models 1–4) were more successful than the lateral-position types (models 5–7) with the exception that model 7 had a larger r_{max} than model 3. Least successful were the STCC models (models 8–10). Since the STCC depends entirely on $\Delta\Phi(t)$, this may be evidence for an important contribution of $\Delta L(t)$ to incoherence detection.

Model 1 had the largest r_{max} for all three listeners and for the averaged data. For the averaged data, $r_{max}=0.87$. The performance of model 2 is very similar to that of model 1, except that the r_{max} is always slightly smaller for model 2.

Table I shows the values of the free parameters that maximized r for the 14-Hz bandwidth modeling. Table I

TABLE I. Values of free parameters that optimize r in modeling the detection results of Experiment 1 with 100 noises with a bandwidth of 14 Hz. Parameter τ is the exponential window time constant. Parameters a and b weight IPD and ILD contributions, with extremes of 1 and 0 equivalent to IPD only and ILD only, respectively. Parameter g is the envelope threshold for discounting IPD. The IPD is ignored if the envelope in either the left or right channel is less than g times the overall rms value. Parameter h is a threshold for models that measure the duration of time the function is greater than the threshold value. Threshold is lateral position for models 4 and 7; it is deviation from perfect coherence for model 10.

Model	Listener	τ (ms)	a,b	g	h
d_1	D	3.0	0.53	0.17	⋯
	M	5.0	0.45	0.21	⋯
	W	0.5	0.53	0.23	⋯
	Ave	4.0	0.50	0.19	⋯
d_2	D	3.0	0.52	0.17	⋯
	M	4.0	0.46	0.23	⋯
	W	0.5	0.54	0.23	⋯
	Ave	3.0	0.50	0.23	⋯
d_3	D	1.0	0.62	0.04	⋯
	M	4.5	0.44	0.15	⋯
	W	6.5	0.51	0.12	⋯
	Ave	3.5	0.50	0.12	⋯
d_4	D	2.0	0.78	0.04	4.00
	M	0.0	0.48	0.11	3.75
	W	0.5	0.59	0.12	3.25
	Ave	0.5	0.66	0.04	3.75
d_5	D	1.0	0.00	0.07	⋯
	M	3.5	0.00	0.12	⋯
	W	3.5	0.00	0.12	⋯
	Ave	3.0	0.00	0.11	⋯
d_6	D	1.0	0.97	0.04	⋯
	M	2.0	0.11	0.15	⋯
	W	0.0	0.42	0.15	⋯
	Ave	0.5	0.91	0.04	⋯
d_7	D	2.0	0.99	0.04	4.00
	M	1.0	0.96	0.04	3.75
	W	0.0	0.54	0.24	2.50
	Ave	1.0	0.88	0.04	3.75
d_8	D	0.0	⋯	0.28	⋯
	M	0.0	⋯	0.28	⋯
	W	0.0	⋯	0.27	⋯
	Ave	0.0	⋯	0.28	⋯
d_9	D	6.0	⋯	0.11	⋯
	M	6.0	⋯	0.11	⋯
	W	6.0	⋯	0.27	⋯
	Ave	6.0	⋯	0.11	⋯
d_{10}	D	7.5	⋯	0.00	0.06
	M	0.0	⋯	0.00	0.06
	W	6.5	⋯	0.00	0.08
	Ave	6.0	⋯	0.00	0.06

shows that the parameters are similar for different listeners over the different types of models (independent-centers, lateral-position, and STCC). Consequently, the fits to the average listener shown in Fig. 4 are meaningful. Table I also shows that fitting parameters that optimize r are similar across models, to the extent that the models permit them to be compared.

2. Optimized parameters for model 1

The most successful model was model 1, and Fig. 5 shows how the free parameters covary to maximize r for the average listener in that model. Figure 5 was generated by varying two parameters and keeping the other constant at the optimum value. Plots of a vs τ assume that $g=0.19$, plots of g vs τ assume that $a=0.50$, and plots of a vs g assume that $\tau=4$.

The results of Fig. 5 can be summarized as follows:

a. Integration time. The greatest r occurs for an integration time of $\tau=4$ ms. However, r was quite insensitive to τ over the 0–10 ms range tested in detail. Apparently, the characteristic stimulus fluctuations, expected to be of order 1/14 s, are slow enough that they do not challenge a system with time constants of this order. Longer integration times will be discussed later.

b. Critical envelope weighting. According to the regression analysis, the best critical envelope weight for model 1 is $g=0.19$, though the r_{max} is insensitive to g in this vicinity (approximately $0<g<0.3$). This result indicates that there is a modest benefit on the average of ignoring phase differences that coincide with a particularly small envelope in one or both of the ears.

A greater benefit from envelope weighting is seen when one tries to predict detection for individual waveforms. The weighting omits large phase fluctuations that occur during the onset and offset of the stimulus, where the temporal shaping is applied and the envelope is small. One would expect that even large phase fluctuations during these times would often be missed by the listener because they occur at the very beginning or end of the stimulus. This benefit can also be seen in the lower panels of Figure 3. In Fig. 3(c) (without envelope weighting) there were nine stimuli to the right of the rightmost dotted line. In Fig. 3(d) (with envelope weighting), two stimuli moved within the dotted line region and the rest of these points moved close to the dotted line.

The envelope weighting applied in Eq. (9) is a simple on/off type. Other envelope weighting functions were tried—linear envelope weighting of the IPD and squared envelope weighting of the IPD—but including these functions led to lower r values than did the on/off type envelope weighting.

c. Relative IPD-ILD contributions. The regression analysis for the average listener data found that the best value of a for model 1 is 0.50, which means that transformed IPD and ILD values contribute equally to the sensation of incoherence. Because the transformed values were scaled by Eqs. (7) and (8), we interpret this equality to mean that the scaling correctly represents the relative perceptual importance of IPD and ILD. Because the scaling was derived from Yost's steady-state sine experiment results, we conclude that it is valid to extend the results of steady-state measurements to the case of slowly fluctuating interaural differences.

3. Optimized parameters for other models

a. Longer integration times. Longer integration times were also tested for models 1 and 3–10 (model 2 was omitted because the results were so similar to model 1). Longer times were tested because the oscillating coherence experiments of Grantham and Wightman (1978) led to binaural time constants as long as 64 ms (−3 dB response at 2.5 Hz), an effect commonly called "binaural sluggishness." MLD experiments using a masker with temporally varying coherence (Grantham and Wightman, 1979) led to time constants that were even longer. Although the phenomenon of binaural

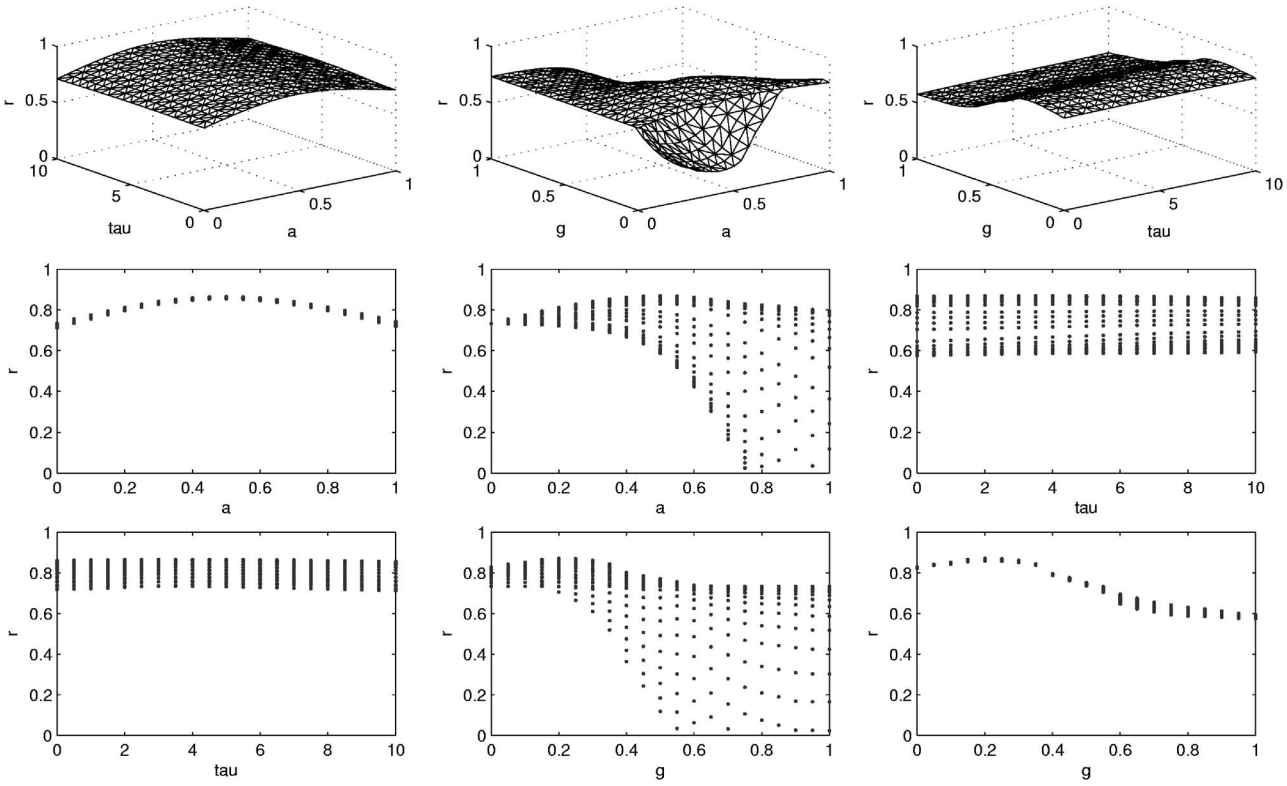

FIG. 5. The free parameter surfaces for fitting model 1 to the average listener data from Experiment 1—100 noises with 14-Hz bandwidth. In the upper-left panel, r is plotted against a and τ for $g=0.19$. In the upper-middle panel, r is plotted against a and g for $\tau=4$ ms. In the upper-right panel, r is plotted against g and τ for $a=0.5$ ms. The two panels below an upper panel flatten one of the free parameter dimensions. The variations of free parameters create smooth surfaces. At this bandwidth, the dependence upon τ is negligible.

sluggishness probably does not indicate an inertia affecting all binaural temporal variations (Hall *et al.*, 1998), we performed spot checks at 50-ms intervals (25 ms for model 1), trying to fit CAS data with fluctuations that had longer values of τ applied to the stimuli. Figure 6 shows that the value of $r_{\max}$ decreases monotonically for increasing values of τ for models 1 and 3–7. For model 1, the value of $r_{\max}$ dropped from $r_{\max}\approx 0.9$ for $\tau=0$ ms, to $r_{\max}\approx 0.6$ for $\tau=150$ ms. We conclude from Fig. 6 that there is no useful role for binaural sluggishness. For models 1 through 7, incorporating sluggishness through large τ leads to worse agreement with the experiment. For cross-correlation models 8–10, increasing values of τ lead to negligible change in agreement. However, we note that the approximation made in Eq. (24) assumed a short-analysis window. The validity of this assumption and the use of large values of τ will be addressed in Sec. VI.

b. Order of operations. The calculations described above applied temporal averaging to the physical stimulus, then applied laterality compression. However, it is not clear that this is the correct order of operations. Therefore, the models were rerun, first applying laterality compression and then temporal averaging. The integration times used were both the fine-scale (0–10 ms in 0.5-ms steps) and the longer times (50, 100, 150 ms). It was found that the order of operations did not matter for τ ranging from 0 to 10 ms in that the value of r changed by less than 0.01. Reversing the order of operations for the longer integration times always led to smaller r values compared to those in Fig. 6. The reduction could be as much as 0.2. Therefore, the best model applies laterality compression to signals that have been temporally averaged at a previous stage of processing.

c. Lateral-position models. Table I shows that lateral-position model 5 favors the transformed ILD over the transformed IPD in fitting the average listener data ($b=0$). The other lateral-position models, 6 and 7, mostly favor the transformed IPD ($b\approx 1$). This is in contrast to the independent-centers models (1–4), that weigh IPD and ILD as equally important. However, a lateral-position model that uses only IPD ($b=1$) or only ILD ($b=0$) is equivalent to an independent-centers model that uses only IPD ($a=1$) or only ILD ($a=0$). For example, an independent-centers model that incorporates IPD fluctuations separately must lead to an r value that is at least as large as the r for the lateral-position model with $b=0$ or 1. Therefore, model 1 must perform as least as well as model 5 (where b equals zero) in Fig. 4.

d. Short-term cross-correlation. The STCC models (models 8–10) correlated least well with the data, possibly because only the IPD is used in these decision statistics. Model 10 produced an interesting result in that the optimum threshold magnitude is approximately 0.06 (Table I), which means that listeners detected brief decorrelations from unity at coherence values of 0.94. This is larger than the jnds found by Gabriel and Colburn (1981). However, that work measured the jnd for the entire duration of the stimulus and not decorrelations over a short time interval. Therefore, 0.94 seems like a reasonable result. Even though the STCC models correlate with the data least well, it appears that the approximation that yields Eq. (24) is not a bad approximation.

4. The advantage of preprocessing

The results of modeling the data with the preprocessing removed (no temporal averaging, no laterality compression,

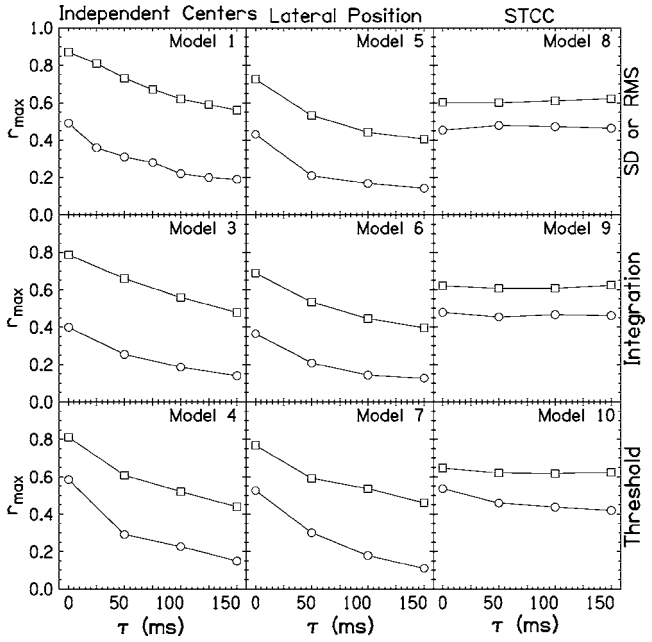

FIG. 6. The results of using long integration times for models 1 and 3–10. The squares show r_{max} for the 14-Hz data (Experiment 1), the circles show r_{max} for the 108-Hz data (Experiment 2). The top row is for standard-deviation models 1 and 5 or rms model 8. The middle row is for models that integrate absolute values. The bottom row is for threshold models.

and no critical envelope weighting) yielded r_{max}=0.69 for the independent-centers model for the averaged listener data to be compared with 0.87 with preprocessing included, i.e., model 1. Thus, these preprocessing assumptions prove beneficial in our modeling attempts. Without preprocessing r was maximized by weighting ILD and IPD fluctuations by the ratio of 0.14 dB/deg. Yost and Hafter (1987) reported that the trading of intensity and phase should be 0.10 dB/deg for interaural phases less than 90° and should be 0.08 or 0.10 dB/deg for interaural phases greater than 90°. Our ratio is higher than that of Yost and Hafter, but not much higher. The difference may arise because our experiment has dynamic fluctuations, whereas Yost and Hafter analyzed static interaural differences.

IV. EXPERIMENT 2: 108-Hz BANDWIDTH

After testing the ten models against the 14-Hz bandwidth noises and coming to some preliminary conclusions, we wondered how the models would perform for the wider bandwidth of 108 Hz.

A. Method

The 100 noises used in Experiment 2 are described in Fig. 7 along with the means, standard deviations, and correlation of interaural parameters. It was the same collection from which particular noises were selected in Article I. The same three listeners participated and the same procedure was used.

B. Results

The results of Experiment 2 are shown in Fig. 8, entirely parallel to Fig. 2 for Experiment 1. For Experiment 2 the

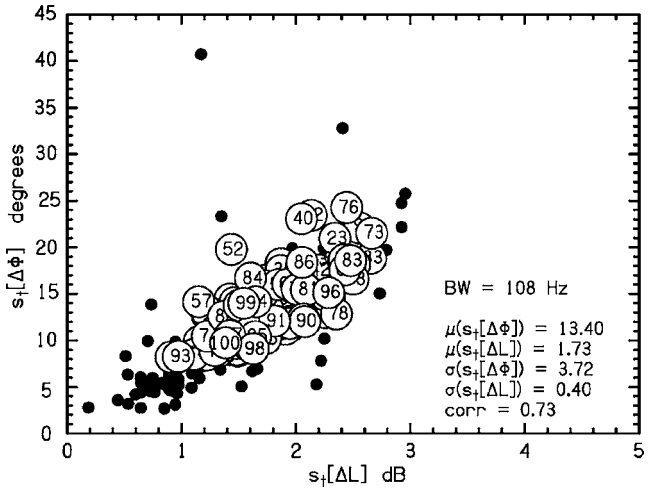

FIG. 7. Fluctuations of IPD vs fluctuations of ILD for the 100 reproducible noises with a 108-Hz bandwidth used in Experiment 2. Each noise is labeled by a serial number. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The means remained about the same as in Fig. 1, but the standard deviations decreased. Closed symbols replot the data of Fig. 1 for comparison.

percentage of correct responses shows a ceiling effect but the CAS does not. Unlike Experiment 1, where the ceiling was reached for particular noises for all the listeners, the ceiling for P_c in Experiment 2 was a factor for listeners D and M, but not necessarily W.

1. Comparison of model types

The ten models tested with Experiment 1 were also tested with Experiment 2. The results of the regression analysis for Experiment 2 are shown in Fig. 9, parallel to Fig. 4 for Experiment 1. The values of r_{max} are smaller than in Fig. 4 because there is less variation in detectability for a band as wide as 108 Hz compared to a band with a 14-Hz width. Model 4 gave the largest r_{max} for all three listeners and for the averaged data. For the averaged data, r_{max}=0.59. It is

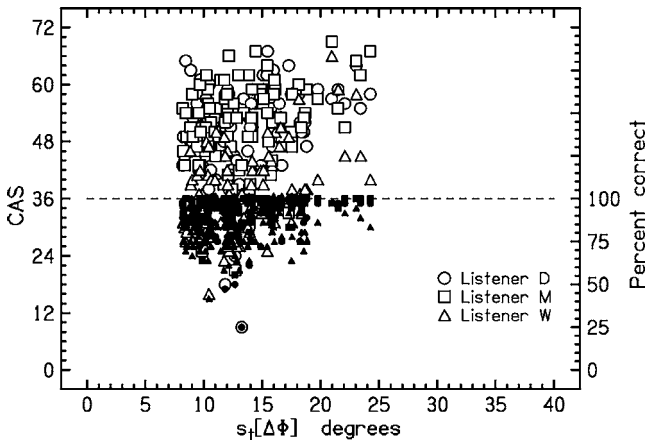

FIG. 8. All the detection data for the 100 noises from Experiment 2 for three listeners, D, M, and W, are plotted twice, once as the number correct—on a scale from 0 to 36, (0 to 100%) and once as CAS—on a scale from 0 to 72. The data are plotted as a function of the standard deviation of the interaural phase as in Fig. 2.

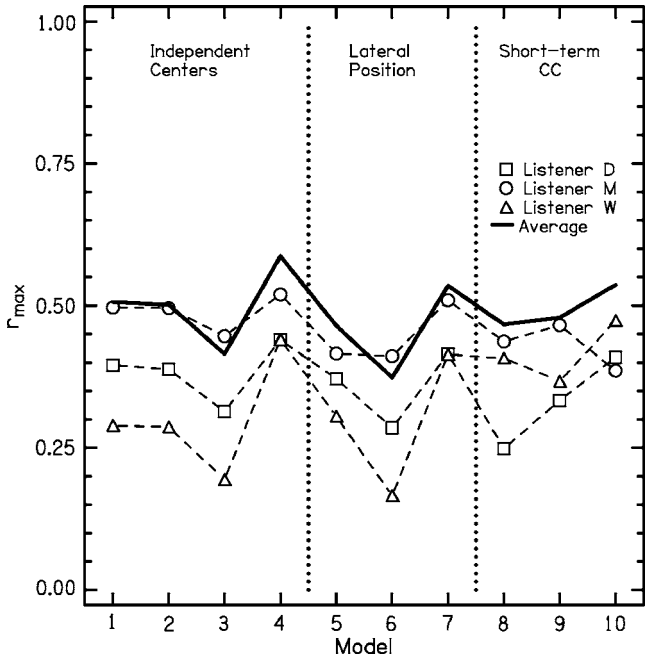

FIG. 9. The comparison of CAS scores for the 108-Hz noises of Experiment 2 with ten models. The value of r_{max} shows the correlation between the experimental CAS scores for the 100 noises and the best fit for each model, optimized by adjusting the model parameters. The solid line represents a fit to the data of the average listener. The values of r_{max} are not as high as for the 14-Hz data shown in Fig. 3.

possible that r values are small because the stimulus bandwidth is wider than the relevant auditory analysis filter. Auditory filtering will be discussed later.

Table II shows the values of the free parameters that maximized r for the 108-Hz bandwidth modeling. Table II shows that the parameters for the models at the 108-Hz bandwidth are mostly similar for different listeners. Consequently, the fits to the average listener shown in Fig. 9 are meaningful, although less convincing than the 14-Hz bandwidth fits.

Figure 9 shows that the ten models all perform about equally well. For the average listener, the highest r value, 0.59, is not much greater than the lowest, 0.37. As for Experiment 1, the most successful model is of the independent-centers type, but the threshold models (4, 7, and 10) outperform the others—even model 1, the most successful model in Experiment 1. The reason for this may be that models 4 and 7 have an extra free parameter, but model 10 has the same number of free parameters as model 1. This could be evidence that a threshold statistic is used for detecting incoherence in larger bandwidth stimuli. Again, the models account better for the average listener than for any individual listener, with a few exceptions.

2. Optimized parameters for model 4

Figure 10 shows how the four free parameters covary in the exploration of model 4, the best model from Experiment 2. Each panel shows how two free parameters change while two others are kept constant. As in Fig. 5, the constant parameters are set equal to the free parameters that yield r_{max}. When kept constant, $a=0.48$, $\tau=0.5$ ms, $g=0.15$, and $h=6.0$.

Figure 10(a) shows a vs τ. It shows that model 4 is fairly insensitive to changes in a, but there is a peak near $a=0.5$, consistent with the modeling from Experiment 1. Also, the calculation leads to positive r values only when τ is rather brief, 1.5 ms or less. Figure 10(b), a vs g, again shows that the model is fairly insensitive to a. Also, the largest values of r occur for g near 0.15.

Figure 10(c) shows that the optimum h is insensitive to different values of a. Also, it is possible to see the sharp drop off of r for values of h greater than 6. Figure 10(d) shows that the best value of τ is 0.5 ms and small values of g lead to the highest r values.

TABLE II. Values of free parameters that optimize r in modeling the detection results of Experiment 2 with 100 noises with a bandwidth of 108 Hz. Parameters are defined in the caption to Table I.

Model	Listener	τ (ms)	a,b	g	h
d_1	D	4.5	0.71	0.02	⋯
	M	0.5	0.50	0.02	⋯
	W	0.5	0.69	0.03	⋯
	Ave	1.5	0.54	0.03	⋯
d_2	D	4.5	0.68	0.02	⋯
	M	0.5	0.46	0.02	⋯
	W	0.5	0.63	0.03	⋯
	Ave	1.5	0.49	0.03	⋯
d_3	D	2.5	0.47	0.06	⋯
	M	1.0	0.29	0.02	⋯
	W	1.0	0.45	0.00	⋯
	Ave	1.5	0.36	0.03	⋯
d_4	D	0.0	0.60	0.18	5.50
	M	0.0	0.42	0.15	6.50
	W	0.0	0.60	0.17	7.00
	Ave	0.5	0.48	0.15	6.00
d_5	D	1.5	0.00	0.10	⋯
	M	1.5	0.00	0.00	⋯
	W	0.5	0.00	0.08	⋯
	Ave	1.0	0.00	0.08	⋯
d_6	D	5.5	1.00	0.03	⋯
	M	1.0	0.00	0.00	⋯
	W	6.0	0.17	0.15	⋯
	Ave	1.0	0.00	0.00	⋯
d_7	D	0.0	0.34	0.19	4.25
	M	0.5	0.38	0.26	4.00
	W	2.0	0.86	0.00	6.50
	Ave	1.0	0.97	0.01	5.75
d_8	D	0.0	⋯	0.04	⋯
	M	0.0	⋯	0.45	⋯
	W	0.0	⋯	0.00	⋯
	Ave	0.0	⋯	0.04	⋯
d_9	D	0.0	⋯	0.18	⋯
	M	0.0	⋯	0.06	⋯
	W	0.0	⋯	0.06	⋯
	Ave	0.0	⋯	0.06	⋯
d_{10}	D	0.0	⋯	0.24	0.13
	M	0.0	⋯	0.09	0.24
	W	0.0	⋯	0.11	0.30
	Ave	0.0	⋯	0.09	0.23

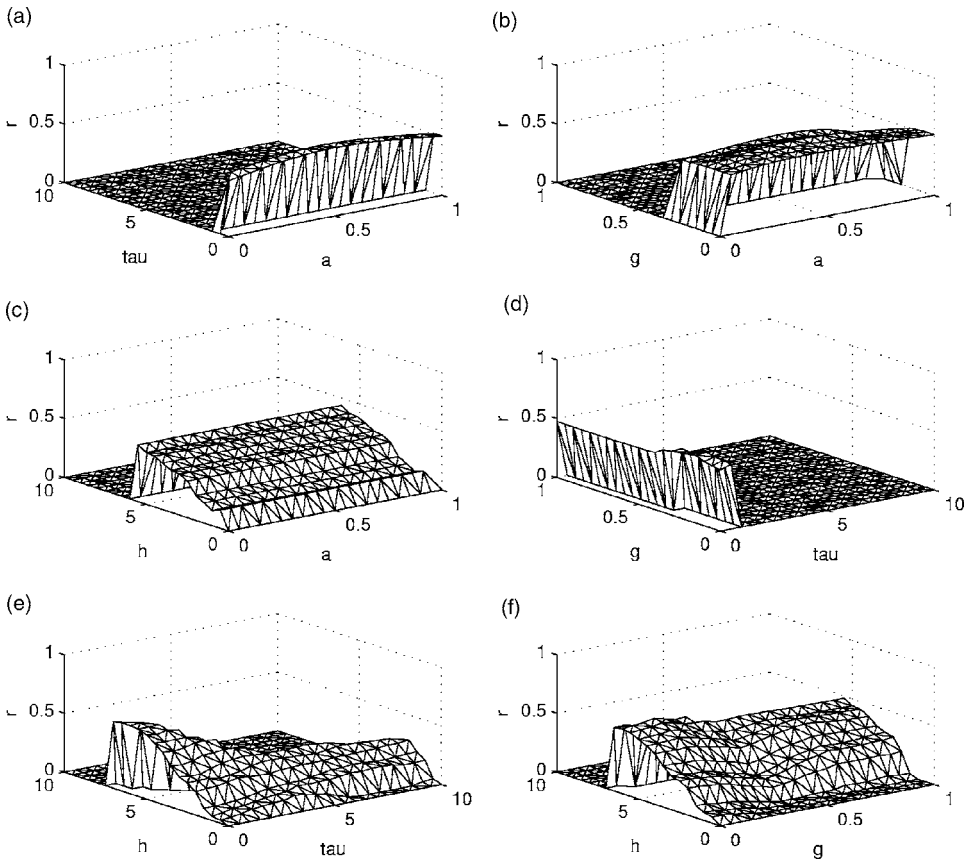

FIG. 10. The free parameter surfaces for fitting model 4 to the average listener data from Experiment 2 with 108-Hz bandwidth. Fixed parameters, not appearing along the axes, were given the optimum values for the average listener for model 4 in Table II. Results for model 4 show a strong interaction between h and g, a strong interaction between h and τ, and insensitivity to a.

Figure 10(e) shows that for no temporal averaging ($\tau=0$), model 4 leads to positive r values for all values of h up to 7.5. However, as the fluctuations become smoother due to larger values of τ (fewer peaks above a high threshold), the model leads to negative r values for h greater than 3.

Finally, Fig. 10(f) shows a rather strong interaction between h and g. The largest values of r occur when h is large and g is near 0.15.

3. Optimized parameters for other models

a. Longer integration times and order of operations. Figure 6 shows r values for Experiment 2 (108-Hz bandwidth) modeled with larger values of τ. As for Experiment 1 (14-Hz bandwidth), the data are described best by integration times less than 10 ms. Also, changing the order of operations showed that the best results occurred for temporal averaging followed by laterality compression, as in Experiment 1.

b. Lateral-position models. Table II shows that the lateral-position models (models 5–7) usually fit the Experiment 2 data best when the values of b are near 0 or 1. A similar result appeared in Table I for Experiment 1. Therefore, these models are most successful when they use only IPD or ILD information. However, a lateral-position model that makes no use of one of the interaural differences is indistinguishable from an independent-centers model. In fact, lateral-position models with $b=1$ are identical to independent-centers models with $a=1$. Therefore, in the case of the 108-Hz bandwidth data, it seems that independent centers may again be the better type of model, even if there is little distinction between models 1–7 by the values of $r_{\max}$.

c. Auditory filtering. A bandwidth of 108 Hz is 8% smaller than a Munich critical band at 500 Hz (Zwicker and Terhardt, 1980), but it is 37% larger than a Cambridge band (Moore and Glasberg, 1983). A possible role for auditory filtering was tested by centering a Cambridge gammatone filter on 500 Hz to filter the noise with 84-Hz bandwidth. Such filtering never increased the r values. Instead the r values decreased by as much as 0.1. As noted in Article I, it does not seem possible to understand our experimental results using a model in which information is confined to a critical band. A similar conclusion in connection with the MLD was reached by Evilsizer *et al.* (2002).

4. The advantage of preprocessing

The preprocessing assumptions of laterality compression, temporal averaging, and critical envelope weighting were removed to gauge their effect in Experiment 2. For model 1 and average listener data from Experiment 2, $r_{\max}=0.47$ without preprocessing can be compared with $r_{\max}=0.50$ with preprocessing included. As for Experiment 1, the preprocessing assumptions improved the agreement between model and data, but the improvement was much smaller than in Experiment 1. The best fit without preprocessing was obtained by weighting ILD and IPD fluctuations in the ratio of 0.08 dB/deg, which impressively matches the ratio suggested in the review by Yost and Hafter (1987), 0.08–0.10 dB/deg.

V. REPRODUCIBILITY

Prior to Experiment 1, a similar experiment was performed with two differences. First, the values of coherence among the 100 noises varied from 0.969 to 0.998, to be compared with 0.992 in Experiment 1. Second, listener D in Experiment 1 was replaced by listener E, female, age 19, and

well-practiced in incoherence detection. This experiment will be called Experiment 0. The distributions of phase and level fluctuations in Experiment 0 were almost identical to those seen in Fig. 1. A regression between the CAS data and the coherence values yielded an $r=0.48$ for the average listener.

The results of Experiment 0 were essentially the same as Experiment 1. Again model 1 emerged as the best model for all listeners and for the average listener with $r_{max}=0.89$, compared to 0.87 for Experiment 1. For Experiment 0, the values of the free parameters were $\tau=0.5$ ms (vs 4.0 for Experiment 1), $a=0.43$ (vs 0.50 for Experiment 1), and $g=0.17$ (vs 0.19 for Experiment 1). The r_{max} values, as a function of the ten models, for Experiment 0 correlated with those from Experiment 1, as shown in Fig. 4, at 0.90. Other notable results were also consistent between Experiments 0 and 1. Again, the independent-centers models outperformed the lateral-position models, which, in turn, outperformed the STCC models. Again, the model fits to data were insensitive to changes in τ, and there was an important advantage to preprocessing. The value added by the results presented in this section is to demonstrate that major results from Experiment 1 were reproducible with stimuli with different values of coherence but similar fluctuation statistics.

VI. SHORT-TERM CROSS-CORRELATION REVISITED

Articles I, II, and this article used noises that have a fixed value of long-term cross-correlation of 0.992. This value was calculated by cross-correlating the physical signals, x_R and x_L, over the entire duration of the stimulus. The major difference between the long-term and short-term cross-correlation is that the long-term cross-correlation calculation yields a single value whereas the short-term cross-correlation calculation yields a function of time. Models 8–10 approximated a STCC model by using the cosine of the IPD.

In this section, the STCC is computed as a function of time directly from an equation of the form of Eq. (22) without use of the cosine approximation. Again, a linear regression is used to compare models and data, leading to correlation coefficients r.

A. Physiological transformations

To bring the STCC model into line with current models of binaural processing, physiologically motivated stimulus transformations were also included in the model. These calculations were performed to test the possibility that a model using some form of STCC, as it appears in the auditory system after peripheral processing, might be able to fit the experimental results as well as our best model based on interaural fluctuations.

1. Auditory filtering

Breebaart and Kohlrausch (2001) found that using a fourth-order gammatone filter to approximate auditory filtering changed the value of coherence of noises with a bandwidth as small as 10 Hz. Therefore, we tested such a filter in our analysis, expecting it to change the effective coherence from a constant value of 0.992. The center frequency of the filter was 500 Hz.

2. Cochlear compression and rectification

Bernstein *et al.* (1999) were able to account for MLD data from Eddins and Barber (1998) and from Bernstein and Trahiotis (1996), by applying envelope compression and square-law half-wave rectification to simulate the cochlea. Bernstein *et al.* (1999) found that a compression exponent of approximately 0.2 could describe the data. This exponent, together with a square-law rectifier, corresponds to an exponent of 0.4 together with a half-wave rectifier, which agrees with the exponent derived by Oxenham and Moore (1995) and used by van de Par and Kohlrausch (1998).

The analysis in this section used envelope compression of the form

$$x'(t) = [E(t)]^{p-1}x(t), \tag{29}$$

where p is the compression exponent, a value between zero and one. Consistent with previous studies we used $p=0.4$ for both the right and left channels.

After cochlear compression, half-wave rectification was applied to the stimulus to represent the response of haircells. In the first calculation, only the positive portions of the waveform were retained; in the second, only the negative portions. The change from positive to negative led to negligible changes in r values, changes of less than 0.002.

3. Temporal averaging

The duration of the rectangular window Δt in the analog to Eq. (22) was varied from 10 to 300 ms. This range encompasses the values of binaural sluggishness that have been found in varied coherence experiments (e.g., Grantham and Wightman, 1979). The short-term cross-correlation function was calculated for each instant, t, of the 500-ms duration of a noise. When t was less than Δt, the window extended back in time only to the beginning of the noise.

4. Decision making

Since the STCC is a function of time, a mathematical operation was needed to calculate a final decision statistic to compare against the psychological data. The calculations used to obtain a statistic were similar to those used in models 8, 9, and 10—rms deviation, integration, and threshold. The threshold values tested in the threshold models were $h=0.001$, 0.01, 0.025, 0.05, and 0.075.

B. Results and discussion

The results of the regression between the STCC statistics, as calculated with model physiology, and the CAS for the average listener can be seen in Fig. 11. The solid lines are for the 14-Hz bandwidth and the dotted lines are for the 108-Hz bandwidth. Values of r are plotted for two values of the threshold level, $h=0.025$ and $h=0.05$, representative parameters for the best performing threshold models.

For the 14-Hz data, the best model was the threshold model with $h=0.05$ ($r=0.78$), shown in the bottom panel of

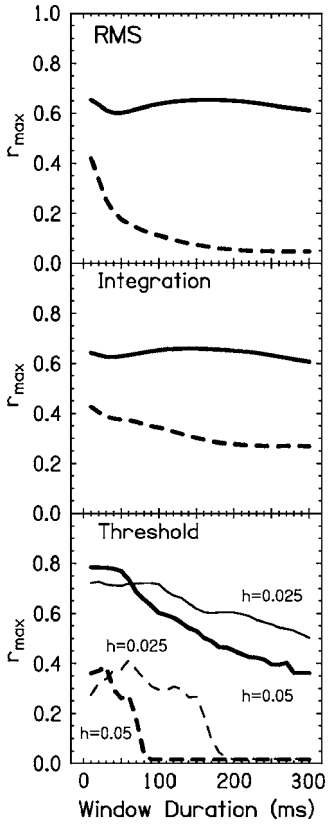

FIG. 11. Performance of the three types of models (rms, integration, and threshold) for the physiologically motivated short-term cross-correlation models for different window durations. Each model incorporates auditory filtering, cochlear compression, and half-wave rectification. The solid lines represent the correlations between the model and the 14-Hz data (Experiment 1). The dashed lines represent the correlations for the 108-Hz data (Experiment 2). There are two threshold calculations for each bandwidth, with $h=0.025$ (thin) or $h=0.05$ (thick).

Fig. 11. However, as noted in Sec. IV, the threshold model might outperform the other models simply because it has an extra degree of freedom. For the 108-Hz data, the best model was the integration model. However, for this bandwidth, there was little difference between the best integration model ($r=0.43$), the best rms model ($r=0.42$), and the best threshold model with $h=0.025$ ($r=0.41$).

As shown in Fig. 11, the best fit to the data used window durations less than 100 ms. The shortest window duration, 10 ms, was optimum in all but one case—the threshold model for the 108 Hz, which had an optimum window of 60 ms for $h=0.025$ and an optimum window of 30 ms for $h=0.05$. Once again, the long integration times that are suggested by binaural sluggishness experiments were not supported by our modeling of incoherence detection data.

In summary, STCC models that include physiologically-based transformations can describe incoherence detection data better than models 8, 9, and 10, which do not have those transformations. However, no STCC model could account for incoherence detection data quite as well as models based on interaural fluctuations. For a bandwidth of 14 Hz, the most successful STCC model produced an r value of 0.78 to be compared with model 1, which led to $r=0.87$. For a bandwidth of 108 Hz, the most successful STCC model produced an r value of 0.43 to be compared with model 4, which led to $r=0.59$.

VII. GENERAL DISCUSSION

The goal of the experiments reported in this article was to discover the signal characteristics and the binaural perceptual operations that enable a listener to detect small amounts of interaural incoherence in bands of noise.

A. Summary

Experiments 1 and 2 used 100 reproducible noises, with bandwidths of 14 and 108 Hz, respectively, to test models of interaural incoherence detection. Each model was rated by varying its parameters to find the best agreement with detection data.

The results of Experiment 1 were the most informative. They showed that the independent-IPD/ILD-centers models outperformed lateral-position models, and that the lateral-position models outperformed the short-term cross-correlation models. The results of Experiment 2 were less informative, mainly because the wider bandwidth led to interaural fluctuations that varied less across different noises. In the comparison of Experiment 2 data with model predictions, the data were less well fitted by the models, the data discriminated among models less clearly, and the best performing models had one or two more parameters than the other models, which may have made the comparison unfair.

Because of the greater power of Experiment 1, with the narrow bandwidth, the general discussion will be mainly concerned with the comparison of models with the results of Experiment 1. The most interesting comparison is between model types.

The first comparison is between models 1–4, which treat fluctuations in IPD and ILD by independent centers, and models 5–7, which consider a fluctuation of the lateral position of the image. Figure 4 for 14-Hz bandwidth shows that the independent-centers models outperform the lateral-position models for all three listeners and for the average listener. Experiment 0 supports this conclusion. Thus, models of the independent processing type are favored unambiguously.

Tables I and II show that values of the IPD-ILD trading parameter b for the lateral-position models were often near 1 or 0, so that only the IPD or only the ILD contributes to detection. That result means that in the optimizing process the lateral-position models become unstable and become equivalent to independent-binaural-centers models. These problems with lateral-position models favor the independent processing model by default. Also, there is some possible support for models that use IPD and ILD independently in the observation that multiple images can be tracked over short durations (Hafter and Jeffress, 1968; Ruotolo *et al.*, 1979).

In a second comparison, the independent-centers models, models 1–4, also performed better than the STCC models 8–10 and physiology-motivated STCC models tested in Sec. VI. This result continues a pattern: Article I showed that the long-term cross-correlation of the physical signals, as

measured over the full, 500-ms stimulus duration, was an inadequate predictor of incoherence detection. Article II showed that the cross-correlation of the physical signals, as measured over short durations of 25–100 ms, was also inadequate. The present article shows that cross-correlation of transformed signals is also relatively unsuccessful compared to fluctuation detection models.

In the end, the best of the best-fitting models was of the following form: The binaural system detects incoherence on the basis of fluctuations in independently processed IPD and ILD channels, as though IPD and ILD were encoded at different centers without regard for the relative timing of their fluctuations. Nevertheless, there is a residual IPD-ILD interaction in that an IPD fluctuation has no effect if the envelope in the left or right channel becomes smaller than about 20% of the rms envelope value. The IPD and ILD fluctuations are temporally averaged by an exponential window with a time constant less than 5 ms. The time-averaged fluctuations are then laterality compressed. The laterality compression factors found in steady-state experiments on lateral position turn out to be adequate to describe the laterality compression of fluctuations. The processing centers register laterality-compressed fluctuations in IPD and ILD as measured by standard deviations over time. The registered fluctuations are added on a laterality scale at a more central site to form a decision statistic used to detect incoherence. For narrow bands near 500 Hz, time-averaged and laterality-compressed IPD and ILD fluctuations are added with approximately equal weight. As the bandwidth grows, different noises with a given interaural coherence have increasingly similar fluctuations, and the detection of incoherence can be predicted with increasing reliability by the value of coherence itself.

The above paragraph specifying the best binaural model is based on a literal interpretation of the correlation coefficient values, although these values often varied little with model parameters. We also doubt that it is really possible to say that combining independent IPD and ILD fluctuations (model 1) is appreciably more successful than combining independent IPD and ILD mean square fluctuations (model 2). (See Fig. 4.) Further, it is always possible that other models, not tested, might do better. Also, model 1 cannot be proved to be the best for a bandwidth as large as, or larger than, a critical band.

The conclusion that incoherence is detected on the basis of IPD and ILD fluctuations contradicts the conclusions of Breebaart *et al.* (1999) and Breebaart and Kohlrausch (2001) concerning the similar problem of NoSπ detection. In those articles it was pointed out that the distributions of the IPD and the ILD do not vary with noise bandwidth. By contrast, NoSπ detection shows considerable bandwidth dependence. Thus, it was argued, detection is unlikely to be mediated by IPD or ILD fluctuations. Our distribution calculations agree numerically in the sense that the mean standard deviations of IPD and ILD, as shown in Figs. 1 and 7, hardly change when the bandwidth is increased by a factor of 8. However, the variance of the fluctuations among different noise samples, as shown in those figures, depends greatly on bandwidth. We think it possible that the large variance in fluctuations across different samples of noise having small bandwidth, particularly the occurrence of especially large fluctuations, is responsible for the observed bandwidth dependence of detection.

This conjecture, based on the width of the *ensemble* distribution, appears to answer the objection to fluctuation models from the work of Breebaart and his colleagues. It may also help to explain the large individual differences observed in NoSπ detection for narrow bands (Bernstein *et al.*, 1998; Buss *et al.*, 2007) because it suggests that good detection performance requires recognizing the signal in atypical epochs of the stimuli.

B. Binaural processing

Like the experiments of Articles I and II, the experiments presented here conclude that long-term coherence inadequately predicts incoherence detection when the bandwidth is narrow. Long-term coherence may be adequate in the wideband limit. Three results from Article I and the present article indicate features of the wideband limit. These three trends with increasing bandwidth are: (1) the variance among different noises of the fluctuations of IPD and the fluctuations of ILD decreases, (2) the ability of listeners to detect incoherence varies less among different noise samples, and (3) different models of incoherence detection make predictions that are increasing similar, consistent with the prediction of Domnitz and Colburn (1976).

An important difference between the experiments with 14-Hz bandwidth and experiments with 108-Hz bandwidth is the speed of the fluctuations. On the basis of our experiments and modeling, we would agree with Zurek and Durlach (1987) about the advantage of slow fluctuations, but we would not agree that binaural sluggishness plays a role. Instead, our calculations suggest that the binaural system responds rapidly, with a time constant of the order of milliseconds. The best fitting model found an insensitivity to τ in the region of 4 ms for the 14-Hz bandwidth. This time constant is not inconsistent with our matched-noises experiment in Article I wherein the slow fluctuations at 14-Hz bandwidth proved advantageous compared to the fluctuations at 108 Hz. It is commensurate with modulation transfer functions seen in such monaural tasks as the detection of amplitude modulation of broadband noise (Viemeister, 1979). Recently, Stellmack *et al.* (2005) measured temporal modulation transfer functions with time constants of 1 ms for monaural and 1.3 ms for interaural modulation.

By contrast, binaural sluggishness is associated with time constants of tens, or even hundreds, of milliseconds. As suggested in the last paragraph of Hall *et al.* (1998), binaural sluggishness seems to arise in situations where both the masker and the signal plus masker contain dynamical interaural cues. If the masker is interaurally stable the binaural system can take advantage of events in brief epochs. The detection of a small amount of incoherence as a contrast to a diotic noise, as in our experiments, is well modeled as a stable masker (No) and a noise-like signal with a different phase relationship. A rapid response for such a task is consistent with other experiments cited by Hall *et al.*

The best integration time less than 5 ms can be compared with the integration time of 300 ms found to be best in

the loudness meter model of localization as calculated by Hartmann and Constan (2002). Thus, it seems that the binaural auditory system is capable of employing either short or long integration times depending on which better suits the task. When the task is to lateralize an image based on a binaural cue (loudness meter) the integration time is long. When the task is to detect rapid fluctuations in binaural cues, as in the present article, the time is short. A similar point of view was taken with respect to monaural listening by Eddins and Green (1995) wherein integration times of several hundreds of milliseconds are possible for detecting the presence of a signal but times as short as several milliseconds are possible for detecting rapid signal variations.

VIII. CONCLUSION

Experiments of this article studied the detection of small amounts of interaural incoherence in noise bands near 500 Hz. The goal of the experiments was to test different binaural detection models: independent-IPD/ILD, lateral-position, and short-term cross-correlation. Several different transformations were included in the analysis: temporal averaging, laterality compression, and critical envelope weighting. The parameters of these transformations were systematically optimized in the tests of the models. The strongest test came from experiments with a 14-Hz bandwidth. There it was found that the best model independently added the standard deviations of transformed IPD and ILD. This model outperformed a variety of plausible short-term cross-correlation models, even when the cross-correlation models incorporated auditory filtering, cochlear compression, and half-wave rectification.

The nature of incoherence detection depends on the stimulus bandwidth. In the limit of extreme narrow bands the interaural parameters vary extremely slowly, and one imagines that listeners can track the lateral positions indicated by the interaural differences, or by their combination, to detect incoherence. At the other extreme, where the bandwidth is considerably larger than a critical band, the incoherence detection data do not distinguish between different models. In the intermediate range, near 10 Hz, home to the most dramatic MLD results, our experiments indicate that the binaural system is not only sensitive to a running average of the interaural differences, as reflected in the perception of lateral position, but also makes use of the separate fluctuations of the IPD and ILD to detect interaural incoherence.

ACKNOWLEDGMENTS

We are grateful to Dr. H. S. Colburn, Dr. N. I. Durlach, Dr. A. Kohlrausch, Dr. S. van de Par, and Dr. C. Trahiotis for useful discussions about coherence. This work was supported in part by the National Institute on Deafness and Other Communicative Disorders, Grant No. DC00181.

[1]This idea was first suggested by Dr. H. S. Colburn in 2004.

Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–316.

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

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). "Inter-individual differences in binaural detection of low-frequency tonal signals masked by narrowband or broadband noise," J. Acoust. Soc. Am. **103**, 2069–2078.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized interaural correlation: Accounting for NoSπ thresholds obtained with Gaussian and low-noise masking noise," J. Acoust. Soc. Am. **106**, 870–876.

Breebaart, J., and Kohlrausch, A. (**2001**). "The influence of interaural stimulus uncertainty on binaural signal detection," J. Acoust. Soc. Am. **109**, 331–345.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.

Buss, E., Hall, J. W., and Grose, J. H. (**2007**). "Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz." J. Acoust. Soc. Am. **121**, 411–419.

Colburn, H. S., Isabelle, S. K., and Tollin, D. J. (**1997**). "Modelling binaural detection performance for individual masker waveforms." in *Binaural and Spatial Hearing*, edited by R. H. Gilkey and T. Anderson (Erlbaum, Englewood Cliffs, NJ).

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. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination. II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1556.

Eddins, D. A., and Barber, L. E. (**1998**). "The influence of stimulus envelope and fine structure on the binaural masking level difference," J. Acoust. Soc. Am. **103**, 2578–2589.

Eddins, D. A., and Green, D. M. (**1995**). "Temporal integration and temporal resolution," *Handbook of Perception and Cognition—Hearing*, 2nd ed., edited by B. C. J. Moore (Academic, San Diego), pp. 207–242.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (**1959**). "Operating characteristics determined by binary decision and by ratings," J. Acoust. Soc. Am. **31**, 768–773.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336–345.

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

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.

Goupell, M. J., and Hartmann, W. M. (**2006**). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," J. Acoust. Soc. Am. **119**, 3971–3986.

Goupell, M. J., and Hartmann, W. M. (**2007**). "Interaural fluctuations and the detection of interaural incoherent. II. Brief duration noises," J. Acoust. Soc. Am. **121**, 2127–2136.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517.

Hafter, E. R. (**1971**). "Quantitative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116–1122.

Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Hall, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). "The masking level

difference in low-noise noise," J. Acoust. Soc. Am. **103**, 2573–2577.
Hartmann, W. M., and Constan, Z. A. (**2002**). "Interaural level differences and the level meter model," J. Acoust. Soc. Am. **112**, 1037–1045.
Isabelle, S. K., and Colburn, H. S. (**1987**). "Effects of target phase in narrowband frozen noise detection data," J. Acoust. Soc. Am. **82**, S109.
Isabelle, S. K., and Colburn, H. S. (**1991**). "Detection of tones in reproducible narrowband noise," J. Acoust. Soc. Am. **89**, 352–359.
Isabelle, S. K., and Colburn, H. S. (**2004**). "Binaural detection of tones masked by reproducible noise: Experiment and models," Report BU-HRC 04–01.
Jeffress, J. A., Blodgett, H. C., Sandel, T. T., and Wood, C. L. (**1956**). "Masking of tonal signals," J. Acoust. Soc. Am. **28**, 416–426.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **50**, 1494–1511.
Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.
Ruotolo, B. R., Stern, R. M., and Colburn, H. S. (**1979**). "Discrimination of symmetric time-intensity traded binaural stimuli," J. Acoust. Soc. Am. **66**, 1733–1737.
Schulman, A. I., and Mitchell, R. R. (**1966**). "Operating characteristics from yes-no and forced-choice procedures." J. Acoust. Soc. Am. **40**, 473–477.
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2005**). "Monaural and interaural temporal modulation transfer functions measured with 5-kHz carriers," J. Acoust. Soc. Am. **118**, 2507–2518.
van de Par, S., and Kohlrausch, A. (**1998**). "Diotic and dichotic detection using multiplied noise maskers," J. Acoust. Soc. Am. **103**, 2100–2110.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Webster, F. A. (**1951**). "The influence of interaural phase on masked thresholds. I. The role of interaural time-deviation," J. Acoust. Soc. Am. **23**, 452–462.
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., and Hafter, E. R. (**1987**). "Lateralization" in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 49–84.
Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.
Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.