Interaural fluctuations and the detection of interaural incoherence. II. Brief duration noises

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

(Received 12 December 2005; revised 20 December 2006; accepted 5 January 2007)

Listeners detected a small amount of interaural incoherence in reproducible noises with narrow bandwidths and a center frequency of 500 Hz. The durations of the noise stimuli were 100, 50, or 25 ms, and every one of the noises had the same value of interaural coherence, namely 0.992. When the nominal noise bandwidth was 14 Hz, the ability to detect incoherence was found to depend strongly on the size of the fluctuations in interaural phase and level for durations of 100 and 50 ms. For the duration of 25 ms, performance did not appear to depend entirely on fluctuations. Instead, listeners sometimes recognized incoherence on the basis of laterality. However, when the nominal bandwidth was doubled, leading to a greater number of fluctuations, detection performance at 25 ms resembled that at 50 ms for the smaller bandwidth. It is concluded that the detection of a small amount of interaural incoherence is mediated by fluctuations in phase and level for brief stimulus durations, so long as such fluctuations exist physically. This conclusion presents a promising alternative to models of binaural detection that are based on the short-term cross-correlation in the stimulus. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2436714]

PACS number(s): 43.66.Ba, 43.66.Pn, 43.66.Qp [AK] Pages: 2127–2136

I. INTRODUCTION

Interaural coherence is a measure of the similarity of signals in the left and right ears. Interaural incoherence occurs when the signals differ in some way apart from a simple delay or attenuation. Incoherence is introduced, for example, when an unrelated noise is added to the signal appearing at one of the ears. In a previous article (Goupell and Hartmann, 2006), to be called "article I," it was concluded that the ability of listeners to detect interaural incoherence in narrowband noises is mediated by fluctuations in the interaural phase and level differences. Those conclusions were based on experiments using hundreds of different reproducible noises, all with the same value of interaural coherence, namely 0.9922. The interaural coherence was defined, as usual, as the maximum value of the cross-correlation function, as computed for the stimulus signals themselves. In these experiments, with no reference interaural delay, the amount of admixed, unrelated noise was so small that the maximum value always occurred for zero lag. Although all the noises had the same coherence, different noises had interaural phase fluctuations and interaural level fluctuations of different size. The experiments showed that these fluctuations were critical in detection. Specifically, incoherence was far more easily detected for noises with large interaural fluctuations than for noises with smaller fluctuations.

The thrust of the conclusions in terms of auditory theory was to emphasize the central importance of fluctuations and to deemphasize the stimulus interaural coherence or cross-correlation function *per se*. It was suggested that interaural coherence is an average signal property that can indicate a range of expected interaural fluctuations, but that it has no further perceptual significance. Said another way, one can expect that when the interaural coherence is close to one, then decreasing the value of coherence normally makes the incoherence easier to detect, but only because this decrease broadens the distributions of the fluctuations in the interaural differences. When the coherence decreases, the standard deviations across different noise tokens differ more from the mean standard deviation. Article I also found that as the bandwidth of the noise increases, the distributions of fluctuations across different noise tokens become more narrow. Therefore, the value of the coherence becomes a better predictor of detection performance because it better predicts the variance over time of the interaural differences.

The experiments of article I used noise stimuli with a duration of 500 ms, and the common value of coherence (0.9922) was computed over the entire 500-ms duration. That relatively long duration opens the experiments and conclusions to a serious challenge because binaural processing can be rapid, with time constants much shorter than 500 ms (Hall *et al.*, 1998). It would seem entirely possible to argue that incoherence detection is fundamentally based on the cross-correlation function after all, but that the relevant cross-correlation is computed over brief time intervals. The objection would continue by suggesting that those particular noise tokens with incoherence that was easy to detect probably had important minima in the short-term coherence, but the experiments of article I would not have been sensitive to that fact because of the excessively long temporal average used in generating and selecting the stimuli.

The purpose of the present article is to address the objection by exploring the effects of duration on incoherence detection. Specifically, the reported experiments will try to determine whether a fixed value of coherence determined over small durations—100 ms and shorter—can predict incoherence detection. In this way, the experiments presented

---

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

here address the above-mentioned challenge, and the results will have implications for the validity of the conclusions reached in article I.

II. EXPERIMENT 1

Experiment 1 begins with the null hypothesis that the detection of incoherence is determined by the short-term cross-correlation as measured over a brief stimulus and not by the interaural fluctuations. The hypothesis predicts that if stimuli are as brief as the binaural computation time for short-term cross-correlation, then detection scores will be predictable from the stimulus cross-correlation peak (coherence) itself. In particular, if every noise in an ensemble has the same value of interaural coherence, as computed over the brief duration, then the incoherence ought to be equally detectable for all noises of the ensemble.

A. Stimuli

1. Stimulus generation

Three collections of 100 two-channel, narrowband noises (i.e., left-right noise-pairs, to be called “noises”) were created for Experiment 1. The different collections were distinguished by the noise durations—100, 50, and 25 ms. The generation of the noises was a multistep process, which is described in the following paragraphs.

Initially, noises were 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. Components below 495 Hz and above 505 Hz were attenuated with a raised-cosine window, which zeroed the amplitudes at 490 and 510 Hz. Consequently, there were nine spectral components that had nonzero amplitudes, and the 3-dB bandwidth was 14 Hz. An orthogonalization procedure guaranteed that the interaural coherence of each noise was precisely 0.9922. Up to this point the stimulus generation procedure was identical to the procedure followed in article I.

Next, the stimuli were given temporal windows with total durations of 100, 50, or 25 ms including 10-ms Hanning edges for attack and decay. After the temporal shaping, the value of the coherence was recomputed. Noises were accepted only if the value of the coherence was 0.992 ± 0.001. In order to obtain 100 noises with a given duration and correct coherence, it was necessary to reject more than ten times that number.

After a collection of 100 noises was created, ten noises were selected to make a *phase set* and ten were selected to make a *level set*. As in article I, the phase set consisted of those five noises with the largest standard deviations in interaural phase, as computed over the duration, plus those five noises with the smallest standard deviation in interaural phase. Similarly, the level set was constructed by selecting noises with the five largest and five smallest standard deviations in interaural level. Occasionally a particular noise would be common to both phase and level sets. These two sets of stimuli were presented to listeners.

2. Stimulus spectra

The goal of the stimulus generation technique was to pack a number of components into a narrow band to create a noise with complicated fluctuations but a brief duration. However, the brief duration resulted in a bandwidth greater than the nominal value of 14 Hz. The spectra of the noises selected for the phase and level sets were measured. Average statistics were as follows: For the 100-ms noises, 90% of the energy was contained in a band 24 (±11) Hz wide, and 99% was contained in a band 76 (±22) Hz wide. For the 50-ms noises, the 90% and 99% bandwidths were 39 (±8) and 106 (±16) Hz. For the 25-ms noises, the corresponding bandwidths were 74 (±4) and 126 (±16) Hz. No correlation could be seen between the bandwidths of individual noises and the sizes of the interaural fluctuations in phase or level.

3. Stimulus interaural values

Figure 1 shows the fluctuations for all the stimuli used in this experiment. Fluctuations in interaural phase and interaural level for each noise are expressed, as in article I, as the standard deviations $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ computed over the stimulus duration. For example, for a given noise, $s_t[\Delta\Phi]$ is computed by taking the square root of the sum of the squared deviations of the instantaneous interaural phase from the mean value. (Although an ensemble-average mean value would be zero, mean values for individual noises can be important for brief durations.) The average value of $s_t[\Delta\Phi]$, namely $\mu(s_t[\Delta\Phi])$, and the standard deviation of $s_t[\Delta\Phi]$, namely $\sigma(s_t[\Delta\Phi])$, are computed over all the noises of the ensemble, i.e., over the collection of 100 noises. This ensemble standard deviation appears in the legends in the figure panels on the left, together with the average value and the standard deviation for $s_t[\Delta L]$ as well as the correlation between phase and level fluctuations. It is interesting to observe that the average values decrease as the duration decreases for this fixed nominal bandwidth. By contrast, it was found in article I, that the average values were roughly constant for a long fixed duration when the bandwidth varied. It is worth noting that the 10-ms edges did not affect the computation of the fluctuations. In our approximations, the calculated interaural phase and interaural level depend on ratios in which the multiplicative Hanning edge is canceled.[1] The left-hand panels of Fig. 1 make it clear that noises with a large fluctuation in interaural phase also tend to have a large fluctuation in interaural level, and vice versa. The correlation between phase and level fluctuations ranges from 0.53 to 0.75.

The right-hand panels of Fig. 1 show the mean and standard deviations of the five stimuli with the greatest interaural fluctuations and of the five stimuli with the least. Circles are for the phase sets; squares are for the level sets. These stimuli were selected from the collections shown in the left-hand panels. These are the ten stimuli that were heard by the listeners. The difference between the larger fluctuations and the smaller fluctuations diminishes dramatically as the duration decreases to 25 ms.

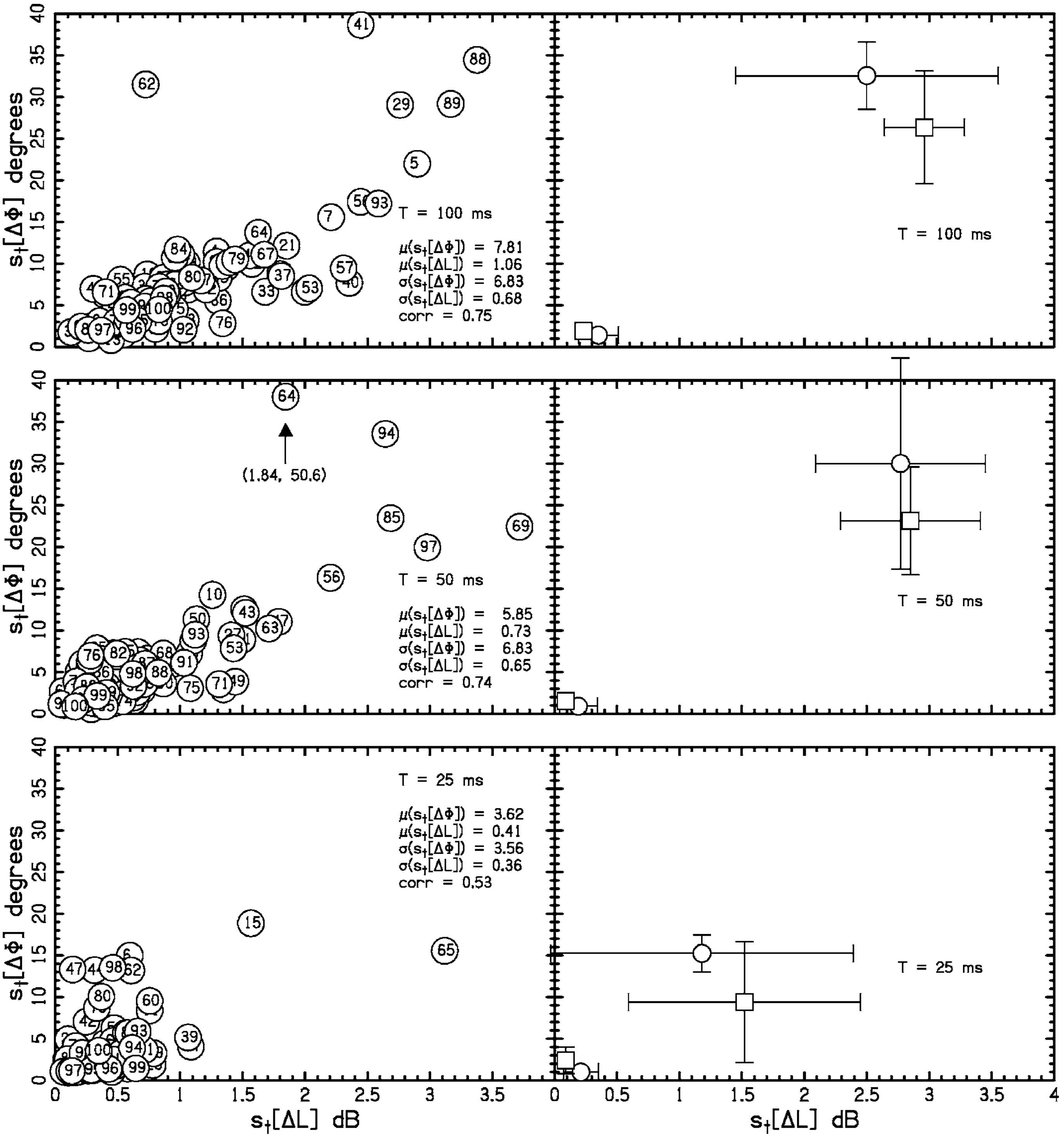

FIG. 1. Panels in the left column show the interaural fluctuations for the three collections of 100 noises. Fluctuations in interaural phase and interaural level for each noise are expressed as the standard deviations, $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, computed over the stimulus duration. Panels in the right column show the mean and standard deviations of the five stimuli with the greatest interaural fluctuations and of the five stimuli with the least. (Circles are for phase sets; squares are for level sets.) These ten stimuli were presented to the listeners. Error bars are 2 s.d. in overall width.

4. Stimulus synthesis

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 8 ksps. The noise was low-pass filtered with a corner frequency of 4000 Hz and a −115 dB/octave rolloff. The noises were presented at 70±3 dB with levels determined by programmable attenuators (PA4) operating in parallel on the two channels. The level was randomly chosen, in 1-dB increments, 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. Each experi-

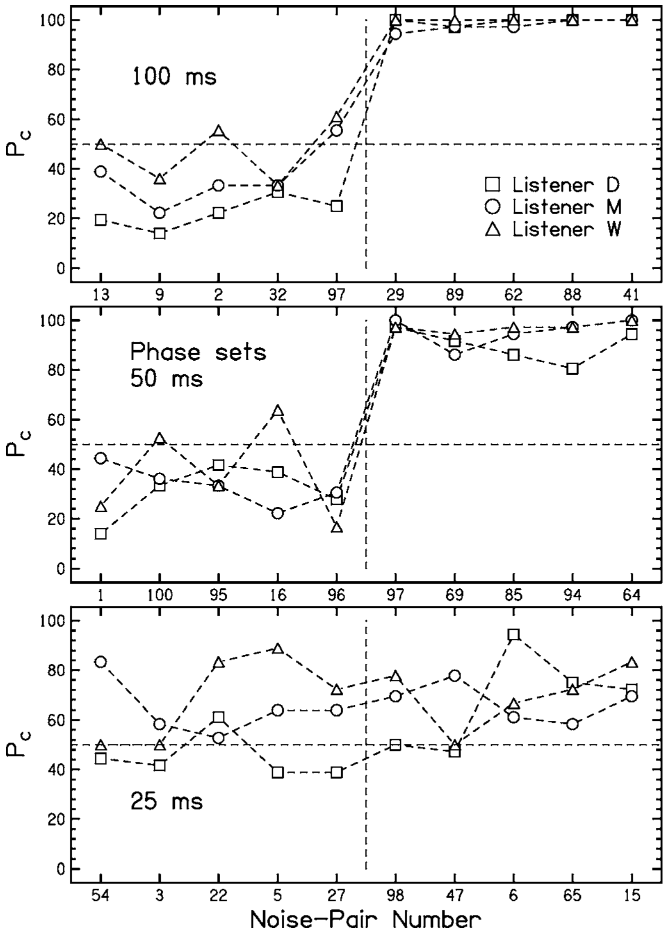

FIG. 2. The percent correct (P_c) for three listeners for the *phase set* from the 25-, 50-, and 100-ms collections. The noises were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noises. The noises are rank ordered by increasing $s_t[\Delta\Phi]$ along the horizontal axis. The vertical dashed line represents 90 unused noises. The horizontal dashed line represents the level of guessing. Several of the 50- and 100-ms noises to the left of the dashed line are below the level of guessing.

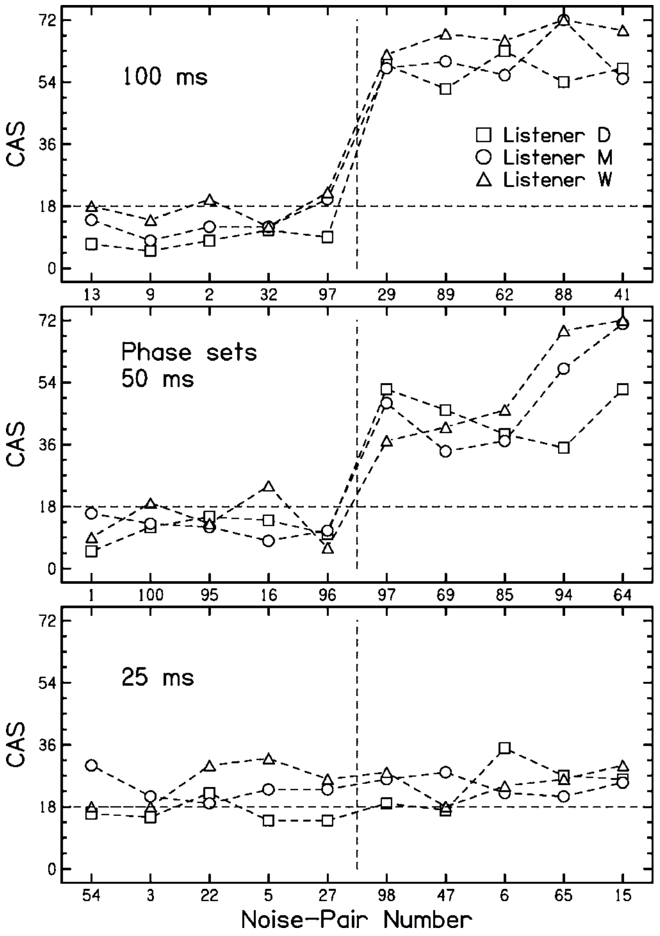

FIG. 3. Same as Fig. 2 except that the CAS values are plotted instead of the percentage of correct responses.

mental run was devoted to either a phase set or a level set. Within a set, the order of the reproducible noises was randomized—differently on each run. Six runs were devoted to the phase set and six to the level set. Listeners completed one set before moving on to the other.

The structure of runs, trials within a run, and the data collection procedure was the same as that in article I. 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 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 noises in the set except that they were required to be different from either channel of 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 article I, listeners were allowed to give a confidence rating. By making a correct choice and indicating confidence, a listener gained two points for the trial instead of just one point for a simple correct response. Making a wrong choice and indicating confidence was discouraged by allowing the listener only one such error in a run. With the second "confident" error, the run terminated and the listener was required to begin it again. This procedure led to a confidence-adjusted score (CAS) with a maximum possible score of 72 for a run of 36 trials. The data collection procedure kept track of both the percentage of correct responses (P_c) and the CAS.

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.

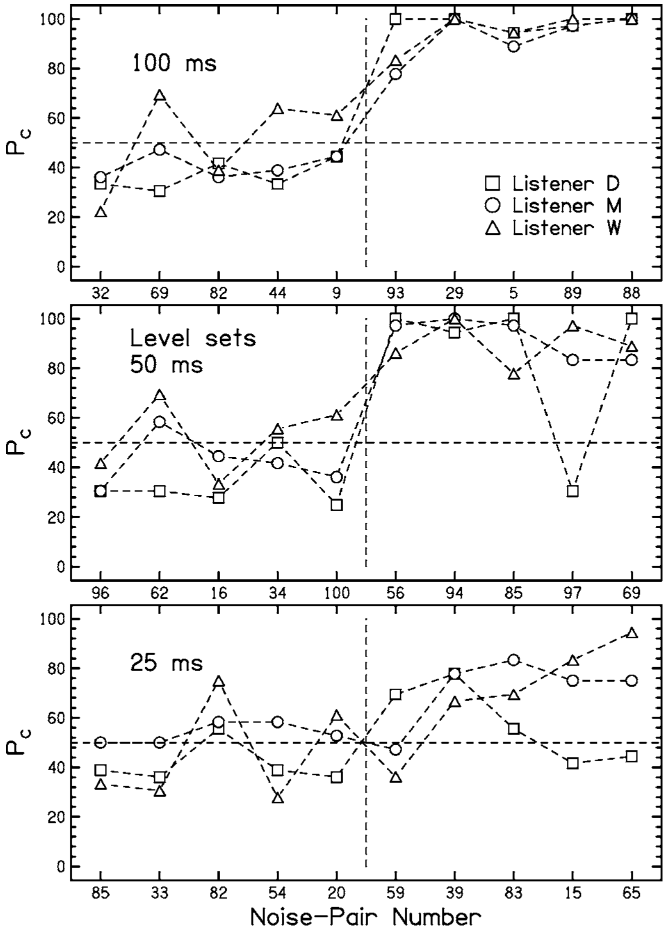

FIG. 4. The percent correct (P_c) for three listeners for the *level set* from the 25-, 50-, and 100-ms collections. The noises were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noises. The noises are rank ordered by increasing $s_t[\Delta L]$ along the horizontal axis. The vertical dashed line represents 90 unused noises. The horizontal dashed line represents the level of guessing. Several of the 50- and 100-ms noises to the left of the dashed line are below the level of guessing.

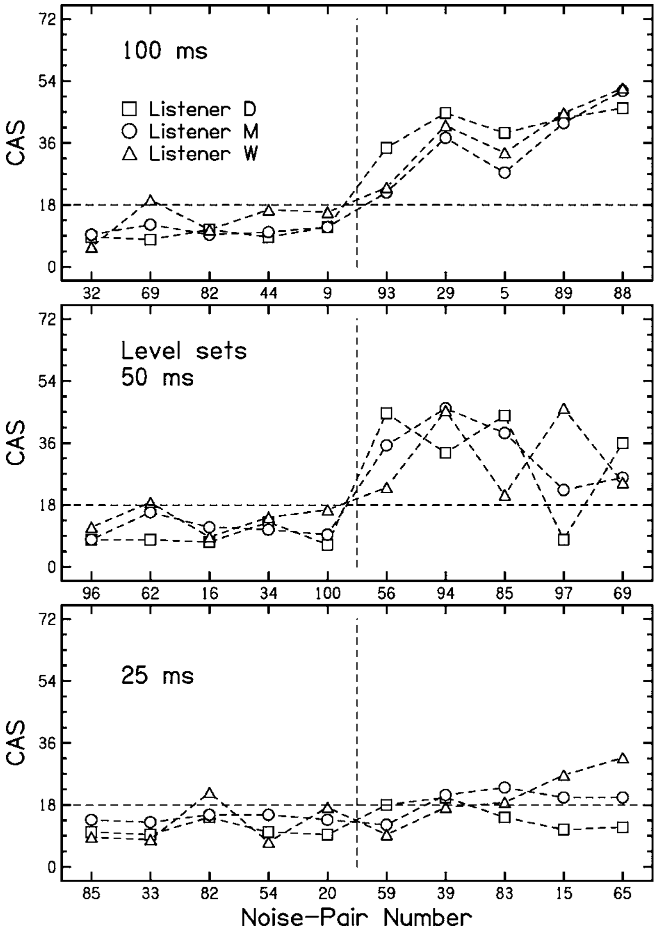

FIG. 5. Same as Fig. 4 except that the CAS values are plotted instead of the percentage of correct responses.

D. Results

The average value of P_c over phase and level sets over all three listeners was: 68% for a duration of 100 ms, 65% for 50 ms, and 61% for 25 ms. These values of P_c can be compared to the value from article I for a duration of 500 ms as obtained for the same three listeners. That value was $P_c = 92\%$, notably larger.

Figures 2–5 show the P_c and CAS values for the phase and level sets for the 100-, 50-, and 25-ms noises. The vertical dashed line shows the division between five noises with the smallest fluctuations and the five noises with the largest fluctuations. Thus, the vertical line represents 90 unused noises. The horizontal dashed line shows the value corresponding to guessing.

In Figs. 2–5, the 100- and 50-ms conditions show higher values of P_c and CAS for the five noises to the right of the vertical dashed line compared to the five noises to the left. These noises with the largest fluctuations have values of P_c and CAS that are near the ceiling ($P_c = 100\%$ or CAS=72). This is different from the noises with a 25-ms duration, which had few noises with P_c values greater than 75% and no noises with CAS values greater than 36.

As in article I, t-tests were performed to determine whether detection of incoherence more frequently occurs in the five noises with the largest fluctuations in phase or level when compared to the five noises with the smallest fluctuations in phase or level. The p values from these t-tests are in Table I. For the 100- and 50-ms sets all 12 p values were significant at the 0.01 level for P_c, as were 11 of 12 for CAS. The significance of these tests is a strong indication that the hypothesis that began this experimental section does not apply for durations of 100 and 50 ms: For these durations, the coherence statistic is inadequate to predict detection performance. Instead, detection performance depends on interaural fluctuations, contrary to the hypothesis that detection is mediated by the interaural coherence of the short stimulus as a whole.

For the 25-ms sets, three of six p values were significant at the 0.05 level for P_c and also for CAS. Clearly, the number of significant p values and the levels of significance for the 25-ms sets did not match those for 100 and 50 ms.

E. Discussion: 50 and 100 ms

The results for noise durations of 50 and 100 ms differ so greatly from those at 25 ms that they are separately discussed.

TABLE I. The p values for the phase and level sets with a nominal bandwidth of 14 Hz and four durations: 25, 50, 100, and 500 ms. The 500-ms p values are taken from article I. Small p values indicate that incoherence is significantly more detectable for the five noises with large fluctuations than for the five noises with small fluctuations.

P_c	25 ms		50 ms		100 ms		500 ms	
Listener	Phase	Level	Phase	Level	Phase	Level	Phase	Level
D	0.029	0.039	<0.001	0.008	<0.001	<0.001	0.095	0.006
M	0.334	0.027	<0.001	<0.001	<0.001	<0.001	0.021	0.047
W	0.457	0.056	0.001	<0.001	<0.001	0.003	0.015	0.014

CAS	25 ms		50 ms		100 ms		500 ms	
Listener	Phase	Level	Phase	Level	Phase	Level	Phase	Level
D	0.027	0.039	<0.001	0.010	<0.001	<0.001	<0.001	<0.001
M	0.306	0.023	0.002	0.002	<0.001	0.004	0.002	0.002
W	0.457	0.065	0.002	0.015	<0.001	0.002	0.001	<0.001

1. Comparison with long-duration stimuli

The tests of significance for 100- and 50-ms durations can be compared with those from article I where noise durations were 500 ms. The CAS data obtained here for 100-ms noises show the same number of significant p values and levels of significance as the 500-ms CAS data from article I. However, compared to the 500-ms P_c data, the 100-ms P_c data show a larger number of significant p values. The reason for the difference in statistics is that when the duration was 500 ms, the P_c ran into a ceiling, which limited its statistical usefulness. Then the CAS measure became essential. However, as noted earlier, average P_c values were normally smaller for durations of 100 ms and shorter as studied here. Therefore, the P_c became statistically useful, and the CAS measure became less necessary. Nevertheless, the CAS does reveal more subtle features. For instance, for 100-ms duration, Figs. 2 and 4 show that incoherence could be detected with approximately equal success in both phase sets and level sets, but Figs. 3 and 5 show that listeners were more confident about their answers in the phase sets.

2. Responses worse than chance

Figures 2 and 4 for the percentage of correct responses show that many of the 100- and 50-ms noises with small fluctuations led to responses below the random guessing limit. For each duration there are 30 data points for small-fluctuation noises (3 listeners × 2 sets × 5 small-fluctuation noises). Both for the 50-ms noises and for the 100-ms noises, 23 of these led to P_c less than 50%. By contrast, in the comparable experiments of article I, with 500-ms noises and 14-Hz bandwidth, only one value of percent correct was less than guessing out of 40 points (4 listeners × 2 sets × 5 small-fluctuation noises).

The reason for this dramatic drop in P_c for the short durations is probably the single-channel envelope fluctuations, as were studied in Experiment 5 of article I. Physically, it is found that noises that have large interaural fluctuations also tend to have large envelope fluctuations. The envelope fluctuations, of course, are prominent even when the noise is heard diotically. In an experiment where it is very hard to hear the interaural fluctuations, as it was for the five selected small-fluctuation noises, a listener may choose the interval with large envelope fluctuation as the interval with the most "action." Apparently, listeners are more apt to confuse envelope fluctuations with interaural fluctuations for the durations of 100 and 50 ms than for a duration of 500 ms. Visual inspection of the envelopes for the small fluctuation noises reveals envelopes that are particularly flat.

Quantitatively, the performance worse than chance can be understood from the following logic: In the limit that noises with large interaural fluctuations always have large single-channel envelope fluctuations, and noises with large single-channel fluctuations are always chosen over noises with small interaural fluctuations in our task, values of P_c should be approximately 25% for the small-interaural-fluctuation noises. The reason is that, on average, dichotic small-interaural-fluctuation noises will be presented against diotic large-fluctuation noises half of the time. This, in turn, would reduce the P_c values of these noises from guessing level, which is 50%, to half of guessing level, which is 25%. As can be seen in Figs. 2 and 4, many small interaural fluctuation noises with a duration of 100 or 50 ms lead to P_c scores near 25%. Similarly, if listeners are misled, as hypothesized, by envelope fluctuations but are never confident about their decisions, this limit predicts a CAS of 9, half the guessing limit. Figures 3 and 5 show a number of scores as low as 9 for durations of 100 and 50 ms.

F. Discussion: 25 ms

The results for noises of 25-ms duration were unlike the results for any other duration because the fluctuations were few in number. Interaural fluctuations of a 14-Hz bandwidth noise are expected to vary with a time scale of $1/14 = 71$ ms. Therefore, it is expected that the interaural phase difference (IPD) and interaural level difference (ILD) would not vary much over a stimulus duration of 25 ms. The 25-ms noises used experimentally had bandwidths larger than this—about 74 Hz ($1/74 = 14$ ms). Nevertheless, fluctuations in the 25-ms noises were rare. Fluctuations can be roughly quantified by examining the number of changes in sign of the interaural phase and interaural level during the course of the stimulus. Out of 20 noises used in the phase and level sets, 3 were common to both sets. For the 17 different noises, 10 had no change in sign for either interaural level or phase, 4 had one change in sign, 2 had two, and 1 had three. This can be compared to the longer duration noises. For the 13 different 50-ms noises, 5 had no change in sign, 3 had one change in sign, 2 had two, and 3 had three or more. For the 15 different 100-ms noises, 3 had no change in sign, 2 had one change in sign, 4 had two, and 6 had three or more.

Noises that were 25 ms in duration led to only 11 values of P_c and CAS below chance level. It will be argued below that these stimuli, with this duration-bandwidth product, are likely to be too short to elicit perceptible fluctuations, monaural or binaural.

1. The laterality cue

All three listeners reported that when the duration was reduced to 25 ms, they sometimes used a laterality cue because there seemed to be little or no perceived width for most of the noises. This change in detection strategy is a likely explanation for the near-chance values of P_c and CAS and for the negative results of the t-tests for the 25-ms sets. To generate the stimulus sets, noises were selected by the s_t statistic, which is a fluctuation statistic associated with a width cue. As a standard deviation, this statistic is unaffected by a constant shift in mean value, which is associated with lateralization. When listeners used a laterality cue on sets that had been sorted by a statistic associated with a width cue, *p* values did not show as many significant differences between the noises.

2. Lateralization experiment

An auxiliary experiment was performed to test the idea that laterality is a salient cue for the 25-ms noises, as suggested earlier. Each of the ten noises from the 25-ms phase set and each of the ten from the 25-ms level set was alternated repeatedly with a diotic noise. The listener's task was to say whether the noise was to the left or right of the diotic standard. Results were compared with predicted values based on a sum of the compressed IPD and ILD lateralization according to a formula fitted to the data of Yost (1981). Out of 20 noises, the responses agreed with prediction on 17 for Listener M, on all 20 for Listener W, but only on 12 for Listener D. The agreement is an indication that laterality is salient, at least for some listeners. It also indicates that lateralization as computed from Yost's sine-tone data can be applied to brief noise bursts. This observation was previously made in connection with the modeling of incoherence detection (Goupell, 2005).

3. Modeling for 25 ms

Models for the detection of incoherence begin with the physical stimulus, and, to a greater or lesser extent, take into account known facts about the human auditory system. Such models are applied to the detection of interaural fluctuations in a future article in this series, where the predictions are compared with the results from experiments using long-duration noises. The present section is dedicated to simple, stimulus-based models for the shortest-duration noises, 25 ms. The models were tested only on the P_c data because confident responses were infrequent for this duration. The results of the tests are shown in Table II.

TABLE II. Linear correlation coefficients, *r*, for 25-ms sets with 14-Hz nominal bandwidth. The data were averaged over listeners. Several different laterality (mean, μ) and fluctuation (standard deviation, s_t) statistics were considered. The use of laterality compression allows IPD and ILD measures to be combined.

Statistic	No Lat. Comp.	Lat. Comp.
$\mu(\Delta\Phi)$	0.25	0.31
$\mu(\Delta L)$	0.35	0.32
$\max(\lvert\mu(\Delta\Phi)\rvert,\lvert\mu(\Delta L)\rvert)$	⋯	0.51
$\mu(\Delta\Phi)+\mu(\Delta L)$	⋯	0.48
$\lvert\mu(\Delta\Phi)\rvert+\lvert\mu(\Delta L)\rvert$	⋯	0.54
$s_t(\Delta\Phi)$	0.58	0.59
$s_t(\Delta L)$	0.54	0.61
$s_t(\Delta\Phi)+s_t(\Delta L)$	⋯	0.63

If $s_t(\Delta\Phi)+s_t(\Delta L)<1$ and $P_c>56\%$ then
$\lvert\mu(\Delta\Phi)\rvert+\lvert\mu(\Delta L)\rvert$
else
$s_t(\Delta\Phi)+s_t(\Delta L)$
	⋯	0.72

Running short-term cross-correlation	0.38	⋯

The first two models assume that incoherence is detected on the basis of displacements from the midline, the average value of the IPD alone, $\mu(\Delta\Phi)$, or the average value of ILD alone, $\mu(\Delta L)$. Table II shows that the correlation with the experimental values of P_c is only about 0.3. It is possible to combine IPD and ILD displacements from the midline if both are put on the same scale of laterality using the compression function that fits Yost's 1981 data. Then choosing the maximum of the two displacements, or the sum of the displacements, or the sum of the absolute values of the displacements all lead to a correlation *r* of about 0.5.

A second set of models ignores average displacements from the midline and assumes that incoherence detection is based on fluctuations alone: phase fluctuations, $s_t(\Delta\Phi)$, level fluctuations, $s_t(\Delta L)$, or a combination. As shown in Table II, these models correlate with P_c data with *r* about 0.6.

Because both laterality models and fluctuation models were positively correlated with the data, there is an indication that a combination of the two models would make predictions that agree even better with the data. It was found that combining fluctuations and laterality conditional upon the P_c score could achieve a value of $r=0.72$. This conditional model says that if the fluctuations are small and yet performance is somewhat better than chance $(P_c>56\%)$ then the model uses the sum of phase and level laterality magnitudes as the decision variable. Otherwise, it uses the sum of the phase and level fluctuations. We suspect that the reason that the correlation between model and data is not higher is that the treatment of the fluctuations omits important facts. These facts are included in the models of a future article.

A final calculation for the 25-ms data assumed that in-

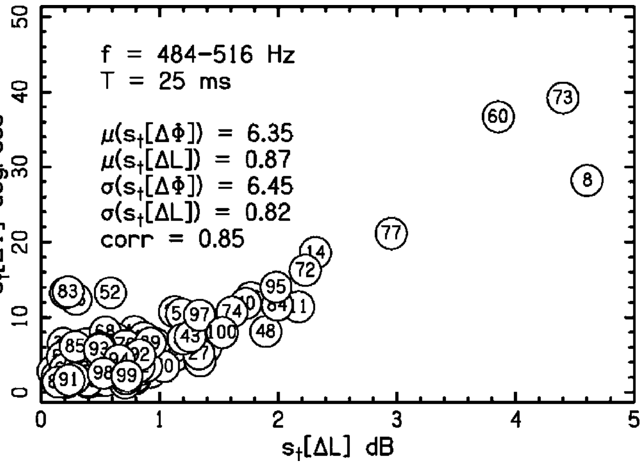

FIG. 6. Standard deviations in interaural phase and level showing the fluctuations for the 100 noises of Experiment 2, with nominal bandwidth of 28 Hz and duration of 25 ms. The scale is changed compared to Fig. 1. The ensemble-average fluctuations (μ) resemble those in Fig. 1 for 14-Hz nominal bandwidth and duration of 50 ms.

coherence is detected on the basis of the running short-term cross-correlation function. The cross-correlation was computed over a rectangular window, as long as 10 ms, for every instant in time. Then the average of this cross-correlation over the duration of the stimulus was used as a statistic to compare with P_c data. This model correlated with the data with $r=0.38$—better than the laterality models but not as successful as the fluctuations models. When the window was increased to 20 ms, the r value remained about the same.

III. EXPERIMENT 2

Experiment 1 found that the detection of incoherence in noises with a nominal bandwidth of 14 Hz depends on interaural fluctuations for durations of 50 ms and longer. Little fluctuation dependence was observed for 25 ms, but it seemed likely that the reason was that interaural fluctuations were too infrequent and too small for this duration, with mean values less than half of those at 100 ms, as shown in the legends of Fig. 1. An alternative possibility is that a duration of 25 ms is simply too short to perceive fluctuations. In order to decide between these two ideas, Experiment 2 again used 25-ms noises but doubled the nominal bandwidth to 28 Hz. The larger bandwidth was expected to lead to faster interaural fluctuations, which would then lead to more fluctuations within the brief duration of the stimulus. If interaural incoherence detection scales perfectly, one would expect that 25-ms noises with a nominal bandwidth of 28 Hz would be comparable to 50-ms noises with a nominal bandwidth of 14 Hz, as tested in Experiment 1.

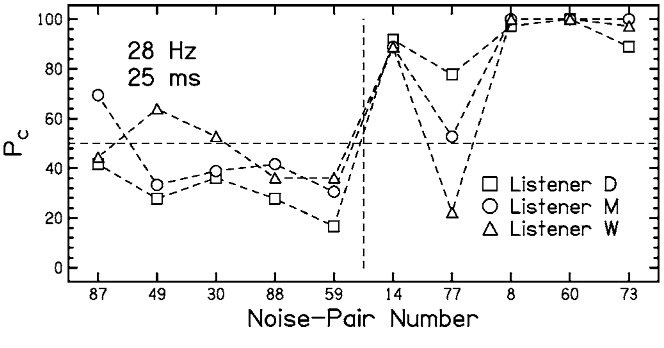

FIG. 7. Percent correct responses as in Figs. 2 and 4 except that the data are from Experiment 2, where the duration is 25 ms and the nominal bandwidth is 28 Hz. Phase sets and level sets are essentially the same for this experiment.

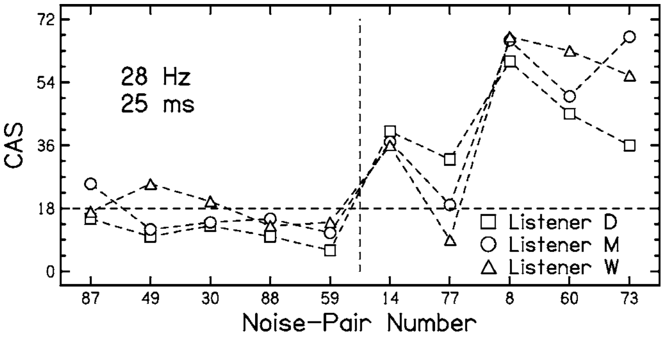

FIG. 8. CAS as in Figs. 3 and 5 except that the data are from Experiment 2, where the duration is 25 ms and the nominal bandwidth is 28 Hz. Phase sets and level sets are essentially the same for this experiment.

A. Method

The noises for Experiment 2 were generated in the same way as for Experiment 1 except for the bandwidth. Noises having a nominal 28-Hz bandwidth per the 3-dB down points were made by using 17 components with finite amplitude from 484 through 516 Hz. The amplitudes of the two components with the lowest frequencies and the two with the highest were shaped as for Experiment 1.

When it came to selecting noises for phase sets and level sets with the five largest and smallest values of interaural fluctuation, it was discovered that the five noises with the largest phase fluctuations also had the largest level fluctuations. Consequently, we thought it adequate to do the experiment with a single set of ten stimuli, presenting the five largest and smallest interaural phase fluctuations from the 100 noises shown in Fig. 6. For the 10 independent noises, 3 had no change in sign for either interaural level or phase, 4 had one change in sign, 1 had two, and 2 had three. Thus, the average number of sign changes per noise was 1.2, compared to 0.64 for the narrower band used in Experiment 1. The spectral distributions were computed for the ten noises of Experiment 2. On average, 90% of the energy was in a band 80 (±10) Hz wide, and 99% was in a band 130 (±15) Hz wide.

B. Results

The percentage of correct responses is shown in Fig. 7, which can be compared with Fig. 2. The CAS values are shown in Fig. 8, and these can be compared with Fig. 3. Both comparisons show that the performance for 28-Hz nominal bandwidth and 25-ms duration most closely resembles the performance for 14-Hz nominal bandwidth and 50-ms duration. One-tailed t-tests of the hypothesis that the percentage of correct responses is higher for the five noises on the right were significant at levels 0.001, 0.002, and 0.040, for Listeners D, M, and W, respectively. Similar tests for the CAS values were significant at levels 0.001, 0.011, and 0.028, for Listeners D, M, and W, respectively. These values can be compared with those for either phase or level sets in Table I.

C. Discussion

As noted in the legends of Fig. 1, the fluctuations for 25-ms, 14-Hz noise are about half of those for 50-ms, 14-Hz noise. However, as shown in the legend of Fig. 6, when the nominal bandwidth is doubled in Experiment 2 the fluctuations for 25-ms, 28-Hz noise are about the same as those for 50-ms, 14-Hz noise. *A priori* one might expect the detection of incoherence for these two noises to be about the same as well. However, although the difference in detectability between large fluctuation noises and small fluctuation noises is clearly significant for 25-ms, 28-Hz noise, it is not quite as significant as for 50-ms, 14-Hz noise. Evidently, the time-bandwidth equality does not lead to identical performance. The reason may be that, by our energetic measures, the physical bandwidths of 25-ms noises with a 28-Hz nominal bandwidth are not exactly twice those for 14-Hz nominal bandwidth.

The most important conclusion from Experiment 2 is that listeners have no difficulty perceiving interaural fluctuations in noises that are 25 ms in duration if the fluctuations are physically present. Listeners can use these fluctuations to recognize interaural incoherence.

IV. CONCLUSION

This article began with a challenge to a previous work, article I (Goupell and Hartmann, 2006). That article had concluded that for 500-ms noises, with bandwidths equal to a critical band or narrower, the detection of interaural incoherence depends on interaural fluctuations in phase and level. Article I specifically denied that the detection could be predicted from the cross-correlation of the entire stimulus. That conclusion was based on experiments using noises, which all had the same interaural coherence. It was found that those noises with larger fluctuations were more readily identified as incoherent. The challenge noted that the stimulus duration of 500 ms is far longer than binaural integration times. The challenge continued by suggesting that coherence measured over 500 ms was possibly perceptually meaningless because the binaural system would likely base its decisions on coherence measured over much smaller time spans.

The present article addressed the challenge by experiments that were the same as those in article I except that the stimuli had durations of 100, 50, and 25 ms—intended to match probable binaural integration times. The coherence measured over those short durations for each of the stimuli was again 0.992. The bandwidth was again nominally 14 Hz, as in article I.

The experiments found that for durations of 100 and 50 ms, incoherence in those noises with large fluctuations was much more readily detected than incoherence in those noises with small fluctuations, even though coherence values were the same. Statistically, the results were even stronger than were found in article I. Consequently the conclusions of article I were upheld for these durations. Coherence, i.e. the peak of the cross-correlation function, is inadequate to predict incoherence detection.

For a duration of 25 ms, it was discovered that there were sometimes too few physical fluctuations to be perceptually useful when the bandwidth was 14 Hz. The small fluctuations for 25-ms, 14-Hz noises make this stimulus unique among our stimuli. The perceived difference between these noises and diotic signals is that these noises are sometimes lateralized. By contrast, all the other stimuli explored in this article lead to salient *fluctuations* in interaural differences. These subjective differences are our explanation for two facts about the 25-ms, 14-Hz noises: (1) Larger fluctuation noises are not better identified as incoherent than are smaller fluctuation noises. (2) Performance, as measured by percent correct or CAS values, is always near chance.

If stimuli are selected on the basis of fluctuations but the most salient cue is laterality, then detection performance differences are unlikely to reflect the selection criterion. The smaller physical range of fluctuations for 25 ms is likely to translate into a smaller perceptual difference between large and small fluctuations. In connection with the second fact, it appears that the lateralization cues that are available for brief stimuli are less salient than the fluctuation cues available for longer stimuli. Consequently, performance suffers.

We suspect that this distinction between laterality and fluctuation is also seen in forward (and backward) fringe experiments in NoSπ detection (McFadden, 1966). A homophasic (No) fringe is especially helpful when the signal is brief (Robinson and Trahiotis, 1972). As noted by Yost (1985), who used a 20-ms signal, a homophasic fringe (No) provides a reference such that an $S\pi$ signal leads to a variation in lateralization. The variation leads to improved detection. A similar observation was made by Gilkey *et al.* (1990), who referred to the sudden change in interaural parameters as an "onset-effect." By contrast, when the signal duration is long, the interaural fluctuations introduced by the out-of-phase signal produce temporal variations of their own, and a fringe becomes less valuable.

To try to understand the incoherence detection data for 25-ms, 14-Hz noises we tested some simple stimulus-based models for detection. Some of the models were based on laterality, defined as an average lateralization, others were based on fluctuation. The most successful was a model that depended on both, depending on circumstances, consistent with the idea that for this (most awkward) condition some noises have more prominent laterality while others have more prominent fluctuations. However, none of the models were highly successful.

Our conclusion that the anomalous results of the experiment with 25-ms, 14-Hz bandwidth noise were due to a lack of fluctuations was validated by a second experiment. This experiment too used 25-ms noises, but with a doubled bandwidth, effectively doubling the number of fluctuations presented to the listener. When the bandwidth was doubled, significant differences in fluctuations appeared amongst different stimuli, and significant differences in detectability of incoherence reappeared in the P_c data.

Assuming, as we do, that the short-term analysis capabilities of the binaural system can be correctly explored by using stimuli with short durations, it follows that there is no sense in which the short-term cross-correlation of the stimulus is, by itself, adequate to predict detection. Instead, the cross-correlation statistic provides a guide to the distribution of interaural fluctuations that can be expected in an ensemble of noises. It is inadequate to predict the fluctuations for any particular noise, and it is these fluctuations that are at the basis of incoherence detection.

The significance of these results for binaural modeling is to clarify the role of the binaural cross-correlator. A model that ascribes incoherence detection to a reduction of stimulus cross-correlation as measured at zero lag (midline) is inconsistent with the results of our research. A model that examines the output of the cross-correlator off the midline, where momentary fluctuations occur, is consistent. In addition, the fluctuations in interaural phase measured by the cross-correlator must be supplemented with the fluctuations in interaural level in order to make a successful detector of incoherence.

Although the short-duration experiments presented here seem to have dealt with the challenge to article I in a satisfactory way, this article together with article I have only established the importance of *stimulus* interaural fluctuations in comparison with the cross-correlation function. This article has not established what properties of these fluctuations, or what transformations, or what combinations of IPD and ILD fluctuations are used by the auditory system in detection. This article has also not compared the fluctuation model with the performances of a model *neural* cross-correlator. A future work will develop a signal-based auditory model that attempts to treat fluctuation detection in a realistic way and will make the necessary comparisons.

ACKNOWLEDGMENTS

This work was partially supported by the NIDCD under Grant No. DC 00181, and partially supported by a dissertation completion fellowship to M.J.G. from the College of Natural Science of Michigan State University. Dr. A. Kohlrausch, Dr. C. Trahiotis, and Dr. H. S. Colburn made useful comments on an earlier version of this article.

---

[1] To determine whether the values of fluctuations in the legends of Fig. 1 were representative of a large population, 30 000 noises were computed for each duration, and interaural fluctuations were calculated. The comparison between the legend and the large-population values showed that means, standard deviations, and correlations for 100- and 500-ms durations differed by less than 10%. For 50- and 25- ms durations, the discrepancy was less than 15% except that the standard deviation of phase fluctuations for 50 ms was 5.63° in the large ensemble instead of 6.83, and the standard deviation of level fluctuations for 25 ms was 0.50 dB in the large ensemble instead of 0.36. Also for 25 ms, the large-ensemble correlation was 0.69 instead of 0.53.

Gilkey, R. H., Simpson, B. D., and Weisenberger, J. M. (**1990**). “Masker fringe and binaural detection,” J. Acoust. Soc. Am. **88**, 1323–1332.

Goupell, M. J. (2005). “The use of interaural parameters during incoherence detection in reproducible noise,” Ph.D. dissertation, Michigan State University, East Lansing, MI.

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.

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.

McFadden, D. (**1966**). “Masking level differences with continuous and with burst masking noise,” J. Acoust. Soc. Am. **40**, 1414–1419.

Robinson, D., and Trahiotis, C. (**1972**). “Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions,” Percept. Psychophys. **12**, 335–338.

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. (**1985**). “Prior stimulation and the masking level difference,” J. Acoust. Soc. Am. **78**, 901–907.