Localizing the sources of two independent noises: Role of time varying amplitude differences

William A. Yost[a)]
Spatial Hearing Laboratory, Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

Christopher A. Brown
Department of Communication Sciences and Disorders, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 25 January 2012; revised 7 January 2013; accepted 24 January 2013)

Listeners localized the free-field sources of either one or two simultaneous and independently generated noise bursts. Listeners' localization performance was better when localizing one rather than two sound sources. With two sound sources, localization performance was better when the listener was provided prior information about the location of one of them. Listeners also localized two simultaneous noise bursts that had sinusoidal amplitude modulation (AM) applied, in which the modulation envelope was in-phase across the two source locations or was 180° out-of-phase. The AM was employed to investigate a hypothesis as to what process listeners might use to localize multiple sound sources. The results supported the hypothesis that localization of two sound sources might be based on temporal-spectral regions of the combined waveform in which the sound from one source was more intense than that from the other source. The interaural information extracted from such temporal-spectral regions might provide reliable estimates of the sound source location that produced the more intense sound in that temporal-spectral region.
 [http://dx.doi.org/10.1121/1.4792155]

PACS number(s): 43.66.Qp [RYL] Pages: 2301–2313

I. INTRODUCTION

Although a great deal is known about the ability of listeners to locate a single sound source (see Blauert, 1997, for a thorough review of the spatial hearing literature), far less is known about listeners' abilities in localizing two simultaneous sound sources. This paper deals with human listener's localization performance when two sound sources in the azimuth plane in the free field produce simultaneous sound.

In order for the auditory system to determine the location of the sources of two simultaneous sounds, the sounds from the sources must differ in some way. Two identical sounds presented from two different azimuth locations interact acoustically to produce the perception of a single source (a phantom source) located midway between the two originating sound sources (see Bauer, 1961). Even when the sound at one source location occurs after that of an identical sound from a different source location, only a single sound source is perceived under many conditions (the effects of precedence; see Litovsky *et al*., 1999).

In this study we used independently generated wideband noise bursts (at least 200 ms in duration) that were presented at the same time from two loudspeakers located at different azimuth angles in the free-field. The noise bursts were identical in all other respects. Two independently generated noise bursts cannot be easily distinguished one from the other, in fact, it is difficult to make same-different discrimination judgments between two independently generated noise bursts that are 200 ms or longer (for instance, see Hanna, 1984). The stimuli in the present study are not ones that allow experience with the sounds to aid in identification (i.e., as might happen if something like speech were used). The goal was to make the task as difficult as possible, with the rationale that it could be made easier if needed by making the stimuli different along another dimensions.

Conditions in which independently generated noises are presented simultaneously from different source locations are often referred to as producing a "diffuse" sound field or the perception of a diffuse sound, i.e., the perception of a sound that has a broad spatial extent (see Gardner, 1969; Santala and Pulkki, 2011). Recently Santala and Pulkki (2011) showed that listeners can locate up to five different loudspeaker locations when independently generated noise is presented from each loudspeaker. While detailed measures of localization acuity were not obtained in that study, their data and that from other studies cited in their paper suggest that listeners can localize two sound sources producing independent samples of noise at the same time. Best *et al*. (2004) showed in a virtual anechoic environment that listeners could determine if there were one or two sources when the sounds were concurrently generated independent noise bursts located in a "virtual" azimuth plane. These authors did not have listeners determine the location of the sound sources. The major purpose of the current paper is to document listeners' performance in localizing two sound sources in the azimuth plane producing simultaneously and independently generated noise bursts and to test a hypothesis of how the auditory system might accomplish locating multiple sound sources.

[a)]Author to whom correspondence should be addressed. Electronic mail: william.yost@asu.edu

Most sound source localization studies that have used multiple sound sources (see for instance Braasch and Hartung, 2002; Croghan and Grantham, 2010; Erno *et al.*, 2001; Good and Gilkey, 1996; Good *et al.*, 1997; Hawley *et al.*, 1999; Kopčo *et al.*, 2007; Lee *et al.*, 2009) have investigated the localization of one sound source in the presence of one or more additional sound sources. These studies indicate that performance for locating a target sound source in the presence of competing sound sources is poorer than when the task is to locate just the target sound source. The amount of the sound source localization performance decrement varies from study to study, most likely due to the fairly large differences in stimuli and sound source location procedures used across studies. There are very few data indicating listener's performance in locating more than just one sound source when multiple sources produce simultaneous sounds, especially if the sounds are independently generated noise bursts. This is the aim of this study.

The stimulus conditions studied in this paper are similar to those used to study spatial release from masking. Spatial release from masking is the reduction in masking that occurs when a target sound is at a different location from a masking sound source as compared to conditions in which the target and masking sounds originate from the same source. In spatial release from masking studies there are two spatially separated sound sources (target and masker) producing simultaneous sounds as is the case for the conditions of the present study. It is usually assumed that some aspect of processing interaural cues (interaural time differences, ITDs, and/or interaural level differences, ILDs) is responsible for spatial release from masking. These are the same cues that one would use to locate azimuthal sound sources in multisource situations. Thus, better understanding sound source localization for multiple sources might provide useful information about spatial release from masking.

In spatial release from masking studies the listener detects, discriminates a difference in, or recognizes/identifies the target sound in the presence of distractor/masking sound sources. The target and masker(s)/distractor(s) stimuli usually differ significantly (e.g., target is a sentence and masker is a speech-shaped noise). The task in this study is sound source location identification and the stimuli are very similar (in some sense as similar as possible, but yet being acoustically different). Thus, caution is warranted in generalizing from the conditions of this study to those used in spatial release from masking studies. The focus of this paper is on sound source localization performance, no aspect of signal detection, discrimination, and/or sound identification was measured in this study.

In spatial release from masking studies, listeners are usually asked to make a response regarding the target sound source, and they do not generally make a response regarding the masker(s). In a sound source localization task the listener could be asked to indicate the source of a target sound in the presence of a distractor sound (similar to spatial release from masking studies and to most multisource localization experiments reported in the literature) or the listener could be asked to indicate the location of both sound sources when two sources produce sound. Both conditions were tested in the present study.

In experiment I listeners were asked to localize in the azimuth plane either a single sound source, a sound source in the presence of another source at a known location, or two sound sources. In the last condition, listeners had no prior knowledge about the location of either sound source, and they were to determine the location of both sound sources.

II. EXPERIMENT I

A. Listening environment

Experiments were conducted in an echo-reduced listening room, 11 ft × 12 ft., lined with 4 in. acoustic foam (Noise Reduction Coefficient-NRC = 0.9) on all six surfaces along with special sound treatment on the floor and ceiling. The room contained a 13-loudspeaker (Boston Acoustics 110 x) array arranged in an arc in the front hemifield 1.67 m away from the listening position, and at the height of the listeners' pinna while seated. Loudspeakers were positioned from −90° to +90° with 15° between each. Loudspeakers 1 and 13 did not produce any sound, but the listeners were not told this. Thus, the loudspeakers that presented sound span the range from −75° to +75°, in 15° spatial separations.

A small control room adjacent to this room contains the control computer, Echo Gina 12-channel DA/AD converters, amplifiers and attenuators for 11 of the 13 loudspeakers, and video monitoring of the subject. Speaker calibration was made at the location of the subject in the room. All loudspeakers were within +−8 dB across 100 to 15 000 Hz across all 11 loudspeakers. Additional digital equalization, done on-line for all experiments, reduced the variation to +− 2 dB across all frequencies and loudspeakers. Reverberation times (RT_{60}) were determined for each of 11 loudspeakers at the location of the subject. Broadband noise bursts (500 ms) and 1-ms transients were used to determine RT_{60}. Broadband RT_{60} ranged from 90 to 122 ms across the 11 loudspeakers and the two measurement signals. On average RT_{60} was 97 ms for the noise and 101 ms for the transient. In an octave band centered at 1000 Hz, RT_{60} for the noise was on average 324 ms, while for an octave band centered at 4000 Hz average RT_{60} was 56 ms.

B. Subjects

Eight listeners who reported having normal hearing, five females and three males all under the age of 30 years, served as listeners. All procedures used in this study were approved by the Arizona State University Institutional Review Board (IRB) for the protection of human subjects.

C. Stimuli

All stimuli were generated in MATLAB and presented to the 12-channel Echo Gina DA system at 44 100 Hz per DA channel. Noise bursts were 200 ms in duration, with 20-ms cosine-squared rise/decay times, bandpassed filtered between 125 and 6000 Hz with an 8-pole (∼48 dB/octave) Butterworth filter, and presented at 65 dBA (measured at the position of the listener with a Type 1 sound level meter using the slow setting) with a +−2-dB random level rove over loudspeakers and presentations (the 4-dB level rove was to

deal with any cues that might be associated with the slight level differences across loudspeakers). A particular noise burst was never presented more than once. Noise bursts were independently generated across trials, across presentations within a trial, and across loudspeakers within a presentation.

D. Procedure

1. General

Listeners were instructed to face straight ahead and look at a red dot fixed to the center loudspeaker (7) at the start of each stimulus presentation, and were monitored via closed-circuit video to ensure compliance. Listeners pressed keys on a computer keyboard to initiate stimulus presentations and make responses. The listeners' task was to identify the loudspeaker or loudspeakers presenting sound, with possible responses in the range of 1 to 13. Each trial consisted of two stimulus presentations, even in conditions containing only one sound source for consistency. No trial-by-trial feedback was provided. One 10-trial practice block was run prior to testing. Listeners were told that when there were two sounds they would always be presented from different loudspeakers.

2. One sound source, one sound source and one source is localized (1S-1L)

After two presentations from the same loudspeaker, listeners indicated the loudspeaker number that corresponded to the perceived sound source location. Then 275 trials were run (25 trials for each of the 11 loudspeaker locations that presented sound) in five, 55-trial blocks.

3. Two sound sources, one source is localized (2S-1L)

Listeners were told that two loudspeakers would present sound at the same time and that one of the sound sources would always be the center (7) loudspeaker. After two presentations, they were to indicate the location of the loudspeaker that presented the other sound (loudspeakers 1 to 13). In addition to loudspeakers 1 and 13 (see above) loudspeakers 6 and 8 (i.e., those immediately adjacent to the center loudspeaker, 7) also did not produce sound, although listeners were not told this. There were 200 trials (25 trials for each of the eight loudspeakers that presented sound; 2, 3, 4, 5, 9, 10, 11, 12), divided into four, 50-trial blocks.

4. Two sound sources, two sources are localized (2S-2L)

For each trial, one of eight combinations of loudspeaker locations (see Fig. 1) was chosen at random, and stimuli were presented twice from this loudspeaker pair. Listeners were instructed to indicate the location of one of one sound after the first presentation, and the other location after the second presentation. There were 200 trials (25 trials for each of the eight loudspeaker pairings) presented in four, 50-trial blocks.

E. Results

The data are plotted as histograms in Figs. 2–4 as the percent of the total trials (across conditions and listeners) in which a perceived loudspeaker location (X axis, 1–13) was indicated (Y axis). Figure 2 shows data for the 1S-1L conditions, Fig. 3 for the 2S-1L conditions, and Fig. 4 for the 2S-2L conditions. Bars with circles indicate correct responses (i.e., the position of the loudspeaker presenting sound). The histogram scale of percent (%) responses is shown on the lower left of the figures. For the 1S-1L condition (Fig. 2) and the 2S-1L condition (Fig. 3) the number of trials was the same as the maximum number of responses since only one loudspeaker location was reported on each trial. However, for the 2S-2L condition (Fig. 4) there were twice as many possible responses as there were trials, as

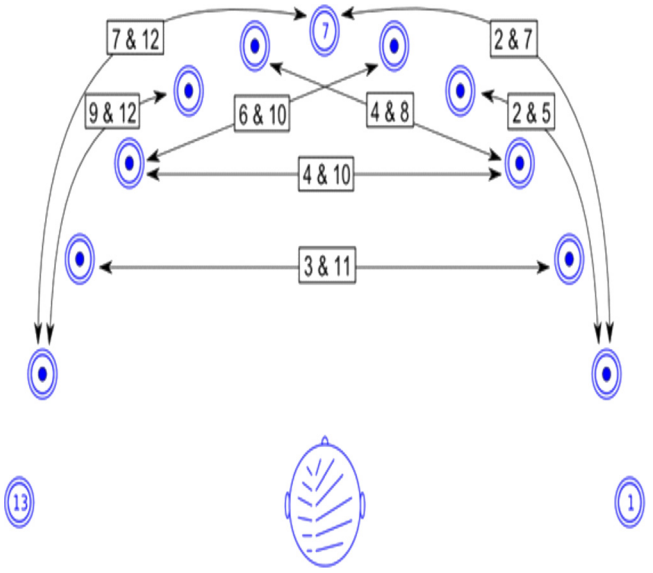

FIG. 1. (Color online) The eight loudspeaker pairs used in experiment I (2S-2L condition) and experiments II and III.

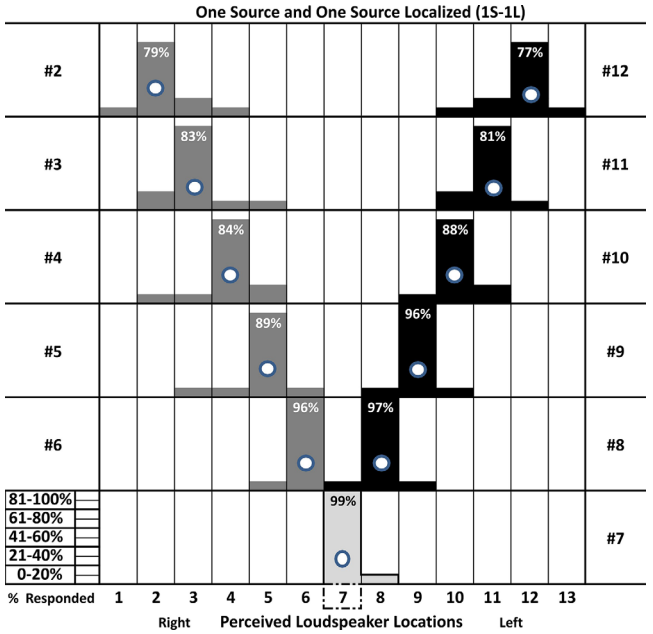

FIG. 2. (Color online) Histogram (percent of responses) of the localization responses across all conditions and listeners for the 13 loudspeaker locations in the 1S-1L condition for each of the actual loudspeaker location. Circles indicate the location of the loudspeaker that presented a sound.

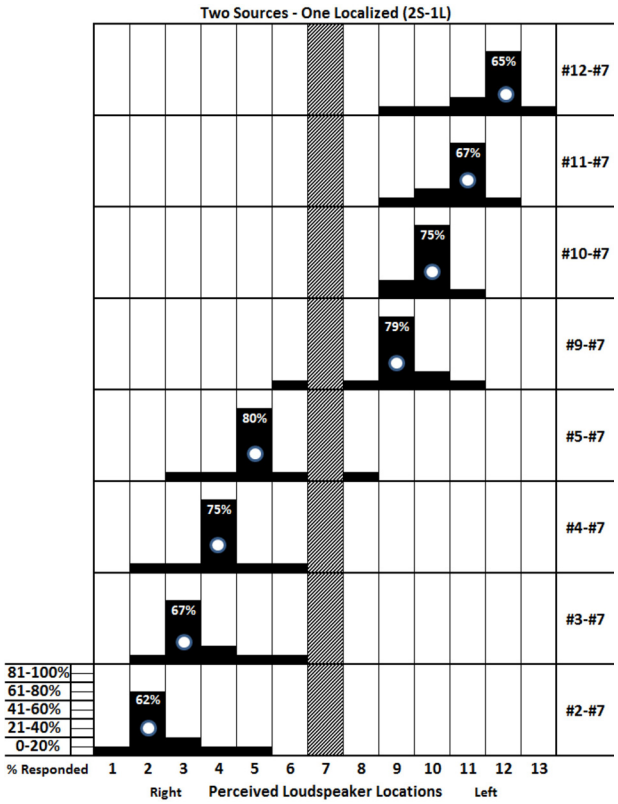

FIG. 3. (Color online) Same as Fig. 2, but for the 2S-1L condition.

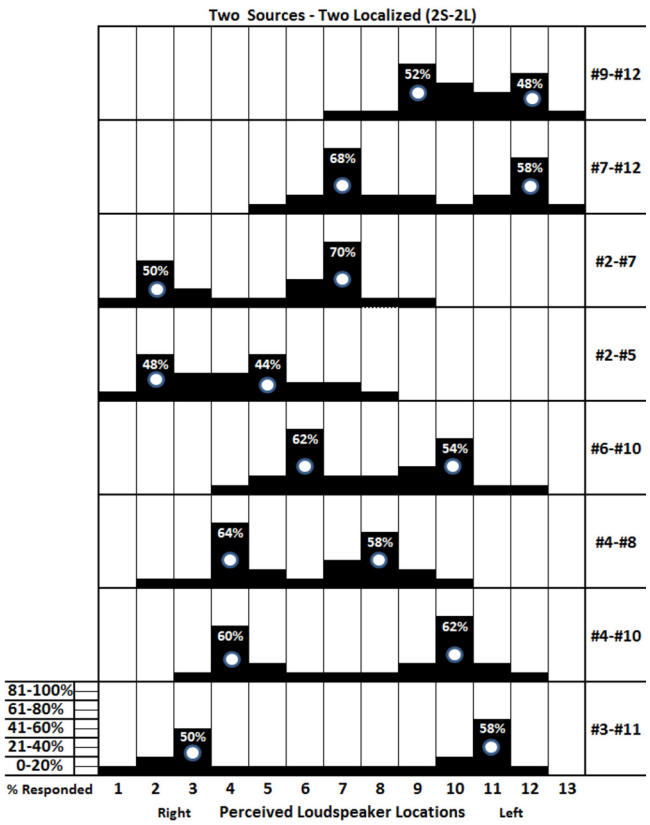

FIG. 4. (Color online) Same as Figs. 2 and 3, but for the 2S-2L condition. The percent (%) responses were calculated by dividing the total number of times a particular perceived location was reported by the total number of trials.

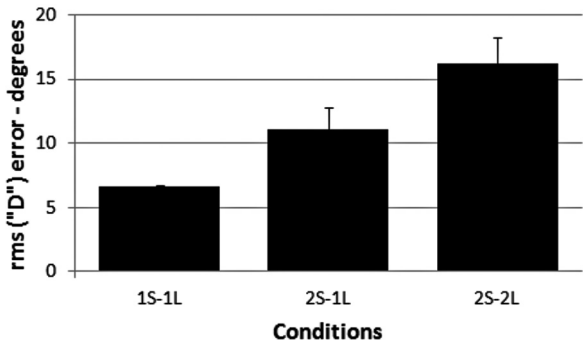

FIG. 5. Mean (across 8 listeners) rms error in degrees as a function of the three conditions of experiment I. Error bars are one standard error of the mean.

there were two responses per trial. In all three conditions (Figs. 2–4) percent (%) responses was calculated and displayed by dividing the total number of responses for any particular perceived location by the number of total trials (not responses) for the particular actual loudspeaker pair. For the 2S-2L condition (Fig. 4) this results in the maximum percent responses for anyone actual loudspeaker pair totaling 200%.

Figure 5 indicates the mean (across subjects and loudspeaker locations) root-mean-square (rms) error (using the "D" calculation of Rakerd and Hartmann, 1986) for the three conditions.[1,2] A one-way repeated measure analysis of variance (ANOVA) with condition as the factor indicated a statistically significant Main Effect [$F(2,5)$; $p \ll 0.01$] and a repeated measures *a priori t* tests indicate that the rms error for 2S-2L was statistically greater than the rms error for 1S-1L ($p \ll 0.01$) and the rms error for 2S-2L was statistically greater than the rms error for 2S-1L ($p < 0.01$). Figure 6 indicates the percent of trials in which listeners correctly located either both loudspeaker locations (both correct) or at least one of the two loudspeaker locations (one correct) in the 2L-2S condition.

F. Discussion

Sound source localization performance for locating a single noise source (1S-1L condition) was similar to that obtained by other investigators using listeners with normal

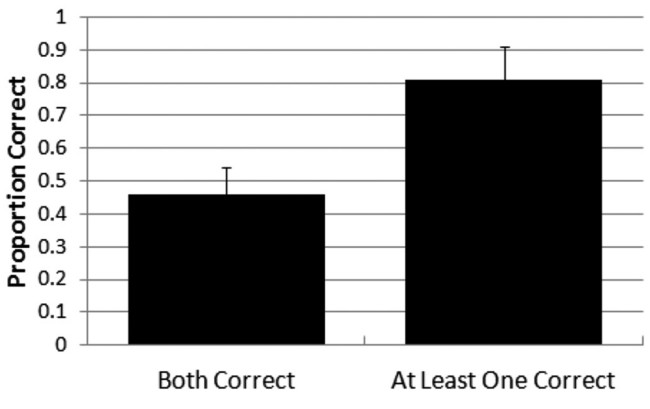

FIG. 6. Mean (across 8 listeners) proportion of correct responses in getting both loudspeaker locations correct (both correct) or at least one loudspeaker location correct (one correct) in the 2S-2L condition. Error bars are one standard error of the mean.

hearing (e.g., Grantham *et al*., 2007; Wightman and Kistler, 1989). The rms errors of approximately 5°–8° and the fact that localization performance is best for locations directly in front of the listener are common findings. All eight listeners performed similarly in the 1S-1L condition.

When two sources presented simultaneous independent noise bursts and the listener knew that one of the sources was the center loudspeaker (2S-1L condition), performance was better than when the listener did not have any prior information about which loudspeaker would be presenting either one of the two sounds (2S-2L condition). In the 2S-1L condition sound source localization performance was worse than when listeners were asked to locate only one sound source (one condition). Thus, it appears that having prior information about the source of one sound when there are two sound sources aids localization of two sound sources. In the 2S-2L condition, on any trial, listeners were able to correctly locate both sources slightly less than half of the time and they located at least one correct sound source slightly more than 80% of the time. These data appear to agree qualitatively with those of Santala and Pulkki (2011) in indicating that listeners can localize the sources of two simultaneous and independent noises in the free field, but not as well as they can localize a single sound source.

In many experiments involving multiple sound sources the sounds from the various sources can be identified making it possible to assign a particular sound to a particular source. For instance, if two words from two different sources were used, the data could be tabulated as the percent perceived loudspeaker location for word one and for word two. This cannot be done in this experiment, since the two sounds (independently generated noise bursts) are barely discriminable (Hanna, 1984), i.e., any one noise burst is not identifiably different from any other noise burst.[2] One of the motivations for using independent noise bursts in this study was to provide for the possibility to evaluate how being able to identify the sound from a source might influence multiple sound source localization. For instance, how does multiple sound source location performance compare for speech versus noise and to what extent are any differences in performance attributable to the ability to identify different speech stimuli but not different noise stimuli?

Given that two simultaneous noise bursts interact acoustically, it might seem surprising that listeners do as well as they do in localizing two independent noise sources. Our results showing that listeners can locate sound sources under these conditions is similar to the results obtained by Santala and Pulkki (2011) in the free-field and by Best *et al*. (2004) in a virtual-listening condition. Several investigators (e.g., Keller and Takahashi, 2005; Meffin and Grothe, 2009; Woodruff and Wang, 2010) have suggested that localization of multiple sound sources might occur because in the combined waveform some proportion of temporal-spectral regions might contain high relative levels of the sound from one of the sources. The interaural differences (ILDs and ITDs) in these temporal-spectral regions may provide reliable information about the interaural differences of the sound from that source. When the level of the sounds from the two sources are about the same within a temporal-spectral region, the interaural cues would not reflect those of either source (the interaural cues would be spurious) in that the interaction of the sound waveforms would obscure the interaural cues associated with the originating sound sources. Perhaps the ability to localize simultaneous sounds from two sources occurs when there are enough temporal-spectral regions in the combined waveform with reliable interaural cues relative to spurious interaural cues. Experiment II was designed to investigate sound source localization when there were differences in level over time from different sound sources.

In experiment II, the two independent noise bursts were sinusoidally amplitude modulated (SAM). In one case the noise bursts presented to both loudspeakers were modulated with the same envelope phase (in phase). In the other condition the modulation at one loudspeaker was 180° out of phase (out of phase) with that occurring at the other loudspeaker. In the out of phase condition, when the overall level at one loudspeaker was high, the level at the other loudspeaker was low. This is in contrast to the in phase condition in which the overall level at both loudspeakers was always the same. In experiment II listeners were asked to determine the location of the two sound sources as was done for the 2S-2L condition of experiment I. Both in phase and out of phase amplitude modulation between the two loudspeakers of independently and simultaneously generated noise were randomly mixed within a block of trials. The goal of experiment II was to determine if the out of phase condition leads to better localization performance than the in phase condition, and if so how performance changes with modulation rate. If level differences in different temporal regions of the combined waveform from two independently generated noise bursts (unmodulated) are a basis for sound source localization, the temporal regions would probably have to be fairly short, since it is unlikely that there would be long periods of time when two independently generated noises had significant level differences. If so, we hypothesized that fairly high SAM rates would produce better localization performance for the out of phase conditions as compared to the in phase conditions.

III. EXPERIMENT II

A. Subjects

Six listeners (four females and two males all under the age of 30) who reported normal hearing were used in experiment II. Two of the subjects (one male and one female) also participated in experiment I.

B. Stimuli

The independent noise bursts were generated as they were in experiment I, except the bursts were 500-ms in duration and shaped with 50-ms cosine-squared rise/decay times in experiment II. The noise burst duration was increased to 500 ms in experiment II to allow for the use of slow modulation rates. The noise bursts were sinusoidally amplitude modulated (SAM) at rates of 5, 50, 200, and 500-Hz (depth of modulation was always 100%). In the in phase condition

the envelope phase for both sounds was 0°, whereas for the out of phase condition one sound was generated with 0° envelope phase and the other with 180° envelope phase. The 50-ms rise-fall times were used to reduce the effect of on-set cues.

C. Procedure

The same procedure used in experiment I for the 2S-2L condition was used in experiment II. The same eight loudspeaker pairs were used as in experiment I and they and the in phase and out of phase conditions were randomly mixed with in a block of 64 trials (4 repetitions of the 8 loudspeaker pairs and two modulation phase conditions). Six of these 64 trial blocks (384 trials) were run providing 24 observations for each loudspeaker pair by modulation phase condition. This same procedure was repeated in separate blocks for each of the four SAM rates (5, 50, 200, 500 Hz). As in experiment I, each pair of noise bursts was presented once and the listener indicated the loudspeaker number (1–13) of the source presenting either one of the sounds and then the noise bursts were repeated (a different set of independent noise bursts) and the listener indicated the other loudspeaker location. The process for monitoring head position at the start of each stimulus presentation and the lack of any feedback was the same as in experiment I. One 10-trial practice block using a SAM rate of 5-Hz was run at the very beginning of experiment II.

D. Results

Figures 7–10 present the histograms of loudspeaker location responses using the same format used for Fig. 4. Figure 7 are data for a SAM rate of 5 Hz, Fig. 8 for a rate of 50 Hz, Fig. 9 for a rate of 200 Hz, and Fig. 10 for a rate of 500 Hz. In Figs. 7–10 the dark bars are the responses for the in phase envelope conditions and the lighter bars for the out of phase conditions. The circles indicate the loudspeaker location of the presenting source for the in phase conditions and the triangles for the out of phase conditions. Figure 11 presents the rms error (see Fig. 5) for the in phase and out of phase conditions as a function of SAM rate. For the data of Fig. 11 a two-factor repeated measures ANOVA [F(1,3,5)] with modulation rate and in phase vs out of phase conditions being the factors indicate that both main effects were significant ($p < 0.01$) and there was a significant interaction ($p < 0.05$). *A priori* repeated measures *t* tests indicated that the in phase errors were statistically greater than the out of phase errors for 5-Hz ($p \ll 0.01$), 50-Hz ($p \ll 0.01$), and 200-Hz ($p < 0.05$) rates of modulation, but not at 500-Hz ($p > 0.05$). Figure 12 indicates for the in phase and out of phase conditions the proportion of correct responses as a function of SAM rate in correctly locating both (both) or at least only one (one) source on each trial (see Fig. 6). For the data of Fig. 12 a three-way repeated measures ANOVA was tabulated with modulation rate, out of phase vs in phase conditions, and both vs at least one correct conditions as the factors [F(1,4,1,5)]. The three main effects were significant ($p < 0.01$), and there was one significant interaction ($p < 0.05$) of modulation rate with in phase vs out of phase modulation. In calculating repeated measures *t* tests, the proportion correct data were averaged across the both and one correct conditions in order to determine how the statistical differences between the out of phase vs in phase conditions

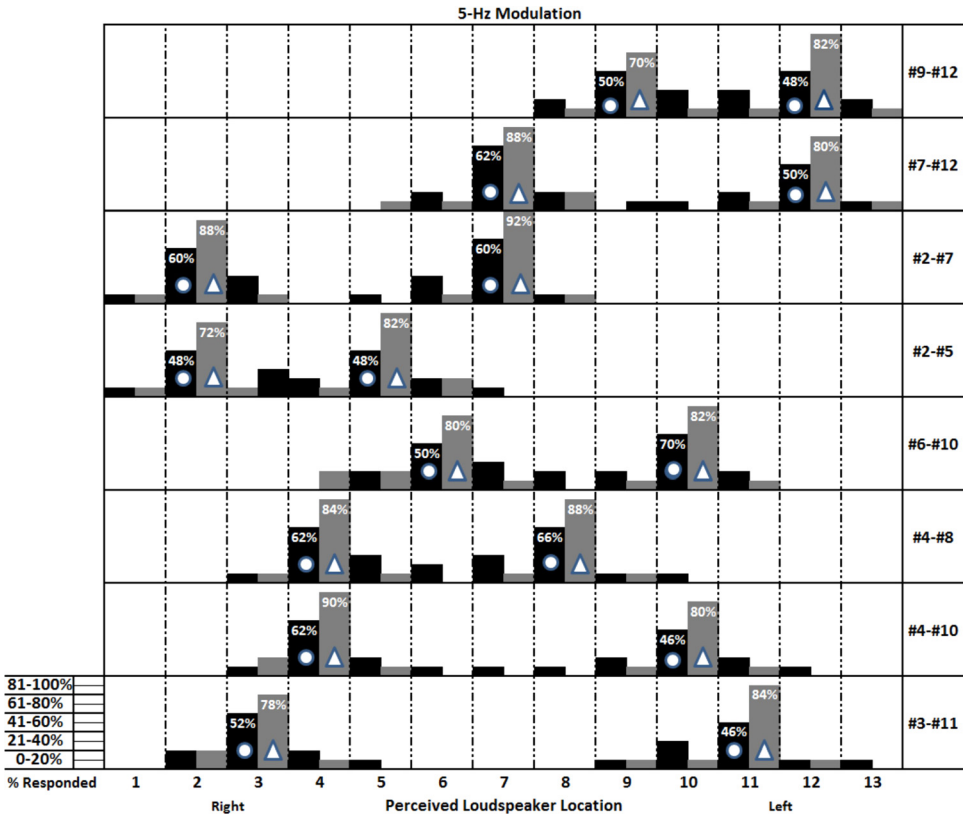

FIG. 7. (Color online) Histogram (percent of responses) of the localization responses across all conditions and listeners for the 8 loudspeaker locations when the SAM rate was 5 Hz. Dark bars are for the in phase conditions and light bars are for the out of phase conditions. Circles indicate the location of the loudspeakers that presented a sound in the in phase conditions, and triangles in the out of phase conditions. The percent (%) responses were calculated by dividing the total number of times a particular perceived location was reported by the total number of trials.

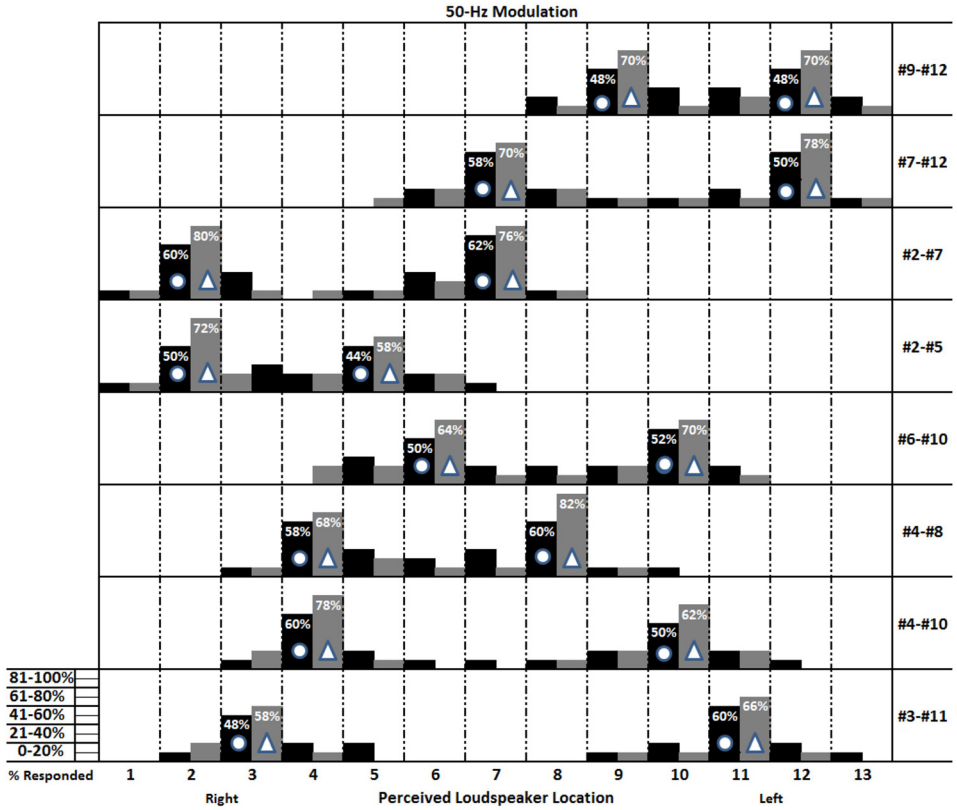

FIG. 8. (Color online) Same as Fig. 7, but when the SAM rate was 50 Hz.

varied with modulation rate. These repeated measures *a priori t* tests indicated that proportion correct responses were statistically greater for the out of phase as compared to the in phase conditions at 5-Hz ($p \ll 0.01$), 50-Hz ($p \ll 0.01$), and 200-Hz ($p < 0.05$) modulation rates, but not at 500-Hz ($p \gg 0.05$).

E. Discussion

All measures of localization performance (Figs. 7–12) suggest that localization is more accurate for the out of phase conditions as compared to the in phase conditions for the 5, 50, and 200-Hz SAM rates. It appears as if by a SAM rate of

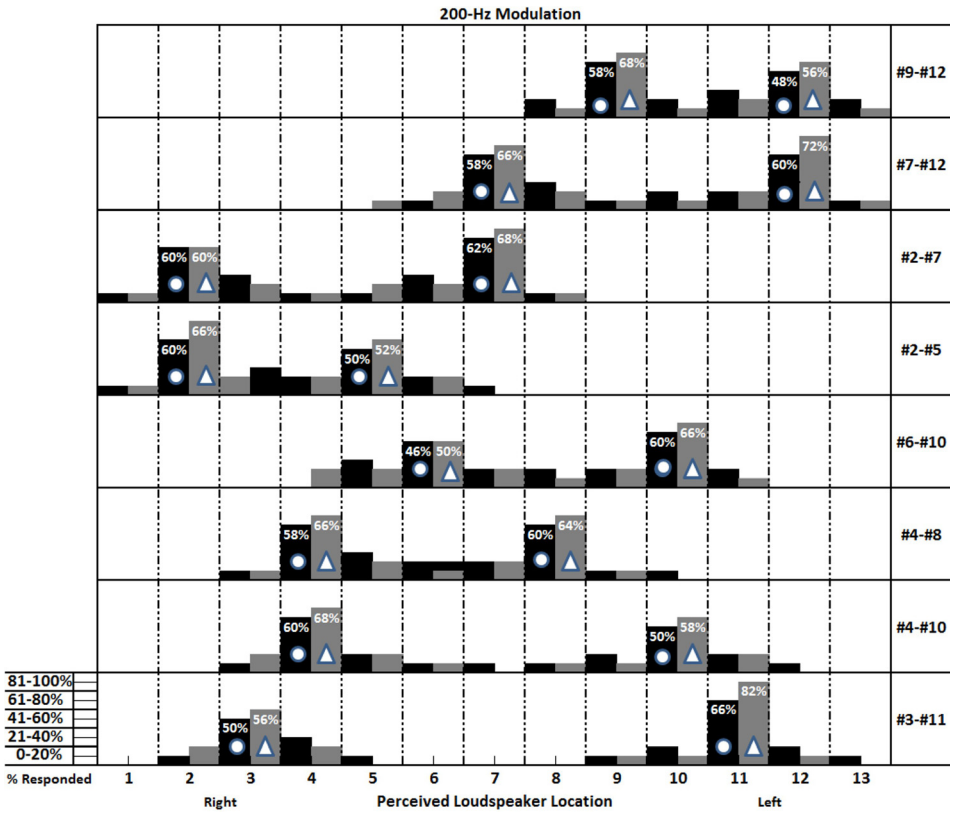

FIG. 9. (Color online) Same as Fig. 8, but when the SAM rate was 200 Hz.

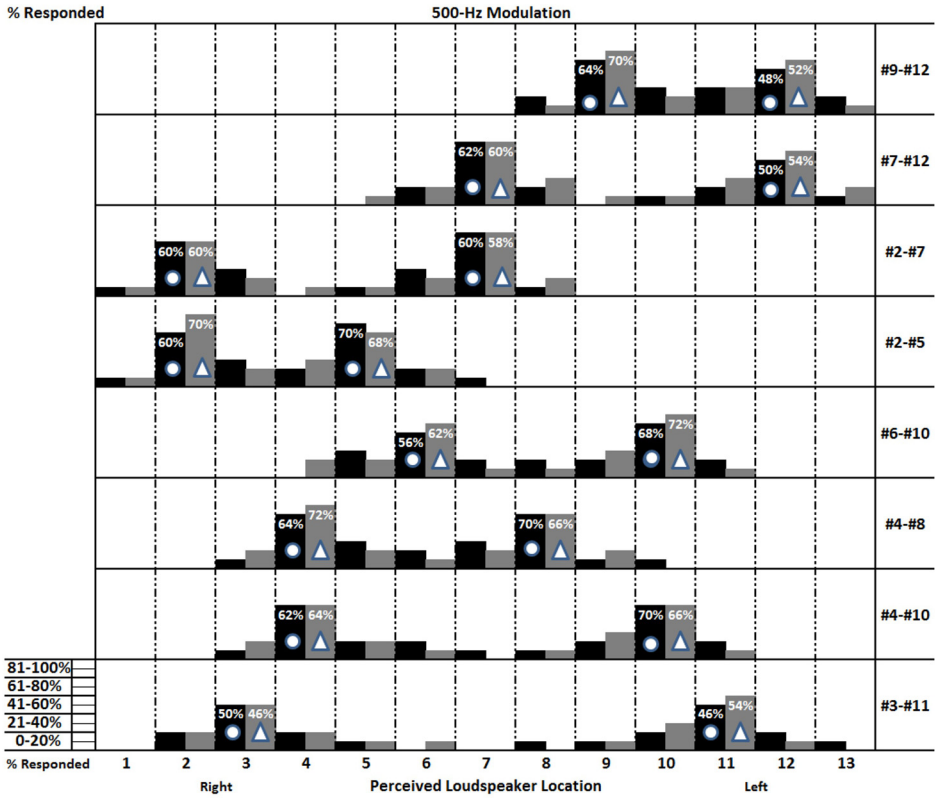

FIG. 10. (Color online) Same as Fig. 9, but when the SAM rate was 500 Hz.

500 Hz performance for locating two sound sources is essentially the same for the out of phase and the in phase conditions, and these data from experiment II at a 500-Hz SAM rate are similar to those obtained in experiment I for the unmodulated 2S-2L condition (see Fig. 5). These results would suggest that at a modulation rate of 500-Hz sound source localization performance is not affected by fast modulations. However, the data for 500-Hz rates of modulation (and to a limited extent the 200-Hz rate data) need to be considered in light of the fact that in the spectral region below 500-Hz (in general below the modulation rate when there is energy in the carrier's spectrum below the modulation rate), components below 500 Hz are not sinusoidally amplitude modulated. As a result, changing the phase of the envelope (e.g., in phase and out of phase as was done in experiment II) may not result in the same amplitude differences over time that occur when the carrier is sinusoidally amplitude modulated. Thus, for spectral components of the carrier with frequencies lower than the modulation rate, the out of phase conditions of experiment II will usually not result in the level from one loudspeaker being more intense at the same time the level from the other loudspeaker is less intense. As a result, there would not be an advantage for localization in the out of phase condition as compared to the in phase condition for spectral components of the carrier that are lower in frequency than the modulation rate. If localization performance in these tasks is dominated by information in the low-frequency region below 500 Hz, then the poor performance at a 500-Hz modulation rate may not be just due to the overall fast modulation of the level from the two sources,

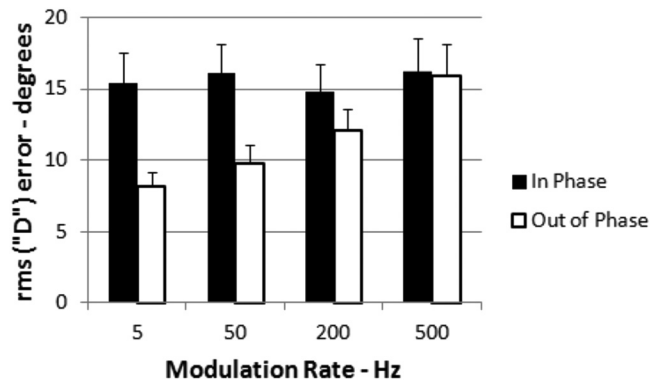

FIG. 11. Mean (across 6 listeners) rms error in degrees as a function of the in phase and out of phase conditions as a function of SAM rate (Hz). Error bars are 1 standard error of the mean.

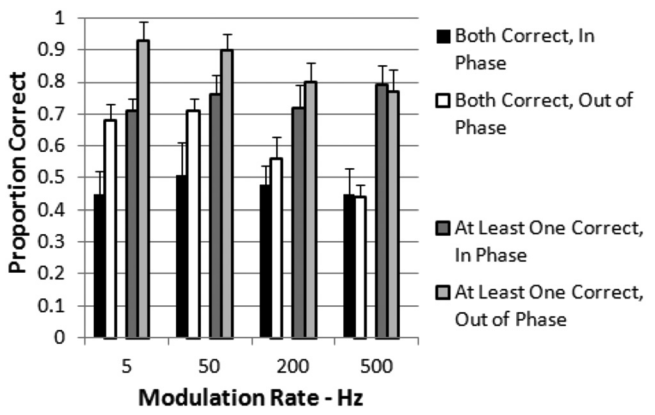

FIG. 12. Mean (across 6 listeners) proportion of correct responses in getting both loudspeaker locations correct (Both) or at least one loudspeaker location correct (one) in the in phase and in the out of phase conditions as a function of SAM rate (Hz). Error bars are one standard error of the mean.

but performance may be due (partially due?) to the lack of sinusoidal amplitude modulation below 500 Hz. Below 500-Hz the main differences in level between the sounds from the two sources in temporal-spectral cells is that due to the carrier noise bursts and is not a function of the additional modulation provided by the 500-Hz sinusoidal modulator. While modulation in level due to the modulator would be present above 500-Hz, the data of Figs. 11 and 12 suggest that modulation in these spectral regions does not assist in sound-source localization when the sounds from the two sources are modulated out of phase at a rate of 500 Hz. But, perhaps the key spectral region for modulation to provide maximum benefit for sound source localization is below 500 Hz.

In order to address this issue and before additional discussion of experiment II is provided, experiment III will be described. In experiment III, the independent noise bursts were high-pass filtered from 1500 to 6000 Hz, and all other conditions used in experiments III were the same as those used in experiments I and II. Since there is very little spectral energy below 500-Hz for these high-pass filtered noise bursts, all spectral components in the high-pass region of the noise burst (above 1500 Hz) would be sinusoidally modulated when the modulator was a 500-Hz sinusoid. Three conditions were tested: sound source localization of unmodulated, high-pass noise in the 2S-2L condition used in experiment I, and sound source localization of 500-Hz sinusoidally amplitude modulated high-pass noise in the out of phase and in phase conditions used in experiment II. The logic of experiment III is that if information below 500-Hz is crucial for sound source localization of two independently generated noise bursts, then sound source localization performance for the high-pass noise bursts should be worse than that for the wideband condition that includes information below 500 Hz. If performance is poorer for the high-pass conditions of experiment III, then the poor performance measured in experiment II for the 500-Hz modulation rate out of phase condition is likely due, or due in part, to the lack of sinusoidal amplitude modulation in the spectral region below 500 Hz for the wideband noise used in experiment II.

IV. EXPERIMENT III

A. Subjects

Four of the six subjects of experiment II, including the two from experiment 1, were used in experiment II. There were two males and two females.

B. Stimuli

All of the stimulus conditions of experiments I and II were used in experiment III with the exception that the noise bursts were filtered between 1500 and 6000 Hz using the same 8-pole Butterworth filter described in experiment I. As was the case in experiment I, the unmodulated noise in experiment III was 200 ms in duration with 20-ms rise-fall times, while the amplitude modulated noises in experiment III were 500 ms in duration with 50-ms rise-fall times as they were in experiment II.

C. Procedure

The same procedures used in experiment I for the 2S-2L condition and those used for the in phase and out of phase conditions of experiment II were used in experiment III.

D. Results

The data for experiment III are shown in Fig. 13 as mean rms error as a function of the three conditions tested in experiment III (white bars) compared to the data for these four listeners in experiments I (unmodulated) and II (in phase and out of phase 500-Hz modulation). Data from experiments I and II are represented by the black bars. A two-way repeated measures ANOVA (experiment III vs experiments I and II and Conditions were the factors) was conducted and there were no statistically significant main effects or interactions at a 0.05 level of significance [F(1,2,3)].

E. Discussion

The data of Fig. 13 suggest that sound source localization performance for locating two sound sources does not differ between noise bursts that are wideband (125 to 6000 Hz) and those that are high-passed filtered (1500 to 6000 Hz). This implies that spectral regions below 1500 Hz do not provide more important information for sound source localization than spectral regions above 1500 Hz. Thus, the data of experiment II at 200 and 500-Hz rates of modulation are probably not overly influenced by the fact that in spectral regions below the modulation rate, the spectral components of the noise carrier would not have been sinusoidally amplitude modulated when the sinusoidal amplitude modulator was applied. Overall the data suggest that, up to rates of at least 200 Hz, listeners are better able to localize two independent and simultaneously presented noise bursts when there are brief moments when the level of the two sources are different (out of phase) as compared to when the levels

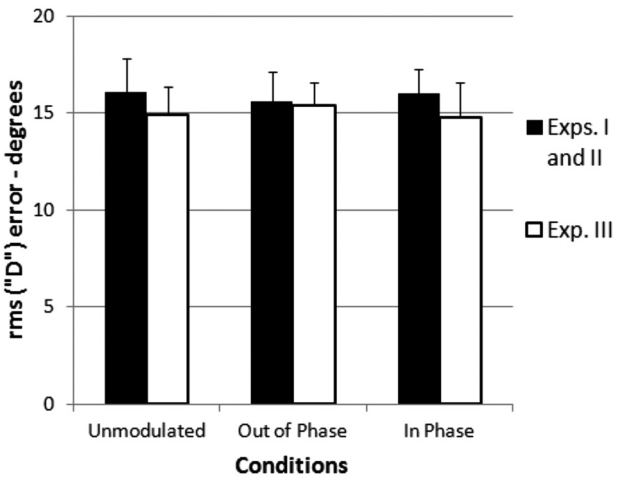

FIG. 13. Mean (across 4 listeners) rms error in degrees is plotted for three conditions of experiment III involving high-pass noise (white bars) and the similar three conditions from experiments I and II involving broadband noise (black bars). The three conditions are the unmodulated, 2S-2L condition of experiment I, and the in phase and out of phase, 500-Hz modulation conditions of experiment II. Error bars are one standard error of the mean.

are the same (in phase). This suggests that time "windows" as brief as 5 ms maybe sufficient to provide reliable estimates of interaural differences used to localize two sound sources. In fact, these temporal windows may be even shorter than 5 ms since the levels were changing as a sinusoidal function of time and localization performance was not measured for rates between 200 and 500 Hz.

Comparing the data from the unmodulated 2S-2 L conditions (experiment I) to the modulated In-Phase conditions (experiments II and III) indicates that there were few, if any, differences in localization performance between conditions in which there was no amplitude modulation (Figs. 4–6) and when the amplitude modulation was in phase at both loudspeakers (Figs. 7–13). Sinusoidally modulating the amplitude of the broadband noise bursts the same at both loudspeakers did not alter localization performance as compared to conditions in which the sounds were not amplitude modulated. This suggests that for these wideband noise bursts, that there is no additional benefit for localization in the free field associated with amplitude modulation (see Bernstein and Trahiotis, 2003, for a discussion of amplitude modulation and binaural processing and Eberle *et al.*, 2000 for evidence that amplitude modulation may not assist sound source localization in the free field).

At slow rates of SAM in the out of phase condition the perceived location of the sound source changes location from one of the loudspeakers presenting the sound to the other loudspeaker. As the intensity of the sound at the two loudspeakers alternates so does the perceived location. We conducted a pilot study to address this, in which four of the listeners from experiment II were asked to indicate if their perception of the position of the loudspeaker presenting a sound changed from one loudspeaker to another during a trial. SAM rates of 3, 6, 9, 12, 18, 21, and 24 Hz were used and all listeners indicated that for SAM rates of 18 Hz or less the perceived source of the sound did appear to change location, 2 out of the 4 listeners indicate the perceived location changed at 21 Hz, and no listener indicated that the perceived location changed at 24 Hz. The finding that perceived change in location does not appear above about 20–25 Hz is consistent with the "binaural sluggishness" literature (see, for instance, Grantham and Wightman, 1978a,b) indicating that changes in perceived laterality due to changes in the ITD and/or ILD cues does not occur at rates higher than approximately 25 Hz. Such a change in perceived location may have aided the listener in locating the two sources in the out of phase condition with a 5-Hz SAM rate, but not at the three other higher (50, 200, and 500 Hz) SAM rates.

In most investigations of sound source localization of a single source, the sound is presented only once. Localizing two sources is a far more challenging task. In pilot work we found that listeners' performance improved when two presentations of the sound sources were provided as compared to only one presentation, and that presenting the sounds more often than twice did not tend to lead to additional improvement. Asking the listeners to respond with one location after the first presentation and the other location following the second presentation resulted in better performance than when the two responses were made after the second presentation. Listeners reported that they had difficulty switching their attention from one source to another when there was only one presentation or when they had to respond with both locations at the end of the second presentation. They reported that the two presentation procedure we used improved their ability to attend to one and then the other source. In our experience, the results of this study would most likely have indicated poorer localization performance for locating two sound sources had only a single presentation been used.

While head movements were not tightly controlled in these experiments, listeners were very good at looking forward at the red dot on the center loudspeaker before each presentation, and they rarely had to be reminded to do so. The few times that reminders were given were during the second presentation in the two-sound source conditions. Once the stimulus was presented the listeners often did move their heads, but with durations less than 500 ms this does not provide enough time to turn the head fully $+-90^{\circ}$ in time to face a sound source while it is presenting sound. However, it is possible that for the SAM conditions in which signal duration was 500 ms that listeners would have performed slightly more poorly had head movements been more stringently controlled.

In experiment I the duration of the noise bursts was 200 ms and the rise-fall times were 20 ms. In experiments II and III the duration was lengthened to 500 ms and the rise-fall times to 50 ms when the noise bursts were amplitude modulated. Pilot work indicated that maximal performance in the out of phase modulation conditions occurred for modulation rates up to around 10 Hz. In order to measure sound source localization performance in the out of phase modulation conditions we wanted to be sure there was more than one period of modulation. This necessitated using durations greater than 200 ms. In the out of phase amplitude modulation conditions, the sound from one source comes on before that at the other source. These temporal conditions can be similar to those used in studies of precedence. In order to avoid the effects of precedence (see Litovsky *et al.*, 1999) having an influence on the results, we used a long (50 ms) rise-decay time, since the precedence literature suggests that effects of precedence are minimal at delays of 50 ms or longer.

As mentioned in Sec. I, extracting ITD and ILD cues from the presentation of two simultaneous sounds from different source locations is related to issues of the spatial release from masking. In spatial release from masking experiments the listener is almost always asked to detect, discriminate, or identify some aspect of a target sound presented from a known target location. The location of masking sound sources may or may not be known to the listener. It is often argued, especially in cases of informational masking (see Freyman *et al.*, 1999), that a listener's ability to localize the sources of the target and/or maskers plays a role in any spatial release from masking that is measured. To the extent that sound source localization is important for spatial release from masking, the data of the present study suggest that in localizing two sound sources, localization

performance is better when a listener knows where one sound source is located than when the listener has no prior information about the location of either sound source.

The data from the current study are consistent with the hypothesis that localization of two sound sources occurs when there are moments in time and regions in the spectrum when the levels of the two sounds are not the same. Meffin and Groth (2009) made this argument and suggested that in the real world the modulation of the sound level due the differences in the sound from the sources would vary at a faster rate than would changes in level that resulted from a source moving in nature or an animal moving. They suggested that a circuit in the dorsal nucleus of the lateral lemniscus (DNLL) might filter out the fast changes in the rate at which interaural cues vary across the sources (changes that are spurious with respect to source location), leaving only the slower rates associated with source or self-movement. Woodruff and Wang (2010) demonstrated that good sound source segregation for speech in a reverberant environment could be achieved when a computational algorithm was used to extract interaural time and level differences in the cells of a temporal-spectral matrix of the target speech mixed with the reverberation. Liu *et al*. (2000) demonstrated that when multiple sound sources are present, using temporal-spectral processing and directional microphones improve localization performance in their proposed hearing aid. There have been other computational approaches to "tracking" multiple sound sources producing speech (e.g., Faller and Merimma, 2004; Dietz *et al*., 2011). In some of these approaches a measure of interaural correlation (coherence) is determined in different spectral-temporal regions. In those regions where the interaural correlation is high, a form or ITD and/or ILD analysis may be carried out to indicate the position of one sound source or another. In a study of detection/identification Kopco and Shinn-Cunningham (2008) suggested using peaks in the modulation envelope to find periods of time when it is likely that the interaural values would be informative of the location of one source or the other in a spatial release from masking context. Thus, the general idea of using interaural cues in spectral-temporal regions of the combined waveform when two or more sources present simultaneous sound has been suggested by several investigators.

Figure 14 indicates that there is useable localization information in cells of a temporal-spectral matrix of the combined waveform of independent noise bursts from two sources presented with a 5-Hz SAM that is 180° out-of-phase between the two loudspeakers (the out of phase, 5-Hz SAM condition of experiment II). In Fig. 14 the temporal-spectral cells had time widths[3] of 20 ms (non-overlapping) and spectral widths (non-overlapping) of 1-ERB (center frequencies from 100 to 8385 Hz). The matrix was extracted for the sound presented alone from each loudspeaker and for the combined waveform when the noise bursts were presented simultaneously. The matrices were extracted from left and right KEMAR channel recordings when the sources were +−45° left and/or right of midline (loudspeakers 4 and 10 in our experimental setup). The ILDs were based on the left and right KEMAR recorded rms levels computed for each temporal-spectral cell for the two sources and for the combined waveform. The ITD was computed by finding the interaural delay that generated the highest cross-correlation value (see Woodruff and Wang, 2010). The time shift resulting in the maximal cross-correlation value for each cell was the estimated ITD for that cell. If the ITD or ILD value in a cell for the combined waveform was within +−10% of that for the corresponding cell for the right sound source that cell in the temporal-spectral matrix of Fig. 14 was made white (right). If there was a +−10% agreement for the left sound source, the cell in Fig. 14 was made black (left). Otherwise the cell was grey. Thus, black and white cells suggest temporal-spectral cells in the combined waveform that contain "reliable" estimates of the ITD and ILD cues associated with the left or right sound sources. Grey cells in Fig. 14 suggest cells with "spurious" ITD and ILD cues. ITD values were only tabulated for spectral regions below about 1350 Hz and ILD values for spectral regions above about 1600 Hz. The out-of-phase SAM waveforms are shown in between the two temporal-spectral matrices. As can be seen the black and white cells in the spectral-temporal matrix align when the level at the appropriate loudspeaker is intense, and spurious cells (gray) occur when the levels from the two sources are nearly the same.

While Fig. 14 does not provide a prediction (or a model) of human listeners' ability to localize two sound sources, Fig. 14 does demonstrate that a temporal-spectral representation of the combined waveform contains interaural information that is similar to that associated with the spectral-temporal interaural cues of one or the other sound source presented alone. Any model of how the human auditory system processes multiple sound source localization based on an analysis of spectral-temporal regions of the combined waveform would need to address several issues. For instance: (1) How do differences across the spectrum (as opposed to across time as measured in the present study) influence sound source location when more than one source produces simultaneous sound? (2) What are the durations and spectral widths of the spectral-temporal regions and are they always the same duration and width for all stimulus conditions? (3) Are all spectral-temporal regions analyzed or is there a process/mechanism that would indicate which regions should be analyzed (e.g., using interaural correlation/coherence)? (4) How is the information in the spectral-temporal regions combined (integrated) across time and/or frequency to arrive at an estimate of the interaural time and level differences that may be associated with one of the sound sources? (5) For a single a broadband sound such as noise, the interaural differences are not constant across time and frequency (this is especially true of interaural level differences measured across frequency, i.e., the result of the head shadow). In order for interaural measures of the combined waveform to be informative of the location of one source or the other, some way must be provided of establishing that the pattern of interaural changes measured for the combined waveform across time and frequency "matches" those of a single source at a particular location. Simply knowing what the ITD or ILD is for a particular spectral-temporal cell cannot always by itself predict the location of a sound source. For instance, head-shadow produces a frequency dependent

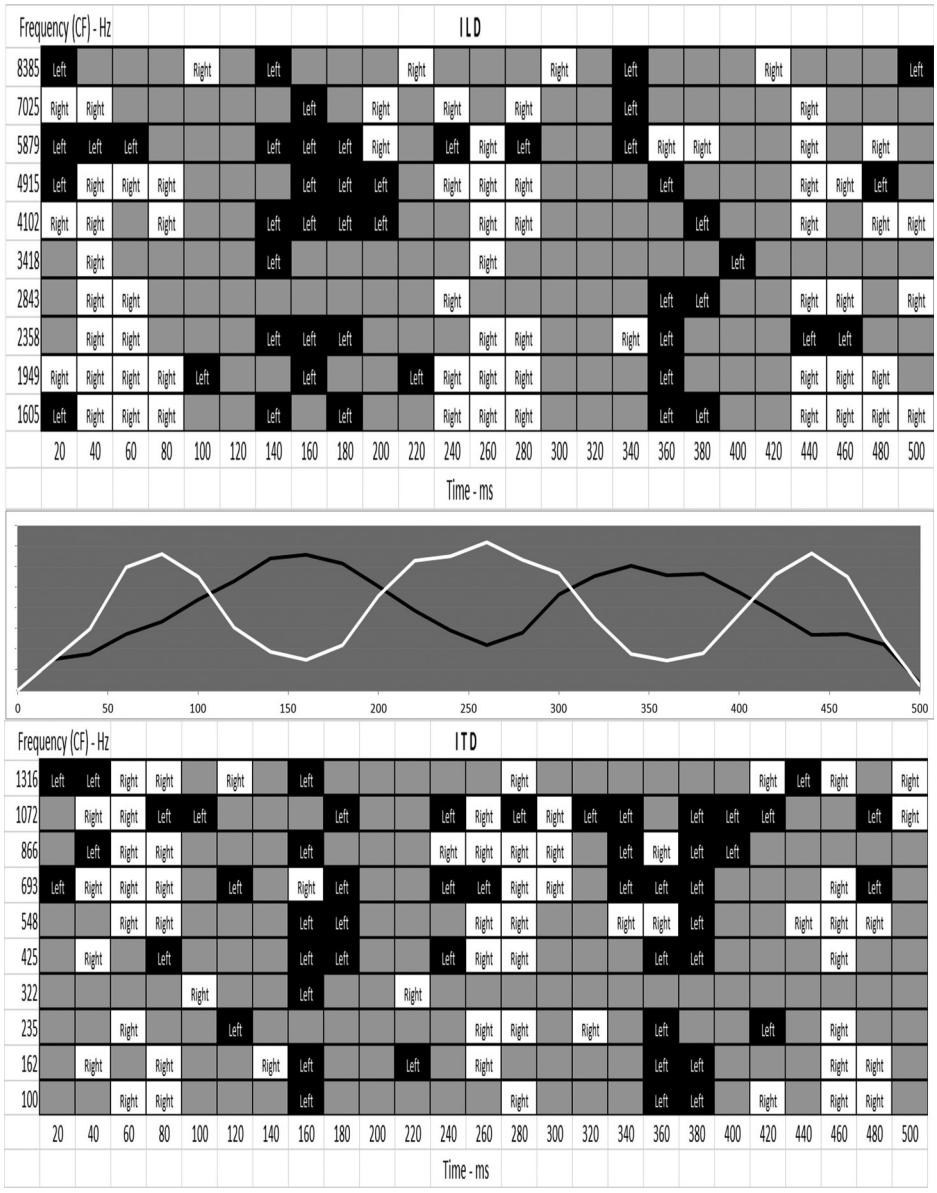

FIG. 14. A temporal-spectral matrix of the combined waveform from out-of-phase, 5-Hz SAM noise presented to loudspeakers 4 and 10 simultaneously. Each cell in the matrix is 20-ms long and one ERB wide. The cells that are black (left) are those with ITD or ILD values within +−10% of those measured in the same cell for the noise presented from just the left loudspeaker. The cells that are white (right) are those with ITD or ILD values within +10% of those for the same cell for the right loudspeaker. Gray cells represent cells with interaural values not within +−10% of the interaural values measured for either source when presented alone. (Bottom) ITD values were only tabulated for cells representing frequencies less than approximately 1350 Hz and (top) ILD values were only computed for cells representing frequencies greater than approximately 1600 Hz. (middle) The two out-of-phase SAM noise waveforms are shown.

ILD, such that multiple azimuth sound source locations can produce the same ILD in different frequency regions (see Kuhn, 1987; Macaulay *et al.*, 2010). Thus, one needs to know information about both the ILD and the spectrum to be able to predict an azimuth location. Answers to most of these questions would be required to provide a model for how human listeners located the sources of noise bursts as used in the current experiments. However, the analysis shown in Fig. 14 suggests that interaural information in spectral-temporal cells of the combined waveform when the level of sound from one source is significantly greater than that from another source may provide reliable information about the interaural values for the same spectral-temporal region of one source or the other.

V. SUMMARY

In summary, listeners can localize the sources of independently and simultaneously generated noise bursts, but not as well as they can localize a noise burst from a single source. Providing prior information about the source of one of the two sounds leads to better localization performance as compared to when no such prior information is provided. This paper presents data based on amplitude modulated noise bursts that support the hypothesis that multiple sound source localization might be based on the use of temporal-spectral regions when the level from one source is greater than that from another source to extract reliable estimates of the interaural values of one source or the other.

ACKNOWLEDGMENTS

This research was partially supported by a NIDCD grant awarded to Michael Dorman; W.A.Y. is an investigator on this grant, by a NIDCD grant awarded to Sid Bacon and C.A.B., and AFOSR grant awarded to W.Y. The hard work of ASU undergraduates Britta Martinez and Brian Shock as well as that of Farris Walling, a former AuD student at ASU,

is much appreciated as is the interaction with Michael Dorman, Tony Spahr, Kate Helms-Tillery, and Sid Bacon.

[1]The D calculation of Rakerd and Hartmann (1986) was used: $D(k) = \sqrt{[A^2/M \Sigma_{i=1}^{M}(ri - k)^2]}$; "A" is angular separation of the loudspeakers, "M" is the number of responses per condition, "r" is the response (1–13) on the *i*th trial and "k" is loudspeaker location (2–12). D is the average over all k loudspeakers (2–12).

[2]In calculating the rms error for the 2S-2L conditions, an assumption was made that errors were minimized. For instance, if the actual loudspeaker locations were 4 and 10 and the responses were 3 and 9, we assumed that perceived location 3 corresponded to actual location 4, and not 10. But, given that the stimuli were unidentifiable independently generated noise bursts, it is not possible to know for sure if response 3 did correspond to actual location 4.

[3]A temporal window of 20 ms was chosen since a SAM rate of 50-Hz (reciprocal of 20 ms) produced excellent localization performance in all conditions (see Fig. 8).

Bauer, B. B. (**1961**). "Phaser analysis of some stereophonic phenomena," J. Acoust. Soc. Am. **33**, 1536–1539.

Bernstein, L. R., and Trahiotis, C. (**2003**). "Enhancing interaural-delay based extents of laterality at high frequencies by using 'transposed stimuli,' " J. Acoust. Soc. Am. **113**, 3335–3347.

Best, V., van Schaik, A., and Carlile, S. (**2004**). "Separation of concurrent broadband sound sources by human listeners," J. Acoust. Soc. Am. **115**, 324–336.

Blauert, J. (**1997**). *Spatial Hearing* (MIT Press, Cambridge, MA), 494 pp.

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distractor and reverberation in the frontal horizontal plane I. Psychoacoustic data," Acta Acust. Acust. **88**, 942–955.

Croghan, N. B. H., and Grantham, W. D. (**2010**). "Binaural interference in the free filed," J. Acoust. Soc. Am. **127**, 3085–3091.

Dietz, M., Ewert, S. D., and Hohman, V. (**2011**). "Auditory model based direction estimation of concurrent speakers from binaural signals," Speech Comm. **53**, 592–605.

Eberle, G., McNally, K. I., Martin, R. L., and Flanagan, P. (**2000**). "Localization of amplitude modulated high-frequency noise," J. Acoust. Soc. Am. **107**, 3568–3671.

Erno, H. A., Langendijk, E. H. A., Kistler, D. J., and Wightman, F. L. (**2001**). "Sound localization in the presence of one or two distracters," J. Acoust. Soc. Am. **109**, 2123–2133.

Faller, C., and Merimma, J. (**2004**). "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3081.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gardner, M. B. (**1969**). "Mage fusion, broadening, and displacement in sound localization," J. Acoust. Soc. Am. **6**, 339–349.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal to noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 349–376.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., and Haynes, D. S. (**2007**). "Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants," Ear Hear **28**, 524–541.

Granthman, D. W., and Wigthman, F. L. (**1978a**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–521.

Granthman, D. W., and Wigthman, F. L. (**1978b**). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1518.

Hanna, T. E. (**1984**). "Discrimination of reproducible noise as a function of bandwidth and duration," Percept. Psychophys. **36**, 409–416.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Keller, C. H., and Takahashi, T. T. (**2005**). "Localization and identification of concurrent sounds in the owl's auditory space," J. Neurosci. **25**, 10446–10461.

Kopco, N., and Shinn-Cunningham, B. G. (**2008**). "Influences of modulation and spatial separation on detection of a masked broadband target," J. Acoust. Soc. Am. **124**, 2236–2250.

Kopčo, N., Best, V., and Shinn-Cunningham, B. G. (**2007**). "Sound localization with a preceding distractor," J. Acoust. Soc. Am. **121**, 420–432.

Kuhn, G. F. (**1987**). "Physical acoustics and measurements pertaining to directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–26.

Lee, A. K. C., Deane-Pratt, A., and Shinn-Cunningham, B. G. (**2009**). "Localization interference between components in an auditory scene," J. Acoust. Soc. Am. **126**, 2543–2555.

Litovsky, R., Colburn, S., Yost, W. A., and Guzman, S. (**1999**)."The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Liu, C., Wheeler, B. C., O'Brien, W. D. O., Bilger, C., Lansing, C. R., and Feng, A. S. (**2000**). "Localization of multiple sources with two microphones," J. Acoust. Soc. Am. **108**, 1888–1905.

Macaulay, E. J., Hartmann, W. M., and Rakerd, B. (**2010**). "The acoustical bright spot and mislocalization of tones by human listeners," J. Acoust. Soc. Am. **127**, 1440–1449.

Meffin, H., and Grothe, B. (**2009**). "Selective filtering to spurious localization cues in the mammalian auditory brainstem," J. Acoust. Soc. Am. **126**, 2437–2454.

Rakerd, B., and Hartmann, W. M. (**1986**). "Localization of sound in rooms III: Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706.

Santala, O., and Pulkki, V. (**2011**). "Directional perception of distributed sound sources," J. Acoust. Soc. Am. **129**, 1522–1530.

Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–887.

Woodruff, J., and Wang, D. L. (**2010**). "Sequential organization of speech in reverberant environments by integrating monaural grouping and binaural localization," IEEE Trans. Audio Speech, Lang. Process. **18**, 1856–1866.