Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression

R. Y. Litovsky[a)]
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

B. G. Shinn-Cunningham
Hearing Research Center, Departments of Biomedical Engineering and Cognitive and Neural Systems, Boston, Massachusetts 02215

(Received 2 January 2000; revised 7 August 2000; accepted 4 October 2000).

Listeners have a remarkable ability to localize and identify sound sources in reverberant environments. The term ''precedence effect'' (PE; also known as the ''Haas effect,'' ''law of the first wavefront,'' and ''echo suppression'') refers to a group of auditory phenomena that is thought to be related to this ability. Traditionally, three measures have been used to quantify the PE: (1) *Fusion*: at short delays (1–5 ms for clicks) the lead and lag perceptually fuse into one auditory event; (2) *Localization dominance*: the perceived location of the leading source dominates that of the lagging source; and (3) *Discrimination suppression*: at short delays, changes in the location or interaural changes2-ral(audial)-(t-eidial)-t;

with the lead–lag delay randomized; the listener reports her subjective impression of whether one or two sounds are heard on each trial. For click stimuli, at short delays (1–5 ms) most listeners report hearing only one sound on 100% of trials; at long delays (8–10 ms) most listeners report hearing two sounds on 100% of trials; at intermediate delays there is a transition in the percentage of trials in which ''two sounds'' are reported. In general, the percentage of ''two sound'' trials increases fairly steeply with delay, although the exact delay at which this sharp transition occurs varies across individuals (e.g., Freyman *et al.*, 1991). This critical delay, known as the *echo threshold*, is usually defined as the delay at which two sounds are reported on some predetermined percentage of trials (usually between 50% and 75%). Echo threshold varies with stimulus conditions, testing situation, and instructions given to the listener (Zurek, 1987; Blauert, 1997). Finally, it should be noted that the fusion task does not measure masking; listeners can detect the presence of the lag even when they do not perceive the lag as a separate auditory event.

Most localization dominance studies have been conducted under headphones using ''adjustment'' protocols. In these experiments, listeners match the position of a reference stimulus by setting interaural parameters (such as time, ITD, or level, ILD) of a test stimulus. This approach provides a quantitative measure of the relative influence of lead and lag binaural cues on lateralization (von Bekesy, 1960; Wallach *et al.*, 1949; Haas, 1951, 1972; Snow, 1954; Leakey and Cherry, 1957; Yost and Soderquist, 1984; Shinn-Cunningham *et al.*, 1993). These studies show that when the delay is a few milliseconds, the heard location of a fused image is much nearer to the position of the lead (presented in isolation) than that of the lag. Localization cues of the lag also contribute to the lateralization; however, when the delay is near or equal to zero, the perceptual influence of the lag increases until it contributes almost equally to the overall spatial impression. Although free-field measurements of localization dominance are less common, these studies also show that the lag contributes relatively little to the perceived location of the fused image (Hafter *et al.*, 1992; Litovsky *et al.*, 1997).

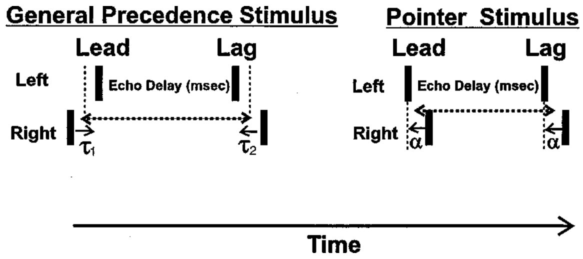

FIG. 1. General precedence stimulus (left) used for all three experiments and pointer stimulus (right) used on the localization dominance pointer task. Stimuli consisted of 1-ms Gaussian noise bursts with a 0-ms rise–decay time. The lead and lag each consisted of a pair of binaural noise bursts presented with a specified interaural time differences (ITDs), denoted as τ_1 for the lead and τ_2 for the lag. The echo delay represents the delay between the lead and lag pairs, defined as the time interval between the midpoints of τ_1 and τ_2. In the general stimulus, τ_1 and τ_2 could have ITD values that were either the same or different. In the pointer stimulus, the lead and lag pairs had the same ITD value.

listeners had previous experience in psychoacoustic tasks (S4, S5). All listeners were given a minimum of 1 h of practice on each of the tasks.

All testing was conducted in a double-walled soundproof booth. Testing was initially conducted on the fusion task, was followed by a randomized sequence of trial blocks for the discrimination and pointer tasks, and ended with a repetition of the fusion measurements.

B. Stimuli

A Tucker-Davis Technologies System II stereo analog interface was used to construct the stimuli. The output was fed through a 16-bit DAC to Sennheiser HD 520 II headphones. The general precedence stimulus (Fig. 1) was used for all three experiments. All stimuli consisted of 1-ms Gaussian noise bursts with a 0-ms rise–decay time. A lead–lag stimulus configuration consisted of two pairs of binaural noise bursts presented with various combinations of interaural time differences (ITDs) for the lead (τ_1) and lag (τ_2) pairs. Within a given interval, lead and lag were identical noise samples with new samples chosen for each interval. Delays varied from 1–15 ms.

C. Test parameters

1. Fusion

On each trial, the general stimulus was presented three times, with interstimulus intervals of 500 ms. The ITDs of the lead and lag were constant within each trial. ITDs for lead and lag were chosen from the set (+400, 0, −400) μs, for a total of nine combinations. For five of the six subjects, eight delays were used (1, 2, 3, 5, 7, 10, 12, and 15 ms). The sixth subject was also tested at longer delays of 20, 30, 50, 70, and 100 ms (see Sec. III). On each trial, the ITDs and delays were randomly chosen. A total of 20 trials were presented at each delay and lead/lag ITD combination for a total of 1440 trials per listener. On each trial, listeners were instructed to report whether they perceived ''one fused auditory event'' or ''two sounds'' on the third interval. Listeners were aware of the fact that two events were always present in each interval. No feedback was provided, since two stimuli were always present. Testing was repeated both prior to (first run) and following (second run) all other experiments.

2. Discrimination suppression

On each trial, the general stimulus was presented three times in an ABX forced-choice task. In this procedure, the ''target'' ITD of the first (A) and second (B) interval differed. The target ITD of the third interval (X) was randomly chosen to equal either that of A or B with equal likelihood. The nontarget ITD and the lead/lag delay were the same in all three intervals of a given trial. Three conditions were tested that differed in the ''target'' ITD. In one condition, the target was the ITD of the lead in the general precedence stimulus (Fig. 1). In the second condition, the target was the lag ITD. The final condition was a control in which only one binaural burst was presented (i.e., the control did not use a precedence stimulus).

An adaptive procedure was used to estimate the jnd in the target ITD at different reference ITDs and delays. In each run, the delay and reference ITD were fixed. The change in the target ITD (around the reference) varied adaptively using a modified 2-down/1-up protocol with 14 reversals (Hawley, 1994). The starting ITD was 400 μs. For the first four reversals the ITD was either increased or decreased by a factor of 2; subsequent changes were by a factor of 1.4. Threshold was estimated by averaging the ITDs of the last ten reversals. Feedback was provided on every trial. Thresholds were obtained at delays of 1, 2, 3, 5, 10, and 15 ms for the two conditions (lead- and lag discrimination) using the general precedence stimulus. The reference target ITD was either 0 (center) or −400 μs (left). In each trial of lead- and lag discrimination, the ITD of the noise burst that was not being discriminated (lag and lead, respectively) was chosen randomly (from a uniform distribution ranging from −500 to 500 μs), forcing listeners to use directional information in the target to perform the task. All delay and stimulus combinations were repeated three times with the order of the conditions randomized.

3. Localization dominance

In the final task, listeners adjusted an acoustic pointer to indicate lateral positions of a target stimulus. On each trial, listeners alternated between listening to the general stimulus (target) and the pointer stimulus (Fig. 1). The pointer stimulus had the same basic structure and temporal characteristics as the general stimulus, except that the lead and lag ITDs were equal. Listeners controlled the ITDs of the pointer by adjusting a potentiometer dial. ITDs could vary between ±1000 μs in steps of 10 μs. Subjects were asked to indicate the perceived location(s) of the lead/lag target by adjusting the pointer ITDs. Since two images are often perceived at the longer delays used in the experiment, measurements were repeated twice for all stimuli, with two separate sets of instructions. On half of the trials listeners were told to match the ''right-most'' image; on half of the trials instructions were to match the ''left-most'' image. If only one image was heard, both instructions should yield identical results. The right-most and left-most trial types remained constant within a block, and the order of the blocks was randomized within each session. The final ITD of the pointer (the subject response) will henceforth be referred to as ‘‘alpha’’ or the ‘‘matched ITD.’’

Stimuli alternated between seven presentations of the target and nine presentations of the pointer. The pointer location could be adjusted while it was being presented. Stimuli automatically alternated between target and pointer until the listener indicated confidence in their match by pressing a button. The ITDs of the lead and lag (τ_1 and τ_2), and the delay (1, 2, 3, 5, 10, and 15 ms) varied from trial to trial, but were held constant within each trial. ITDs of both lead and lag were chosen from the set {+400, 0, −400 μ

for the conditions at each delay. Error bars show the standard error around the means across six repetitions (three per condition). Performance depended strongly on delay for five of the six listeners and weakly for S6 (the remaining subject). At short delays lag discrimination was poor, evidenced by large ITD jnd's. In contrast, lead discrimination performance was relatively good at the short delays, as evidenced by much smaller ITD jnd's. Analyses of variance tests examining the effect of the two reference conditions (0 μs and -400 μs) found no significant difference between the conditions ($p > 0.05$), as expected from the results shown in Fig. 3.

The results show that at short delays, listeners were able to use directional information in the lead much more readily than directional information in the lag. This presumably reflects the fact that for precedence effect conditions, the lead carried more perceptual weight in localization than the lag (e.g., Zurek, 1980; Shinn-Cunningham *et al.*, 1993). As delays increased, lag discrimination improved so that by 10 ms, lead and lag performance was roughly equal. This result suggests that precedence was no longer effective by 10 ms. For some listeners (S1, S3, S5), lead discrimination was actually worse than lag discrimination at delays greater than 10 ms. This reversal suggests that at these long delays (and for these subjects), the lag interfered with the lead ITD information more than the lead interfered with the lag ITD information. Finally, intersubject differences were large. For instance, the difference between lead and lag conditions was greater for three listeners (S1, S2, S3), primarily due to better lead discrimination at the shortest delays. In contrast, results for S6 suggest that lead and lag interact strongly at all delays, as evidenced by poor discrimination in both the lead and lag conditions for all measured delays.

C. Localization dominance

Figure 4 shows a sample data set for the pointer task. For brevity, we will refer to the various experimental conditions in the pointer task using two letters to denote the lateral positions (right, R; center, C; and left, L) of the lead and lag, respectively. The instructions are denoted by which letter is bold (recall that listeners were instructed to match either the right-most or left-most image). The bold letter denotes which of the bursts in the target was farther to the side indicated by the instructions. For instance, in the **R**-C condition, the lead ITD was $+400$ μs (right) and the lag ITD was 0 (center)..

with the C stimulus, the R is bold when the instructions were to match the right-most. In R-**C** the C is bold because the instructions were to match the left-most. We henceforth refer to a condition such as **R**-C as one for which ‘‘the instructions were to match the lead’’ (and, similarly, R-**C** as a condition for which ‘‘the instructions were to match the lag’’), even though the instructions were always to match either the left- or right-most sound image.

In the example in Fig. 4, four conditions are shown: two with lead on right and lag at center (**R**-C and R-**C**), and two with lead on left and lag at center (**L**-C and L-**C**). Closed symbols denote cases in which the instructions were to match the lead, and open symbols denote cases in which the

Results show a strong effect of delay and a dependence on the relative ITDs of the lead and lag for all listeners. In this figure, the open and closed symbols should differ if listeners hear two distinct positions. For instance, the open symbols in the left column would fall at 0 μs if listeners matched the position of the lag (independent of the lead). Similarly, the closed symbols should remain at either $+400$ or -400 if listeners matched the position of the lead (independent of the lag).

At short delays, regardless of instructions, all listeners placed the pointer near the ITD of the lead, suggesting that they perceived one location near the lead. As delay increased, different instructions elicited different responses for the same stimulus, although not all listeners perceived two images at longer delays. In addition, the likelihood of perceiving two distinct images depends on the relative ITDs of the lead and lag. Listeners S1–S4 generally heard two separate images for delays equal to or greater than 15 ms. However, some results are asymmetric, most notably for listeners S2 and S4, who heard an image near the lag ITD when the lead was on the right, but not when the lead was on the left. Even at the longest delays measured, listeners S5 and S6 did not appear to hear two separate images. For these subjects, results are roughly independent of instructions: the open and closed symbols are near the lead ITD at short delays and are approximately midway between the lead and lag ITDs at longer delays.

When the lead was at center and lag lateral (to either the right or left; second column) three listeners (S1–S3) heard one image for delays ranging from 1 to 5 ms and two images at longer delays. The other three listeners (S4–S6) heard one image whose location was near the lead at short delays and midway between the lead and lag at longer delays.

Finally, when the lead and lag were on opposite sides (±400 μs; third column), four listeners (S1–S4) localized two distinct images at the longer delays. The matched positions of the two images were essentially equal to the locations at which the lead and lag bursts would be perceived when presented in isolation, indicating that the lead and lag images did not interact for these subjects and conditions. Listener S5 showed some asymmetry. S5 matched two distinct images when the lead was on the right or left, but the spatial separation of these images was much smaller when the lead was on the left. Listener S6 never matched two distinct locations.

D. Match performance near echo threshold

The ability of listeners to locate two distinct images does not seem to be directly related to their subjective reports of whether one or two images are present. Fusion data (Fig. 2) show that many of the listeners reported hearing two sounds at delays near 5 ms; however, at these delays the same listeners matched a single location near the lead, independent of instructions (Fig. 5). Thus, it appears that localization dominance persists to longer delays than fusion.

To illustrate this point, Fig. 6 plots estimated matched ITD at the fusion echo threshold delay (found by interpolating matched ITDs across delay). Each plot shows data from one listener. For every lead/lag ITD and instruction combination, the matched position is plotted as a function of lead ITD. The symbol and fill indicate whether instructions were to match the lead (squares) or the lag (open circles). Filled circles are used for matches in which lead and lag ITD were equal and instructions were expected to have no effect. In Fig. 6, if the lead ITD completely dominated perception, the data would fall along the diagonal, independent of instructions or lag ITD. In other words, the matched ITD would be roughly equal to and highly correlated with lead ITD, independent of instructions. If two locations were perceived, the squares would generally be expected to fall nearer the diagonal and the open circles to be independent of lead ITD value.

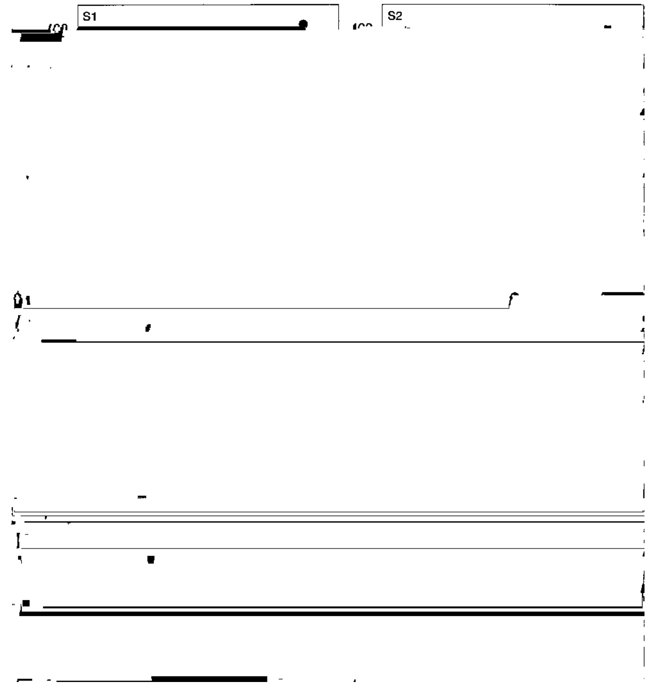

FIG. 6. Estimated matched ITD at the fusion echo threshold delay (found by interpolating matched ITDs across delay). Each plot shows data from one listener. For every lead/lag ITD and instruction combination, the matched position is plotted as a function of lead ITD. The symbol and fill indicate whether instructions were to match the side closer to the lead ITD (squares) or the lag ITD (open circles). Filled circles are used for matches in which

torons. T3-3713814 03ruction.9.0309 50 G50 J(cij the99 w 10 M []0 d 1 i [(FIG.)15 TDo Q

Table II shows correlation values between lead or lag ITD and match ITD at fusion echo threshold when instructions were to match lead or lag. For some subjects, the correlation with lead ITD was quite high regardless of instruction. For other subjects, these correlations were more modest. For all subjects, correlations were low between lag ITD and matched ITD regardless of instructions. These results suggest that, at fusion echo threshold, listeners were primarily utilizing directional cues contained in the lead.

The data are replotted in Fig. 7 as a function of lag ITD to further illustrate this point. If data fell along the diagonal, it would indicate that subjects heard a single location near the lag ITD, independent of lead ITD. If subjects heard two independent images at the lead/lag locations, open circles would fall on the diagonal (be highly correlated with the lag ITD) and squares would show little dependence on (be essentially uncorrelated with) lag ITD. Both the lack of structure in the data in the plot and the low correlation between matched ITD and lag ITD (Table II) further confirm that precedence is strong at echo threshold.

For three listeners (S1, S2, S3) the lead was clearly dominant, with the correlation between lead ITD and matched ITD close to 1.0 regardless of instructions (see Table II). Listener S4 had high correlations (a) between lead ITD and matched ITD when instructions were to match the lead, and (b) between lag ITD and match ITD when instructions were to match the lag. This result suggests that S4 was able to match the location of either source. Both S5 and S6 showed only moderate correlations with either lead or lag ITD. S5 showed some asymmetry, with matches dominated more by the lead when the lead ITD was to the right (+400 μs) than to the left (-400 μs).

E. Model estimate of precedence weight based on pointer results

The metric c (described in Shinn-Cunningham *et al.*, 1993) was calculated to quantify the relative influence of the lead and lag in localization. According to the model, the value of c is estimated by

$$c=(\alpha_p-\tau_2)/(\tau_1-\tau_2),$$

where α_p is the matched ITD and τ_1 and τ_2 are the lead and lag ITDs, respectively, for a given condition. A c value of 1.0 indicates that precedence is complete and that the lead dominates lateralization entirely. A c value of 0.5 indicates that the lead and lag both contribute equally to localization perception. A c value of 0 indicates that the lag dominates lateralization completely. In our study, instructions varied, and listeners were told to match either left or right images (see Figs. 4 and 5 for details). When told to match the lag, a c value of 0 would be expected if listeners heard two distinct images, one near the lead ITD and one near the lag ITD. If listeners were told to match the lead and a distinct image was heard near the location at which the lead would be heard in isolation, a c value of 1 is expected. Finally, if the lead and lag form a single image, then c should fall between 0 and 1 and be independent of instructions.

In Fig. 8, c values for each listener are shown as a function of delay for combinations of conditions in which the lead was lateral (

stronger) when the lead–lag separation is 800 μs compared to 400 μs. This effect is especially pronounced at longer delays. This finding suggests that interference from the lag on the lead image is greater when the lead and lag are spatially close. However, when listeners were instructed to match the lag, there was no consistent difference between *c* values for the 800- and 400-μs lead–lag separations, suggesting that the strength of the interference of the lead on the primarily lag image was independent of spatial separation.

These observations were confirmed statistically. Left–right symmetry was assumed in a statistical analysis of the *c* values (e.g., by combining data for conditions **L**-R and **R**-L). For each subject, the pairs of conditions whose differences are plotted in Fig. 9 were compared using one-tailed, paired

performed either using headphones (Zurek, 1980; Shinn-Cunningham *et al.*, 1993) or in free field (Leakey and Cherry, 1957; Snow, 1954; Haas, 1951; Litovsky *et al.*, 1997). Although there are few existing parametric data for comparison, current results are generally consistent with previous reports: localization dominance is strongest at delays of 1–5 ms and weakens thereafter (

mechanisms that are involved in ongoing assessment of room acoustics (Clifton and Freyman, 1997). Yang and Grantham (1997a) found that fusion is more susceptible than discrimination to the build-up of precedence and concluded that the mechanisms mediating these two aspects of precedence are different.

Our study was not aimed at investigating aspects of the build-up effect. Both discrimination and fusion experiments presented three lead/lag intervals in each trial; however, in the fusion experiment, all three intervals were identical, while in the discrimination (

gration will be detrimental if spatial information from a lagging source is combined with information from a leading source, particularly if the spectral content of the lead and lag differs.

One interpretation of these results is that the precedence effect is a general process that enables robust localization not only in the presence of echoes, but whenever any competing information from a second source arrives before the direction of a previous source has been computed. This view suggests that echo suppression is a special case of a more general computational mechanism in the spatial auditory pathway for suppressing any information that could be disruptive to spatial auditory perception. In addition, the results suggest that the mechanisms underlying the three phenomena described here might have some general commonality, not merely at the initial stages of processing, but at later stages as well.

The current results lend further support to this view. Results from the localization dominance experiment indicate that the strength of the precedence effect as measured in a localization dominance task varies with spatial separation of lead and lag, consistent with a general mechanism for improving sound localization. Although there are links among fusion, discrimination, and localization dominance, further work is necessary to quantify how these measures relate to one another.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R29-DC03083 to R. Y. Litovsky and ONR MURI award No. Z883402. The authors are grateful to Adam Stein and Daniel Singer for their invaluable help in programming and data collection. Dr. Miriam Valenzuela, Dr. D. Wesley Grantham, and two anonymous reviewers provided helpful feedback on an earlier draft of this paper. Portions of these data were presented at the meeting of the Association for Research in Otolaryngology, 1998.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization*, revised ed. (MIT Press, Cambridge, MA).

Blauert, J., and Divenyi, P. (**1988**) ''Spectral selectivity in binaural contralateral inhibition,'' Acoustica