Spatial consistency as a cue for segregation and localization


*Corresponding author's address: Ball Aerospace & Technologies Corp., Fairborn, OH 45324, eric.thompson.ctr@wpafb.af.mil

Many real-world auditory scenes are dynamic and complex, with multiple sounds that may change location over time. In this experiment, we examined the ability of listeners to localize a spatially-consistent target sound in a dynamic, spatially-varying auditory scene. The target and masker stimuli were composed of sequences of 60-ms bursts of uncorrelated noise (2 to 16 bursts in duration) and differed only in their degree of spatial consistency. Specifically, each target burst within a sequence came from the same spatial location (which varied from trial to trial), whereas each masker burst within a sequence came from a different, randomly chosen spatial location. The listener's task was to localize the spatially-consistent sequence. Localization errors decreased by approximately 11° with each doubling of the sequence duration, and approached quiet performance with 16-burst sequences. Adding a second masker increased localization errors by approximately 14° overall. These results suggest that spatial information can be combined across multiple observations over time to identify and localize a spatially-consistent target in a dynamic auditory scene. These data will be discussed in terms of the information obtained from each burst and the manner in which the information is combined across bursts.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Listeners in the real world are routinely charged with the task of interpreting meaning from complex, dynamic acoustic environments containing multiple concurrent sounds. In order to accomplish this task, listeners combine information across a number of stimulus dimensions, and this information may be linked across time (streamed) to form auditory objects that can then be segregated from competing sounds in the environment (e.g., Bregman, 1990). Consistency of information in one particular dimension (e.g., spectral content, modulation rate) across time facilitates the formation of auditory objects and serves as a segregation cue, and this may be particularly true when other elements in the environment are varying along that same dimension (see, e.g., Kidd et al., 1994). In this study, we examined the degree to which spatial consistency could be used to both define a target stimulus and support the segregation of this stimulus from a background of spatially-varying maskers in order to facilitate the localization of the target.

METHODS

Subjects

There were a total of 6 listeners (3 males, 3 females), ages 20-30 (mean = 24) years, drawn from a panel of long-term, part-time (20 hr/wk) listeners, who participated in this study. All had clinically normal hearing (i.e., thresholds ≤15 dB HL from 0.125-8.0 kHz) and all were paid for their participation in this study. Each of the listeners had previously participated in other experiments on sound localization.

Apparatus

The study was conducted in the Auditory Localization Facility (ALF) at the Air Force Research Laboratory at Wright-Patterson Air Force Base (Figure 1A). This facility consists of an anechoic chamber, the walls, floor, and ceiling of which are covered with 1.1-m thick fiberglass wedges to reduce echoes. Housed within this chamber is 4.3-meter diameter geodesic sphere with 277 Bose 11-cm, full-range loudspeakers mounted on its surface; however, only those loudspeakers located above -45° in elevation were employed in this study. A cluster of four light-emitting diodes (LEDs) was mounted on the front of each loudspeaker; these LEDs were used to provide pre-trial feedback regarding the orientation of the listener’s head and post-trial response feedback indicating the actual target locations. Situated in the center of the sphere is a bench with a chinrest (used to stabilize the listener’s head position during stimulus presentation), fastened to a height-adjustable platform, allowing the experimenter to position the listener’s head in the center of the loudspeaker array (Figure 1B). An ultrasonic tracking system (Intersense IS-900) was used to measure the position and orientation of the listener’s head as well as the position of a handheld wand, which was used by the listener as a pointing/response device.

FIGURE 1. A) The geodesic sphere in the Auditory Localization Facility at Wright-Patterson Air Force Base. Loudspeakers mounted in metal enclosures are positioned at each of the vertices. B) Listener seated on bench in the center of the sphere with his head held stationary on the chinrest. The listener is wearing a headtracking device and used a handheld position-tracking wand to point to the perceived location of the sound.
Stimuli

The target stimulus was composed of a sequence of 2, 4, 8, or 16 independently-drawn broadband (.2-14.5 kHz) noise bursts. Each burst was 60 ms in duration, with 5-ms $\cos^2$ on/off ramps. A total of 64 loudspeaker locations, roughly evenly distributed throughout the sphere, were used as target locations. (Note: Due to a technical issue, only 63 loudspeakers were used in the second experiment). The masker stimuli were composed of similar sequences of noise bursts with the same bandwidth, duration, and temporal structure as the target, and all masker bursts had onsets and offsets that were simultaneous with those of the target. The masker bursts could come from any of the loudspeakers employed in the study. The target and masker stimuli differed only in their degree of spatial consistency. Specifically, each noise burst within a target sequence came from the same spatial location (which varied from trial to trial), whereas each noise burst within a masker sequence came from a different, randomly chosen spatial location. All stimuli were presented at the same level (i.e., the target-to-masker ratio was equal to 0 dB).

Procedure

On each trial, the listener task was to locate the loudspeaker from which a noise burst was presented on each interval of the sequence (i.e., the target was the spatially consistent sound). At the beginning of each trial, the listener placed her/his head in the chinrest, oriented to a loudspeaker at 0° azimuth, 0° elevation, and pressed a button on the wand indicating a readiness to begin the trial. The stimulus was then presented, after which the listener indicated the perceived location of the target by pointing the handheld wand at that location and pressing a button on the wand to register a localization response. Trial-by-trial feedback was provided to the listener by activating the LED at the actual target location.

On each block of trials, the condition was selected randomly, but only one sequence duration (2, 4, 8, or 16 bursts) and one masker number condition (1 or 2 maskers) was tested in each block. A total of 320 trials were collected in each condition (5 trials at each of the 64 target locations) for each listener. Data in the target-alone (0-masker) condition were collected after completing data collection in the target-plus-masker(s) conditions.

RESULTS

In Figure 2, overall angular localization errors, averaged across listeners, are plotted as a function of the number of noise bursts in the stimulus sequence for the 1-masker (white circles) and 2-masker (black squares) conditions. In addition, data are plotted for the baseline target-alone condition (gray triangles). The average localization error in this baseline condition was approximately 10°, independent of the number of bursts in the sequence. This level of performance is comparable to previous results from our laboratory (e.g., Brungart & Simpson, 2009) and suggests that when the target is presented in isolation, listeners can obtain sufficient spatial information from just two noise bursts to accurately localize the sound.

Localization performance in the masked conditions varied substantially as a function of sequence length. In the 1-masker condition, localization errors were found to decrease systematically (i.e., by roughly 9-12° degrees) with each doubling of sequence length, approaching performance found in the baseline condition with a 16-burst sequence. However, even with only two bursts, listeners were performing much better than chance. Similarly, performance in the 2-masker condition improved systematically with increasing sequence duration at a similar rate. However, errors in this condition were elevated by roughly 14° overall relative to the 1-masker condition. These results suggest that listeners are able to accumulate information about the auditory scene over time in order to both identify the target (i.e., spatial consistency is its only defining feature) and estimate its location. As the number of intervals increases, the listener is presumably better able to distinguish between the spatial consistency of the target and the spatial variability of the masker. In addition, the listener gets “multiple looks” at the target (Swets, Shipley, McKey, and Green, 1959) and is thereby able to localize it more accurately. Although the relatively flat curve in the target-alone condition suggests that the multiple looks is not a major factor, it is likely that a ceiling effect limited the improvement in this condition.
FIGURE 2. Overall angular localization errors plotted as a function of the number of noise bursts in the stimulus sequence. The parameter in the figure is the number of maskers presented (0, 1, or 2). Error bars represent ± 1 standard error of the mean across listeners.

To illuminate the relative importance of the two factors that may be contributing to the improvement in performance found with increasing sequence duration (i.e., more chances to compare the relative spatial consistency of the target and the masker, and more “looks” at the target location), a second experiment was conducted in which we fixed the number of bursts in a sequence at 8, employing the same target structure as before (i.e., a sequence of spatially-consistent noise bursts). However, in this experiment we varied the number of masker intervals, keeping the overall sequence duration constant. Specifically, a two-interval masker condition was one in which the first four bursts of the sequence came from one location and the subsequent 4 bursts came from a second location. Similarly, a four-interval masker consisted of 4 pairs of bursts, each pair coming from a different spatial location. In this way, the number of “looks” at the target was fixed, but the number of masker locations within the sequence varied (and thereby the opportunities to estimate its spatial variability).

As was the case in Experiment 1, the experimental condition was selected randomly on each block of trials, but only one masker interval condition (2, 4, or 8 masker intervals) and one masker number condition (1 or 2 maskers) was tested in each block. And, as before, 5 trials were collected in each condition for each of the possible target locations.

In Figure 3, overall angular localization errors, averaged across listeners, are plotted as a function of the number of masker intervals in the stimulus sequence for the 1-masker (white circles) and 2-masker (black squares) conditions. In both conditions, localization errors were found to decrease systematically as the number of masker intervals increased, even though the overall sequence duration (and therefore the number of looks at the target) remained constant. That is, the only stimulus characteristic that was changing was the spatial variation of the masker(s). Thus, it appears that the relative target-masker spatial consistency is a strong segregation cue.
FIGURE 3. Overall angular localization errors plotted as a function of the number of masker intervals. The parameter in the figure is the number of maskers presented (1 or 2). Error bars represent ± 1 standard error of the mean across listeners.

DISCUSSION

Despite the fact that the target and masker stimuli employed in this task were very similar, differing only in their relative spatial consistency over time, listeners nevertheless demonstrated a clear ability to perform the task. Specifically, they were able to identify a target based on its spatial consistency and report its location. Experiment 1 demonstrates that this ability improves reliably as the number of bursts in the sequence increases. Experiment 2 suggests that this improvement is related not to the fact that the listener gets more opportunities to estimate the target location, but rather to the fact that listeners have more chances to compare the target and masker spatial consistency. One issue that makes these data difficult to interpret is the fact that the one defining feature of the target, a consistent location, is also the response in the task (i.e., localize the target). Future studies will attempt to address this issue by employing a detection response and varying the spatial variability of the target and the masker.

ACKNOWLEDGMENTS

This work was sponsored by a grant from the Air Force Office of Scientific Research.

REFERENCES