

EFFECTS OF TEMPORAL FINE STRUCTURE ON THE LOCALIZATION OF BROADBAND SOUNDS: POTENTIAL IMPLICATIONS FOR THE DESIGN OF SPATIAL AUDIO DISPLAYS

Douglas S. Brungart, Brian D. Simpson

Air Force Research Laboratory, WPAFB, OH
douglas.brungart@wpafb.af.mil

ABSTRACT

In the design of symbology for spatial audio displays, a great deal of emphasis has been placed on the importance of bandwidth for achieving robust localization accuracy. However, there are some cases where the temporal characteristics of a audio signal can have a large impact on its localizability. In this study, we examined localization accuracy for three different wideband stimuli: a broadband noise, a 100-Hz click train, and a 100-Hz click train with randomized phase. When the stimulus presentation level was low (40 dB SPL), localization performance was reasonably good for all three stimuli. However, as the presentation level increased to 70 dB or higher, localization performance degraded dramatically for the click train stimulus but remained roughly constant for the other two stimuli. The results suggest that display designers must consider some factors other than bandwidth when they design the symbology for real-world spatial audio displays.

1. INTRODUCTION

As spatial audio displays continue to make the inevitable migration from research systems in laboratories to practical applications in operational systems, the design of auditory symbology will become an increasingly important topic for audio scientists and engineers. In real-world applications, this will almost certainly require designers to select relatively large sets of sounds that are both highly localizable and easily distinguished from one another. Thus, it will be vitally important for display designers to have some guidelines about the acoustic properties that contribute to the localizability of a particular type of sound.

At this point, virtually all auditory researchers would agree that the most important factor in determining the localizability of a sound is the bandwidth of the stimulus. If all other factors such as stimulus duration and presentation level are held constant, then there is a general understanding that accurate auditory localization requires a sound source with a bandwidth of at least 12-13 kHz [1]. Indeed, a substantial portion of the limited research that has been done in

the area of spatial auditory symbology has focused on extending the bandwidth of current auditory warning tones to allow improved auditory localization [2, 3]. However, beyond this requirement of sufficient bandwidth, few guidelines are available for predicting the localizability of an arbitrary sound. Thus, a reasonable approach to designing auditory symbology would be to select sound sources that have equivalent bandwidth but are as perceptually distinct from one another as possible.

In a recent study in our laboratory, we were examining a listener's ability to localize one of two simultaneous sounds. Such a task would likely be required in most complex environments where auditory displays are employed. For our two stimuli, we selected two sounds that clearly met the criteria of high bandwidth and perceptual distinctness: a broadband noise, and a 100-Hz click train. *A priori*, we expected these two stimuli to produce similar localization performance when presented in isolation. However, contrary to our expectations, we found that localization of the click train stimulus was much worse than localization of the noise stimulus. Initially, we suspected that a methodological problem with the stimulus generation may have accounted for the discrepancy. However, after a series of careful acoustic measurements failed to identify any discrepancies, we began to wonder if some property of the stimulus itself might account for the results.

Although few studies have explicitly compared localization performance with click trains and noise, a small number of studies have shown that the localization of very brief sounds, such as individual clicks, can be severely impaired when the level of the stimulus is very high [4, 5, 6]. These results suggested to us that perhaps the performance difference observed between the localization of click trains and noise might be related to the level of the stimulus. In order to address this question, an experiment was conducted that compared the localization of noise, click trains, and random-phase click trains as a function of the overall level of the stimulus.



Figure 1: Auditory Localization Facility (ALF) where the experiment was conducted

2. METHODS

2.1. Participants

A total of 9 paid volunteer listeners participated in the experiment. All had normal audiometric thresholds, and their ages ranged from 22 to 23.

2.2. Apparatus

The experiment was conducted in the Auditory Localization Facility (ALF) at Wright-Patterson Air Force Base (see Figure 1). This facility consists of a geodesic sphere (4.3m in diameter) with 277 Bose 11-cm, full-range loudspeakers mounted on its surface, and a small cluster of 4 LEDs mounted on the front of each loudspeaker. The sphere is housed within an anechoic chamber, the walls, floor, and ceiling of which are covered in 1.1-m fiberglass wedges.

2.3. Stimuli

Figure 2 shows the time and frequency domain representations of the three stimuli used in the experiment. The left panels show the 'Burst Noise' stimulus. This stimulus was a white noise signal that was bandpass filtered between 200 Hz and 16 kHz. The middle panels show the Burst Click stimulus, which consisted of a 100-Hz sequence of 22.6 μ s pulses. The right panel shows the 'Random-Phase Click' stimulus. This stimulus was generated by using an FFT to randomize the phase of the 'Burst Click' stimulus. This produced a signal with exactly the same frequency content

as the Burst Click stimulus, but a much different temporal structure. Each stimulus was designed to be 250 ms in length, with 25 ms \cos^2 rise and fall ramps.

Prior to each stimulus presentation, these signals were processed with a 128-pt equalization filter to correct for the individual differences in the loudspeakers in the ALF. Then they were scaled to produce one of five different overall levels at the location of the listener's head (40, 50, 60, 70, or 80 dB SPL). In order to verify that there were no level-related distortions in the acoustic signal generated by the ALF speaker system, a probe microphone (Etymotic ER-7C) placed in the center of the sphere was used to make a recording of each stimulus waveform at each presentation level used in the experiment. This analysis indicated that there was no significant harmonic distortion in any of the stimulus types until the level of the acoustic presentation exceeded 80 dB SPL. Figure 3 shows a time-domain recording of the burst click stimulus presented at an overall level of 80 dB SPL.

2.4. Procedure

The localization judgments were collected with the listener standing in the center of the ALF. Prior to the start of each trial, a headtracker (Intersense IS-900) mounted on an adjustable headband worn by the listener was used to measure the orientation of the listener's head and ensure that the listener's head was in a stationary position. Then a short stimulus was presented from a randomly selected speaker in the ALF facility. The listener was then asked to point a handheld wand, also tracked by the IS-900 tracker, in the direction of the perceived location of the sound. A cursor, slaved to this handheld wand, turned on the LEDs at the speaker location to which the wand was pointed. The listener was therefore able to use the wand to move the LED cursor to the perceived speaker location of the stimulus, and press a button to make a response. Feedback was then provided by a second LED that was then turned on at the actual location of the stimulus. Once the location of the feedback LED was determined, the listener pressed a button on the LED to start the next trial.

Each listener participated in a minimum of 40 trials for each combination of stimulus type and presentation level, for a minimum of 600 total trials for each of the nine listeners.

2.5. Results

Figure 4 shows the results in terms of the mean great circle error between the actual location of the stimulus and the location of the response in each trial. From these data, it is apparent that the overall level had a strong influence on the localizability of the Burst Click stimulus, but that level had very little effect on the localization of the other

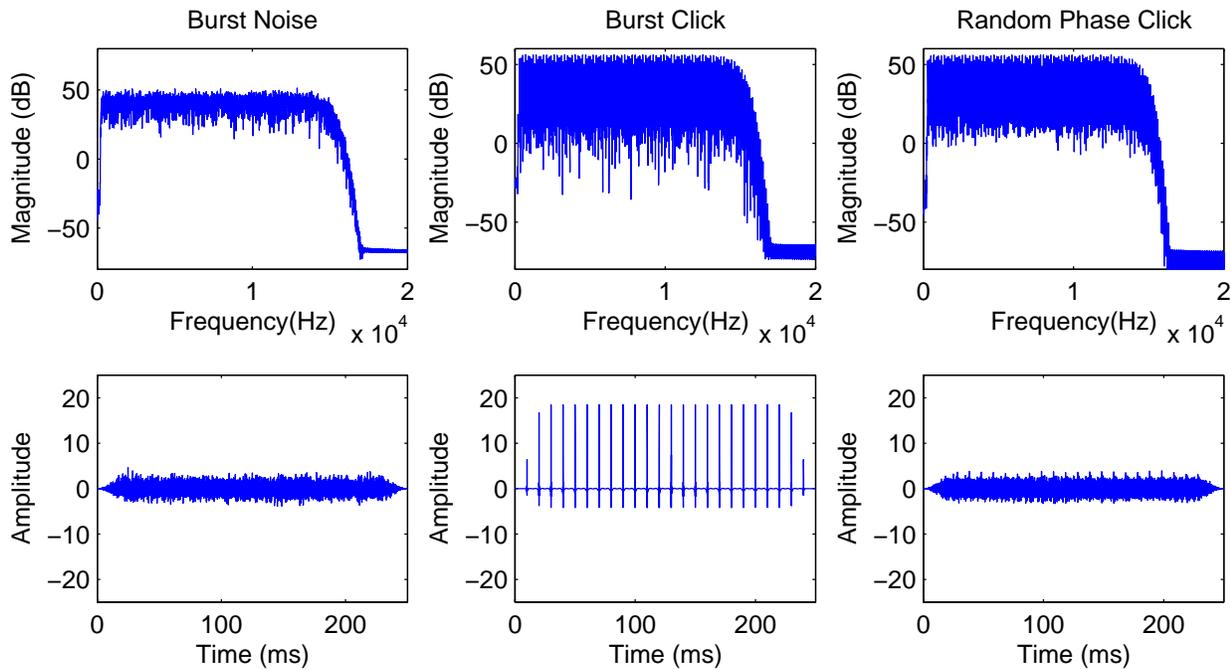


Figure 2: Comparison of magnitude spectra (top panels) and time-domain waveforms (bottom panels) of the three stimuli used in the experiment.

two types of stimuli. When the stimulus presentation was 40 dB SPL, the mean great circle error was lowest for the noise stimulus (roughly 12° mean error), slightly higher for the random click stimulus (roughly 15° error), and highest with the Burst Click stimulus (roughly 20° error). Thus, we see that, even at very low presentation levels, performance was substantially better for the Random-Phase stimuli than for the Burst Click stimulus. However, as the presentation levels increased, the difference in performance between the Burst Click stimuli and the other two stimuli increased dramatically. Even more surprisingly, these increases occurred at relatively modest sound levels that would be expected to occur frequently in the everyday experience of almost all listeners. By the time the stimulus reached a level 60 dB SPL (approximately the level of a quiet conversational voice), the localization error with the Burst Click stimulus had increased almost 25% relative to the 40 dB SPL condition. When the stimulus level reached 80 dB SPL (a level approximating that of a busy street), the mean angular error for the Burst Click condition was nearly 3 times as large as that in the Burst Noise condition.

Often in localization tasks it is useful to incorporate an analysis that distinguishes between localization errors in the left-right dimension, which are based on interaural difference cues and tend to be relatively robust even for low-frequency stimuli, and those in the up-down or front-back

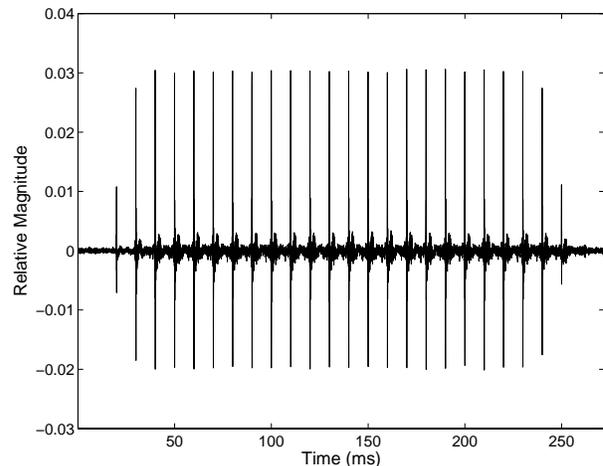


Figure 3: Time domain recording of a burst click stimulus measured at the center of the ALF facility at a presentation level of 80 dB SPL.

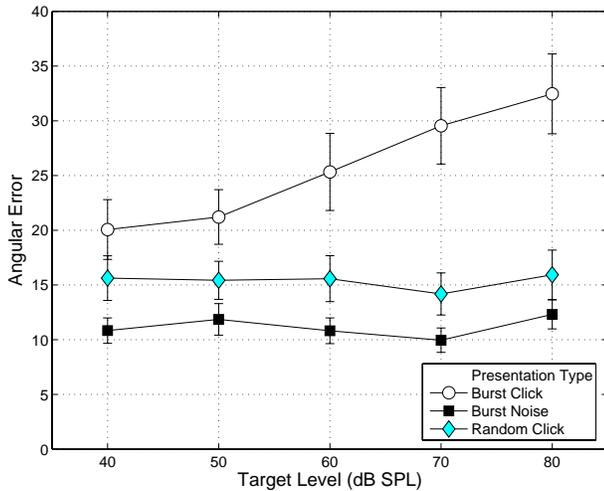


Figure 4: Overall great circle error for each combination of stimulus type and presentation level in the experiment. The error bars indicate the 95% confidence intervals for each condition.

dimension, which are based on high-frequency pinna cues and tend to be more sensitive to stimulus bandwidth. Figure 2 5 shows performance in the task in the horizontal polar coordinate system, as defined by Macpherson [5]. The left panel of the figure shows performance in terms of the mean angular error in the left-right dimension on each trial, and the right panel in the figure shows performance in terms of the mean vertical-polar angle on each trial. The small diagrams to the right of each panel illustrate how these dimensions were calculated, based on the location of the target and response locations relative to the position of the listener’s head at the start of each trial.

These two panels clearly show that the level-dependent degradation in localization performance that occurred for the Burst Click stimulus is almost entirely the result of an increase in error in the vertical-polar dimension. In that dimension, a substantial degradation in performance occurred for the Burst Click stimulus when the presentation level increased from 40 dB SPL to 50 dB SPL, and it continued to increase at a rate of approximately 0.5° per dB SPL until it reached 42° at a presentation level of 80 dB SPL. In the left-right dimension, the differences in performance across the different stimulus conditions were relatively much smaller and there was only a very minor level dependence in the error for the Burst Click condition.

3. CONCLUSIONS

From the results of this experiment, it is apparent that bandwidth alone is not a sufficient metric for predicting the lo-

calizability of an arbitrary sound. Indeed, the results from the Burst Click and Random-Phase Click conditions clearly show that even two stimuli with nearly identical spectral content can result in quite different levels of localization accuracy.

It is also clear that these differences are not limited to extremely loud stimulus presentations. Large differences in performance across different stimulus waveforms can start to occur at presentation levels that are no louder than a noisy conference room. When one considers that many of the proposed application areas for spatial audio displays involve inherently noisy environments like aircraft cockpits, it is clear that the operational utility of a spatial audio display could be severely compromised by the selection of a seemingly innocuous signal such as a periodic click train.

Unfortunately, it is not clear at this time how to generalize the results of this experiment much beyond the specific click-train stimuli tested in the study. Certainly there is substantial evidence to suggest that acoustic stimuli based on clicks should be avoided to the greatest extent possible in the design of spatial audio symbology. However, without more research, it is hard to predict what other kinds of stimuli might also result in this kind of level dependent effect. At this point, it appears that auditory display designers must be cautious when selecting novel broadband sounds that have not been explicitly tested for localization accuracy, particularly those sounds that consist of a periodic train of short burst or clicks.

4. REFERENCES

- [1] R.B. King and S.R. Oldfield, “The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional auditory displays,” *Human Factors*, vol. 39, pp. 287–295, 1997.
- [2] R.D. Patterson and A.J. Datta, “Extending the frequency range of existing auditory warnings,” in *AGARD Conference Proceedings 596: Audio Effectiveness in Aviation*. 1997, pp. 8:1–8:8, Copenhagen, Denmark.
- [3] R.L. Martin, R.A. Parker, K. McAnally, and S.R. Oldfield, “The abilities of listeners to localise defense research agency audio warnings,” in *AGARD Conference Proceedings 596: Audio Effectiveness in Aviation*. 1997, pp. 9:1–8:7, Copenhagen, Denmark.
- [4] W. M. Hartmann and B. Rakerd, “Auditory spectral discrimination and the localization of clicks in the sagittal plane,” *Journal of the Acoustical Society of America*, vol. 94, pp. 2083–2092, 1993.
- [5] E.A. Macpherson and J.C. Middlebrooks, “Localization of brief sounds: Effects of level and background noise,”

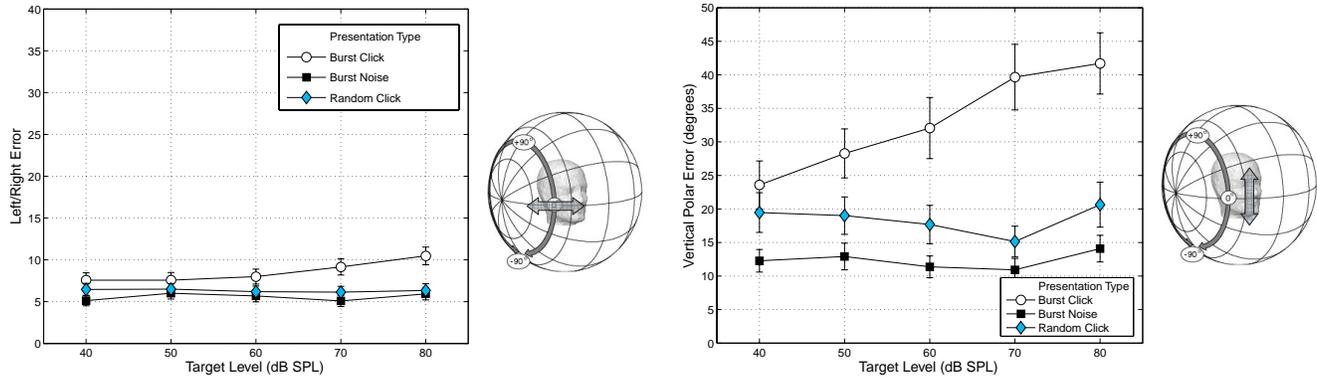


Figure 5: Left-right error (left) and vertical-polar error (right) for each combination of stimulus type and presentation level in the experiment. The error bars indicate the 95% confidence intervals for each condition.

Journal of the Acoustical Society of America, vol. 108, pp. 1834–1849, 2000.

- [6] P. M. Hofman and J. V. Opstal, “Spectro-temporal factors in two-dimensional human sound localization,” *Journal of the Acoustical Society of America*, vol. 103, pp. 2634–2648, 1998.