OPERATIONAL CONSTRAINTS ON THE UTILITY OF VIRTUAL AUDIO CUEING

Mark A. Ericson, Robert S. Bolia, & W. Todd Nelson
Air Force Research Laboratory, Wright-Patterson AFB, OH 45433

The potential of virtual audio display technology to provide operators with veridical spatial cues may be substantially constrained by factors that are common in many operational settings – i.e., high noise level, limitations in the bandwidth of the audio source and/or display. The purpose of this study was to examine the effects of varying the bandwidth of a virtual sound source in the presence of broadband noise in a reverberant environment. Specifically, the signal-to-noise ratio (SNR) was varied from +50 dB to -10 dB, and the signal was low pass filtered at 1.6 kHz, 4 kHz, 8 kHz, and 15 kHz. Correlational analyses comparing actual and perceived sound source location revealed that both signal bandwidth and signal-to-noise ratio influenced auditory localization acuity, and that even under optimal bandwidth and noise conditions (15 kHz and +50 dB) localization in elevation was extremely poor. These findings have numerous implications for the design of spatial audio displays, especially those meant to be used in noisy environments.

INTRODUCTION

King and Oldfield (1997) systematically varied the bandwidth of signals to determine the minimal frequency composition of a signal, which provides optimal localization acuity. Signals with low pass cutoff frequencies of 13 kHz enabled listeners to best determine the azimuth and elevation of signals and reduce the number of front to back reversals. Signals with a high pass cutoff frequency of 9 kHz minimized the front to back reversal rate in azimuth. Ideally, such wide bandwidth signals should be used when possible to optimize performance with directional audio displays. While King and Oldfield's results have important implications for the design of effective spatial audio displays, their empirical study did not address the potential effects of noise and/or the interaction of noise and reduced bandwidth on virtual audio cueing.

Most free-field binaural masking experiments have used a single directional masker and a single directional signal in an anechoic environment (Gilkey and Good, 1996). The presence of a directional masker tends to "push" or "pull" the perceived location of the signal. A single masker is the least complex of all possible masking conditions. At the other extreme of masking complexity is the case of an infinite number of maskers presented all around a listener. A reverberation chamber can approximate a spatially diffuse masking condition over a wide range of frequencies, typically from 100 Hz to 8 kHz. Hirsh (1950) measured auditory localization acuity of human listeners in highly reverberant environments. The directional signals and masker(s) were located together in the reverberant environment. In general, such signals are more difficult to localize and are more easily masked than in free-field listening conditions. Most real-world listening environments fall between anechoic and highly reverberant conditions.

The advent of virtual audio technology provides as many opportunity to present directional signals over headphones while a listener is immersed in some ambient noise environment. The purpose of present study was to assess the effects of ambient masking noise, as found in many operational environments, on a listener's ability to identify the direction of a virtual sound source presented over headphones. Additionally, these data are meant to be compared to empirical results reported in the literature, as summarized in Table 1.

Method

Participants. Three male and three female participated in these experiments. Participants had normal hearing threshold levels, localization acuity within 30 degrees precision. Participants were paid for their participation.

Experimental Design. Two experiments were conducted with the same design. The first study used a 300 ms noise stimulus with no head tracking. The second experiment used a continuous noise source with head tracking. The following description applies to both experiments. Six listeners participated in this within subjects, factorial design. Five signal-to-noise-ratios, (~50, 10, 0, -5, -10 dB), were employed and four low pass filter cutoff frequencies at 1.6, 4, 8 and 15 kHz values were employed. A total of 37 locations were used. Twenty-four of the 37 target angles were equally distributed, eight each, along the median, frontal and transverse planes. Five orthogonal vertexes of front, back left, right, and zenith were chosen. The eight remaining locations were symmetrically distributed at the locations of ± 45° azimuth and ± 37° elevation. The volunteer listeners were randomly assigned to one of six blocks to reduce order effects due to possible learning of the task. Each listener responded to 5 signal to noise ratios X 4 cutoff frequencies X 37 angles X 5 repetitions = 3,700 data points.

Apparatus. All experiments were conducted in facilities of the Air Force Research Laboratory's Aural Displays and Bioacoustics Branch at Wright-Patterson Air Force Base. The 8000 ft^3 reverberation chamber and sound system of the voice

Author(s)	Stimulus	Response Method	Acuity (°) Dispersion1/K Reversal (%)	Listening Environment	Virtual Audio Synthesis Methods		
					Hardware	HRTF's	Headtracker/ Head Motion
Gilkey, Good, Ericson, Brinkman & Stewart	100 Hz-25µs pulse train, 268 ms long, .53 – 11 kHz	Sphere Point	r-squared .989 - .993 Az .75 -. 83 El	Anechoic Chamber	**NA**	NA	None/ Bite bar
Gilkey and Anderson	Clicks .4 – 11 kHz	Sphere Point	15° L/R 22.6° F/B 20.6° U/D	Anechoic Chamber	NA	NA	None/ Bite bar
King & Oldfield	Filtered white noise	"gun" point	Graphical representation	Anechoic	NA	NA	None/ Instructed to hold head still
Ricard and Meiers	Speech	Verbal coordinates	16° Az 19% F/B	Masking noise over headphones	Convolvotron	SDO	Polhemus 3-Space/ full head motion
Oldfield & Parker	White noise	"gun" point	9.1° Az 8.2° El	Anechoic	NA	NA	None/ Chin Rest start-Unrestricted
McKinley, Ericson and D'Angelo	Pink Noise .1 – 10 kHz	Nose point	6 - 7° Az <1% F/B	Reverberant masking noises	ALCS	KEMAR @ 1° azimuth	Polhemus 3-Space/ full head motion
Valencia, Calhoun, Ericson	Pink Noise .1 – 10 kHz	Circle Point	18° Azimuth 31.8% F/B	Quiet laboratory	ALCS	KEMAR @ 1° azimuth	None
Wenzel , Arruda, Kistler, Wightman	8, 250 ms noise bursts	Verbal coordinates	.53-.79 Az .33-.77 El 31% F/B 18 % U/D	Headphone	Aerial DSP-16	SDO, Non-individual	None/ Instructed to hold head still
Wightman & Kistler 1989	8, 250 ms noise bursts	Verbal coordinates	.95-.99 Az .43-.94 El 11% F/B	Anechoic	VAX-11/750	Individual	None/ Head rest
Wightman & Kistler 1990	8, 250 ms noise bursts	Verbal coordinates	21.9° .07 13.9%	Anechoic	VAX-11/750	Individual	None/ Head rest

Table 1. Summary of existing empirical localization data.

communications research evaluation system (VOCRES) facility were used to generate an ambient noise field at 95 dBA SPL. Two General Radio 1382 random noise generators produced the non-correlated ambient masking and directional target noises. The masker noise was played over the VOCRES sound system. The directional target noise was played through a Hewlett-Packard low pass filter set, Wilsonics programmable attenuators, directionally encoded either by a Tucker-Davis Technologies (TDT) Digital Signal Processing System (Experiment 1) or by an Air Force Research Laboratory 3-D Auditory Display Generator (3-D ADG; see McKinley, Ericson and D'Angelo, 1994 for detailed description) (Experiment 2), and presented to listeners over Sennheiser HD 560-II headphones.

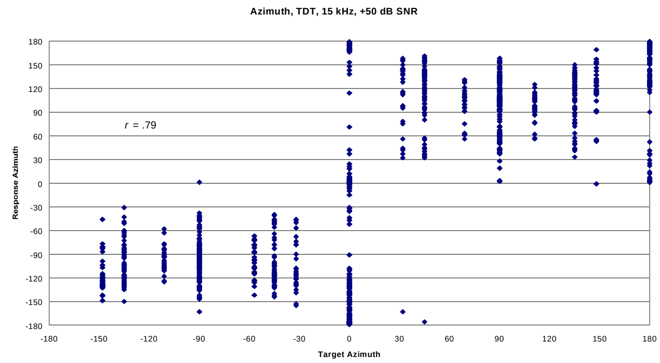
Head motion cues were provided via a Fast-Trak electro-magnetic tracking system coupled to the 3-D ADG. The Fast-Trak measured the orientation of the listeners' head 120 times per second with 8 ms latencies to enable a space-stabilized auditory image of the target noise. Listeners reported the perceived direction of the target noise using the GELP (Gilkey et al., 1995) spherical pointing technique. The GELP technique employs a hand held stylus to record azimuth and elevation responses automatically via the Fast-Trak system.

Procedure. The participants were seated inside the VOCRES reverberation chamber and provided with the headphones, head tracking sensor and stylus. Listeners were instructed to keep their head level and facing towards the front before stimulus presentation. For both experiments, head motion was unrestricted while pointing with the stylus. After responding, listeners were instructed to return to the starting position. Each session included 185 stimulus presentations and lasted approximately twenty minutes.

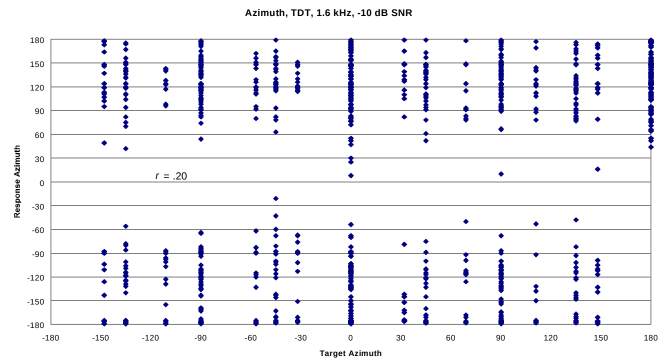
Results

Results from both experiments are displayed in Figures 1 and 2. In Figure 1, response azimuth is plotted as a function of target azimuth for the extreme SNR and bandwidth conditions in each experiment. Figures 1a and 1b represent data collected under conditions of a +50 dB SNR and a 15 kHz bandwidth, while Figures 1c and 1d represent data collected under conditions of a –10 dB SNR and a 1.6 kHz bandwidth. In Figure 2, response elevation is plotted as a function of target elevation for the same conditions. Pearson product moment correlation coefficients were computed for all eight conditions, and are displayed in each of the eight scatter plots.

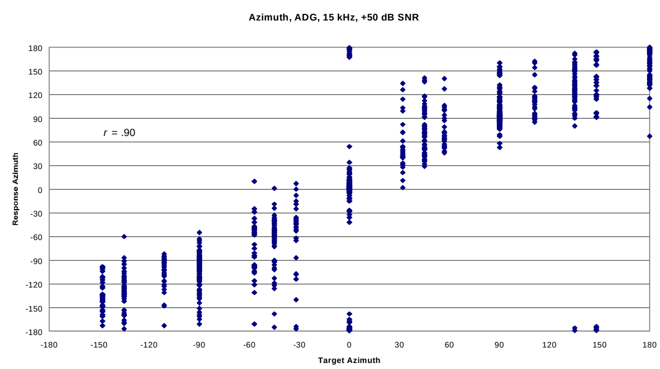
It is evident in Figures 1 a) and c) – broadband signal, no noise - that perceived and actual sound source locations were highly correlated, with the exception of sound sources that were located on the median plane (i.e., 0° and 180°). This latter result can be explained by noting that the data illustrated in these figures were NOT corrected for front-back confusions. In contrast, Figures 1 b) and d) demonstrate the deleterious effect of limited bandwidth and noise on



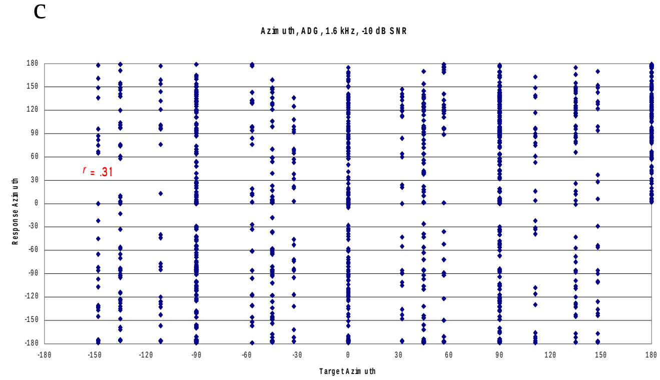
a



b



c



d

Figure 1. Response azimuth as a function of target azimuth (Figs 1a-b represent data from Experiment 1; Figs 1c-d represent data from Experiment 2).

localization in azimuth. Localization in elevation, on the other hand, was inferior in all conditions, as evidenced by Figures 2 a-d.

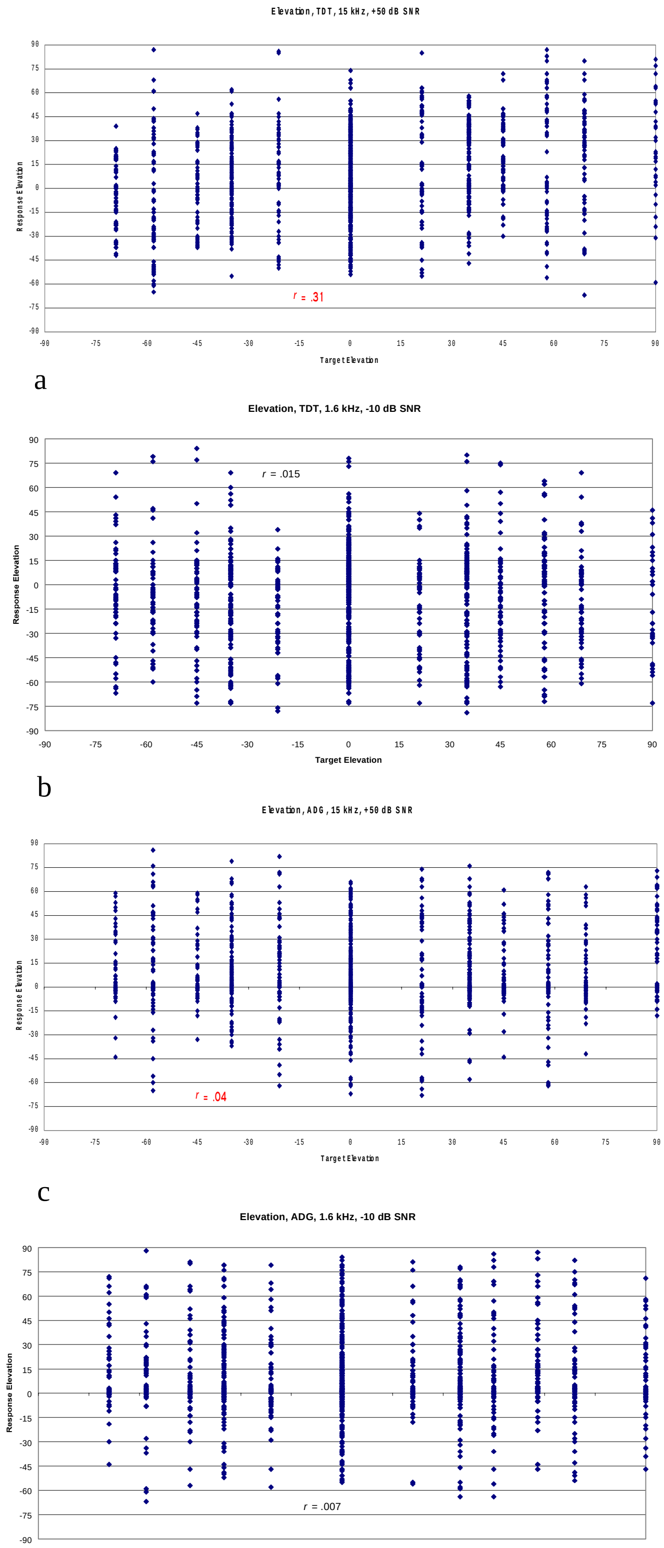


Figure 2. Response elevation as a function of target elevation (Figs 2a-b represent data from Experiment 1; Figs 2c-d represent data from Experiment 2).

CONCLUSIONS

The results of the experiments described herein clearly demonstrate the efficacy of existing virtual audio displays for the presentation of veridical cues to the locations of sound sources in the horizontal plane. What is equally clear is that this is not the case in the vertical plane. Indeed, the distribution of responses and the range of correlation coefficients (.31 $\leq r \leq$ .007) indicates that participants localized poorly in elevation even under optimal noise and bandwidth conditions. Given that the displays employed in these investigations represent the state of the art in spatial audio technology, it makes sense for designers to consider, pending further technological developments, constraining their displays such that only the azimuth is cued.

References

Doll, T., Gerth, J., Engelman, W. and Folds, D. (1986). Development of simulated directional audio for cockpit applications (AAMRL-TR-86-014). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Gilkey, R., Good, M., Ericson, M., Brinkman, J. and Stewart, J. (1995). A pointing technique for rapidly collecting localization responses in auditory research, *Behavioral Research Methods, Instruments and Computers*, **27**,1-11.

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

Gilkey, R. and Anderson, T. (1995). The accuracy of absolute localization judgments for speech stimuli, *Journal of Vestibular Research*, **5**, 487-497.

Hirsh, I. J. (1950). The relation between localization and intelligibility, *J. Acoust. Soc. Am*., **22**, 196-200.

King, R. B. and Oldfield, S. R. (1997). The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays, *Human Factors*, **39(2),** 287-295.

Oldfield, S. and Parker, S., Acuity of sound localisation: A topography of auditory space. 1. Normal Hearing conditions. *Perception*, **13**, 581-600, (1984).

McKinley, R, Ericson, M. and D'Angelo, W. (1994). 3-dimensional auditory displays: Development, applications, and performance, *Aviation Space and Environmental Medicine*, **65**(5, Suppl.): A31-A38.

Ricard, G. and Meiers, S. (1994). Intelligibility and localization of speech from virtual directions. *Human Factors*.**36**, 120-128.

Valencia, G., Calhoun, G., Ericson, M. and Agnew, J. (1990). Localization performance with synthesized directional audio, (AL-TR90-025) Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Wenzel, E., Arruda, M., Kistler, D. and Wightman, F. (1993). Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, **94**(1), 111-123.

Wightman, F. and Kistler, D. (1989). Headphone simulation of free-field listening II: Psychophysical validation. J. *Acoust. Soc. Am*., **85**, 868-878.

Wightman, F. and Kistler, D. (1990). Field measurement of head related transfer functions. (AAMRL-TR-90-019) Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.