3-D Audio: Military Applications and Symbology

Mark A. Ericson, Brian D. Simpson, and Richard L. McKinley
Air Force Research Laboratory (AFRL/HECB)
2610 Seventh Street, Bldg 441
Wright-Pattterson, AFB, OH 45433-7901

mark.ericson@wpafb.af.mil

3-D AUDIO SYMBOLOGY

Five experiments were conducted to study the acoustic attributes that enable the accurate identification and localization of rudimentary spatial warning sounds. In each experiment, two sounds were played simultaneously over loudspeakers at various azimuths and elevations. The stimuli consisted of pure tone complexes with a 13 kHz bandwidth. The fundamental frequencies, the amplitude modulation rate of the complex, the harmonicity of the carrier and modulation frequencies, and the coherence of the carrier and modulation phase were varied in the experiments. The combination of all the cues provided the best localization and identification performance. When and only when all cues were used, the subjects were able to accurately localize and identify the target sound.

1.0 INTRODUCTION

When warning sounds are played simultaneously over a single loudspeaker, they tend to combine into a single, jumbled warning clamour. In addition, the cacophony often tends to distract and annoy rather than inform an operator [1]. Alarm systems tend to be turned off by the operator, or never turned on, to avoid distractions under high-stress, high-workload conditions. The situation will potentially become more intractable if the current warning sounds, designed for single channel or monaural displays, are presented via spatial auditory displays.

Coding techniques have typically been employed to facilitate the identification of multiple warning sounds during monaural presentation over a headphone or over a single loudspeaker. Repetition rate has been found to be the most salient feature of warning sounds for purposes of identification [2] and for encoding urgency [3]. When the repetition rates of two sounds are similar, frequency coding and modulation can be used to segregate multiple warning sounds. However, such encoding techniques require the user to remember the association among a repetition rate, a frequency, a modulation and the airborne event that triggered the warning. Most pilots do not recall these associations accurately when asked to identify pre-recorded warning tones [4], [2], [5]. Presumably, identification performance would decrease further under high-workload conditions [6].

Naish [7] was one of the first to report the potential benefits of combining aircraft warning sounds with auditory lateralization cues. Using interaural phase and level differences, reaction times to left and right signals were found to be significantly faster when the perceived locations of the warning sounds correlated with the verbal direction of the sounds. Over the past sixteen years, many improvements have been made in 3-D auditory display technology. The accurate measurement and implementation of head-related transfer

Ericson, M.A.; Simpson, B.D.; McKinley, R.L. (2005) 3-D Audio: Military Applications and Symbology. In *New Directions for Improving Audio Effectiveness* (pp. 22-1 – 22-10). Meeting Proceedings RTO-MP-HFM-123, Paper 22. Neuilly-sur-Seine, France: RTO.
Available from: http://www.rto.nato.int/abstracts.aps.

functions provides many opportunities for integrating warning sounds into spatial aircraft display systems. Of particular importance to the military is the rapid and accurate display of threat warning information to the pilot.

Radar warning receivers typically display missile threats at one of four priority levels: 1) search, 2) scan, 3) alert, and 4) launch. Warnings are commonly coded by one or two tones. Search tones are usually a single pure tone with a relatively low frequency. Scan tones often sweep up and/or down in frequency. Alert tones, which signify a radar lock, alternate in pairs. Missile launch tones are represented by a single, high-frequency pure tone. Although radar warning receiver sounds are different for all aircraft, they can be roughly categorized by the preceding descriptions [8].

Experience with the integrated helmet auditory visual system (IHAVS) symbology demonstrated that pilots could effectively identify threat and priority levels for only one monaurally presented threat [5]. Two to four simultaneous radar warning receiver (RWR) sounds were not reliably identified. Spatial separation and the addition of uncorrelated noises enabled pilots to localize and identify up to four simultaneous warning sounds.

These spatial warning sounds were used in flight tests at Edwards Air Force Base to measure in-flight performance [9]. Although pilots could determine the location of four sounds in the laboratory, they reported being able to utilize only two sounds in flight because of the high workload. Daniels, Ericson, and French [10] described other airborne applications of combined 3-D audio visual displays.

Other techniques have been explored for extending the frequency range of current warning sounds for application in spatial auditory displays. Martin, Parker, McNally, and Oldfield [11] measured localization performance with rotary- and fixed-wing warning sounds. The sounds with the greatest number of pulses were more easily localized than the wide bandwidth sounds with fewer pulses. Therefore, the temporal structure of sounds is generally more important than signal bandwidth for high localization accuracy.

Patterson and Datta [12] extended existing warning sounds for use in spatial auditory displays. Three aspects of sounds were extracted and applied to wider bandwidth representations of similar sounds. These techniques included envelope filtering, Nyquist whistling, and fine structure doubling. Extending the frequency range of current warnings to 12 kHz was found to significantly improve localization accuracy.

In the current study, the amplitude modulation envelope was varied to impose a type of repetition rate onto the signals. The values chosen represent four regions: steady-state (0 Hz), slow temporal envelope (3 Hz), trill (10 Hz), and roughness (100 Hz). There are four parameters to vary in the amplitude modulation formula. These are the carrier frequency and phase and the modulation frequency and phase. This has the equivalent effect of adding noise to the spatial frequency, affecting the low-frequency resolution more than the high-frequency resolution. Unlike adding random noise, this technique maintains optimum signal strength for use in high noise environments.

2.0 METHODS

2.1 Equipment

The auditory localization facility at Wright-Patterson Air Force Base was used to conduct the experiments. The auditory localization facility was used for the presentation of free-field stimuli for multiple sound source localization and identification. The large set of loudspeaker locations in this facility reduced the possibility of location choice biases due to set size.

The God's-Eye Localization Procedure (GELP) [13] was used for collection of localization responses. In this technique, the subject responded after each stimulus presentation by positioning the tip of an electro-magnetic stylus at a point on the surface of a 20-cm plastic spherical model of auditory space to indicate the perceived direction of the auditory image. The subject's chin was restrained by a chin rest to reduce head motion cues.

The stimuli consisted of pure tone complexes with a 13 kHz bandwidth. The fundamental frequencies were in the range of 500 to 1000 Hz and are described for each experiment. The harmonicity of the carrier and modulation frequencies was either purely harmonic or randomized over a plus-or-minus five percent range. The coherence of the carrier and modulation phase was either completely in-phase or uniformly randomized over a two Pi range. All sounds were generated with Tucker-Davis-Technology System II equipment and a personal computer with Pentium II processor and a Windows 98 operating system.

2.2 Subjects

Ten naïve subjects were recruited from the general population. All had normal hearing and normal or corrected normal vision. The paid volunteers participated in all experiment conditions of the first and second experiments. Each volunteer subject had a normal hearing threshold levels and consented to participate in various listening experiments.

2.3 Procedures

The subjects were instructed to localize the center of the "auditory image" using the GELP technique. Two sequential presentations of the threat warning sounds were played before each pair of simultaneous sounds. Subjects were instructed to respond by indicating the locations in the order in which the sequential sounds were played immediately before the simultaneous sounds.

3.0 RESULTS

3.1 Steady State Pure Tone Complexes

In the first experiment, the two complexes differed in their fundamental frequencies by no difference, a musical fifth, a random non-musical interval, and a musical octave. The components were either regularly spaced (harmonic) or irregularly spaced (inharmonic). Data for this experiment are shown in Figure 1 below.

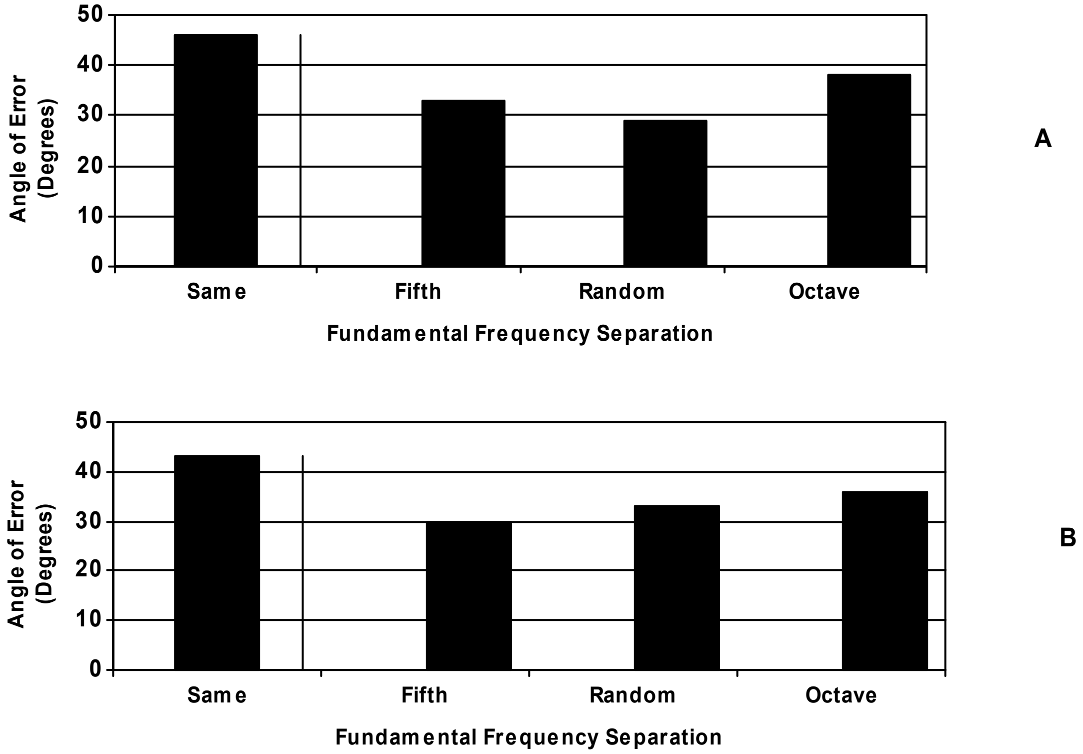

Figure 1. Angle of error versus fundamental frequency separation for (A) harmonic spacing and (B) inharmonic spacing of frequency components

3.2 Amplitude Modulation Rate

In the second experiment, the two complexes differed in their rate of amplitude modulation. The amplitude modulation rates included 0 (no modulation), 3, 17, and 100 Hz. Localization performance was relatively poor for these signals. Performance was worst when the target and masker were of the same amplitude modulation frequency. The 3 Hz stimulus was generally the worst target and the worst masker. The low modulation rate caused an apparent motion effect between the target and masker, which contributed to poor localization performance. Data for this experiment are shown in Figure 2 below.

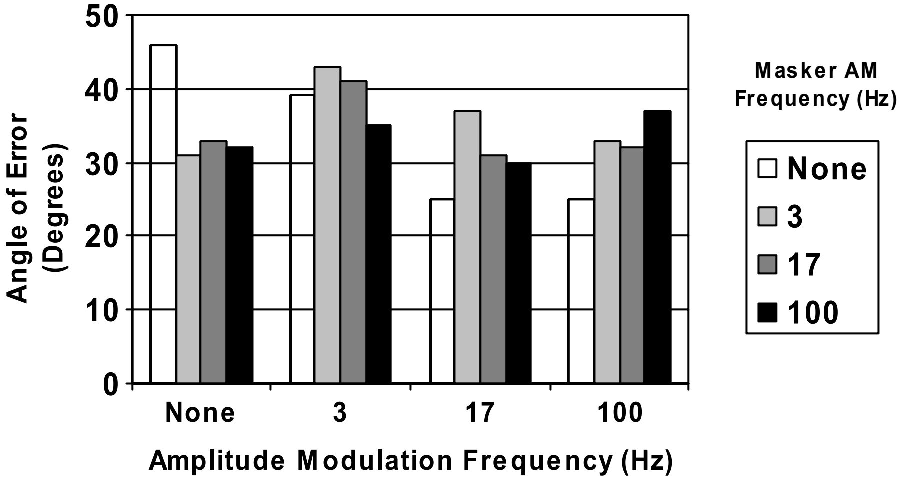

Figure 2. Angle of error versus amplitude modulation frequency of target sound

3.3 Amplitude Modulation Rate with Modulation Rate Randomized +/- 5%

In the third experiment, the two complexes differed in their rate of amplitude modulation as in experiment 2. However, the rate was randomized for each frequency component over a +/- 5% range. Localization performance was better than in experiment 2. The 3 Hz stimulus was still the worst target and the worst masker. The addition of the randomized modulation frequency greatly reduced the apparent motion effect between the target and masker. Data for this experiment are shown in Figure 3 below.

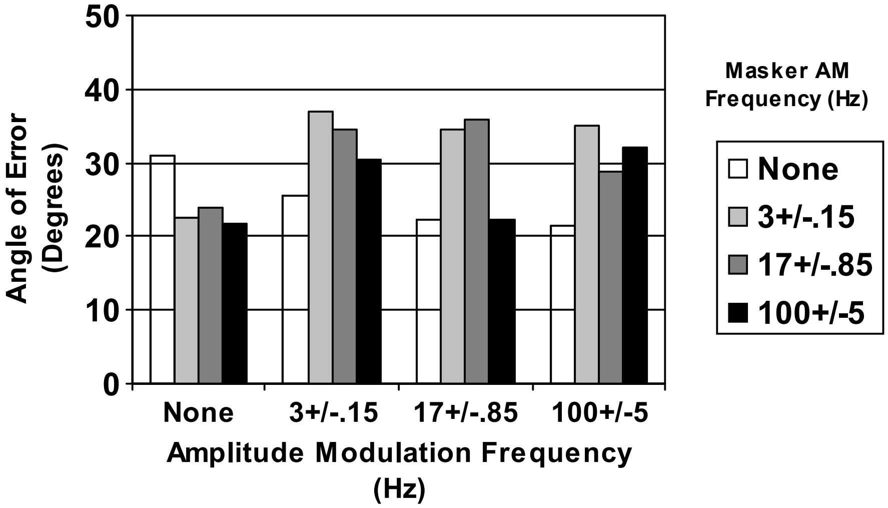

Figure 3. Angle of error versus amplitude modulation frequency of target sound

3.4 Amplitude Modulation Rate with Randomized Modulation Rate and Carrier Frequency

In the fourth experiment, the two complexes differed in their rate of amplitude modulation and had randomized modulation rates as in experiment 3. In addition, the carrier frequencies were also randomized over a plus-or-minus five percent range. Localization performance improved over those measured in experiments 2 and 3. The addition of the randomized carrier frequency to the randomized modulation frequency nearly negated the apparent motion effect between the target and masker. Data for this experiment are shown in Figure 4 below.

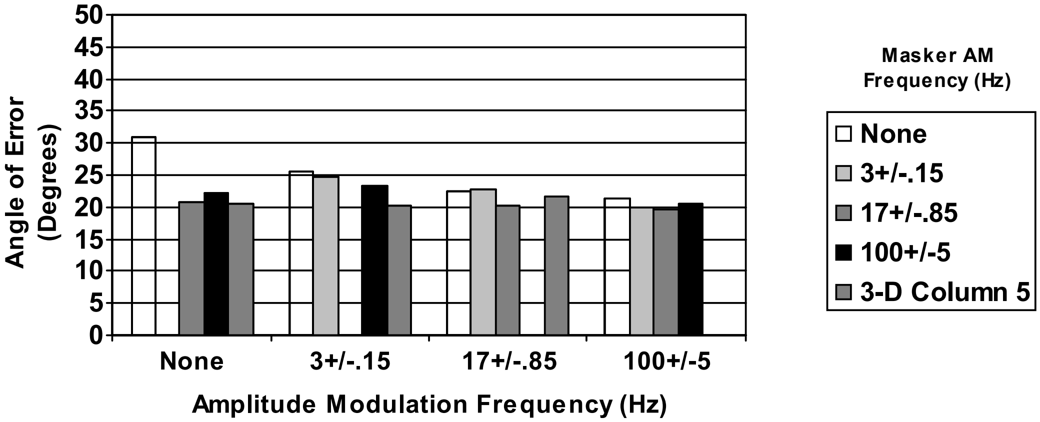

Figure 4. Angle of error versus amplitude modulation frequency of target sound

3.5 Amplitude Modulation Rate with Randomized Frequencies and Phases

In the fifth experiment, the two complexes differed in fundamental frequency (500 Hz and 571 Hz) and all four possible ways of creating amplitude modulated signals. The carrier frequencies and modulation frequencies were randomized over a plus-or-minus five percent range. The phases of the carrier and modulation frequencies were randomized over a 2 Pi range. Data were collected on the most difficult amplitude modulated signals to localize, the 3 and 17 Hz cases. Localization performance was close to localization acuity without a masker. Acuity averaged about fifteen degrees and correct identification performance about 86%. Some of the identification errors were due to loss of attention over the course of the twenty-minute sessions. Data for this experiment are shown in Figure 5 below.

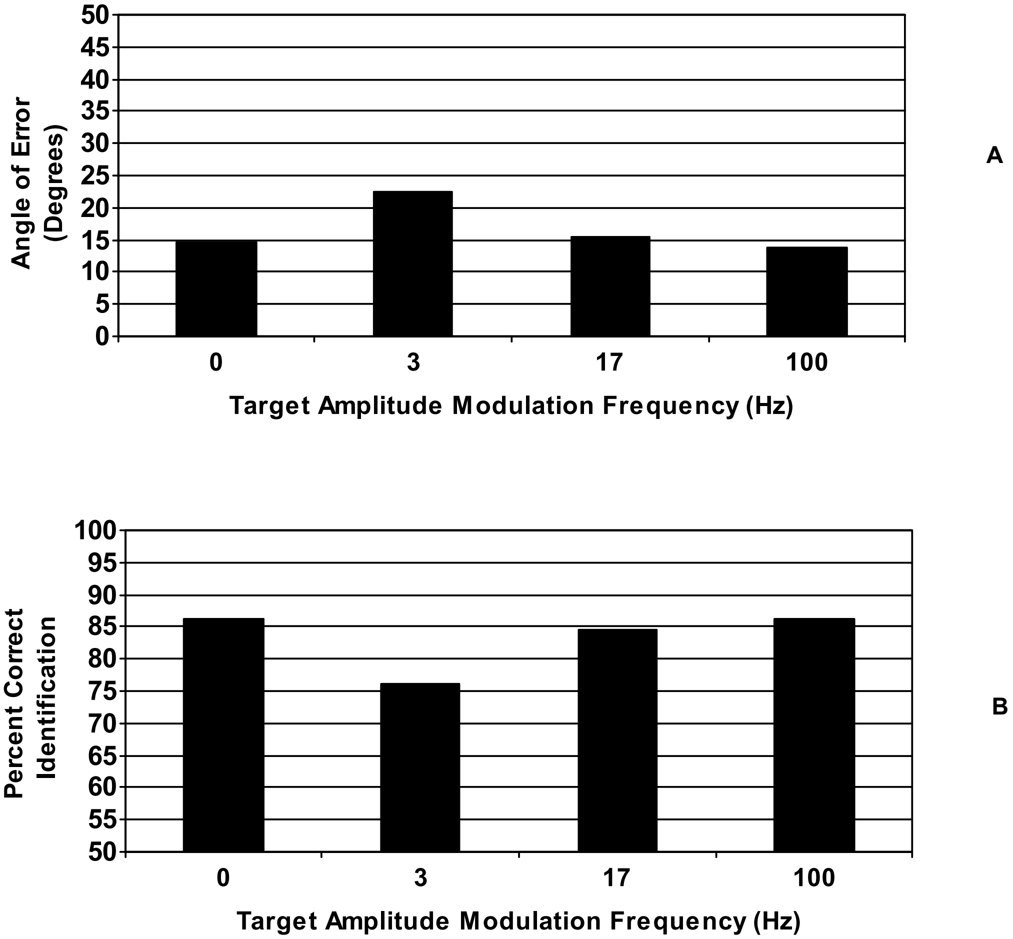

Figure 5. (A) Angle of error versus amplitude modulation frequency of target sound with a 3 Hz masking sound and (B) percent correct identification versus target sound with a 3 Hz masker

4.0 DISCUSSION

There are special challenges for designing warning sounds for military applications. One is that multiple threats of the same type may need to be presented at the same time. Another challenge is the harsh acoustic environment in which the sounds are typically presented, usually over headphones. One advantage of the current approach to warning sound design is that it maximizes the signal-to-noise-ratio without loss of localization acuity or identification performance. Perturbations to the component frequencies and relative phases do affect the noisiness of signals, which affects urgency but only a small amount. Repetition rate and loudness are the most salient cues for perceived urgency; harmonicity is third most important.

The introduction of spatial warning sounds can potentially improve military and non-military aviation safety. Spatial auditory displays for collision avoidance would reduce reaction times of pilots to take evasive manoeuvres and provide a more intuitive display of impending trouble. In military aircraft, integration of spatial warning sounds with radar warning receivers should greatly improve their utility by allowing the pilot to hear the location of the threat and keep his head and eyes free to attend to other tasks. Pilots would improve their situation awareness by hearing the location of their team members' voices and threat locations. Onboard aircraft warning sounds could be displayed from the location of the failure. In command and control centers, the locations of friendly and enemy forces could be simultaneously monitored. The laboratory research on spatial warning sounds will continue to provide a framework for their integration into military auditory displays.

5.0 CONCLUSIONS

Several experiments were conducted on the ability of trained subjects to localize and identify pairs of sounds. All sounds were presented over loudspeakers in anechoic space. Presentation of multiple sounds over headphones in virtual auditory space can potentially degrade performance further if the virtual auditory cues are not properly implemented. The major experimental findings are summarized in the statements below.

1) Pure tone, in-phase, harmonic complexes were poorly localized. Auditory images tended to fuse into a single large source, especially when located close to each other and when their fundamental frequencies were equal or at octave multiples of each other.

2) Any single cue manipulated in the experiments, e.g. fundamental frequency, harmonicity of the carrier and modulation frequencies, phase of the carrier and modulation frequencies, and amplitude modulation rate, was relatively ineffective in improving localization acuity.

3) The combination of all the cues provided the best localization and identification performance.

4) When applying these findings to the design of threat warning sounds, consideration should be given to their effects on the perceived urgency of the warning sounds.

6.0 REFERENCES

[1] Edworthy, J. and Haas, E. (1995). Toward a safer, more efficient auditory warning signal, *J. Acoust. Soc. Am. Newsletter,* **5**(2).

[2] Patterson, R. D. (1982). Guidelines for auditory warning systems on Civil Aircraft, *CAA Paper 82017*.

[3] Edworthy, J. (1994). Urgency mappings in auditory warning signals*, Human Factors and Alarm Design,* Ed. Neville Stanton, 15-30.

[4] Duross, S. H. (1978). Civil aircraft warning, systems – A survey of pilot opinion with British Airways. *Tech. Report 78056,* Royal Aircraft Establishment, Farnborough.

[5] Gerth, J. M., Folds, D. J., Fain, W. B., McKinley, R. L., and D'Angelo, W. R. (1995). Development and evaluation of auditory symbology for representation of simultaneous threats, *NAECON 95*, Dayton, Ohio, 8.

[6] Tun, P. and Wingfield, A. (1994). Speech recall under heavy load conditions: Age, predictability and limits on dual-task interference. *Aging, Neuroscience, and Cognition.* **1**(1) 29-44.

[7] Naish, P.L.N. (1988). The addition of localisation cues to auditory warnings. *Technical Memorandum MM2,* Royal Aerospace Establishment.

[8] Folds, D. J. (1990). Advanced audio displays in aerospace systems: Technology requirements and expected benefits, *NAECON*, Dayton, Ohio, 739-743.

[9] McKinley, R. L. and Ericson, M. A. (1997). *Flight demonstration of a 3-D auditory display*, in Binaural and Spatial Hearing in Real and Virtual Environments, Gilkey, R. and Anderson, T., Eds. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

[10] Daniels, Reginald, Ericson, Mark A., and French, Guy A. (2002). Improved performance from integrated audio video displays, *Proceedings of SPIE* ***4712****, 113-119.*

[11] Martin, R., Parker, S., McNally, K., and Oldfield, S. (1996). The abilities of listeners to localize defense research agency auditory warnings, *AGARD-CP-596*, 9-1 - 9-7.

[12] Patterson, R. D. and Datta, A. J. (1996), Extending the frequency range of existing auditory warnings, AGARD-CP-596, 8-1 - 8-8.

[13] Gilkey, R. H. Goode, M. D., Ericson, M. A., Brinkman, J., and Stewart, J. M. (1995). A pointing technique for rapidly collecting localization responses in auditory research, *Behavioural Research Methods, Instruments, and Computers* **27**, 1-11.