N94-33618

Laboratory and In-flight Experiments to Evaluate 3-D Audio Display Technology

Mark Ericson and Richard McKinley
Armstrong Laboratory, USAF

Marion Kibbe and Daniel Francis
Naval Air Warfare Center, USN

ABSTRACT

Laboratory and in-flight experiments were conducted to evaluate 3-D audio display technology for cockpit applications. A 3-D audio display generator was developed which digitally encodes naturally occurring direction information onto any audio signal and presents the binaural sound over headphones. The acoustic image is stabilized for head movement by use of an electromagnetic head-tracking device. In the laboratory, a 3-D audio display generator was used to spatially separate competing speech messages to improve the intelligibility of each message. Up to a 25 percent improvement in intelligibility was measured for spatially separated speech at high ambient noise levels (115 dB SPL). During the in-flight experiments, pilots reported that spatial separation of speech communications provided a noticeable improvement in intelligibility. The use of 3-D audio for target acquisition was also investigated. In the laboratory, 3-D audio enabled the acquisition of visual targets in about two seconds average response time at 17 degrees accuracy. During the in-flight experiments, pilots correctly identified ground targets 50, 75, and 100 percent of the time at separation angles of 12, 20, and 35 degrees, respectively. In general, pilot performance in the field with the 3-D audio display generator was as expected, based on data from laboratory experiments.

INTRODUCTION

Virtual audio display generators are being developed for aerospace and non-aerospace applications. Until the mid 1980s, acoustic manikins and loudspeaker arrays were required to simulate 3-D audio environments (Ericson, 1993). Other technological improvements, such as head tracking devices and digital signal processors, have aided in the realization of electronic virtual audio display generators for headphone applications. Many possible applications exist for virtual audio displays. Some aerospace applications include threat warning, collision avoidance, navigation beacons for landing at night and in bad weather, and spatially separated communications. These displays are created by encoding binaural cues onto an audio input signal.

Directional cues are contained in the head related transfer function (HRTF). The HRTF is the difference between the sound field at the entrance to a listener's ear canals and those same points in space in the absence of a listener's body. A more detailed discussion about HRTFs can be found in Blauert (1983) and Genuit (1992). In some applications, especially those in which distance cues are important, the inclusion of auralization or environmental cues becomes critical. Auralization cues include the reflections and reverberation characteristics of a particular listening environment (Lehnert, 1992). However, the experiments presented in this paper only involve directional encoding of audio signals.

All experiments discussed in this paper used virtual audio display generators developed at the US Air Force Armstrong Laboratory (McKinley, 1988, and McKinley , 1993). Two types of applications were explored by measuring human performance with virtual audio displays. One set of experiments explored visual target acquisition using virtual audio over headphones. The other experiments measured the intelligibility of spatially separated speech communications. For each application, experiments were first conducted in the laboratory followed by in-flight tests in a two seat AV-8B Harrier aircraft.

OBJECTIVE/PURPOSE

The objectives and purposes of the four experiments are described below. 1) The objective of the laboratory target acquisition experiment was to measure visual and auditory target acquisition response time and accuracy while performing a secondary compensatory tracking task. The purpose was to determine the effect, if any, of spatially correlated auditory information on visual target acquisition performance. 2) The purpose of the in-flight acquisition experiments was to determine if virtual audio cues could be used to distinguish ground targets in non-maneuvering and maneuvering environments. 3) The objective of the laboratory communication experiment was to measure the intelligibility of diotic, dichotic, and spatially separated speech presentations over headphones. Diotic refers to identical signals at each ear with the perceived location of the sound in the center of the head. In the dichotic presentation, one talker was presented through the left earcup and the other talker through the right earcup. Spatially separated speech was output from the 3-D audio display generator and perceived to come from different directions in the horizontal plane outside the listener's head. The purpose was to determine the relative intelligibility of diotic, dichotic and spatially separated speech messages. 4) The objective of the in-flight communication experiment was to determine if a pilot can better comprehend spatially separated speech messages compared to diotically presented speech messages.

METHODS FOR TARGET ACQUISITION EXPERIMENTS

METHODS FOR THE LABORATORY EXPERIMENT

PROCEDURE - Twenty-four LED displays were placed at fifteen degree separations on a seven foot radius horizontal ring at the level of a subject's head. Directional information was presented either visually on a 3" by 5" monitor directly in front of the subject, binaurally over headphones, or with a simultaneous presentation of visual and auditory binaural information. While waiting for the random targets to appear, the subjects performed a compensatory tracking task using a game joystick and a 14" diameter VGA monitor. The subjects were instructed to find the number zero on the horizontal ring that surrounded them. Once the LED target was presented, the subject turned his/her head towards the "zero" target on the ring and pressed a button switch on the joystick. Random false alarm targets were intermixed with the real targets 2% - 8% of the time to help ensure an honest response. Response time, the interval between presentation of the LED target and pressing of the joystick button, was the primary performance measure. Head pointing accuracy and tracking accuracy were secondary performance measures.

EXPERIMENTAL DESIGN - A balanced, repeated measures design was used in which each subject participated in all test conditions. Zero targets from each of the 24 directions were presented twice to each subject for each condition. Each subject participated in the auditory only, visual only, and combined visual and auditory conditions. Presentation orders of the three conditions were randomized across subjects to reduce order effects. Eight subjects participated in the experiment.

SUBJECTS - Eight volunteer, paid subjects participated in the experiment. Four males and four females ranged from 18 to 25 years in age with a mean age of 20. All had normal hearing sensitivity and function. All had normal (or corrected to normal) vision.

METHODS FOR THE IN-FLIGHT EXPERIMENT

During the in-flight tests, the forward pilot performed a series of passes, some straight and level and some maneuvering, on a path towards 3 ground targets: a bullseye, a tower, and an F-4 bunker. At a distance of 0.3 nautical mile (nmi) from the target, corresponding to 20 degrees of angular separation between the targets, the forward pilot randomly selected one of three targets which produced a 3-D audio beacon for five seconds. The task for the aft aviator was to report which of the three targets had produced the sound. If the response at 20 degrees was correct, then on the next pass, the audio beacon was presented at 0.1 nmi, corresponding to twelve degrees of angular separation between targets. If the response at 20 degrees was incorrect, then on pass two the beacon was presented at 1.0 nmi, corresponding to 35 degrees of angular separation between targets. All non-maneuvering passes were made before all maneuvering passes. The performance measure for this test was accuracy in identifying the correct target by the aft aviator. While there were a total of eleven 3-D audio flights, not all tests were completed for every flight. In the maneuvering condition, six tests were completed at 20 degrees, five runs at twelve degrees, and four runs at 35 degrees. In the maneuvering condition, five runs were made at 20 degrees, four at twelve degrees, and none at 35 degrees.

RESULTS FOR TARGET ACQUISITION EXPERIMENTS

LABORATORY EXPERIMENT

Results from the laboratory experiment are plotted in Figure 1. Response times in the audio, visual, and combined conditions were very similar across presentation angle, with the audio being slightly longer. In the audio condition, response times ranged from 1.6 to 2.4 seconds. Response times for the visual and combined conditions ranged from 1.5 to 2.2 seconds. There were no significant differences at $p=.01$ for response times. Head pointing accuracy was also very similar across conditions. There was no significant differences at $p=.05$. For the audio condition, there were individual differences in the amount of difficulty with which one could use the directional audio to determine the target direction.

IN-FLIGHT

Pilots reported that directional audio information enabled faster acquisition of the visual targets, with an approximate accuracy of fifteen degrees. On the completed tests, accuracy and the number of runs were sometimes given as approximations by the pilots. Thus, results are given as estimates of accuracy and not as precise figures. In the non-maneuvering passes, approximately 85% were accurate at 20 degrees of separation between targets, 50% at twelve degrees, and 85% at 35 degrees. For maneuvering approaches, there were fewer passes, and estimates of accuracy were 100% correct at 20 degrees and 40 % at 12 degrees. Pilots reported that at all angles of separation they were able to eliminate one of the three targets, but they had more difficulty in determining with confidence which one of the two remaining targets had produced the audio cue. They felt that in general 3-D audio complemented the visual displays and reduced target acquisition times.

METHODS FOR COMMUNICATION EXPERIMENTS

METHODS FOR THE LABORATORY EXPERIMENT

PROCEDURE - The competing messages experiment was conducted in the voice communications research and evaluation system (VOCRES) (McKinley, 1986) facility. Each of two talkers was prompted to simultaneously read messages of similar structure and content. Each message consisted of a call sign (ringo or baron), a color (red, white, blue, or grey), and an integer (one through eight). The message choices were randomized, however the order of call sign, color, and number were kept constant. Two listeners heard the messages presented diotically over headphones and two listeners heard the messages presented spatially at various angles of separation. Each listener was assigned a call sign, either ringo or baron. The listeners were to respond to the color and number spoken after their call sign. There were two diotic listeners and two spatial listeners, with a baron and a ringo listener in each group. A correct response required reporting all the information correctly about the call sign, color, and number. Scoring was measured automatically by computer, and no correction for guessing was employed.

EXPERIMENTAL DESIGN - This experiment used a balanced, within subjects design. Four ambient noise levels (75, 95, 105, and 115 dB SPL) were generated to simulate typical cockpit listening environments. The coordinate response measure *was used to measure the speech intelligibility for one of two competing messages in noise. Three talker pairs participated in the experiment. Each pair consisted of either a two males, two females, or a male and a female. Two groups of four listeners each participated in all the conditions. The spatially separated speech was presented at five angular separations (0, 45, 90, 135, and 180 degrees). Dichotically presented speech was realized by presenting one talker in the left ear and the other talker in the right ear.*

SUBJECTS - A total of twelve subjects, 6 male and 6 female, were paid to participate in the experiment. Two of the male talkers doubled as listeners. The subjects ranged from 18 to 43 years of age with a mean age of 23. All subjects had normal hearing sensitivity and function.

METHODS FOR THE IN-FLIGHT EXPERIMENT

The communication separation feature of the 3-D audio display generator was evaluated on the return trip from the target acquisition experiments. For this test, the communication (COMM) switch position was selected on the 3-D cuer control panel (Figure 3). Presentation levels of COMM-1 and COMM-2 were adjusted according to user preference. The aft pilot listened to two competing messages, which sounded as if they were coming from 315 degrees and 45 degrees bearing, and at 45 degrees elevation. Two persons on the ground using separate radio frequencies read separate messages; a nine-line brief and an emergency check procedure. The messages were received over two radios, COMM-1 and COMM-2. The aft pilot's task was to determine whether he could better distinguish these dual messages using the 3-D audio display generator than he could under the normal COMM-1 and COMM-2 modes. A total of seven pilots participated in communications separation experiment.

RESULTS FOR COMMUNICATION EXPERIMENTS

LABORATORY

Data from the laboratory experiments are plotted in Figure 4. Separations as small as 45 degrees provided a large improvement (over 25%) in speech intelligibility. Above 80% intelligibility is considered acceptable by flying personnel. Between 70 and 80% is marginal performance, and below 70% is considered unacceptable. The female talker pair tended to mask each other more than the other talker pairs. Dichotic (left/right) presentation provided the greatest intelligibility.

IN-FLIGHT

The communication separation feature of the 3-D audio display generator worked well. Most pilots felt that the spatial separation of speech communications improved the mutual intelligibility of each message. One pilot commented that spatial separation seemed to help a lot. However, the task of listening to one communication while two were broadcast simultaneously was still difficult.

DISCUSSION

Several differences between the laboratory and in-flight testing conditions may explain the relatively better performance with the 3-D audio system while in-flight than in the laboratory. There were only three targets to attend to in-flight, whereas there were 24 targets in the laboratory experiments. Pilots typically flew below 500 feet of altitude at 400 knots equivalent air speed while surrounded by mountains. The in-flight task was more stressful then the laboratory task and required a higher level of attention. In this situation, the 3-D audio display tended to complement the visual display since the pilot was often busy looking out of the cockpit for the targets and not looking down at the visual display. The 3-D audio display reduced workload by making the target acquisition task easier for the pilot to accomplish.

The 3-D audio display could be used for several other visual target acquisition applications. An auditory beacon could be used to help a pilot navigate towards a runway, especially at night or in bad weather. Auditory buoys *could warn pilots of possible collisions with either other airborne objects or with the ground, and thereby help the pilot to avoid collision. Possible military applications include threat warning with radar warning receivers, off-boresight missile targeting, and aerial refueling. An audio beacon could be used to find and track one's wingman in air to air combat. 3-D audio could improve many target acquisition and communication tasks.*

3-D audio displays may be a better modality for alerting a pilot as to the location/direction of a threat. 3-D audio encompasses all space around the person in azimuth and elevation, where visual displays are mostly limited to a person's line of gaze (fovea vision). Current threat warning visual displays are two dimensional and do not map 1 for 1 with the 3-D environment around the person. However, the 3-D audio display was spatially correlated to the ground targets, which provided a much more natural man-machine interface.

Laboratory and in-flight experiments showed spatially separated speech communications to be more intelligible than diotically presented speech. Two factors contributed to the relative success of spatially separated communications. These were the HRTF encoding and the head motion cues. The HRTF consists of magnitude and phase cues. The magnitude portion of the HRTF provided spectral filtering and the phase portion provided time of arrival differences between the two ears (Bronkhorst, 1992). People use these cues to unmask speech from noise. Head motion cues helped to space stabilize the direction of speech presentation. HRTF and head motion cues caused the speech communications to be spatially separated and easier to understand.

The success of the spatially separated speech communication experiments suggests that communication systems have room for improvement. Most of all, a pilots safety would be improved if he/she could better understand multiple communications from on board radios. Critical messages would probably not be misunderstood or have to be repeated as often. If the speech were spatially correlated with the source locations, then a pilot's situational awareness would be greatly improved. Spatially correlated communications would benefit pilots in formation flying situations. The laboratory and in-flight data support the inclusion of 3-D audio technology in airborne communication systems.

Many airborne applications for spatially separating speech communications exist. Any person that receives more than one speech communication at one time could benefit from the spatial separation of speech messages. Any command-control-communication post could benefit from this technology. Armored personnel carriers and submarines are visually blocked from their environments and their operators would probably have better situational awareness with 3-D audio displays.

CONCLUSIONS

Several conclusions can be drawn from the laboratory and in-flight experiments with target acquisition and spatially separated speech communications. They are listed below without any particular rank ordering.

1) 3-D audio cues were equally effective as visual cues for finding targets in the laboratory.

2) 3-D audio improved target acquisition tasks in-flight by reducing acquisition times.

3) 3-D audio improved multiple speech listening tasks up to 25% intelligibility in the laboratory and also worked well in-flight.

4) 3-D audio was reported to improve situational awareness in target acquisition and speech communications tasks without increasing workload.

BIBLIOGRAPHY

Blauert, Jens, "Spatial Hearing", MIT Press, 1983.

Bronkhorst, A.W., and Plomp, R., "Effect of Multiple Speechlike Maskers on Binaural Speech Recognition in Normal and Impaired Hearing," J. Acoust. Soc. Amer., Vol. 92, No. 6, December, 1992, pp. 3132-3139.

Ericson, Mark and D'Angelo, William, "Applications of Virtual Audio," pp. 604-611, NAECON, Dayton, OH, May, 1993.

Ericson, Mark, "Auditory Localization and Cue Synthesis and Human Performance," pp.718-725, NAECON, Dayton, OH, May, 1989.

Genuit, Klaus, and Burkhard, Mahlon, "Artificial Head Measurement Systems for Subjective Evaluation of Sound Quality," Sound and Vibration, Vol. 26, No. 3, 1992, pp. 18- 23.

Lehnert, Hilmar, and Blauert, Jens, "Principles of Binaural Room Simulation," Applied Acoustics, Vol. 36, No. 3, 1992, pp. 293-306.

McKinley, Richard, "Voice Communications Research and Evaluation System,", AFAMRL-TR-012-86, USAF, WPAFB, OH, 1986.

McKinley, Richard, "Concept and Design of an Auditory Localization Cue Synthesizer," Air Force Institute of Technology Master's Thesis, AFIT/GE/ENG/88D-29, 1988.

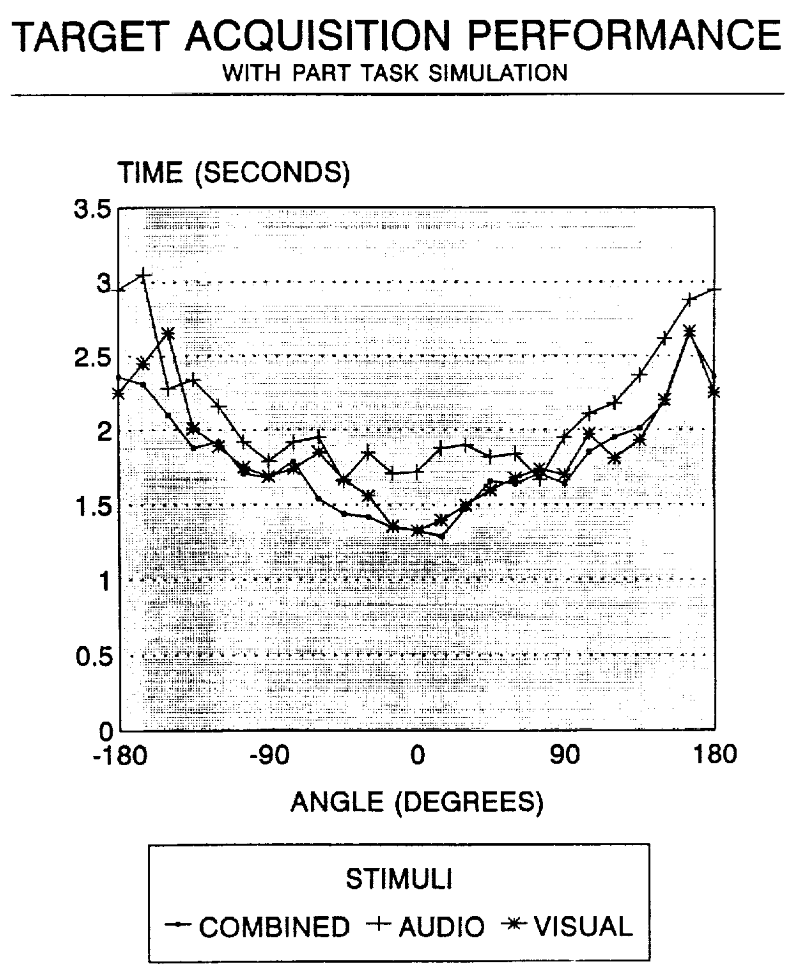


Figure 1

'get Acquisition Accuracy VS Angular Separation

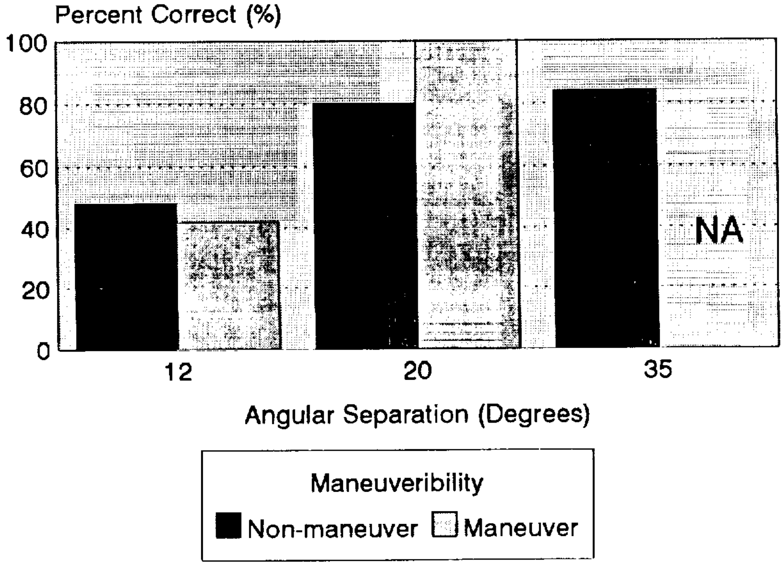


Figure 2

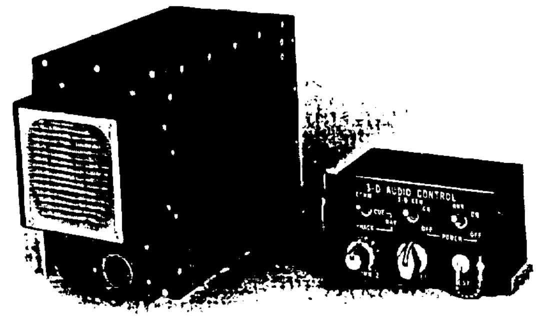
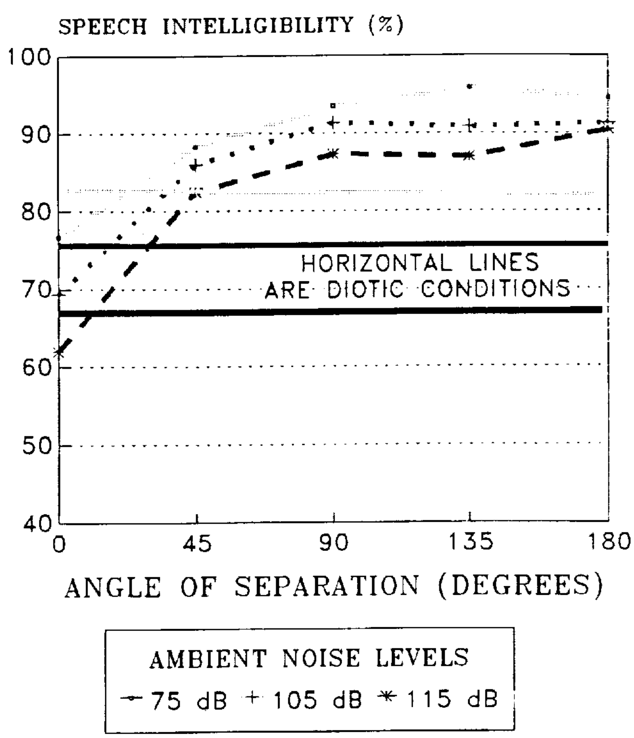


Figure 3

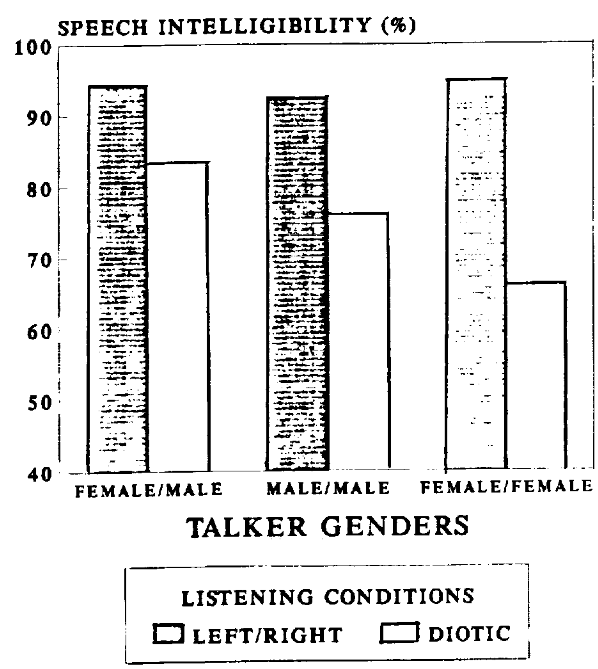
COMPETING MESSAGES
SPATIALLY SEPARATED VS DIOTIC



3 PAIRS OF 2 TALKERS, 4 LISTENERS
MOST COMFORTABLE PRESENTATION LEVEL

Figure 4a

COMPETING MESSAGES
LEFT/RIGHT VS DIOTIC
115 DB OF NOISE



3 PAIRS OF 2 TALKERS, 4 LISTENERS

Figure 4b