Image Cover Sheet CA010621

CLASSIFICATION UNCLASSIFIED	**SYSTEM NUMBER** 515432

TITLE

An evaluation of three-dimensional audio displays for use in military environments

This page is left blank

This page is left blank

Research article / Article de recherche

AN EVALUATION OF THREE-DIMENSIONAL AUDIO DISPLAYS FOR USE IN MILITARY ENVIRONMENTS

G. Robert Arrabito
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, ON M3M 3B9

ABSTRACT

A relatively new technology, the three-dimensional (3-D) audio display, has been proposed for improving operator performance. However, many of the studies supporting 3-D audio displays have been conducted in controlled laboratory settings and it is not clear if all gains in performance will transfer into corresponding real-world applications. The Canadian Department of National Defence (DND) and the defence forces in allied nations are presently investigating 3-D technology for use in operational environments. The implications of the use of 3-D audio technology in real-world applications and recommendations for further research are discussed in this paper.

RÉSUMÉ

On a proposé qu'une technologie relativement nouvelle, l'affichage audio tridimensionnel (3D), puisse améliorer le rendement des utilisateurs. Cependant, comme un bon nombre des études à l'appui de l'affichage audio 3D ont été contrôlées en laboratoire, on ne peut affirmer que tous les gains de rendement se réaliseront en milieu réel. Le ministère canadien de la Défense nationale (MDN) et les forces de défense de certains pays membres de l'Alliance se penchent sur l'emploi possible de la technologie 3D en situation opérationnelle. Dans cet article, il est question des répercussions de l'application de la technologie audio à trois dimensions en milieu réel et des recherches plus approfondies que l'on recommande.

1 INTRODUCTION

A sound source that is presented over headphones can be made to appear as though it would be perceived in the listener's natural free-field environment. The technology used to create this perception is a three-dimensional (3-D) audio display. The effectiveness of a 3-D audio display depends on the listener's ability to localize and discriminate between various sources of information in virtual auditory space. The spatial synthesis of an auditory signal is accomplished by digitally filtering the signal with head-related transfer functions (HRTFs). These HRTFs encode the binaural and spectral cues used in sound localization and discrimination. They are derived via a series of impulse measurements performed at the ears of an observer in response to a sound source placed at various locations in the vicinity of the head (see Wightman and Kistler (1989), for example). While the concept of a 3-D audio display is not new, present technology makes it more feasible and practical to implement.

Investigators have demonstrated that the techniques used for creating and presenting virtual auditory sources have limitations. These include the methods employed to measure HRTFs, the difference in localization performance in virtual auditory space compared to the free-field, and the externalization of the sound image outside of the listener's head (Wightman & Kistler, 1989; Begault, 1991, Wenzel, Arruda, Kistler & Wightman, 1993). It is not known whether these limitations will significantly affect the practical advantages of a 3-D audio display because little experience has been gained with applications in real tasks. Furthermore, it has also been reported that the advantages of 3-D audio are highly task-specific (Perrott, Cisneros, McKinley & D'Angelo, 1996) and that there may be no performance advantages for 3-D compared to 2-D for discriminating messages from two simultaneous talkers (Arrabito, McFadden & Crabtree, 1996) Thus the costs and benefits of equipping future systems with 3-D audio displays versus R&D efforts aimed at enhancing 2-D presentations of auditory information need to be evaluated more closely before the technology is adopted.

The Canadian Department of National Defence (DND) and the defence forces in allied nations are presently investigating 3-D technology for use in military environments. For example, aircrew are often subjected to high workload and have to maintain awareness of a complex, rapidly changing situation while making quick decisions and prompt responses. Other environments include dispersed man-on-the-move communications aboard ships or on the battlefield, where each team member needs to remain aware of the location and status of one or more other team members. In these situations the 3-D audio system must be reliable, effective, durable, flexible and immune to the effects of environmental conditions. To date there has been no known review of 3-D audio technology that addresses these requirements. These issues are discussed in this paper with respect to the use of 3-D audio display technology.

2 RANGE OF APPLICATIONS THE TECHNOLOGY CAN SUPPORT

Three-dimensional audio displays are being explored for improving operator performance in a variety of applications. These include cueing for visual search tasks, improved auditory signal detection and speech intelligibility, localization of signals, and creation of a realistic audio environment for relaxation. Calhoun, Jameson and Valencia (1988) compared the effectiveness of 3-D audio cues for directing subjects' visual attention to peripheral targets. Subjects were engaged in a visual tracking task while periodically responding to cues to identify visual targets at four peripheral locations. The authors found that subjects' response time with 3-D audio cueing was on the average 245 ms faster than non-spatialized auditory cueing, and for the left-right locations only 126 ms longer than the visual cue.

Begault (1993) evaluated a "head-up" spatial auditory air traffic collision avoidance system (TCAS) in a flight simulator. The positions of the visual targets corresponded to aircraft that would activate a TCAS aural advisory. Begault found that the use of a 3-D audio display resulted in an average of 2.2 seconds faster reaction time compared to a single earpiece baseline condition. Enhanced target detection and identification in a flight simulator task was also observed with the use of a 3-D audio display as investigated by Bronkhorst, Veltman and van Breda (1996). These investigators found that search time decreased on average by two seconds with the use of a 3-D audio display compared to visual search with a radar display; they found an additional average two seconds improvement when similar target information was simultaneously presented by both displays. During an in-flight study, pilots reported that a 3-D audio display decreased target acquisition time and visual workload while increasing communication capability and situational awareness (McKinley & Ericson, 1997).

The potential of a 3-D auditory display for improving auditory signal detection and speech intelligibility has also been demonstrated. Arrabito, Cheung, Crabtree and McFadden (2000) measured auditory thresholds when subjects were under sustained +3Gz positive acceleration. They found that subjects reached an average of 6.8 dB lower auditory threshold when a pulsed signal was spatialized at a static position of 90° azimuth on the horizontal plane in virtual auditory space compared to the baseline diotic (the same sound presented to both ears) presentation. In research motivated by the need of aircrew to monitor simultaneous channels of communication (e.g., air traffic control, air traffic information system and company), Ericson and McKinley (1997) investigated the contribution of 3-D audio. They reported improved speech intelligibility when more than two simultaneous talkers were spatialized at different virtual positions when listeners performed a complex divided attention task.

Extended applications of 3-D audio displays have been suggested to support situational awareness and spatial orientation by providing veridical spatial cues to the positions of targets, threats, and beacons (Doll, Gerth, Engelman & Folds, 1986, Furness, 1986). Applications could include auditory cueing to the location of allies (for example, aircraft wingman), or threats. Such applications could be extended from the 'real world' to virtual worlds such as navigation aids and sonar displays. In the area of team performance, the ability of a 3-D audio display to enhance speech intelligibility, especially in the presence of multiple simultaneous talkers, is being explored. Each talker could be spatialized at a different virtual position relative to the listener's head.

A somewhat different benefit was shown in the presentation of a relaxing 3-D audio sound to aircrew during in-flight and layover sleep periods which resulted in enhanced quality and quantity of sleep (Allsten, Downey & Jackson, 1995).

3 PROJECTED IMPACT OF THE TECHNOLOGY ON CURRENT AND FUTURE SPACE LAYOUT

Operator space layout could potentially be changed from a workstation configuration to one that has no physical constraints. For example, a 3-D audio display coupled with head tracking and man-on-the-move communications could improve the situational awareness of command or combat teams that presently are tied to their location because of headset communications systems. The ability of a 3-D audio system to support dispersed teams has been partially demonstrated by Bryden, at the Communications Research Center in Ottawa (Bryden, personal communication, February 10, 2000). Bryden assembled portable workstations consisting of a notebook computer, wireless LAN transceiver, differential GPS receiver (a higher level of accuracy than standard GPS), aviation type headset/microphone and head-mounted electronic compass, in order to conduct outdoor field trials on the effectiveness of directional virtual sound sources as a means of increasing situational awareness. Using this equipment, multiple simultaneous radio communication channels could be presented to listeners in a 3-D audio display.

The spatialization of the sound was created using a combination of HRTFs that were measured from the Knowles Electronic Mannequin for Acoustic Research (KEMAR), and a parametric model. The virtual sound sources were presented on the horizontal plane. Four subjects were each equipped with a portable workstation and were dispersed so that they could not see each other thereby simulating an infantry peacekeeping scenario. Bryden reports that subjects had a sense of situational awareness with respect to the location of each other based on the apparent direction of the virtual position of the radio communication channels. There were two major limitations of the equipment that negatively affected situational awareness: compass instability and latency to head movements.

4 TRANSITION POTENTIAL OF THE TECHNOLOGY

The transition potential of 3-D audio displays from the laboratory setting into real-world applications will depend on several technical factors, as well as issues of cost, availability and reliability. These issues include the robustness of the technology, the effectiveness of HRTFs, the characteristics of the sound to be localized, and the integration of the 3-D audio system with head tracking and underlying communications systems.

4.1 Robustness of the Technology

It should be noted that the above studies have been conducted in controlled environments. The conditions of these trials may not translate well to real-world applications. For example, Perrott et al (1996) found that the greatest advantage of 3-D aurally aided cueing to the detection and discrimination of visual targets occurred when targets were presented in the rear and peripheral regions of the frontal hemi-field. These investigators used high contrast targets, which are unlikely in real-world situations. Findings such as these suggest the need for caution in the adoption of 3-D audio displays.

To date there have been few reports of the feasibility of a 3-D audio display in field trials The flight trial of McKinley and Erickson (1997) is one example. In that study, pilots reported that a 3-D audio display decreased target acquisition time and visual workload while increasing communication capability and situational awareness. In another trial, NORAD CMOC crew members reported improved speech intelligibility when evaluating a 3-D audio system in a Ballistic Missile Defence Organization exercise (North and D'Angelo, 1997). Although the investigators of both of these studies collected only qualitative data, their results are encouraging and suggest that a 3-D audio display could improve operator performance in real-world applications

Some of the issues in transferring laboratory research into real-world applications are exemplified in the debate about the contribution of spatial hearing to the enhancement of speech intelligibility of multiple simultaneous talkers. Cherry (1953) investigated the listener's ability to focus his/her attention on a single sound source or signal in the presence of multiple competing signals and interfering noise He termed this the "cocktail-party problem" Cherry suggested that spatial separation of sound sources was the major contributor for solving the cocktail party problem. This provides support for the application of 3-D audio displays to improve the intelligibility of multiple competing talkers.

A recent review by Yost (1997) argues that spatial hearing may not be the major cue to solve the cocktail-party problem in real-world situations. Yost notes that, over the past 40 years, data on the benefits of binaural listening over monaural listening have been mostly collected via headphones or in simplified free-field studies, which do not represent real-world listening environments. He points out that these studies have been devoted to measures of selective attention, where the listener is asked to focus on a particular signal source and ignore all others. There is very little information on situations of divided attention, where the listener must attend to several or all of the sound sources in the acoustic environment. Yost suggests that there are seven physical attributes of sound that might be used as a basis for sound source determination These are spectral separation, spectral profile, harmonicity, spatial separation, temporal separation, temporal onsets and offsets, and temporal modulations. This suggests that the greatest binaural advantage is found for detection tasks in noise, or measures of discrimination, or recognition tasks conducted at very low signal-to-noise ratios. The binaural advantage decreases rapidly with increasing signal level above threshold and is very small when the signal is relatively easy to detect.

The improvement of speech intelligibility of multiple talkers using a 3-D audio display may depend on the number of simultaneous talkers. For example, Arrabito et al. (1996) found that there was no improvement in speech intelligibility when two simultaneous talkers were spatially presented in a 3-D audio display compared to a dichotic presentation. Ericson and McKinley (1997) found that the greatest benefits of spatializing the speech signals of a communication headset may occur when there are more than two talkers or signals to attend to simultaneously The results of Ericson and McKinley (1997) may be explained by the observation of Cherry (1953). He noted that, in a free-field listening environment, the signal-to-noise ratio at the listener's ears will vary when the noise source is spatially displaced and will be different across the two ears. For example, when the noise source is displaced laterally relative to the listener, the noise level increases in the ipsilateral ear (i.e., the ear closest to the sound source) and decreases in the contralateral ear (i.e., the ear furthest away from the sound source). When there are several sound sources surrounding the listener, these multiple signals reduce the speech to noise ratio at the ear closest to the desired talker. Hence the overall intelligibility level is reduced by the unwanted but necessary binaural signals. Given that the greatest binaural advantage is realized at very low signal-to-noise ratios, it is expected that an increase in speech intelligibility should be observed when more than two competing talkers are presented in a 3-D audio display.

The extension of the research of Ericson and McKinley (1997) has been transferred into a real-world application A Canadian company has developed and is marketing a 4-channel shipboard communication terminal. Each of the four communication channels are spatialized at different static virtual positions relative to the listener. A company representative reported that users of this system have remarked improved speech intelligibility of multiple simultaneous talkers. The system is primarily targeted as a device for channel separation rather than the display of position information. This latter capability may be utilized in the ensuing future as reported by the company representative.

4.2 Head-Related Transfer Functions

Head-Related Transfer Functions are the digital filters used for spatializing a sound and thus are essential to 3-D audio Measurement techniques of HRTFs differ across laboratories and are motivated by the different goals of the investigators (see references 2, 3, 5-27 of Moller, Sorensen, Hammershoi & Jensen, 1995). Some of the parameters that vary significantly in the measurement of HRTFs are·

- type of test stimulus (e g., sinusoidal tones or noise bursts),
- the point in relation to the ear canal where the measurement is made (e.g., at the blocked ear canal or a point somewhere along the ear canal), and
- the number of source positions.

It has been argued that a listener's ability to localize a virtual sound is more accurate when using HRTFs measured from his/her own head ("personal") compared to HRTFs measured from a different head ("generic") (Wightman & Kistler, 1989; Wenzel et al , 1993, Bronkhorst, 1995) Investigators have also shown that generic HRTFs significantly contribute to more reversals (i.e., perceiving the mirror image of the presented sound source) in perceived position of a sound compared to personal HRTFs (Wenzel et al, 1993) Perceiving the mirror image of the sound source may lead to an inaccurate sense of situational awareness. If virtual sources are to be used in a general-purpose 3-D audio display under mission critical conditions, such as those encountered by military personnel, then the HRTFs should be optimized for the targeted application. However, it is presently not practical or affordable to measure the HRTFs for each potential listener. In light of these findings, methods need to be developed to quickly and accurately select a generic HRTF for the targeted application. One vendor of a virtual listening home theater entertainment system requires that the listener select the best available set of HRTFs from a repository measured on many individuals. The selection is made on the basis of a simple virtual sound localization test performed over headphones. This test could be further refined to ensure more accurate spatial synthesis by selecting a localization task emphasizing conditions typical of generic HRTF listening such as front-back reversals and median plane errors (Wenzel et al , 1993).

The criterion for selecting generic HRTFs need not be based exclusively on individuals who are "better localizers" than others due to physiological differences. Good localizers are subjects whose free-field localization performance is better than average and whose headphone localization performance in virtual auditory space closely matches his/her free-field localization performance. F L Wightman (personal communication, March 2, 1997) reported that his laboratory has been unsuccessful in documenting any relation between HRTF characteristics and localization performance despite suggestions made in an earlier study (Wightman & Kistler, 1989). While it is clear that some subjects may have less spectral detail to work with because their pinnae are smooth, it is not clear that this translates into poor performance. With several cues to work with, some individuals seem simply to emphasize one or more cues depending on their own physical characteristics.

4.3 Headphones

The spatial synthesis process of virtual sound sources is not solely dependent on the selection of HRTFs (personal versus generic). The headphones contribute to the total transmission and require equalization for the correct reproduction or synthesis of binaural signals (Wightman & Kistler, 1989; Moller, 1992). Headphone equalization is a digital filtering procedure to cancel the distortion caused by the headphones and the resonance effects on the listener's ears. The equalization is specific to the headset and the end-listener. In the past, the equalization procedure has often been overlooked (Moller, 1992). It should be noted that headphone placement on the listener's head is rarely the same and, hence, the headphone equalization step may not always be exact It may thus be argued that the equalization step might be more detrimental than useful. Proper fitting of the headphone on the listener's head will ensure that spatialization is not degraded. It is critical that both ear cups are over the pinnae, and that the right ear cup is on the right ear and the left ear cup is on the left ear. It should be noted that the author has found that it is common practice for operators, in a ship's operation room for example, to remove the headset from one ear to improve direct communication transmission to those nearby In such settings, operators will be required to be fitted with headsets that allow them to hear the voices of nearby crew members while providing high fidelity communications The selection of a headset equipped with a single knob to control the signal volume under both ear cups is preferred over a headset equipped with dual controls to independently adjust the volume in the left and right ears. Independent volume controls interfere with the correct reproduction of spatial cues from the virtual audio sources.

4.4 Stimulus and Bandwidth

The properties of the stimulus for use in virtual auditory space need to be factored into consideration with respect to user performance. For example, localization performance will be affected by the characteristics of the sound to be spatialized. Broadband impulsive sounds are easier to localize than low frequency sounds that have slow amplitude envelopes (Begault, 1991). Begault and Wenzel (1993) found that localization performance was not as accurate when the broadband stimulus used in a similar study by Wenzel et al. (1993) was replaced by speech stimulus. It should be noted that the bandwidth of speech is less than the bandwidth of white noise and this difference might be a factor in accounting for the different results.

The cues for range in virtual environments also need to be factored into consideration with respect to user performance Begault (1991), for example, reported that these cues are not as well understood as the cues for azimuth and elevation. The signal can also be made to appear that it is outside of the listener's head by adding reverberation cues. This gives the listener a more natural listening advantage. However, the localization error between the presented and the perceived virtual sound source increases with the presence of reverberation (Begault, 1992). Furthermore, speech intelligibility in a virtual environment is poorer with the presence of reverberation cues, as demonstrated by Vause and Grantham (1998).

The bandwidth of the communication system has a typical upper cutoff frequency between 3.5 and 4 kHz (Ericson & McKinley, 1997; King & Oldfield, 1997) which could potentially affect localization performance. If the sole cues afforded by the bandwidth of the communication system are the differences in the time and level of arrival, one could correctly assume the presence of front-back reversals. Front-back reversals are largely resolved with the presence of spectral cues. Spectral cues are contained in the frequency region above 4 kHz and are encoded by the head, pinnae and upper torso (Blauert, 1983) The upper frequency limit of the communication system will be governed by the location of the virtual audio sources in the intended application. In the case of virtual sources positioned on the horizontal plane, an upper frequency limit of 4-5 kHz should be adequate (Blauert, 1983). In the case of sources distributed both in front and at the back, in the median plane, or elevated from the horizontal plane, an upper frequency limit of 8 kHz or more may be required (King & Oldfield, 1997) However, some communication systems have a hardware imposed upper cutoff frequency of approximately 4 kHz as reported by the representative of the Canadian company that is marketing a shipboard communication terminal. Such a limitation may impose a restriction on the types of 3-D audio applications.

4.5 Head Tracking

The need for head tracking is a significant requirement if 3-D displays are to be implemented for other than the most basic applications. Head tracking can assist the listener with exploratory head movement relative to the apparent direction of the virtual sound source provided that the duration of the sound is sufficiently long This in turn can assist the listener with disambiguating between sound sources such as front-back reversals which are located in the so-called "cone-of-confusion", i.e., directions that cause similar interaural differences in level and arrival time (Blauert, 1983). For example, Bronkhorst (1995) found that front-back reversals were low when listeners oriented their head to a long-duration sound source but high when listeners localized a short-duration sound source without head movements

The spatial sound image coupled with a head tracker significantly depends on the number of spatial locations measured for the HRTFs. Smooth motion of a virtual auditory signal (head movement, moving sound source) requires high spatial resolution (equal to or less than 5°) (Hartung, Braasch & Sterbing, 1999). Having a smaller number of measured sound source positions requires the between-source interpolation procedure to be more precise. The advantage of measuring HRTFs with a lower spatial resolution results in minimizing the measurement time. Approximately one hour was required to measure HRTFs with a 10-15° spatial resolution of the whole sphere (Hartung et al , 1999). Minimizing the measurement time of HRTFs will necessitate appropriate interpolation techniques which yield high spatial resolution and synthesis. Hartung et al (1999), for example, found that subjects could distinguish between measured and interpolated HRTFs for some of the tested directions based on location and/or timbre. In a related study Cheng and Wakefield (1999) found that listeners' localization performance was better using interpolated rather than non-interpolated

HRTFs. Although the tasks asked of the subjects in these two studies are not directly comparable (discrimination versus localization), it is interesting to note that a possible explanation for the different findings may be attributed to the source of the measured HRTFs. The HRTFs in the study by Hartung et al (1999) were measured from the KK1412 acoustic mannequin developed by Head Acoustics while those in the study by Cheng and Wakefield (1999) were measured on humans. The results of these studies suggest that the number of measured spatial positions and the source of the measured HRTFs could affect listener performance in the targeted application

A range of head tracking technologies are available, most of them developed from head-mounted display systems Technologies include mechanical linkages, inertial, magnetic and electro-magnetic, electronic, infra-red and ultrasonic Head tracking in user-on-the-move applications, especially on dispersed personnel, cannot be accomplished using conventional head trackers such as electro-magnetic or infra-red trackers. For example, the Polhemus 3Space Fastrak tracker consists of a transmitter emitting a magnetic field which is detected by the receiver This device requires a workspace that is not much larger than the operator workspace and thus cannot be used for man-on-the-move applications. Hence it may be necessary to require reliance on the global positioning system (GPS). GPS provides satellite signals that can be processed in a GPS receiver, enabling the receiver to compute position, velocity, and time estimates. The accuracy of the GPS is categorized by "standard", "precise" and varying levels of precision between these two extremes. Standard GPS is provided without charge or restrictions to the end-user. Precise positioning is more accurate than standard GPS and can be accessed only by authorized users with specially equipped receivers. In the man-on-the-move application, the receiver can be mounted on the helmet. The required degree of situational awareness will dictate the accuracy of the GPS. In the trial of dispersed personnel conducted by Bryden (personal communication, February 10, 2000) the electronic compass exhibited instability and high latency in reacting to rapid subject head movements. There were also drop-outs in the differential GPS and during these instances, the bearing inaccuracies were unacceptably high. Given these results, it may be necessary to decrease the head compass latency and use higher precision in the GPS receiver

4.6 Environmental Conditions

The transfer of user performance from a laboratory setting into a real-world application has been shown to degrade due to environmental factors. For example, McKinley and Ericson (1997) observed a gradual degradation in localization performance from the laboratory baseline conditions to the flight trial. During the baseline study conducted in the laboratory, subjects could point their heads to a sound source presented in the free-field with an accuracy of 4-5° versus 6-7° for the same task but in the presence of 115 dB SPL ambient pink noise The minimal audible angle (MAA) that subjects could discern targets was 12° azimuth under conditions of low level and high speed flight However, it should be noted that McKinley and Ericson (1997) do not clearly indicate if any of the subjects who participated in the baseline study also participated in the flight trial. Furthermore, the head pointing and MAA tasks are not directly comparable Nevertheless, it is probable that the environmental conditions could have affected localization performance despite the aforementioned observations.

Ambient noise and electromagnetic interference are examples of environmental conditions that need to be factored into consideration when exploring the use of 3-D audio In settings where the level of the ambient noise is relatively low, such as that found in the operations room of a ship (typically less than 80 dB(A) based on levels measured by DCIEM personnel), the ambient noise should not affect performance when using a 3-D audio display. On the other hand, the cockpit noise levels of military aircraft range from 95 to 115 dB SPL under normal operating cruising conditions (Ericson & McKinley, 1997). McKinley and Ericson (1997) demonstrated that a 3-D audio display could increase user performance in high levels of ambient noise such as that found in cockpits. However, the author of this paper has empirical data for sound localization in virtual auditory space that was collected in quiet (71 dB(A)) and in the presence of ambient Leopard tank noise (approximately 110 dB(A)). Localization performance was poorer in the latter condition compared to the former

The surrounding metallic structure (e g , overhead beams, frames, instrument panel, etc.) found within personnel compartments (e g., cockpit, armored vehicle, operations room, etc) could significantly degrade the performance of commercial-off-the-shelf electromagnetic head trackers. The head tracker used in the flight trial reported by McKinley and Ericson (1997) was a modified Polhemus 3Space Fastrak magnetic head tracker which reduced the amount of electromagnetic interference Mechanical trackers would be spared from electromagnetic interference; however, the mobility of the end-user would be smaller compared to electromagnetic trackers and thus could impede the operator's performance. The most likely technology for a head tracker in an operational space would be one of the small inertial tracker or an ultrasonic tracker that have recently come on the market. These latter technologies are reported to be immune from the effects of electromagnetic interference and can cover a large user space

5 IMPACT OF THE TECHNOLOGY ON PERSONNEL WORKLOAD AND SAFETY

Three-dimensional audio should reduce the workload of operators. For example, McKinley and Ericson (1997) reported that a 3-D audio display decreased visual workload Increased communication capability when listening to simultaneous messages has been demonstrated by Ericson and McKinley (1997), and McKinley and Ericson (1997). The time taken to react to visual warnings using spatial auditory cueing could be reduced 10-50 percent as demonstrated by Perrott et al. (1996)

Increased operator safety may also be realized with the use of a 3-D audio display, based on reduced response time to alerting signals, improved operator performance, and reduced risk of hearing loss. At present, the level of auditory warnings presented over the communication system used by aircrew is frequently too loud (Patterson, 1982). Many of the existing auditory warnings disrupt thought and verbal communication amongst crew members (Patterson, 1982). In addition, continuous loud sounds hold the crew's attention beyond the point where the problem has been identified, often incapacitating aircrew (Patterson, 1982). Auditory warnings in a 3-D audio display may be detected as much as 6 dB lower compared to a diotic presentation Based on the results of Arrabito et al. (2000). This represents an approximate 50% reduction in the acoustic amplitude of the stimulus. Thus the overall amplitude level of the communication system can be reduced without sacrificing detectability. Lower headphone amplitude could reduce the risk of hearing loss.

It should not be assumed that all operators can benefit equally from the use of a 3-D audio display. In fact operators who have a hearing impairment may impose a safety hazard on themselves and/or fellow crew members due to disrupted spatial cues which could lead to inaccurate localization judgements. Hearing impairment may be broadly categorized as either conductive or sensorineural (Moore, 1989). Conductive hearing loss usually occurs when there is a defect in the middle ear. Normally this results in hearing loss across all frequencies. Sensorineural hearing loss most commonly arises from a defect in the cochlea, but may also result from defects in the auditory nerve or higher centres in the auditory system. Often, the extent of the loss increases with frequency, especially in the elderly.

The above two types of hearing losses have been studied with respect to their effects on localization performance in the free-field (Noble, Byrne & LePage, 1994). In that study 87 bilateral hearing impaired individuals (mean age 65 6 years, s.d. 13.8 years) participated. Of these, 66 had sensorineural hearing loss and 21 had conductive or mixed hearing loss. Localization judgements were tested in four spatial regions relative to the listener· frontal horizontal plane (FHP), median vertical plane (MVP), lateral horizontal plane (LHP) and lateral vertical plane (LVP). The investigators found that both hearing impaired groups exhibited poorer localization performance compared to the tested control group of normal hearing individuals In particular, the sensorineural group was more accurate than the conductive/mixed group in the FHP The difference in performance can be attributed to the inability of individuals with conductive/mixed hearing loss to access interaural time difference cues, which dominate FHP judgements In the MVP, localization performance was poor for both hearing impaired groups but to a greater extent for the sensorineural group. The loss of higher frequencies lead to the collapsing of vertical locations onto the horizontal plane, which suggests that the poorer performance of those subjects with sensorineural hearing loss was attributable to the absence of high-frequency pinnae-based cues. Both groups performed poorly in the LHP with performance being mostly attributed to high-frequency hearing loss, which caused difficulty in front/rear discrimination. These investigators concluded that hearing impaired listeners rely on the same auditory cues as do normally hearing listeners, and that there is no indication that impaired listeners can learn to compensate for their hearing deficiencies by using other types of information (e g., head movement) to process localization cues.

The effect of sensorineural hearing loss on localization performance in virtual auditory space has also been studied (Smith-Olinde, Koehnke & Besing, 1998). The stimulus was spatialized using the KEMAR HRTFs and localization judgements were tested in the FHP These investigators reported that localization for the hearing impaired group was poorer compared to the normal listeners.

Based on the results of Noble et al. (1994) and Smith-Olinde et al. (1998), hearing impaired listeners may partially benefit in improved performance with the use of a 3-D audio display However, the range of applications must not require conditions of high situational awareness Furthermore, it may be prudent that users of a 3-D audio display be audiometrically screened on an on-going basis to ensure normal hearing acuity. This is particularly vital for personnel with advancing age who might have acquired presbycusis (the loss of hearing sensitivity at high frequencies with increasing age). Moreover, the hearing threshold levels at specific frequencies may also serve as a predictor of localization performance for certain positions in auditory space (Noble et al., 1994).

A final point that merits attention from a safety perspective is the treatment and occurrence of localization judgements that result in reversals. Some present reports are misleading in their treatment of data on reversals In localization studies reversals are commonly resolved by coding the subjects' response as if it were indicated in the correct hemisphere (Oldfield & Parker, 1984; Wightman & Kistler, 1989, Wenzel et al., 1993). Clearly, resolving reversals in this manner for critical mission applications could be fatal. For example, there is no tolerance for reversals if virtual sources are to serve to cue aircrew to the spatial location of a potential lethal threat such as another aircraft.

Investigators have also reported greater occurrence of reversals in virtual auditory space regardless of the choice of HRTFs (personal or generic) compared with localization in the free-field. Wightman and Kistler (1989) found that the percentage of front-back reversals on average when using personal HRTFs was almost twice as high for virtual auditory space as for free-field (11% versus 6%). Wenzel et al. (1993), who performed a similar experiment to Wightman and Kistler (1989), found that front-back reversals were higher in virtual auditory space than in free-field (31% versus 19%) when generic HRTFs were used. The cause of this imperfection in the simulation of virtual sound sources is not yet fully understood. Bronkhorst (1995), for example, attributes this imperfection to an incorrect simulation of high frequency spectral cues above 7 kHz probably caused by a distortion introduced by the HRTF measurements performed with probe microphones in the listener's ear canals. However, Martin, McAnally and Senova (submitted) recently evaluated a HRTF measurement technique from their laboratory They compared virtual and free-field localization performance across a wide range of sound-source locations for three subjects. For each subject, virtual and free-field localization performance was found to be indistinguishable, as indicated by both front-back reversal rates and average localization errors. The development of a system of such high fidelity is a significant milestone in the maturation of virtual audio technology.

6 CONCLUSIONS AND RECOMMENDATIONS

Operator performance has been demonstrated to improve with the use of a 3-D audio display in a range of laboratory studies, as discussed in this paper The benefit of increased performance can result in reduced workload and increased situational awareness. A 3-D audio display can also assist with the reduction of the overall amplitude of the communication system without sacrificing detectability and thus ensuring safe headphone levels of auditory warnings which at present are frequently too loud (Patterson, 1982). Quicker reaction time with the use of a 3-D audio display compared to conventional aural presentation has also been reported. Improved reaction time could be extremely critical in emergency situations where appropriate and evasive action must be made quickly and correctly. Further gains in decreased workload and/or quicker reaction time may result if information is presented in a bimodal display and thus it is recommended that this be further investigated.

Real-world applications of 3-D audio are likely given its apparent benefits as demonstrated by the outcome of the field trials discussed in this paper The development by a Canadian company of a commercial shipboard communication product for the spatialization of up to four static talkers indicates that industry is seriously investigating the performance gains with a 3-D audio presentation Before the implementation of 3-D audio into real-world applications, it is essential that investigators first determine the nature and extent of "gaps" in research and technical knowledge bases relative to this new technology For example, Arrabito et al (1996) demonstrated the lack of advantages found for 3-D audio compared to a 2-D audio display. In this instance it appears that R&D efforts should be focused on a 2-D audio display when listeners need to attend to only two competing talkers.

In a real-world application such as one of a complex divided attention task, the binaural synthesis of the signals must provide accurate localization performance to the end-listener in order to maintain a precise and consistent sense of situational awareness. This type of task is the most stringent as far as the quality of the binaural synthesis of signals is concerned, but potentially offers the greatest benefit. More fundamental research is needed to quantify the real advantage gained by spatializing the communication signals for such tasks, in terms of speech intelligibility, total information transfer and/or situational awareness (Ericson & McKinley, 1997).

Furthermore, in spite of the support for 3-D audio technology, performance will be ultimately dependent on the hearing ability of the end-listener. It has been demonstrated that the loss of hearing results in poorer localization performance compared to normal hearing listeners. Poor localization ability may be fatal in critical mission applications and thus hearing loss needs to be factored into consideration. Further research into the effects of hearing loss on localization performance in virtual auditory space with the use of personal and generic HRTFs needs to be conducted Testing should be performed in free-field and virtual auditory space in a similar way to the methods reported by Wightman and Kistler (1989), and Wenzel et al. (1993).

At present, several critical factors are impeding the transition potential of 3-D audio technology into real-world applications. These include the choice of personal versus generic HRTFs, environmental factors and communication bandwidth. In particular, performance in virtual auditory space is more accurate and results in fewer localization reversals with personal HRTFs compared to generic ones. However, personal HRTFs are traditionally derived from binaural measurements in the ears of the end-listener seated in an anechoic chamber (a room without reverberation cues down to a specific cutoff frequency). This requires a substantial investment in infrastructure and equipment, and is presently impractical in most applications. Methods thus need to be developed to quickly and accurately select and/or modify a generic HRTF for the targeted application. To implement man-on-the-move applications will require further investigation into head tracking systems and interpolation techniques of HRTFs. The effect of ambient noise on user performance using either personal or generic HRTFs also needs investigation. The hardware limitation imposed on the communication bandwidth needs to be addressed. Until the above issues are more fully understood and resolved it may be prudent to proceed cautiously before the adoption of a 3-D audio system. Advances in these issues are being made as demonstrated by the results of Martin et al. (submitted) which suggest that the imperfection in the simulation of virtual sound sources is an obstacle that may have been surmounted. This represents a significant achievement in the reproduction of spatial synthesis. In the meantime, it is suggested that defence forces examine archived audiograms given that there is a partial relationship between hearing threshold levels and localization performance as reported by Noble et al. (1994). This would serve as a criterion for evaluating the benefits of a 3-D audio display.

7 ACKNOWLEDGEMENTS

Gratitude is extended to David Beevis and Sharon McFadden who generously offered valuable suggestions The author thanks two anonymous reviewers for their comments on an earlier draft of this paper Technical assistance was provided by Phiroz Dastoor, Garry Dunn, Carolyn Michielsen and Alexandra Ostaniewicz. Special thanks goes to Bill Nugent who motivated the foregoing discussion of 3-D audio for the Technical Co-operation Program Technical Panel HUM TP-9 to identify the human factors issues of advanced technologies.

8 REFERENCES

Alsten, C.R., Downey III. R., & Jackson, T B (1995). Preliminary evaluation of a "sleep training" program to enhance the quality and quantity of aircrew sleep. Aviat. Space Environ Med., 66(5), 500.

Arrabito, G R , Cheung, B , Crabtree, R B., & McFadden, S.M. (2000) Release from masking in virtual auditory space during sustained positive acceleration Aviat Space Environ. Med., 71(5), 536-542.

Arrabito, G R., McFadden, S M., & Crabtree, R B (1996) The effectiveness of binaural auditory displays as a means for improving intelligibility of dichotic messages. In. *Proceedings of the Human Factors Association of Canada 28th Annual Conference*, (pp 47-51). Toronto, Canada. Human Factors Association of Canada

Begault, D R. (1993) Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation Hum. Factors, 35(4), 707-717

Begault, D R. (1992) Perceptual effects of synthetic reverberation on three-dimensional audio systems J Audio Eng. Soc , 40(11), 895-904

Begault, D R. (1991). Challenges to the successful implementation of 3D sound J. Audio Eng Soc , 39(11), 864-870

Begault, D R., & Wenzel, E M (1993) Headphone localization of speech Hum. Factors, 35(2), 361-376.

Blauert, J (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization* J Allen (Trans), Cambridge, MA MIT Press

Bronkhorst, A W (1995). Localization of real and virtual sound sources J. Acoust Soc Am., 98(5), 2542-2553.

Bronkhorst, A.W , Veltman, J.A., & van Breda, L. (1996) Application of a three-dimensional auditory display in a flight task Hum. Factors, 38(1), 23-33

Calhoun, G.L., Janson, W P , & Valencia, G (1988). Effectiveness of three-dimensional auditory directional cues. In: *Proceedings of the Human Factors Society 32nd annual meeting*, (pp 68-72). Santa Monica, CA. Human Factors Society.

Cheng, C I , & Wakefield, G H. (1999). Spatial frequency response surfaces (SFRSs) An alternative visualisation and interpolation technique for head-related transfer functions (HRTFs). In. *Proceedings of the AES 16th International Conference Spatial Sound Reproduction*, (pp 147-159) Rovaniemi, Finland: Audio Engineering Society

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. J Acoust. Soc. Am , 25(5), 975-979

Doll, T J , Gerth, J.M., Engelman, W R., & Folds, D J. (1986) Development of simulated directional audio for cockpit applications. (Report no AAMRL-TR-86-014) Ohio. Wright-Patterson Air Force Base, Armstrong Aerospace Medical Research Laboratory

Ericson, M A , & McKinley, R.L. (1997). The intelligibility of multiple talkers separated spatially in noise In: R.H. Gilkey and T R. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments*, (pp. 701-724). New Jersey: Lawrence Erlbaum Associates.

Furness, T.A. (1986). The super cockpit and its human factors challenges. In *Proceedings of the Human Factors Society 30th annual meeting*, (pp 48-52) Dayton, OH: Human Factors Society

Hartung, K., Braasch, J , & Sterbing, S.J (1999) Comparison of different methods for the interpolation of head-related transfer functions. In: *Proceedings of the AES 16th International*

Conference Spatial Sound Reproduction, (pp 319-329) Rovaniemi, Finland: Audio Engineering Society

King, R B , & Oldfield, S R. (1997) The impact of signal bandwidth on auditory localization Implications for the design of 3-dimensional audio displays Hum. Factors, 39(2), 287-295

Martin, R L , McAnally, K I , & Senova, M (submitted) Free-field equivalent localisation of virtual audio J Audio Eng Soc

McKinley, R L., & Ericson, M.A (1997) Flight demonstration of a 3-D auditory display In: R.H Gilkey and T R Anderson (Eds), *Binaural and Spatial Hearing in Real and Virtual Environments* (pp 683-699) New Jersey Lawrence Erlbaum Associates.

Moller, H (1992). Fundamentals of binaural technology Appl Acoust., 36(3/4), 171-218.

Moller, H., Sorensen, M F, Hammershoi, D , & Jensen, C B (1995) Head-related transfer functions of human subjects. J. Audio Eng. Soc., 43(5), 300-321

Moore, B.C.J. (1989). An Introduction to the Psychology of Hearing (3rd ed.). San Diego. Academic Press.

Noble, W , Byrne, D., & LePage, B (1994) Effects on sound localization of configuration and type of hearing impairment. J Acoust. Soc. Am., 95(2), 992-1005

North, D A., & D'Angelo, W R. (1997). 3-Dimensional audio ergonomic improvement project for the NORAD CMOC. (Report no. AL/CF-TR-1997-0170). Ohio. Wright-Patterson Air Force Base, Armstrong Laboratory.

Oldfield, S R , & Parker, S P.A. (1984) Acuity of sound localisation A topography of auditory space I Normal hearing conditions Perception, 13(5), 581-600

Patterson, R (1982) Guidelines for auditory warning systems on civil aircraft (Report No CAA Paper 82017) London, UK Civil Aviation Authority

Perrott, D R., Cisneros, J , McKinley, R L , & D'Angelo, W R, (1996) Aurally aided visual search under virtual and free-field listening conditions Hum Factors, 38(4), 702-715

Smith-Olinde, L , Koehnke, J , & Besing, J (1998) Effects of sensorineural hearing loss on interaural discrimination and virtual localization J Acoust. Soc. Am., 103(4), 2084-2099

Vause, N L , & Grantham, D W (1998) Speech intelligibility in adverse conditions in recorded virtual auditory environments In *Proceedings of the International Conference on Auditory Displays*, (pp. 1-13). Glasgow, UK British Computer Society

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

Wightman, F L , & Kistler, D J (1989) Headphone simulation of free-field listening. II. Psychophysical validation. J Acoust Soc. Am., 85(2), 868-878.

Yost, W A. (1997). The cocktail party problem. Forty years later In: R H. Gilkey and T R. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments* (pp 329-347) New Jersey· Lawrence Erlbaum Associates

#515432