Limitations of 3D Audio to Improve Auditory Cues in Aircraft Cockpits

C.W. Johnson and W. Dell; Dept of Computing Science, Univ. of Glasgow, Scotland, G12 9QQ.

Keywords: Accident Reporting, Media, Causal Analysis, Accident Investigation.

Abstract

Several organizations, including the FAA and NASA, have sponsored research projects into the use of spatialised audio as a means of improving warnings in commercial aircraft. Other research groups, including the Australian Defence Science and Technology Organisation, have explored the use of these techniques within military aviation. A common factor across all of this work is an apparent enthusiasm for the potential benefits of this technology. In contrast, this paper describes the problems that were encountered when attempting to derive empirical evidence for the benefits of 3D auditory cues in aircraft cockpits. Benefits were identified for the stereo presentation of auditory information but it was far harder to demonstrate any additional support for the use of more sophisticated techniques. The second half of this paper extends our investigation beyond the laboratory to examine pragmatic barriers that frustrate the introduction of this technology. These range from the problems of integrating cockpit warning systems through to an apparent confusion in the recommended practices for headphone use in Europe and North America.

Introduction

3D audio provides means of reducing the clutter that affects many visual displays in safety critical systems (ref. 1, 2, 3). A series of civil projects have been conducted in this area sponsored by the FAA and by NASA. There has also been considerable interest in military applications. For example, the Australian Defence Science and Technology Organisation "has been working on ways to enhance the performance of aircrew in combat situations by giving them sound cues that help to rapidly detect the direction and types of threats and targets. Synthesised three-dimensional sound, which mimics the sounds of the real world, is developed from recordings taken from within the 'ears' of individual crewmembers, and then reproduced digitally to provide the critical directional information. 3D sound technology will reduce the high visual workloads experienced by aircrew and provide them with more time to respond to threats or targets, giving them an operational advantage and a safety edge" (ref. 4).

3D audio systems provide the impression that sounds emanate from different positions within physical space. There are several proposed benefits. Ideally, it should be possible for users to selectively attend to audio information from particular locations within 3D space. This supports the everyday filtering that people use to attend to a single speaker in a crowded room. Conversely, it should also be possible to monitor several different channels of information from audio sources in different positions. This resembles the everyday experience of listening to several different conversations at once without attending to one in particular. This paper presents the results of an investigation to determine whether these benefits might be extended to support flight deck operations.

In order to understand the potential strengths and weaknesses of 3D audio in safety-critical applications it is first necessary to introduce the underlying technology. Most auditory displays warn users about changes in an underlying system or remind operators to perform necessary actions. Existing applications are almost exclusively monaural. Users have the impression that all sounds come from the same location. This is because a single speaker is used to generate them. Monaural reproduction can be used to implement a form of spatialised audio by playing sounds through dedicated speakers in different locations. Unfortunately, space constraints limit the use of this approach in many applications. For example, we might like a sound to emanate from a location that is already occupied by a visual display or by a window that enables operators to directly observe application processes. A variety of techniques can, however, be used to create the impression that sounds emanates from positions between two or more speakers. For instance, two speakers can be used to provide the impression that sounds emanate anywhere on a line between them, as illustrated in Figure 1b, by altering the volume of a sound in each headphone or speaker.

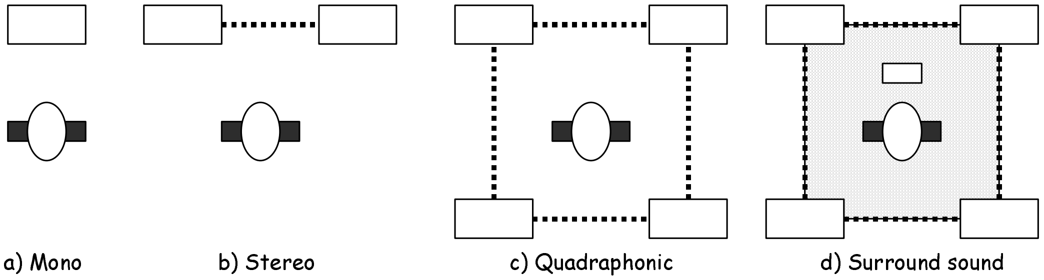

Figure 1 – Locating Sound with Speaker Arrangements

As the name suggests, quadraphonic reproduction extends the stereo approach to four speakers. This technique can be used to place the sound anywhere in the perimeter illustrated by Figure 1c. As before, this relies upon changes to the volume of a particular sound as each speaker reproduces it. Surround sound builds on the quadraphonic configuration with an additional speaker positioned close to the listener to provide the impression that the sound originates from any position within the bounding box of Figure 1d. All of these approaches are restricted to a single horizontal plane unless further speakers are added above and below the listener.

In contrast to the speaker arrangements illustrated in Figure 1, spatialised 3D-audio simulates what happens to sound as it travels from the source to the listener’s ears. Computational techniques can be applied to simulate the delays and changes in amplitude that occur when the same sound is played in different locations. For example, the distance that sound has to travel is different for each ear. This creates a delay that people exploit to determine the direction of a sound source. The interaural time difference is not apparent for sounds directly in front or behind the listener but for sounds positioned to the far left or right it can be up to 0.63 milliseconds. Computers can simulate this delay through standard headphones to provide the impression that the sound originates from particular positions within physical space. Similar techniques can simulate interaural intensity differences. These differences in volume can be created by the ‘head shadow’; that is the difference in amplitude created because the listeners’ head is between the source and an ear. Spatialised audio systems have also been developed to account for sounds that are reflected from the listeners’ shoulders; this provides important cues about the elevation of a sound source. Computational models that simulate echoes and reverberation provide further cues. There are fewer echoes if a source is closer to the user than if it is far away.

The modelling and reproduction of these different cues is a research area in its own right. Collectively, however, they support the development of Head Related Transfer Functions (HRTF). These represent mathematical transformations on the spectrum of a sound that simulate the cues mentioned above. The particular nature of the HRTF depends on the location of the sound being simulated. More than a thousand different functions are needed to cover a sufficient number of directions and distances to adequately simulate even a primitive 3D space. In practice, HRTFs can be constructed by placing microphones in the auditory canals of a dummy. Sounds are then played from fixed locations to analyse the effects of location on different frequencies. Increasingly, however, research is moving from generic HRTFs towards functions that are specifically tailored to the physiological characteristics of particular individuals. These individual models may be required if users are to make fine-grained judgements about the origin of a particular sound. Standards are being developed that will enable researchers to exchange 'head maps'. These record the many different HRTFs that together characterise the way in which an individual will hear sounds from different locations. Further research is also investigating dynamic HRTFs that account for the changes as we move our heads to attend to particular sound sources. In order for computers to simulate these effects, location trackers are being incorporated into headphones so that the system can respond to changes in the users' orientation with respect to the source.

The introduction of electronic display units in 767/757 and A310s has done much to reduce visual clutter in the cockpit. However, British Airways continues to record a higher frequency of situation awareness incidents in its 'glass display' fleets compared to aircraft with more conventional, if cluttered arrangements, such as the B747-200 (Ref 5.). As a consequence, 737-700 and 800 offer electronic analogues of previous displays (ref. 6). Moving more information to auditory channels offers an attractive alternative to systems that over-simplify or hide important information from the flight crew. However, in some cases "traditional visual instrumentation has been replaced with bells, beepers and electronic tones to a point where auditory clutter is almost as bad as visual clutter" (ref. 3). This has reached a stage where many pilots cannot recall the meanings for the non-speech sounds that are presented to them. Doll et al (ref. 7) illustrate further problems in their analysis of the F-16. Audio alarms are used for ground proximity and angle-of-attack warnings. One indicates a need to raise the nose and the other to lower the nose yet they both use an 800 Hz tone. Using speech-based displays that provide content cues about the nature of the warning can reduce some of these recognition problems. However, they can create problems if they distract pilots from other critical tasks. Pilots often criticise speech-based warnings as "noisy, strident and intrusive" (ref. 3). The limitations with existing speech and non-speech aviation displays motivated an investigation into whether 3D audio technology might address some of these problems.

The Experiment

Our study focuses on the use of a simulator. Users were presented with a primary task that involved controlling an artificial horizon, illustrated in Figure 2. Participants had to ensure that the circle was over the cross hair in order to stabilise the horizon. The position of the user in relation to the horizon was changed periodically. They then had to move the circle towards the cross hairs in order to retain control. The intention was that this primary task would require constant interaction by the user. One of the claimed benefits of 3D audio is that it supports interaction during such high-workload situations (ref. 1). Users can gain additional cues based on the location of the sound without devoting the additional cognitive resources that might otherwise have to be used to attend to more complex warning messages, for instance about the location of a conflicting aircraft.

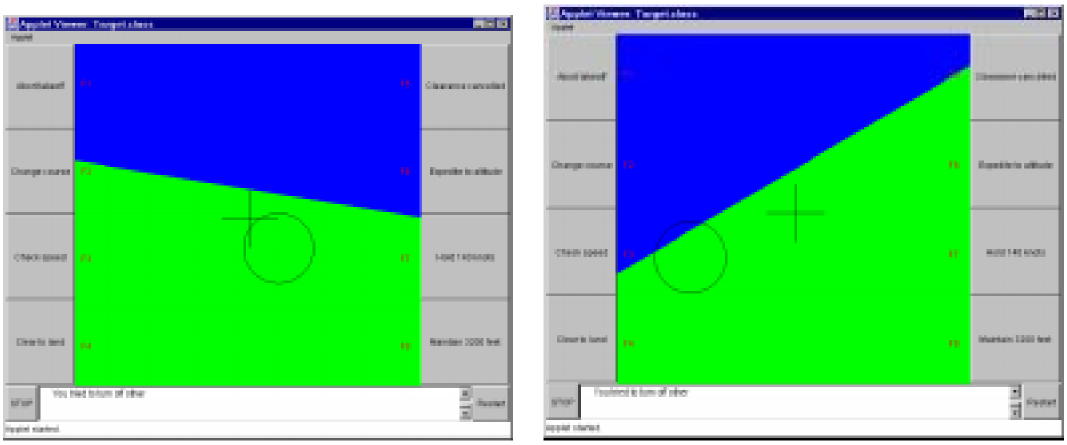

Figure 2 – Simulator Display Showing Primary and Secondary Tasks

The secondary task was acknowledging eight different speech-based Air Traffic Management advisories that were obtained from the Federal Aviation Authority. The user had to wear headphones to hear either monaural, stereo or spatialised forms of these warnings. The monaural advisory was played at the same volume in both ears. The stereo advisory was reproduced at full volume in either the left or the right ear but was silent in the other ear. The spatialised mode mapped each of the eight advisories to a position equidistant from the user's head but spaced at 45 degrees on a horizontal plane around their head. This design was chosen because preliminary studies indicated that participants had particular problems in discriminating sounds at different locations in the Z plane. The position cues offered by the Air Traffic Management advisories could be used to locate the function keys that were used to confirm each instruction.

Twenty subjects completed a brief questionnaire before beginning the experiment. We were keen to identify any biases in our sample population. For instance, the questionnaire revealed that most of our participants played a computer game approximately once every month rather than daily or weekly. 14 out of the 20 participants did not play a musical instrument. We ensured that those individuals who did play an instrument were fairly distributed between the different experimental groups, described below. We relied upon a counter-balanced single-factor, within groups experimental design. The independent variable was the sound mode; mono, stereo or spatialised. We measured dependent variables including response times from the advisory being raised to when it was acknowledged. We also used a tracking system to record performance with the primary task in terms of the time that the circle was over the crosshair. We recorded 'error' rates in terms of whether users correctly acknowledged the alarm. We also conducted a workload analysis using NASA's TLX measures (ref. 3).

Each subject heard twenty monaural advisories. They then heard ten stereo and ten spatialised warnings. The order of the advisories was randomised and the ordering of the modes was counter-balanced to reduce learning effects. Our prediction was that the response times would be lower for the spatialised mode compared to the stereo advisories. We also hypothesised better primary task performance, in terms of the time that the circle was over the crosshair; lower error rates and improved learnability for the spatialised condition. We also expected that the six TLX measures will reflect the lowest overall workload for the spatialised avisories, then stereo and finally monaural.

	Minimum (ms) (with subject)	Maximum (ms) (with subject)	Range (ms)	Mean (ms)	Standard Deviation (ms)
Mono	2520 (B)	5184 (S)	2664	3973	641
Stereo	2189 (A)	4712 (N)	2523	3360	663
3D	2489 (K)	4782 (M)	2293	3479	735

Table 1 – Response Times for Monaural, Stereo and Spatialised Modes.

	Minimum (ms) (with subject)	Maximum (ms) (with subject)	Range (ms)	Mean (ms)	Standard Deviation (ms)
Mono	56.2% (G)	99.6% (I)	43.4%	88.4%	9.8%
Stereo	72.7% (G)	100% (I)	27.3%	94.5%	6.3%
3D	72.8% (G)	99.4% (D)	26.6%	94.1%	6.1%

Table 2 – Performance for Monaural, Stereo and Spatialised Modes

Table 1 presents response times from the evaluation. The letters after the minimum and maximum times refer to individual subjects. This shows that no user was consistently faster or slower across all modes. Although the response times for stereo alarms are faster than 3D alarms (Stereo mean=3360ms, 3D mean=3479ms), a t-test highlighted that this difference was not statistically significant. Table 2 presents the performance results from the evaluation calculated in terms of the percentage of time that the cross hairs were over the target area. Again, although the mean performance was slightly better for stereo than 3D (Stereo mean=94.5%, 3D mean=94.1%) t-test results showed that this was not significant.

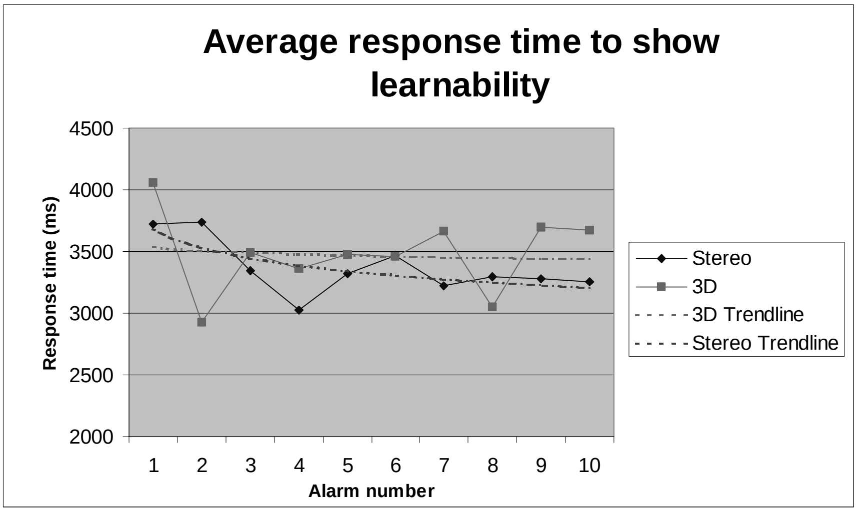

Figure 3 - Graph showing ‘Learnability’ of the Stereo and 3D Simulations.

Figure 3 shows the mean response times of the twenty subjects for the first advisory, then the second advisory, then the third advisory and so on. As can be seen from the trend lines, stereo has a steeper downward gradient than 3D. However, given the lack of statistical significance of the overall results, further work is required before any firm conclusions can be reached about the relative benefits of these modes in terms of performance improvements over time. The participants made remarkably few errors. The mean was 0.2 across the twenty subjects for both the stereo and 3D simulations. More errors were made for the monaural condition. However, these errors may have been the result of a learning effect. To help minimise this effect we chose to run the monaural trials before the counterbalanced stereo and spatialised conditions.

NASA's Task Load Index (TLX) was used to assess the workload associated with the monaural, stereo and spatialised conditions. This instrument has been widely applied in many similar studies and hence provides a common point of reference for work in this area. Each participant was asked to complete a TLX questionnaire. This involved rating six factors that contribute to overall workload: mental demand; physical demand; temporal demand; effort, performance and frustration level. Table 3 illustrates the format of the scalar question for mental demand.

Mental Demand	Low 1	2	3	4	5	6	7	8	9	High 10
How much mental and perceptual activity was required (e.g., thinking, deciding, calculating etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?										

Table 3 – Example of a TLX Scalar Question

The majority of the participants rated stereo as requiring the least amount of mental demand, then 3D and then monaural. The mean ratings were 6.2 for mono, 5.1 for 3D and 4.2 for stereo. To establish if the ratings could have been due to chance effects, a within groups, paired two sample for means t-test was carried out for each category. A confidence level of 95% was used so that a t-value of 2.09 was necessary to conclude that the independent variable was responsible for an effect. Three t-tests were carried out for each category to compare mono versus stereo (t-value of 4.3), mono versus 3D (t-value of 2.9) and stereo versus 3D (t-value of 3.3). For each, the t-value was greater than 2.09 confirming that the stereo simulation required the least amount of mental demand, then the 3D simulation and then the monaural simulation.

The physical demand section of the TLX instrument refers to: "How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?" The mean rating given by the subjects was 5.5 for mono, 4.3 for 3D and 4.1 for stereo. The two t-tests comparing mono against stereo and 3D showed that mono required the most physical demand (t-values of 3.78 and 3.61 respectively). This result was expected. We used the monaural simulation for the initial familiarisation with the primary task. Hence, this stage lasted twice as long as the stereo or spatialised simulations. It would, therefore, arguably have required more physical demand. The t-test comparing stereo to 3D enables us to conclude that the difference could have been caused by chance (the t-value was only 0.62 below the required 2.09 level).

The temporal demand question in TLX refers to: "How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?" The mean rating was 4.6 for mono, 4.1 for 3D and 3.6 for stereo. The two t-tests comparing mono against stereo and 3D show that the monaural advisories imposed the greatest temporal demands (t-values of 3.16 and 2.52 respectively). However, the difference between stereo and 3D could have been due to chance (the t-value was 1.81).

The TLX questionnaire assesses participants' effort by asking "How hard did you have to work (mentally and physically) to accomplish your level of performance?" The mean was 6.2 for mono, 4.4 for 3D and 4.6 for stereo. The two t-tests comparing mono against stereo and 3D show that mono required significantly more effort (t-values of 4.59 and 4.15). However, the t-test comparing stereo to 3D again suggests that the difference could have been caused by chance (the t-value was 0.72).

Performance is assessed by "How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?" The mean rating was 4.9 for mono, 5.6 for 3D and 6.4 for stereo. The t-tests comparing stereo against mono and 3D showed that the subjects thought they were the most successful with stereo (t-values of 3.22 and 3.13 respectively). However, the t-test comparing mono to 3D enables us to conclude that the difference could have been caused by chance (the t-value was 1.63).

TLX measures frustration by asking "How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?" The mean rating was 4.5 for mono, 3.7 for 3D and 3.2 for stereo. The two t-tests comparing mono against stereo and 3D conclude that the subjects were the most frustrated with the mono experiment (t-values of 2.63 and 2.10 respectively). However, the t-test comparing stereo to 3D enables us to conclude that the difference could have been caused by chance (the t-value was 1.21).

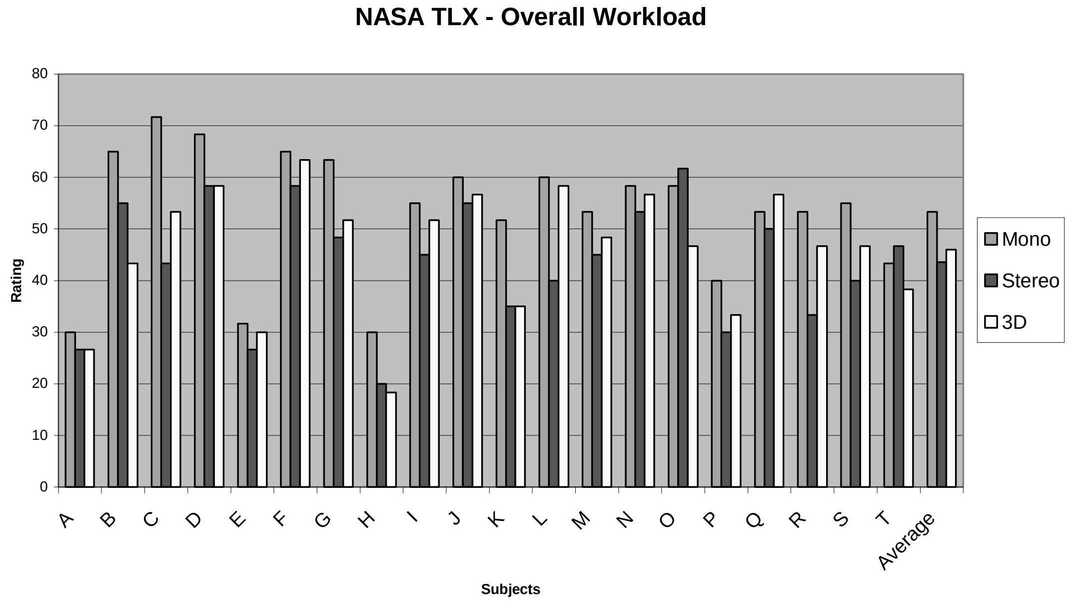

Figure 4 – Overall Workload Using NASA's TLX Metrics

Overall workload is calculated from the six factors described in previous paragraphs. The mean workload was 5.3 for mono, 4.6 for 3D and 4.3 for stereo. As might be expected, the two t-tests comparing mono against stereo and 3D conclude that the subjects had the highest overall workload with the mono experiment (t-values of 5.57 and 5.12 respectively). However, the t-test comparing stereo to 3D enables us to conclude that the difference could have been caused by chance (the t-value was 1.39).

Analysis: What are the Real Barriers to 3D Audio in the Cockpit?

A number of different explanations can be proposed for the differences between our results and those of previous researchers working on the application of 3D audio to safety-critical systems. Firstly, it can be argued that the design of our simulation did not effectively utilize the best potential of spatialised sound. Our sources were placed at 45 degrees of separation, equidistant from the listener in a single horizontal plane. This was justified by a series of preliminary studies that pointed to the problems our participants experienced in discriminating between sources placed at different positions on the z-plane. We are currently working to replicate our study with the Air Traffic Advisories in more complex spatial configurations.

Our results may also stem from particular hardware and software limitations. The use of esoteric computational and audio systems has arguably hindered research in this area. Labs have chosen to develop in-house systems that make it very difficult to replicate any results. This prevents human factors researchers from making direct comparisons between different designs for spatialised audio systems. To address this, we have placed all of our materials on a public web site as a Java application that will run on most desktop PCs (ref. 3). In consequence, designers can compare this set-up and alter the location of our sources in the code to identify the differences that might account for their success and our failure.

A second explanation of our broadly negative results is we relied upon generic HRTFs. Much of the current military work in this area focuses on individual 'head maps' that offer significant improvements in localised sound reproduction. This raises numerous issues about the feasibility of 3D audio in commercial aviation. The development of HRTF's for individual operators is currently a painstaking process. Certification might, therefore, depend upon the development of a suitable spatialised display as well as a set of standards for the recording of individual 'head maps'. It would then be necessary to ensure that these maps were preloaded on a flight deck and correctly recalled, for instance using biometric data prior to flight. All of this is technologically feasible given existing technology. Unfortunately, it is a long way from current practice in commercial aviation. The following reports are intended as a useful 'reality check' for research in this area. The first comes from the US Aviation Safety Reporting System (ASRS) and describes a pilot's ad hoc solution to noise cancelling headsets:

> **"**[I] was unable to hear gear warning horn because of new noise cancelling headsets. I recommend removal of one earpiece in landing phase of flight to audible warning devices to be heard by pilot. The noise-cancelling headsets were tested by three people on the ground and all three noted that with the headsets active that the gear warning horn was completely masked by the headsets." (ASRS Callback, Issue 247, January 2000).

The lack of international standards and operating procedures governing the use of headsets is illustrated by a second quotation from the UK equivalent of the ASRS known as CHIRP:

> "I would like to highlight a common practice by some airlines, including my employer, which I feel is a significant risk to flight safety: namely the practice of not using flight deck intercom systems in favour of half wearing a headset over one ear... I, and other colleagues of mine, have had to ask for the other crewmember to repeat things because of aircraft noise in one ear, and ATC in the other with the volume turned high enough not to miss a call... Myself and others have raised this point several times to our training and safety departments, all of which has fallen, pardon

> the pun, onto deaf ears. The stock answer is that there are no written down Standard Operating Procedures on intercoms, and common agreed practice rules.” (CHIRP Feedback, No: 51 July 1999)

Conclusions

This paper has described an experimental study that contradicts many previous investigations into the use of 3D audio in safety-critical systems. The only significant differences that we found between spatialised and stereo sources showed that 3D advisories imposed fewer mental demands on the users. However, participants felt that they performed significantly better with stereo sources even though this was not confirmed by the outcome measures. Differences in terms of actual performance, error rates, physical demand, temporal demand, effort and frustration were not significant. However, significant differences were found between spatialised/stereo reproduction and the monaural advisories. These differences may have stemmed from the use of monaural training in our experimental design.

Our negative results do not undermine the general utility of 3D audio. Improvements can be expected from individualised HRTFs, from alternative layouts for sound sources across several horizontal planes and from the integration of spatialised audio with head tracking. However, we have established that some configurations do not offer the benefits that many would claim for 3D audio. Our results also indicate that a more complex form of analysis is required. In particular, we would like to know why participants thought they performed better with stereo advisories even though they actually achieved similar results with spatialised sound sources. Such results have particular significance if pilots are to accept spatialised audio on the flight decks of the future.

The closing sections of this paper have discussed a number of directions for future research. In particular, it has been argued that spatialised audio systems may also require some form of biometric recognition if they are to successfully exploit individualised HRTF’s. We have also argued that such proposals are a long way from current practice in commercial aviation. In anticipation of these developments, we must continue to operate in noisy cockpits for which there are few agreed operating procedures governing the use of headphones during critical stages of flight.

References

1 H.O.W. Carlander, Localisation of sound in a horizontal 3D-audio environment, Technical report, Swedish Centre for Human Factors in Aviation Linköping Institute of Technology. 2002.

2. E.M. Wenzel, M. Arruda, D.J. Kistler and F.L. Wightman, Localisation using Non-Individualized Head-realed Transfer Functions. Journal of the Acoustical Society of America, 94:111-123, 1993.

3. W. Dell, The Use of 3D Audio to Improve Auditory Cues in Aircraft, Technical Report, Department of Computing Science, University of Glasgow, 2000. Available with experimental software from http://www.dcs.gla.ac.uk/research/gaag/dell/

4. Australian Defence Science and Technology Organisation, Sixty Years of Aeronautical Research in Australia: Research Overview. Available at http://www.dsto.defence.gov.au/ March 2003.

5. M. O'Leary, Situation Awareness and Automation. In K. Abbott, J.-J. Speyer and G. Boy (eds.) HCI-Aero 2000, 73-79, Cepadues-Edition, Toulouse. France, 2000.

6. C.E. Billings, Aviation Automation. Lawrence Erlbaum Associates, New Jersey, USA, 1997.

7. T.J. Doll, J.M. Girth, W.R. Engelman and D.J. Folds, D.J., Development of Simulated Directional Audio for Cockpit Applications, USAF Report AAMRL-TR-86014, 1986.

Biography

Prof. C.W. Johnson, MA, MSc, DPhil, CEng, Glasgow Accident Analysis Group, Dept. of Computing Science, University of Glasgow, Glasgow, G12 9QQ, Scotland, UK, tel. - +44 141 330 6053, fax. - +44 141 330 4913, johnson@dcs.gla.ac.uk, http://www.dcs.gla.ac.uk/~johnson

Chris Johnson is Professor of Computing Science at the University of Glasgow. He heads one of the largest research teams specifically devoted to new generations of accident analysis techniques. He helped to author European guidelines for mishap reporting in Air Traffic Management. He has developed an incident analysis scheme for the UK Health and Safety Executive that is specifically designed to support the investigation of adverse events involving programmable systems across the process industries. In 2002, he held a NASA/ICASE fellowship analysing a series of mishaps, including the SOHO mission interruption.