Improving 3-D Audio Localisation Through the Provision of Supplementary Spatial Audio Cues

John Towers[*,1], Robin Burgess-Limerick[2] and Stephan Riek[3]

[1]*Boeing Defence Australia, Boeing Research & Technology Australia, Brisbane 4001, Australia*

[2]*The University of Queensland, Centre for Sensorimotor Neuroscience, School of Human Movement Studies, St. Lucia 4072, Australia*

[3]*The University of Queensland, Minerals Industry Safety and Health Centre, St. Lucia 4072 Australia*

Abstract: This study examined whether azimuth localising performance for non-individualised 3-D audio without integrated head tracking can be improved through the provision of supplementary reference signals. Twenty-two participants attempted to determine the location of spatial sounds developed through a non-individualised head-related transfer function (HRTF) while performing a visual distractor task. Localising sounds were randomly presented at 0-degrees elevation for each 10-degree increment about the azimuth. Three audio conditions were tested, two of which included different supplementary cues in the form of stationary and transient sounds that were spatially positioned to aid localising reference toward the midsagittal plane and interaural axis. The supplementary cues decreased errors in front-back perception; however, they did not significantly aid azimuth localising performance, and occasionally were reported to distract and disorient some participants. Supplementary audio cues have the potential to improve localising performance but should be more closely associated with the presented sound to lessen distraction and disorientation.

Keywords: 3-D Audio, Audio Display, Audio Localisation, HRTF.

INTRODUCTION

In and of themselves, sound waves travelling through free space carry no spatial localising cues. The Duplex Theory of Localisation identifies interaural time differences (ITD) and interaural intensity differences (IID) as the primary cues associated with localising sounds about the azimuth [1]. The ability to differentiate between locations where ITD and IID are effectively equal, such as variations in elevation and across similar front and back angles (e.g. 10^o and 170^o), is attributed to spectral cues that appear in the form of frequency distortions in the sound waveform above 4 kHz. They occur when sound is distorted upon impact with the physical shape of the pinnae, head, and upper torso [2]. Synthesising spatial cues for presentation through binaural headphones is made possible by utilising audio filters called Head Related Transfer Functions (HRTFs), which are created from signal recordings at the eardrum or ear canal of an individual situated within an anechoic chamber [3].

Developing individualised HRTFs is both costly and time consuming, often influencing design engineers to shun the use of 3-D audio displays completely, or towards adopting one set of HRTFs as a generalised filter for all operators. Non-individualised HRTFs offer less identifiable localising cues than individualised transfer functions, resulting in degraded localising accuracy and increased front-back errors [4-6].

Address correspondence to this author at the Boeing Defence Australia, 363 Adelaide St, Brisbane 4001, Australia; Tel: +61 7 3306 3527; Fax: +61 7 3306 3123; E-mail: john.towers@boeing.com

The absence of integrated head tracking within 3-D audio displays has been well documented as providing poor externalisation of sound and degraded localising performance [7, 8]. Presenting non-individualised 3-D audio without head tracking constrains its use predominantly about the azimuth and to cases where the individual does not move his or her head. Since cost and environmental constraints may make it impractical to always integrate head tracking in a display, this study has attempted to further knowledge of the use of 3-D audio without head tracking.

The objective of this study was to determine whether the sound localising performance for a non-individualised HRTF display without head tracking could be improved through the introduction of dynamic supplementary reference sounds. Supplementary sounds were expected to provide the listener with time and position based spatial reference toward the midsagittal plane and interaural axis, where interaural cues are at their most discernable [9]. Wightman & Kistler [7] suggest that dynamic cues only reduce font-back errors when under the direct control of the listener. By actively varying head or sound source movement, the listener may determine a known direction relative to changes in ITD. Because the relative direction to produce an increase or decrease in ITD differs between hemifields, such dynamic cues can be used to provide an unambiguous indicator for determining a sound's correct hemifield. The supplementary sounds introduced within this study are expected to provide improved relative spatial reference to the listener, thereby optimising localising performance without requiring listener control over the cues. Since participants established prior knowledge regarding the characteristics of the supplementary sounds through training, the design is expected to provide unambiguous spatial reference cues that

1875-9343/12 2012 Bentham Open

aid with the localising of concurrent spatially positioned sounds.

This study also considered whether additional localising error would be introduced through the mismatch between a horizontally oriented audio display and a vertically aligned visual display that also presented a secondary task. Two different input orientations for localising estimates were employed to assess this issue.

MATERIALS AND METHODOLOGY

Apparatus

Fig. (**1**) provides a pictorial reference for spatial audio terminology and regional segmentation used throughout this study.

Spatially positioned 220 Hz square wave sounds were pre-recorded as '.wav' files using the Slab3D audio rendering system with its default non-individualised HRTF [10]. The sounds were virtually positioned at each 10-degree interval about the azimuth, resulting in 36 different stimulus positions. Distance was set at 0.5 m and an elevation of 0-degrees, with a recording duration for each sound of two seconds.

Fig. (**2**) provides a pictorial representation of the three different sound conditions used throughout the study. The S*table* condition provided a baseline and consisted simply of a stationary sound. The *Swing* condition, previously designed by Kudo *et al*. [11], oscillated across four-degrees in azimuth either side of the intended bearing. The *Sweep* condition introduced supplementary 100-msec square wave accent sounds at the midsagittal plane and the ipsilateral side of the interaural axis. The accent sounds varied subtly by 1 Hz, with a 102 Hz sound presented at 0^o; a 101 Hz sound at 90^o; and a 100 Hz sound at 180^o. In addition to the accent sounds, a 220 Hz square wave sweep sound would transit at 90 deg/s in a 0.5 m diameter arc, alternating its direction from front (0^o) to back (180^o) for the duration of each localising trial. The sweep sound initiated the momentary onset of each accent sound every time it transited through the 0, 90, and 180 degree positions. The two-second localising sound was activated in a similar manner, but only once per trial.

The experiment employed a spatial distractor task in the form of a 2-D flight simulator, shown in Fig. (**3**). The participant was required to position the earth within a square alignment region displayed in the center of the screen. A Logitech Attack 3 joystick provided first-order control over the spaceship, which enabled the participant to guide the spaceship toward the earth. The simulator activity was intended to increase workload for spatial perception to a degree that effects between sound conditions would become more observable. No performance data were collected for the activity. A Hewlett-Packard xw6200 desktop computer ran the simulator and audio presentation, with audio being delivered through Bose® TriPort binaural headphones.

Input mode for registering audio localising estimates and associated confidence was varied between groups to explore the possible effects on localising due to differing orientations between the audio display and visual interface/input device. One input mode group utilised a Wacom® Bamboo™ touch pad graphics tablet oriented in the horizontal plane, while a second group used a desktop mouse for input with the graphical user interface oriented in the vertical plane on a computer monitor, as shown to the right of Fig. (**3**). The localising interface included a circle with consecutive lines indicating 45^o increments in bearing from the midsagittal plane, with each 90^o increment labelled in degrees. Localising confidence estimates were input through a vertically aligned slider bar, which was labelled low to high and logged values from 0 to 10. The tablet and computer monitor displayed identical graphics for the input of localising and confidence estimates, while the distractor task was only displayed on the computer monitor.

A dexterity study was previously undertaken by Towers [12] to ensure that the graphics tablet would not introduce extraneous localising error through poor touch pad sensitivity. The study found a mean error of less than 1^o for participants' accuracy when touching predetermined points about the circumference of an overlaid circle.

Participants

Twenty-two Boeing Defence Australia employees, aged between 24 and 50 (M = 36) participated in the experiment on a voluntary basis. The sample comprised 19 males and three females who were randomly assigned to one of two

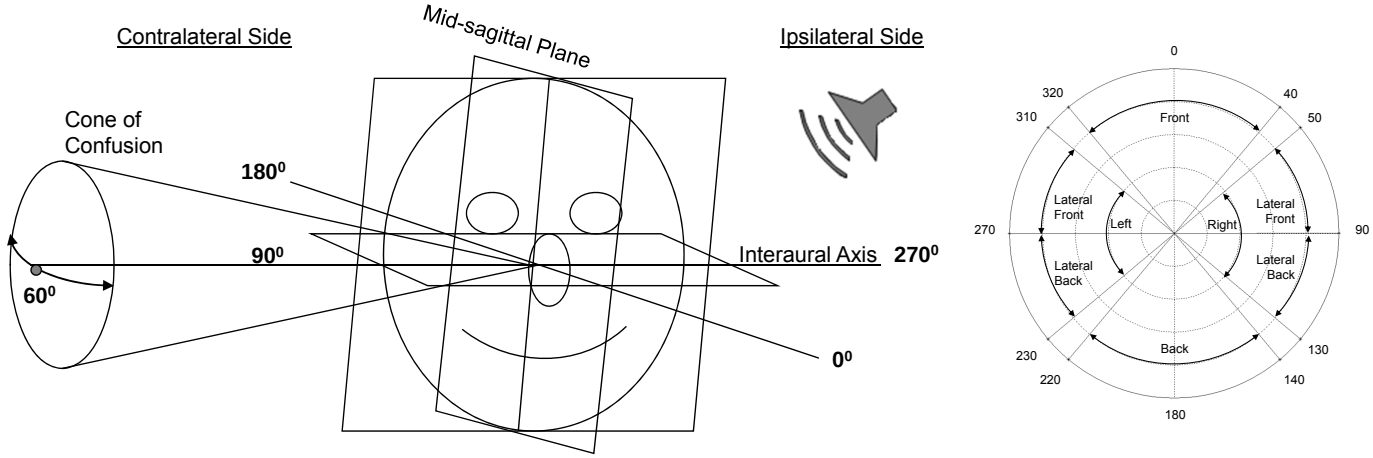

Fig. (1). Spatial Audio Terminology and Regional Dimensions.

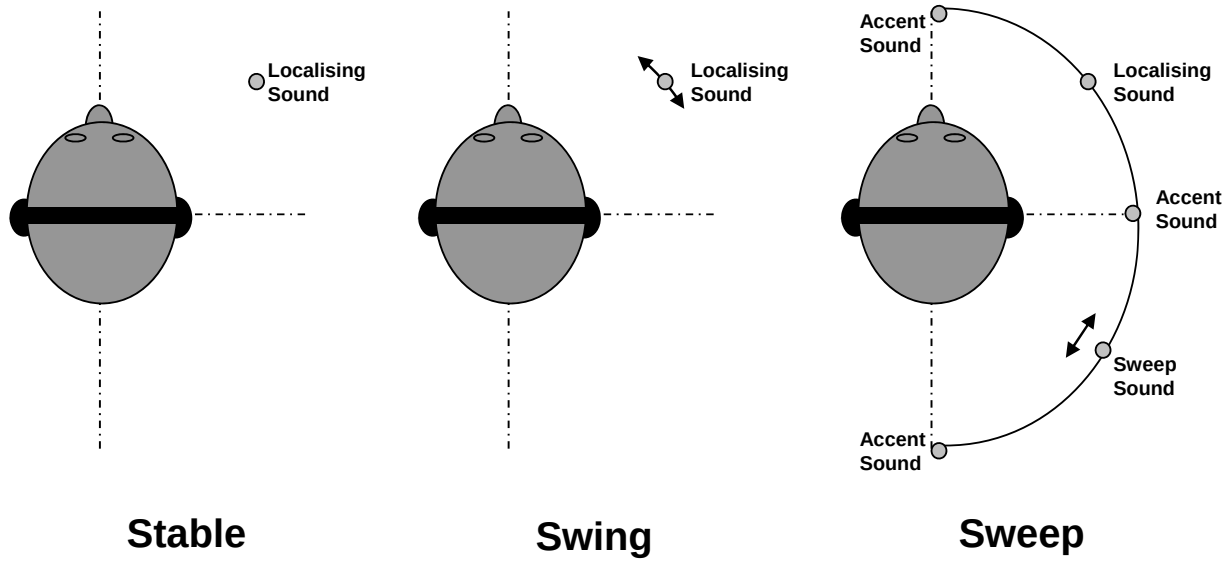

Fig. (2). Sound Conditions.

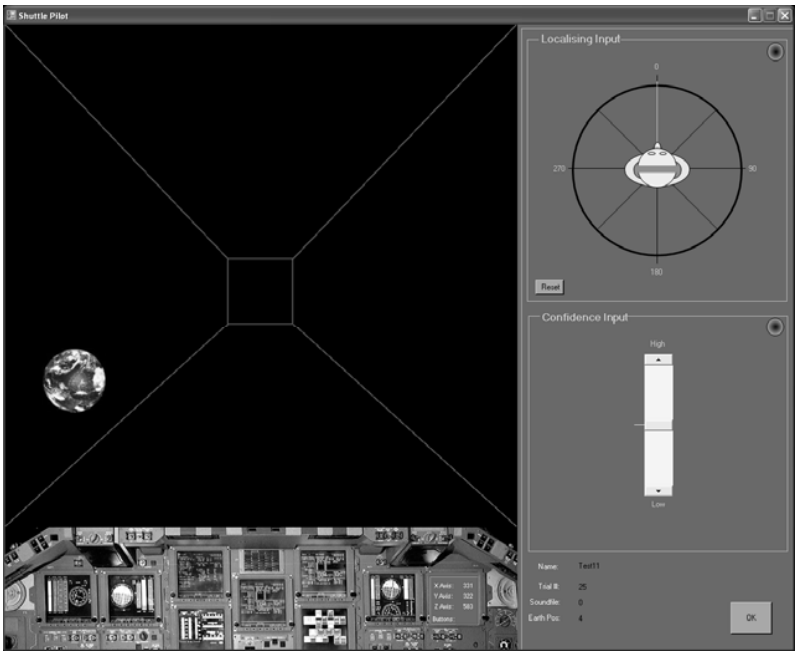

Fig. (3). Simulator Interface.

groups. Hearing screening tests were conducted on all participants prior to undertaking the experiment. Equal loudness tests were conducted on both ears of each participant across a frequency range of 30 Hz to 16 kHz and compared to a standardised curve and dBA weighted curve. An additional four volunteers were found to have abnormal hearing and subsequently did not take further part in the study. No participants had significant experience with spatial audio localising prior to the experiment. This study has been cleared in accordance with the ethical review guidelines and processes of the School of Human Movement Studies, The University of Queensland ethics committee (HMS09/0605).

Procedure

The experiment was constructed as a mixed three-way factorial design. The independent variables for spatial region and sound condition were repeated measures, while input mode was added as a grouping factor for the GUI and graphics tablet independent variables. Each input group comprised 11 participants. Dependent variables were sound localising error, front-back error, and localising confidence.

The experiment was conducted in three phases. During the first phase, participants completed a hearing screening test and spatial audio familiarisation session. An overview of audio localising and spatial audio processing fundamentals was provided to each participant, along with a detailed overview and demonstration of each sound condition used throughout the study. Participants then had approximately 20-minutes practice listening to spatial sounds and flying the simulator.

Phases two and three were data collection phases, where participants attempted to localise the point of origin about

the azimuth for each presented sound while concurrently performing the flight simulator tracking task. Participants sat at a desk with the joystick positioned on the table directly in front of them. Group 1 participants also had the graphics tablet positioned in front of the joystick. Participants were instructed to sit still with their head upright facing forward for the duration of each trial. By pressing the joystick trigger, participants initiated the earth-tracking task and subsequent presentation of the audio file two-seconds later. Upon completion of the sound presentation, the interface paused and the participant was prompted to input localising and confidence estimates in the mode determined by their group allocation.

Participants performed the two data collection phases over two nonconsecutive days within a five-day period. Each data collection phase comprised the presentation of the three sound conditions at each of the 36 target locations, resulting in 108 unique stimuli. Two repetitions were conducted for each stimuli resulting in a total of 216 trials that were initially randomised in order and presented to each of the participants in a consistent order. This set of trials was then repeated again during the following data collection phase with a different randomised order of presentation. The total number of localisation trials for the study was therefore 432. Each phase lasted approximately 50 minutes, with a short break being taken midway through each phase. Any biases due to practice effects were assumed to be averaged out due to the presentation of localisation trials that were independently randomised across sound conditions and repetitions. To balance any potential spatial perception effects that may have been introduced through the tracking task, the earth was reset to a different corner of the screen during each of the four repeated sound presentations. After completing the experiment, each participant was asked to comment on the different sound conditions and their ability to localise sounds.

ANOVA was conducted on front-back error, azimuth localising error, and confidence estimates. Subsequent post-hoc analyses were undertaken for all significant main effects using Bonferroni adjusted alpha levels. A 0.05 significance level is used throughout the analysis. Definitions for spatial regions referred to throughout this section are illustrated in Fig. (**1**).

RESULTS

Input Mode Equivalence

ANOVA conducted on localising error between input modes did not indicate the presence of any main effects $F(1, 20) = 1.10$, $p = .307$. Further analysis was undertaken to determine if the two input modes might be considered equivalent by employing a test for practical equivalence developed by Snow, Reising, Barry, & Hartsock [13]. The Tablet condition was identified as the control for the equivalency interval because the sound presentation and input response axis were correspondingly oriented about the horizontal plane. Results indicate that mean localising error in the tablet and GUI input mode conditions are considered practically equivalent ($\alpha = .05$) for all conditions with the exception of the front stable condition, which is illustrated in Fig. (**4**). It can be seen that the lower 0.90 confidence interval for the GUI input mode extends slightly beyond the shaded area, which indicates the lower equivalency interval.

Upon determining that results for the input factor were practically equivalent, all statistical analysis was once again conducted with the grouping factor removed. No significant differences were observed other than those reported with the input grouping factor in place. Therefore, only those results obtained with the input grouping factor in place have been included.

Fig. (4). Input Equivalency for Front Stable Condition. Error bars denote 0.90 confidence intervals.

Front-Back Errors

Front-back errors, often referred to as reversals, occur when a sound is localised toward the incorrect hemifield with respect to the interaural axis. These errors often occur due to poorly generalised spectral cues not providing adequate discriminating features between points where IID and ITD are similar, especially about the midsagittal plane and cones of confusion. Participants reported experiencing a varying degree of *internalisation* [2]. This refers to the presented sound being perceived to originate within the head, rather than at a point some distance from the listener. Internalisation makes it more difficult to distinguish the hemifield of a sound's point of origin. Participants did report consistency in regard to the regional effects of internalisation, with the greatest effect occurring in the front region, followed by the lateral, and then the back region.

Fig. (**5**) shows total front-back errors occurring across each 10^{o} in azimuth as grouped by sound condition. All participants' data for both input modes are plotted, resulting in a maximum of 88 possible front-back errors at each of the target locations.

Within this study, it was found that the stable condition produced the most reversals, followed by the swing, and then finally the sweep condition, which produced the least number of errors consistently across all regions. When compared to the baseline stable condition, the sweep condition reduced front-back errors by 35% in the front; 48% and 19% respectively in the lateral-front and lateral-back; and 14% in the back.

ANOVA was conducted on front-back error data across the front, back, left, and right regions. Fig. (**6**) shows the

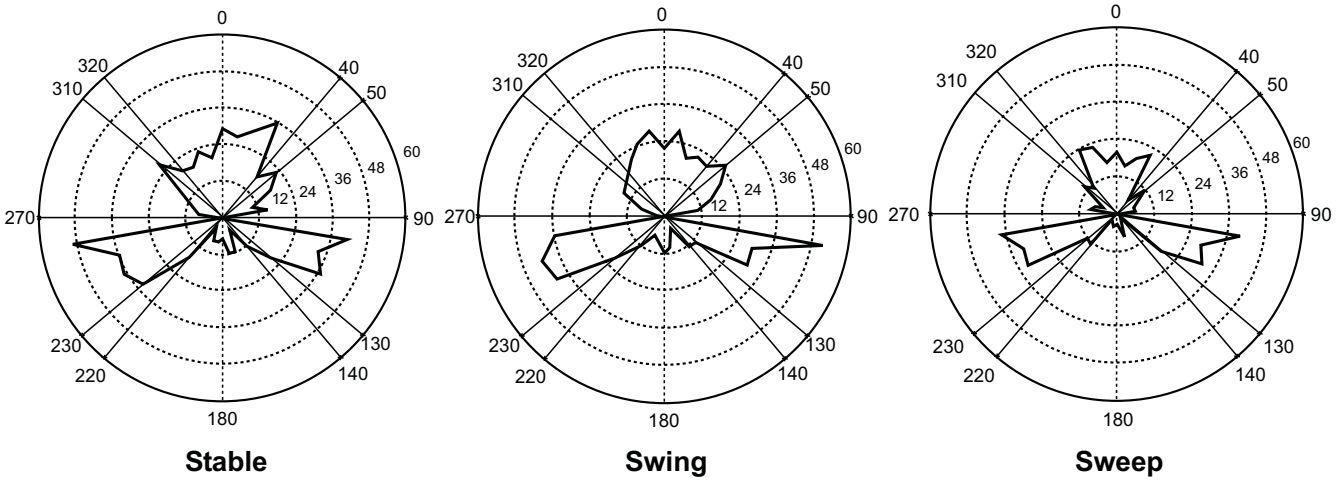

Fig. (5). Total Front-Back Localising Errors Grouped by Sound Condition.

mean front-back errors for each sound condition grouped by region.

Results indicate a significant main effect for sound, F(2, 40) = 8.64, p = .001 and region, $F(3, 60) = 6.68$, $p = .001$. No interactions were observed for sound by region $F(6, 20)=1.87$, $p = .091$, region by input $F(3, 60) = .728$, $p = .538$, sound by input $F(2, 40) = 1.72$, $p = .191$, or region by sound by input $F(6, 120) = .887$, $p = .507$. Post-hoc analysis indicated that the sweep condition produced significantly less front-back errors than the stable ($p = .001$) or swing ($p = .033$) conditions. There were no simple effects observed between the stable and swing conditions. Front-back localising errors occurred significantly less in the back region than any other region (front: $p = .001$, left: $p=.007$, right: $p=.007$).

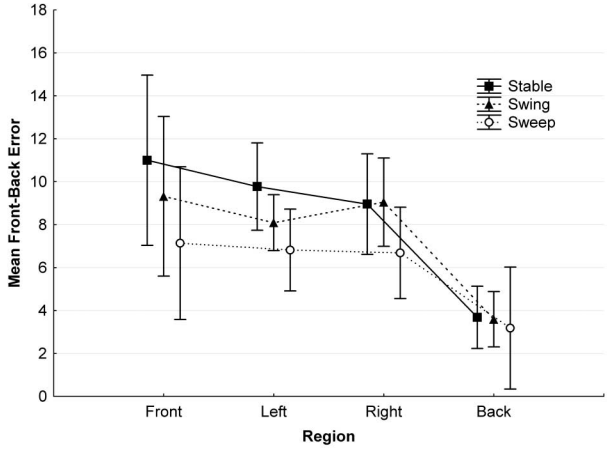

Fig. (6). Mean Front-back Errors for Sound Grouped by Region. Error bars denote 0.95 confidence intervals.

To further explore the front–back error effects within the left and right regions, these data were grouped into lateral-front and lateral-back regions. Fig. (**7**) shows the mean front-back errors for each sound condition within those regions.

ANOVA was conducted on the lateral regions and indicated the presence of a significant main effect for region $F(1, 20) = 14.66$, $p = .001$, and sound $F(2, 40) = 8.24$, $p = .001$. Post-hoc analysis for regional effects indicate that the lateral front contained significantly fewer reversal errors than the lateral back ($p=.001$), while the effect for sound indicated that the sweep condition produced fewer reversals than both the stable ($p = .001$) and swing ($p = .027$) conditions.

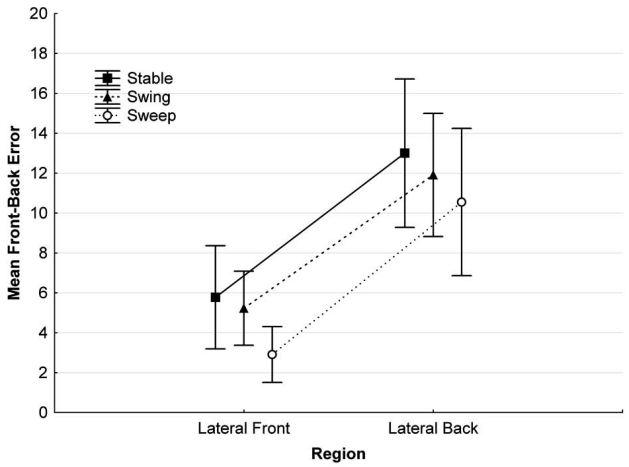

Fig. (7). Mean Front-back Errors for Sound Grouped by Lateral Region. Error bars denote 0.95 confidence intervals.

Azimuth Localising Accuracy

Front-back errors were corrected prior to undertaking localising error analysis. The rationale for correcting front-back errors is that they are generally caused by inadequate spectral cues, which are not considered to be key discriminators when localising azimuth bearing. Front-back corrections were made by subtracting incorrectly localised estimates from 180^o (Oldfield & Parker, 1984a; Wenzel *et al*., 1993; Wightman & Kistler, 1989).

Fig. (**8**) shows the mean localising error for sound condition as grouped by region. No localising main effects were observed for sound or region, however, there was a significant interaction observed for sound by region, $F(6, 120) = 2.50$, $p = .026$. Subsequent Bonferroni post-hoc testing did not detect any significant differences between specific conditions.

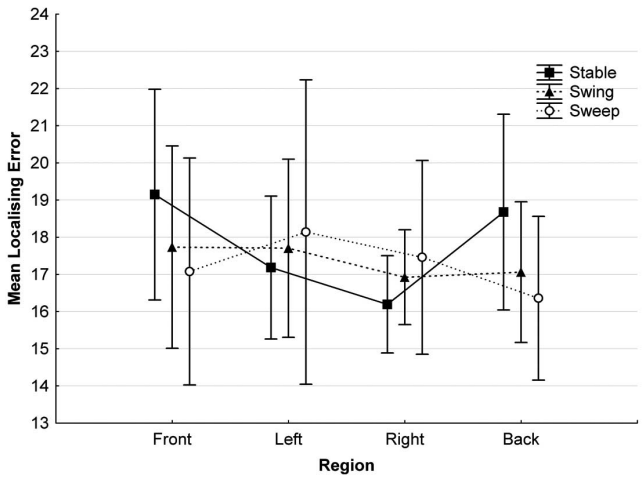

Fig. (8). Mean Localising Error for Sound Grouped by Region. Error bars denote 0.95 confidence intervals.

Fig. (**9**) shows the spread in localising estimates for azimuth bearings as grouped by sound condition. Each circular plot comprises a sample of only 12 of the 36 possible target locations in order to minimise clutter. Changes in the distance from the center for each plot were adopted simply to minimize overlap between adjacent data. The summary data for each target location are plotted in a configuration representative of a curved box-and-whisker plot, with a solid line spaning the quartile range; a hollow circle indicates the mean; and the outer points indicate the 10 and 90 percentile range. A common trend with localising data across all conditions and regions can be observed whereby the estimates tend to consistently bias toward the interaural axis on the ipsilateral side of the midsagittal plane.

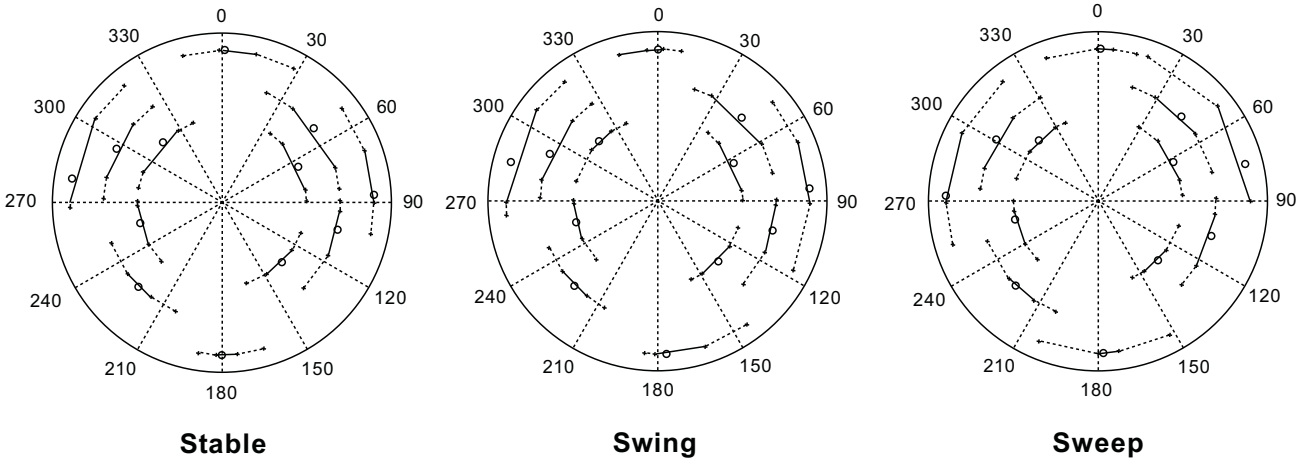

Fig. (9). Spread in Localising Estimates Grouped by Sound Condition.

Confidence Estimates

ANOVA conducted on confidence estimates found a significant main effect for region, $F(3, 60) = 4.39$, $p = .007$, and a significant interaction between region and sound, $F(6, 120) = 4.46$, $p = .001$. There were no main effects observed for confidence estimates relating to input factors. Nine of the 22 participants remarked that the sweep condition was occasionally distracting and disorienting.

Post-hoc testing for the region effect indicated that participants were significantly less confident when localising in the front region compared to both the left ($p = .012$) and right ($p = .026$) regions. Confidence was higher for the sweep condition in the back region than the front ($p = .001$). Post-hoc analysis for sound by region found no effects within the left or right regions. No significant differences between the stable and swing conditions were found within each region. Participants were found to be more confident when localising the sweep condition in the front region than both the stable ($p = .003$) and swing ($p = .002$) conditions. Similarly, participants also felt more confident estimating the sweep condition than both the stable ($p = .001$) or swing ($p = .001$) conditions within the back region. Fig. (**10**) shows the mean confidence estimates for each condition grouped by region.

Similar lateral groupings were adopted for the confidence estimates as were undertaken for front-back errors. ANOVA indicated no significant differences for confidence estimates between those lateral regions $F(2, 40) = 0.13$, $p = .874$.

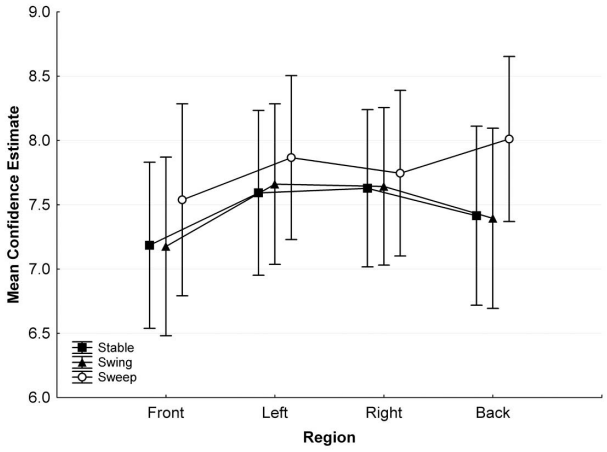

Fig. (10). Mean Confidence Estimates for Sound Grouped by Region. Error bars denote 0.95 confidence intervals.

A Pearson correlation coefficient was computed to assess the relationship between participants' confidence and localising performance for each sound condition and region, along with

confidence as it related to front-back errors within each region. There were no strong correlations found between confidence estimates and localising performance or front-back errors throughout each region.

DISCUSSION

The results indicate that supplementary spatial audio cues in the form of stationary and transient sounds do have potential to aid localising performance by reducing front-back errors throughout all azimuth regions. However, the cues provided in this experiment did not significantly improve localising accuracy and were reported to distract some listeners. No bias in localising error was found to be introduced through the mismatch in orientation between the horizontally presented audio display and vertically oriented visual input display.

Contrasting Display Orientation

Visual interfaces that convey bearing information, such as a Horizontal Situation Indicator (HSI) within an aircraft cockpit, often have a vertical orientation. Within a non-individualised HRTF audio display, however, a horizontal orientation offers the most accurate localising performance due to the utilisation of reliable ITD and IID cues, rather than the more fallible spectral cues relied upon when perceiving elevation. Using a horizontally oriented audio display to supplement information presented through a vertically oriented visual display has the potential to degrade localising perception by introducing extrapolation errors when the listener transitions attention between the dissimilar polar axis of the two displays. The possibility that extraneous errors in localising might be introduced through contrasting display orientation was considered by comparing the difference in results between groups utilising vertical or horizontally oriented input modalities. Although a slight nonequivalence was found in the front stable condition, the conservative equivalency range of $\pm 5^{o}$ tends to support a general conclusion that both vertical and horizontal modes of input are practically equivalent.

These findings support the cautious use of disparate display orientations within future research. This study does not provide any quantitative insight into the degree of workload imposed on an operator when extrapolating information presented through disparately oriented displays, which should possibly be considered in future studies.

Front-Back Errors and Localising Accuracy

Oldfield & Parker [14] occluded the pinnae of participants' own ears during free field localising trials of sounds presented over loudspeakers. They found that in the absence of spectral cues, reversals occur across all azimuth bearings, with the worst being 55% of signals at 0-degrees azimuth. Wenzel *et al.* [5] similarly found that 31% of signals initiated reversal errors, 25% of which occurred from the front, and 6% from the rear. Spectral cues are the most difficult cues to perceive within non-individualised HRTFs, so it is not surprising that we see a similar distribution of front-back errors occurring within the current study.

Participants who found the sweep condition to be distracting and disorienting reportedly adopted similar strategies when listening to the sound. They would initially listen to the sweep sound to determine the front-back origin of the localising sound, and then focus solely on the localising sound at the expense of gaining any relative positional cues that the transient sweep sound may continue to offer. They appeared to lack an ability to concurrently derive cues from the sweep sound while attending to the localising sound. This finding to some extent diminishes the effectiveness of supplementary cues in their current form as they were intended to aid localising for a novice listener.

The swing sound was previously developed by Kudo *et al.* [11] to optimise spatial audio localisation for non-individualised HRTF systems without head tracking ability. When localising the ambiguous position of a sound in free space, turning the head improves localising performance by providing localising cues that change over time. For this reason, it was suggested that establishing a dynamic localisation cue is important when using non-individualised HRTFs. Previous experiments have been supportive of this claim, with head movements being shown to reduce front-back errors by varying localising cues [7, 8]. The sweep condition utilised within this study was intended to build upon this theory and further reduce front-back errors by establishing a robust perceptual framework utilising relative spatial position and time based cues. The significant reduction in front-back errors provided by the sweep condition supports the use of supplementary reference cues to aid localising performance and mitigate issues of localising perception associated with the use of non-individualised HRTFs. Further research is required to explore the workload cost incurred while attending to supplementary cues. In their current form, the design may be too disparate and cluttered to effectively integrate with a more detailed audio display.

Internalisation and Localising Accuracy

A well documented phenomenon associated with non-individualised HRTFs is termed internalisation [2]. This occurs through poor spectral cue generalisation, whereby the listener fails to perceive adequate spatial distance for a sound. Internalisation causes sound to appear to reside more internally along the interaural axis. Participants in this study reported experiencing a varying degree of internalisation, which was markedly more prevalent in the front region, as illustrated by diagram 'a' within Fig. (**11**).

Muller & Bovet [15] found that head movements provide a 10% increase in azimuth localising ability. A significant improvement in localising performance was not observed within this study, which was hoped to be achieved through the supplementary cues contained within the sweep condition. As stated previously, within the sweep condition this lack of improvement may be due to participants ignoring the supplementary spatial cues once the front-back region was identified. In the case of the swing condition, perhaps localising performance was not improved given the absence of prior learning [16].

The observed localising bias, which appeared to skew the estimates toward the interaural axis, is possibly accounted for by the use of an upper percentile interaural time difference (ITD). The Slab3D default HRTF is a particular person's HRTF measurements converted to minimum phase HRTFs. The minimum and maximum ITDs are -784 (left 90 degrees, 0

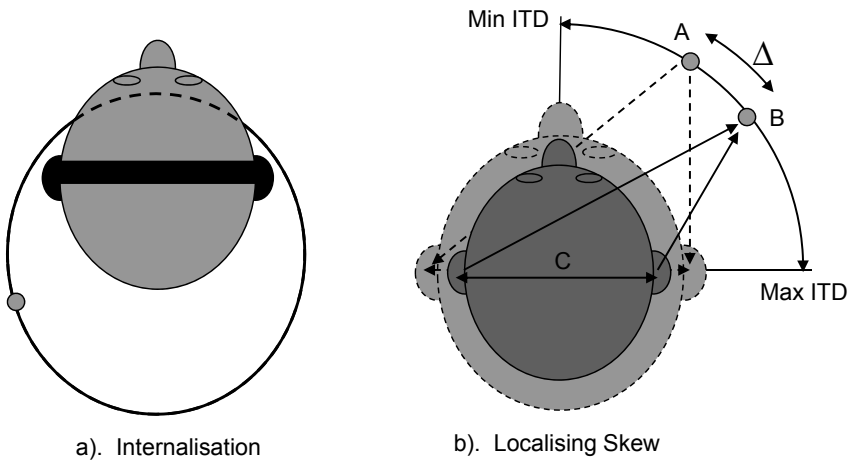

Fig. (11). Anomalies in Azimuth Localising Perception. Δ = skewed localising bias; A = Modelled spatial position of sound; B = Listener's skewed localisation of sound; C = Listener's smaller interaural dimension.

degrees elevation) and +945 microseconds (right 90 degrees, 0 degrees elevation), respectively, computed using a cross-correlation method. This compares to a maximum ITD of about 690 microseconds as shown in Feddersen, *et al.* [17] and as calculated for a spherical head model using a radius of 8.75 cm [18]. The diagram labelled 'Localising Skew' in Fig. (**11**) above, shows a sound synthesized at position 'A'. Position 'B' shows the bias delta (Δ) in azimuth position that a spatially synthesized sound presented at the same virtual distance would be perceived at if the ITD was mismatched by a larger modelling of the interaural distance at 'C'. In contrast, a smaller mismatch in timing onset differences between the HRTF and the listener's expected cues would be expected to skew the perceived position left toward the midsagittal plane. Oldfield & Parker [9] observed the same localising bias occurring under normal hearing conditions, which has possibly been exaggerated within this study due to the use of a non-individual HRTF.

These same trends were identified in a previous study by Towers [12] utilising the same HRTF and were expected to be reduced within this study through the introduction of the sweep condition. If all participants had implicitly adopted the same previously mentioned strategy in dealing with the sweep sound, it could be expected that localising estimates would be based solely on the cues provided within the stable sound, thereby reintroducing the bias in localising that is prevalent within the stable condition.

Confidence

Findings that sounds presented within the front region produced the least confidence in localising estimates could be attributed to the reportedly elevated internalisation observed within that region. The sweep condition facilitated significantly more confidence in localising performance than did other conditions within the front and back regions. Increases in confidence within those areas may be solely attributed to the introduction of significantly better front-back localising performance provided by the sweep condition. Front-back confusion was reportedly a conscious dilemma when localising. The sweep provided the listener with enough additional cues to form a confident front-back determination, which may have been the sole factor for elevated confidence. Further research needs to be conducted into operator trust pertaining to issues surrounding the use of non-individualised HRTFs and audio localising in general. The finding that confidence did not correlate with front-back errors or localising performance suggests that a less than optimal perceptual framework has been established by the supplementary cues.

Application

This research may help facilitate the development of a diverse range of 3-D audio applications, particularly in support of dual-task situations that require head-up monitoring of information. Wickens' Multiple Resource Theory [19] suggests that by diversifying the modality of presentation, independent sensory and cognitive processing channels may be more effectively employed, thereby accommodating otherwise potentially excessive workload demands. Several studies claim that multisensory displays improve dual-task performance and increase sensory perception [20-22], while often facilitating more immersive situation awareness [23, 24]. Multisensory displays may benefit form this research as it attempts to enable the cost-effective use of spatial sonification, which may be useful when presenting information relating to psychomotor activity and the monitoring of discrete variables such as distance or error.

Due to the absence of head tracking, this research is limited to audio displays that do not require spatial alignment with the environment, which would often be the case for navigational displays. Perhaps in some instances alignment with the environment may even prove superfluous and possibly degrade spatial perception. For example, applications where error is represented by the displacement of a sound about the azimuth may benefit with a constant alignment toward the midsagittal plane rather than a forward facing point in the environment, particularly if head movements are occurring regularly between different regions. For such displays, maintaining an alignment toward the external environment may increase operator workload for sensory perception given the additional requirement to consider head orientation relative to the spatial alignment of the display.

The current research may find application within interfaces that control the remote operation of robotic and unmanned platforms. Operators of such systems are often deprived of sensory cues that convey information regarding the state of

elements within the system, which are normally obtained through direct interaction with the operational environment [24]. The effective operation of systems such as unmanned aerial systems, bomb disposal vehicles, and surgical robots, often require the constant monitoring of operational variables. These types of systems may benefit from the use of 3-D audio displays that present sounds that have been scaled in their spatial position about the azimuth to represent variables of interest. The position and movement of 3-D sounds could be used to monitor the value, speed, and direction of remote variables, such as distance, torque, and bearing, therefore allowing the operator to allocate more attention to other visual surveillance or mission planning activities.

Developing cost-effective solutions that overcome requirements for individualised HRTFs and head tracking is considered an important enabler for broadening the use of 3-D audio displays within industry. Establishing robust design paradigms within disparate applications may help optimise operator performance and encourage the use of 3-D audio within future systems.

CONCLUSION

This study introduced supplementary spatial audio cues for 3-D sound delivered through a non-individual HRTF in an attempt to provide the listener with a more robust framework for spatial perception. It was hoped that additional cues would mitigate performance errors such as front-back confusions and degraded localising accuracy, which are commonly associated with generalised HRTF filters. The study found that front-back errors were significantly reduced through the introduction of static and transient supplementary sounds in the sweep condition. Localising accuracy about the azimuth was not significantly improved and the additional cues tended to occasionally disorient and distract some participants. Confidence was elevated for signals containing the supplementary cues, possibly due to improvement with front-back perception. However, the lack of correlation between confidence and localising performance suggests that an appropriate allocation of trust has not been effectively established. Gaining a deeper understanding of the associated workload demands imposed through supplementary cues and how to establish effective trust were identified as important findings that require further attention.

ACKNOWLEDGEMENT

Declared none.

CONFLICT OF INTEREST

Declared none.

REFERENCES

[1] Middlebrooks JC, Green DM. Sound localization by human listeners. Ann Rev Psychol 1991; 42(1): 135.

[2] Yost WA. Fundamentals of Hearing: An Introduction. 5th ed. Burlington, MA: Academic Press 2007.

[3] Wightman FL, Kistler DJ. Headphone simulation of free-field listening. I: Stimulus synthesis. J Acoust Soc Am 1989; 85(2): 858-67.

[4] Wightman FL, Kistler DJ. Headphone simulation of free-field listening. II: Psychophysical validation. J Acoust Soc Am 1989; 85(2): 868-78.

[5] Wenzel EM, Arruda M, Kistler DJ, Wightman FL. Localization using nonindividualized head-related transfer functions. J Acoust Soc Am 1993; 94(1): 111-23.

[6] Moller H, Sorensen MF, Jensen CB, Hammershoi D. Binaural technique: Do we need individual recordings? J Audio Eng Soc 1996; 44(6): 451-69.

[7] Wightman FL, Kistler DJ. Resolution of front-back ambiguity in spatial hearing by listener and source movement. J Acoust Soc Am 1999; 105(5): 28-41.

[8] Iwaya Y, Suzuki Y, Kimura D. Effects of head movement on front-back error in sound localization. Acoust Sci Technol 2003; 24(5): 322-4.

[9] Oldfield SR, Parker SP. Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions. Perception 1984; 13(5): 581-600.

[10] Miller D, Wenzel EM. Recent developments in SLAB: a software-based system for interactive spatial sound synthesis: proceedings of the international conference on auditory display. Kyoto, Japan 2002.

[11] Kudo A, Higuchi H, Hokari H, Shimada S. Improved method for accurate sound localization. Acoust Sci Technol 2006; 27(3): 134-46.

[12] Towers JA. Localising Synthesised Spatial Audio Filtered through a Generalised HRTF: Proceedings of the Human Factors and Ergonomics Society of Australia 44th Annual Conference: Adelaide, Australia. November 17-19, 2008.

[13] Snow MP, Reising JM, Barry TP, Hartsock DC. Comparing new designs with baselines. Ergon Des 1999; 7(4): 28-33(6).

[14] Oldfield SR, Parker SP. Acuity of Sound Localisation: A Topography of Auditory Space. II. Pinna Cues Absent. Perception 1984; 13(5): 601-17.

[15] Muller BS, Bovet P. Role of pinnae and head movements in localizing pure tones. Swiss J Psychol 1999; 58(3): 170-9.

[16] Honda A, Shibata H, Gyoba J, Saitou K, Iwaya Y, Suzuki Y. Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. Appl Acoustics 2007; 68(8): 885-96.

[17] Feddersen WE, Sandel TT, Teas DC, Jeffress LA. Localization of high-frequency tones. J Acoust Soc Am 1957; 29: 988-91.

[18] Woodworth RS. Experimental psychology. Holt: New York 1938.

[19] Wickens CD. Multiple Resources and Performance Prediction. Theor Issues Ergon Sci 2002; 3(2): 159-77.

[20] Veltman JA, Oving AB, Bronkhorst AW. 3-D Audio in the Fighter Cockpit Improves Task Performance. Int J Aviation Psychol 2004; 14(3): 239-56.

[21] Veltman JA, Oving AB, Bronkhorst AW. Effectiveness of 3-D Audio for Warnings in the Cockpit. Int J Aviation Psychol 2004; 14(3): 257-76.

[22] Santangelo V, Spence C. Multisensory cues capture spatial attention regardless of perceptual load. J Exp Psychol-Hum Percept Perform 2007; 33(6): 1311-21.

[23] Hopcroft R, Burchat E, Vince J. Unmanned aerial vehicles for maritime patrol: human factors issues. Melbourne; 2006. Defence Science and Technology Organisation, Contract No.: DSTO-GD-0463.

[24] McCarley JS, Wickens CD. Human Factors Implications of UAVs in the National Airspace. Savoy; 2005. University of Illinois at Urbana-Champaign, Contract No.: AHFD-05-05/FAA-05-01.

Received: January 18, 2012 Revised: February 6, 2012 Accepted: March 6, 2012