



THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

**Enabling the Effective Application of Spatial Auditory Displays in
Modern Flight Decks**

John Arthur Towers
Bachelor of Science (Psychology)

*A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2016
Human Movement and Nutrition Sciences*

Abstract

Modern aircraft are fitted with sophisticated technologies that support or fully automate tasks that were once performed solely by the pilot. This means that pilots now spend much of their time monitoring instruments and managing the automation rather than manually manipulating flight controls. While modern flight decks are extremely safe, pilots do occasionally experience high visual workload conditions that may degrade their ability to effectively monitor flight instruments. This thesis describes the design and evaluation of spatial auditory displays that are intended to improve a pilot's ability to monitor flight deck instruments under conditions of high visual workload. The aims broadly focus on developing design features that enable a pilot to perform head-up monitoring of an aircraft's navigation readouts while concurrently attending to verbal dialogue.

Four studies were undertaken to develop and evaluate an auditory display comprising spatially positioned sonifications that were encoded with information from multiple interrelated aircraft navigation displays. The auditory display also supported the spatial positioning of concurrent verbal communications that delivered navigation instructions. The studies were designed with four broad aims: (1) determine the sound localising performance for listeners using SLAB3D and its non-individual HRTF compared with other free field listening studies; (2) understand how supplementary auditory cues might improve localising accuracy and mitigate front-back hemisphere localising confusions; (3) develop an aircraft flight navigation auditory display that supplements existing visual readouts in order to facilitate increased head-up time and improved navigation accuracy; (4) determine the most accommodating spatial position for verbal navigation instructions that compete with concurrent sonifications for right cerebral hemisphere processing resources.

The results support the use of concurrent spatial sonifications to convey interrelated aircraft navigation information normally attended to through visual displays. Building on established design guidelines, the experiments provide additional knowledge regarding techniques that enhance localising performance, such as through the use of supplementary sound localising cues. The auditory navigation display enabled participants to fly the aircraft more accurately and devote more head-up time to an out of flight deck visual search task. Verbal navigation instructions were found to be most effectively delivered to the left ear, or along the midsagittal plane, rather than the forward left, forward right, or right position. These findings demonstrate a significant left ear advantage in the processing of verbal navigation instructions while in conditions of competing attention with sonified spatial navigation data.

The results outlined in this thesis support the use of spatial auditory displays within flight decks to improve a pilot's situation awareness for the changing state of systems information. While this thesis employs experimental designs involving manual flying as a means to extract performance measures, variants of the resultant design features are expected to readily leverage into more automated flight modes. Spatial auditory displays are not expected to replace visual instruments, but will likely improve the pilot's awareness of priority information when conditions periodically constrain the pilot's ability to attend to visual readouts. A need exists for further research to be undertaken into the applied use of such displays within different flight modes. Knowledge gained through this thesis is also expected to encourage further use of spatial auditory displays within a variety of other high workload command and control environments, such as air traffic management and nuclear/hydro power plants, where operators are similarly required to monitor complex real-time system information under constrained visual conditions.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Towers, J. A. (2008). *Localising synthesised spatial audio filtered through a generalised HRTF*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society of Australia 44th Annual Conference;, Adelaide, Australia.

Towers, J., Burgess-Limerick, R., & Riek, S. (2012). Improving 3-D audio localisation through the provision of supplementary spatial audio cues. *The Ergonomics Open Journal*, 5(1-9).

Towers, J., Burgess-Limerick, R., & Riek, S. (2014). Concurrent 3-D sonifications enable the head-up monitoring of two interrelated aircraft navigation instruments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(8), 1414-1427. doi: 10.1177/0018720814536443

Publications included in this thesis

Publication 1:

Towers, J. A. (2008). *Localising synthesised spatial audio filtered through a generalised HRTF*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society of Australia 44th Annual Conference;, Adelaide, Australia.

Incorporated as Chapter 2, Study 1: Localising Synthesised Spatial Audio Filtered through a Generalised HRTF

Contributor	Statement of contribution
Towers, J. A. (Candidate)	Designed experiments (85) Wrote the paper (100%)
Burgess-Limerick, R. (Supervisor)	Designed experiments (5%) Reviewed paper (50%)
Riek, S. (Principal Supervisor)	Designed experiments (10%) Reviewed paper (50%)

Publication 2:

Towers, J., Burgess-Limerick, R., & Riek, S. (2012). Improving 3-D audio localisation through the provision of supplementary spatial audio cues. *The Ergonomics Open Journal*, 5(1-9).

Incorporated as Chapter 3, Study 2: Improving 3-D Audio Localisation through the Provision of Supplementary Spatial Audio Cues.

Contributor	Statement of contribution
Towers, J. A. (Candidate)	Designed experiments (90%) Wrote the paper (100%)
Burgess-Limerick, R. (Supervisor)	Designed experiments (5%) Reviewed paper (50%)
Riek, S. (Principal Supervisor)	Designed experiments (5%) Reviewed paper (50%)

Publication 3:

Towers, J., Burgess-Limerick, R., & Riek, S. (2014). Concurrent 3-D sonifications enable the head-up monitoring of two interrelated aircraft navigation instruments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(8), 1414-1427. doi: 10.1177/0018720814536443

Incorporated as Chapter 4, Study 3: Concurrent 3-D Sonifications Enable the Head-up Monitoring of Two Interrelated Aircraft Navigation Instruments

This publication won the Alan Welford Award for the best human factors paper published in a peer review journal for the year 2014 as judged by the Human Factors & Ergonomics Society of Australia.

Contributor	Statement of contribution
Towers, J. A. (Candidate)	Designed experiments (90%) Wrote the paper (100%)
Burgess-Limerick, R. (Supervisor)	Designed experiments (5%) Reviewed paper (50%)
Riek, S. (Principal Supervisor)	Designed experiments (5%) Reviewed paper (50%)

Contributions by others to the thesis

William Cheung contributed to this thesis as follows:

1. Configured software to integrate a Polhemus FastTrack head tracker into the audio rendering software SLAB3D.
2. Wrote software to integrate SLABD and Microsoft Flight Simulator.
3. Provided guidance and assistance to the candidate during the development of auditory display software that was written by the candidate.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

I would firstly like to thank my supervisors, Professor Stephan Riek and Professor Robin Burgess-Limerick. Thank you both for providing me with an opportunity to realise my full potential. Your friendship, guidance, and support over the years has facilitated an effective and enjoyable learning experience that I am extremely grateful for.

To my colleague William Cheung, thank you for your assistance in writing software applications for the studies conducted throughout this thesis. I understand the pain that you must have endured while correcting errors in my code. Your patience and coaching is very much appreciated.

To my mother, Jeanette. Thank you for always standing by me. From an early age you encouraged me to chase my dreams and believe in myself, which has served me well in life.

To my girls Emily and Ruby who are a source of motivation for everything that I do. I hope that my achievements provide an inspiration for you to strive to be your best in whatever you choose to do in life. Never let anyone tell you that you can't achieve your dreams, and always find satisfaction with your best effort.

Finally and most importantly, to my beautiful wife Cordelia. Thank you for providing valuable family support for me throughout my studies. I could never have achieved this milestone without your support. I look forward to having more time to invest into our family and future.

Keywords

spatial audio, multimodal display, sonification, situation awareness, human factors

Australian and New Zealand Standard Research Classifications (ANZSRC)

<http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/6BB427AB9696C225CA2574180004463E?opendocument>

ANZSRC code: 080602, Computer-Human Integration, 50%

ANZSRC code: 170112, Sensory Processes, Perception and Performance, 50%

Fields of Research (FoR) Classification

<http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/6BB427AB9696C225CA2574180004463E?opendocument>

FoR code: 0806, Information Systems, 50%

FoR code: 1701, Psychology, 50%

Table of Contents

Abstract	ii
1. Introduction	19
1.1 Research Rationale	19
1.2 Thesis Aims and Overview	19
2. Background	23
2.1 Impact of Technology on Operator Performance	23
2.1.1 <i>Task Automation</i>	23
2.1.2 <i>Information Processing</i>	25
2.1.3 <i>Operator Workload</i>	27
2.2 Audio Localising	28
2.2.1 <i>Terminology and Spatial Reference</i>	28
2.2.2 <i>Interaural Cues</i>	29
2.2.3 <i>Spectral Cues and Head Related Transfer Functions</i>	30
2.2.4 <i>Front-back Localising Ambiguity</i>	31
2.2.5 <i>Audio Localising Through Head Movement Cues</i>	32
2.2.6 <i>Binaural Speech</i>	33
2.3 Auditory Stream Segregation	35
2.3.1 <i>Sequential Integration</i>	35
2.3.2 <i>Simultaneous Integration</i>	36
2.3.3 <i>Schemas</i>	37
2.3.4 <i>Gestalt Principles for Auditory Perception</i>	37
2.4 Spatial Audio Rendering	39
2.4.1 <i>Internalisation of Sound</i>	40
2.4.2 <i>Localising Skew</i>	41
2.5 Auditory Displays	42
2.5.1 <i>Multisensory Displays</i>	42
2.5.2 <i>Aural Alerts</i>	43
2.5.3 <i>Earcons</i>	44
2.5.4 <i>Sonification</i>	44
2.6 Handedness and Gender Effects on Lateralisation	47
3. Study 1: Localising Synthesised Spatial Audio Filtered through a Generalised HRTF	49

3.1	Abstract	50
3.2	Introduction	50
3.3	Method.....	51
3.3.1	<i>Apparatus</i>	51
3.3.2	<i>Participants</i>	52
3.3.3	<i>Procedure</i>	52
3.3.4	<i>Data Analysis</i>	53
3.4	Results.....	54
3.4.1	<i>Input Accuracy</i>	54
3.4.2	<i>Azimuth Localisation Accuracy</i>	54
3.4.3	<i>Front-back Localising Errors</i>	60
3.4.4	<i>Confidence</i>	66
3.5	Discussion	67
3.6	Conclusion.....	69
4.	Study 2: Improving 3-D Audio Localisation through the Provision of Supplementary Spatial Audio Cues	71
4.1	Abstract	72
4.2	Introduction	72
4.3	Materials and Methodology.....	74
4.3.1	<i>Apparatus</i>	74
4.3.2	<i>Participants</i>	77
4.3.3	<i>Procedure</i>	77
4.3.4	<i>Data Analysis</i>	78
4.4	Results.....	79
4.4.1	<i>Input Mode Equivalence</i>	79
4.4.2	<i>Front-back Errors</i>	80
4.4.3	<i>Azimuth Localising Accuracy</i>	84
4.4.4	<i>Confidence</i>	85
4.5	Discussion	90
4.5.1	<i>Contrasting Display Orientation</i>	90
4.5.2	<i>Front-back Errors and Localising Accuracy</i>	91
4.5.3	<i>Internalisation and Localising Accuracy</i>	92
4.5.4	<i>Confidence</i>	93
4.5.5	<i>Application</i>	93

4.6	Conclusion.....	94
5.	Study 3: Concurrent 3-D Sonifications Enable the Head-up Monitoring of Two Interrelated Aircraft Navigation Instruments	95
5.1	Abstract	96
5.2	Introduction.....	97
5.3	Method.....	100
5.3.1	<i>Participants.....</i>	<i>100</i>
5.3.2	<i>Flight Simulator</i>	<i>100</i>
5.3.3	<i>3-D Auditory Display.....</i>	<i>103</i>
5.3.4	<i>Experimental Design</i>	<i>108</i>
5.3.5	<i>Procedure.....</i>	<i>109</i>
5.3.6	<i>Data Analysis</i>	<i>110</i>
5.4	Results.....	111
5.4.1	<i>Course Deviation.....</i>	<i>112</i>
5.4.2	<i>Waypoint Angle (WP) Error</i>	<i>113</i>
5.4.3	<i>Head-up Dwell Time</i>	<i>114</i>
5.4.4	<i>Flare Detection</i>	<i>116</i>
5.5	Discussion	117
5.5.1	<i>Concurrent Monitoring of Interrelated Audio Signals</i>	<i>117</i>
5.5.2	<i>Head Tracking.....</i>	<i>118</i>
5.5.3	<i>Auditory Display Design</i>	<i>118</i>
5.5.4	<i>Future Research.....</i>	<i>119</i>
5.5.5	<i>Application.....</i>	<i>119</i>
5.6	Conclusion.....	120
5.7	Key Points	120
6.	Study 4: Spatial Positioning of Verbal Audio within a Multi-Source 3-D Auditory Display.....	121
6.1	Abstract	122
6.2	Introduction.....	123
6.3	Method.....	126
6.3.1	<i>Participants.....</i>	<i>126</i>
6.3.2	<i>Experimental Design</i>	<i>126</i>
6.3.3	<i>Working Memory Navigation Task.....</i>	<i>127</i>
6.3.4	<i>Flight Simulator</i>	<i>127</i>

6.3.5	<i>Auditory Display Design</i>	128
6.3.6	<i>Procedure</i>	131
6.3.7	<i>Data Analysis</i>	132
6.4	Results.....	132
6.5	Discussion	137
6.5.1	<i>Reliability of Participant Reporting Methodology</i>	140
6.5.2	<i>Future Research</i>	140
6.5.3	<i>Application</i>	141
6.6	Conclusion.....	142
7.	Summary and Discussion	143
7.1	Aims and Scope of the Thesis	143
7.2	Summary of Key Findings.....	144
7.2.1	<i>Key Findings from Study One</i>	144
7.2.2	<i>Key Findings from Study Two</i>	145
7.2.3	<i>Key Findings from Study Three</i>	146
7.2.4	<i>Key Findings from Study Four</i>	146
7.3	Limitations of Current Research	147
7.3.1	<i>Localising in Elevation</i>	147
7.3.2	<i>HRTF Individualisation</i>	147
7.3.3	<i>Context of Information</i>	148
7.4	Future Research and Practical Implications	148
7.4.1	<i>Misplaced Confidence in Localising with Non-Individual Cues</i>	149
7.4.2	<i>Supplementary Cues</i>	149
7.4.3	<i>Workload</i>	150
7.4.4	<i>Speech Processing</i>	151
7.4.5	<i>Utilising Spatial Auditory Displays in Pilot Upset Recovery</i>	152
7.5	The Future of Spatial Auditory Displays.....	152
8.	References	154
9.	Appendices	170
9.1	Appendix A: Participant Hearing Test	171
9.2	Appendix B: Study 1 Participant Briefing	172
9.3	Appendix B: Study 1 Tablet Overlay	173
9.4	Appendix B: Study 2 Participant Information Sheet	174
9.5	Appendix B: Study 2 Participant Informed Consent Form.....	175

9.6	Appendix B: Study 2 Participant Briefing	176
9.7	Appendix B: Study 2 Interview Protocol.....	177
9.8	Appendix C: Study 3 Participant Information Sheet	178
9.9	Appendix C: Study 3 Participant Informed Consent Form	180
9.10	Appendix C: Study 3 Sound Evaluation Sheet.....	181
9.11	Appendix C: Study 3 Participant Briefing	182
9.12	Appendix C: Study 3 Interview Protocol.....	183
9.13	Appendix D: Study 4 Participant Information Sheet	184
9.14	Appendix D: Study 4 Participant Informed Consent Form	186
9.15	Appendix D: Study 4 Subjective Evaluation of Primary Task.....	187
9.16	Appendix D: Study 4 Briefing Protocol.....	188
9.17	Appendix D: Study 4 NASA-TLX Interface.....	189
9.18	Appendix D: Study 4 Edinburgh Handedness Inventory Form.....	190
9.19	Appendix D: Study 4 Interview Protocol.....	191

List of Figures

Figure 1-1. Summary of thesis studies, artefacts, and implications.	22
Figure 2-1. Spatial terminology used throughout the thesis	28
Figure 2-2. Spatial segmentation used for grouping localising estimates in azimuth.	29
Figure 2-3. Duplex Theory of Localisation.	30
Figure 2-4. Front-back hemisphere localising ambiguity.....	31
Figure 2-5. The cone of confusion	32
Figure 2-6. Resolving front-back localising ambiguity through head movements	33
Figure 2-7. Modal similarities for the Gestalt law of similarity	38
Figure 2-8. Modal similarities for the Gestalt law of good continuity	39
Figure 2-9. The internalised perception of sound caused by poorly matched spectral cues	41
Figure 2-10. Azimuth localising error resulting from poorly matched HRTF interaural cues.	42
Figure 2-11. Sonified feedback of simulator data.....	46
Figure 3-1. Graphics tablet overlay for recording participants' localising estimates and confidence.....	52
Figure 3-2. Mean azimuth localising error grouped by azimuth region.	55
Figure 3-3. Mean azimuth localising error grouped by elevation.....	56
Figure 3-4. Mean azimuth localising error grouped by azimuth region and elevation	56
Figure 3-5. Between-study comparison of mean azimuth localisation error.....	59
Figure 3-6. Sample localisation variability plot at 20-degrees elevation.....	60
Figure 3-7. Total reversal errors plotted by azimuth bearing and grouped by 20-, 0-, and - 20-degrees elevation.....	61
Figure 3-8. Mean front-back localising errors grouped by azimuth region	62
Figure 3-9. Mean front-back localising errors grouped by elevation	63
Figure 3-10. Mean front-back localising errors grouped by azimuth region and elevation.	65
Figure 3-11. Mean confidence in localising estimates grouped by region.....	67
Figure 4-1. Configuration of sound conditions	75
Figure 4-2. Study interface.....	76
Figure 4-3. Input equivalency results for front stable condition	79
Figure 4-4. Total front-back localising errors grouped by sound condition.....	80
Figure 4-5. Mean front-back localising errors grouped by region.....	82

Figure 4-6. Mean front-back localising errors grouped by sound condition.....	83
Figure 4-7. Mean front-back localising errors for sound conditions grouped by region.....	83
Figure 4-8. Mean localising error for sound grouped by region.....	84
Figure 4-9. Variance in localising estimates plotted by sound condition	85
Figure 4-10. Mean confidence estimates grouped by region	86
Figure 4-11. Mean confidence estimates grouped by sound condition	87
Figure 4-12. Mean confidence estimates grouped by sound and region.....	88
Figure 5-1. Flight simulator.	101
Figure 5-2. G1000 glass cockpit.	101
Figure 5-3. Flight plan.	102
Figure 5-4. Primary flight display (PFD)	102
Figure 5-5. Auditory display spatial configuration	104
Figure 5-6. Waypoint audio signal component sound configuration.....	106
Figure 5-7. CDI visual interface to auditory display configuration mapping.....	107
Figure 5-8. Mean course deviation by audio condition.....	113
Figure 5-9. Mean waypoint angle by audio condition	114
Figure 5-10. Total out of window gaze dwell time by audio condition	115
Figure 5-11. Mean out of cockpit gaze dwell time by audio condition	116
Figure 5-12. Mean number of detected flares by audio condition	117
Figure 6-1. Flight simulator with GPS display	128
Figure 6-2. Auditory display navigation signal layout	129
Figure 6-3. Spatial positioning of working memory verbal instructions.	130
Figure 6-4. Mean time attending to GPS by verbal audio spatial position.....	133
Figure 6-5. Mean CDI error by verbal audio spatial position	134
Figure 6-6. Mean waypoint error by verbal audio spatial position	135
Figure 6-7. TLX scores by verbal audio spatial position.....	135
Figure 9-1. Online equal loudness hearing test.	171
Figure 9-2. The Fletcher-Munson graph of equal loudness contours.....	171
Figure 9-3. Tablet overlay for input of sound localising and confidence estimates in experiment 1.	173
Figure 9-4. Sound evaluation sheet used for sound selection task in experiment 3.	181
Figure 9-5. Verbal instruction spatial positions for experiment 4.....	187
Figure 9-6. Computer Based NASA-TLX Interface	189

List of Tables

Table 1. List of Abbreviations.....	xviii
Table 3-1. ANOVA and post-hoc results for azimuth localising error (degrees) occurring by region and elevation	57
Table 3-2. Post-hoc analysis and effect size results for azimuth localising error (degrees) occurring by region and elevation.	58
Table 3-3. ANOVA and post-hoc results for mean number of front-back errors occurring by region and elevation.	62
Table 3-4. Post-hoc analysis results for mean number of front-back errors occurring by region and elevation.	64
Table 3-5. Effect size results for mean number of front-back errors occurring by region and elevation.....	65
Table 3-6. Summary analysis results for mean azimuth localising confidence estimates.	66
Table 4-1. Summary statistics for mean front-back errors by sound condition.....	81
Table 4-2. Summary statistics for mean confidence estimates.	86
Table 4-3. Post-hoc analysis of participants' confidence estimates for sound conditions by region.	89
Table 5-1. Auditory display signals	104
Table 5-2. Summary statistics for audio conditions grouped by performance measure ..	112
Table 6-1. Mean correct working memory task responses by verbal audio spatial condition.....	133
Table 6-2. Summary statistics for verbal audio conditions grouped by performance measure	136
Table 6-3. Summary statistics for equivalence between verbal audio conditions grouped by performance measure	137

List of Abbreviations

Table 1.

List of Abbreviations

Acronym	Meaning
AFAIC	Adaptive Function Allocation for Intelligent Cockpits
ANOVA	Analysis of Variance
API	Application Programming Interface
ASA	Auditory Scene Analysis
CDI	Course Deviation Indicator
ECAM	Electronic Centralised Aircraft Monitoring
EEG	Electroencephalography
EMG	Electromyography
GPS	Global Positioning System
GUI	Graphical User Interface
HRTF	Head Related Transfer Function
HT	Head Tracking
HUD	Head-up Display
IID	Interaural Intensity Difference
ILD	Interaural Level Differences
ITD	Interaural Timing Difference
LCD	Liquid Crystal Display
LEA	Left-Ear Advantage
REA	Right-Ear Advantage
LQ	Laterality Quotient
MFD	Multi-Function Display
PFD	Primary Flight Display
REA	Right-Ear Advantage
RM-ANOVA	Repeated Measures Analysis of Variance
SLAB	Sound Lab
TAWS	Terrain Awareness Warning System
TCAS	Traffic Collision Avoidance System
VR	Virtual Reality
WP	Waypoint

1. Introduction

1.1 Research Rationale

It is not uncommon for operators of high workload systems to be exposed to an inordinate amount of information that subsequently degrades performance. This situation occurs in part because modern systems obtain and present data from a multitude of sources with such ease that the operator's ability to process the available information becomes impaired. Exposing an individual's capacity limited information processing faculties to excessive workload is known to degrade performance (Wickens, 2008). In order to mitigate excessive workload, it is therefore important that interfaces present concise information to the operator at the correct time and in an easily interpreted form.

Systems design engineers face difficulties when determining which tasks should be automated and which should remain under the direct control of the operator. The methods employed by interface design engineers to mitigate this workload problem and facilitate effective systems operation is varied and multifaceted. Diversifying the modality in which information is presented can overcome some of the sensory and cognitive processing bottlenecks associated with human limitations experienced when monitoring information, and is acknowledged as a means toward managing workload and improving situation awareness (Hopcroft, Burchat, & Vince, 2006; McCarley & Wickens, 2005).

Further research is required into the use of 3-D auditory displays to enable more effective monitoring of complex real time data, particularly within high visual workload environments. This thesis will explore the benefits of alleviating the visual sensory and cognitive processing resources from an often overburdened role when interacting with systems interfaces. The use of spatial auditory displays as a supplementary means for communicating systems state information is expected to facilitate more effective and safer operation of modern systems.

1.2 Thesis Aims and Overview

The overarching intent of this thesis, as supported by the four studies detailed within, was to develop new designs that enable the use of a spatial auditory display to communicate concurrent voice communications and sonified aircraft navigation data. Following this introduction, chapter two contains a review of the relevant literature, which informs the body of research detailed throughout this thesis. The review highlights that interfaces require design improvements to accommodate the changing role of the

operator. These changes are primarily brought about by new technology that facilitates an increased capacity for automation. Subsequent challenges regarding information processing are discussed, along with an overview of why spatial auditory displays are considered an effective enhancement to modern systems interfaces. The section also provides an overview of the current theories of audio perception and localising, which form the basis of new display design rationale that is elaborated on throughout each study.

Chapter three details an investigation into the effectiveness of the spatial audio rendering tool, SLAB3D. Utilising a default, non-individualised *head related transfer function* (HRTF), the aim was to determine the ability of SLAB3D to enable localising performance similar to previous free field and binaural listening studies (Wenzel, Arruda, Kistler, & Wightman, 1993; Wightman & Kistler, 1989b) and therefore adequately support subsequent studies.

Chapter four contains an overview of a study conducted into the use of supplementary auditory cues to enhance audio localising. The aim was to determine how front-back hemisphere localising confusions occurring in the absence of head tracking cues might be mitigated through supplementary design enhancements to the presented audio. The second aim was to determine if supplementary audio cues could improve azimuth localising performance. Effect on sound localising performance through the introduction of a reference 'sweep' sound that traversed 180-degrees back and forth in azimuth about the listener was compared with a previously studied 'swing' method, which produced a 10-degree oscillation in azimuth position (Kudo, Higuchi, Hokari, & Shimada, 2006).

Chapter five details an experiment where auditory display design knowledge obtained from the previous two studies was incorporated into the design of an aircraft flight navigation auditory display. The aim was to improve a pilot's head-up situation awareness for changes in the state of the aircraft's flight navigation readouts as the aircraft deviated from its intended flight path. This study successfully utilised concurrent sonified navigation information to improve the pilot's head-up time and navigation accuracy. The aim was to develop design solutions that enable the concurrent use of multiple, spatially discrete sonifications within the same auditory display. Supplementary localising cues in the form of broad-band instrument notes, termed *carrier sounds*, were successfully co-located with the sonified signals. Those cues reportedly improved inter-stream segregation and further reduced participants' front-back localising uncertainty. The auditory display successfully facilitated improved head-up monitoring of interrelated navigation readouts.

Chapter six describes the final experiment in which verbal communications were integrated into the auditory display described in chapter five. The verbal communications provided instructions for a primary task involving spatial reasoning that was unrelated to aircraft navigation. Both the sonifications and speech audio conveyed spatial information and were therefore in competition for right hemisphere cognitive processing resources. The intent of this experiment was to determine if the participants' ability to attend to the sonifications and verbal audio would be effected by altering the spatial position of the verbal audio. Ear effects were anticipated, where the delivery of audio toward one ear is known to support or constrain the optimal neural pathway to the cerebral hemisphere best suited to process information of a particular context. A left ear advantage was observed when compared with right ear presentation, with positions along the midsagittal plane providing improved performance over front left and front right positions.

Chapter 7 provides a summary of the key findings and implications of the current research. An outline of the limitations is provided, along with proposed considerations for future research. Figure 1-1 below shows the sequence of studies conducted throughout this thesis and how they contributed to the development of the final auditory display. Implications to the relevant fields of research are also identified.

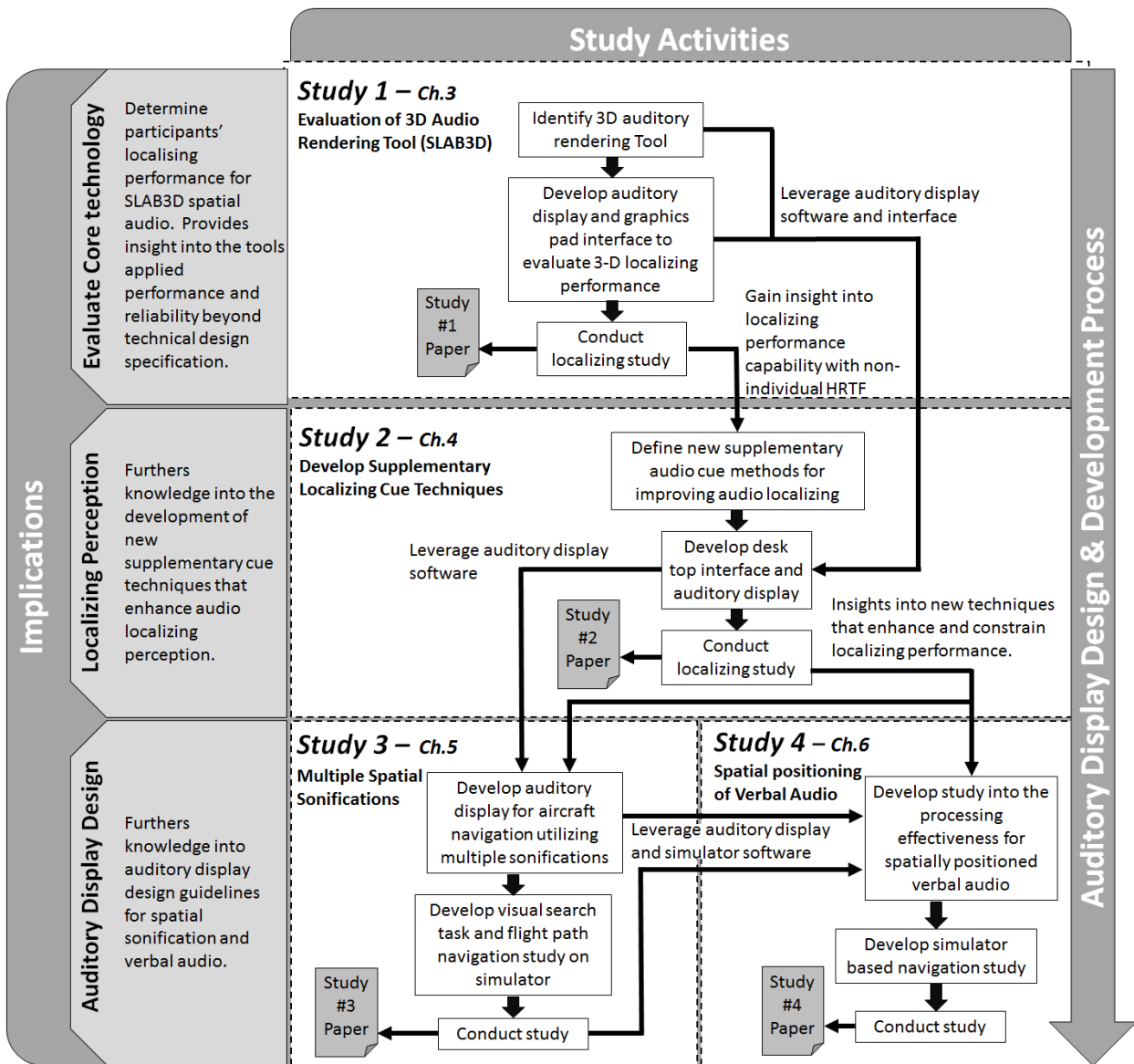


Figure 1-1. Summary of thesis studies, artefacts, and implications. Tasks conducted as part of this thesis are identified, along with their relationship to subsequent activities. The implications of study findings are summarised.

2. Background

2.1 Impact of Technology on Operator Performance

2.1.1 Task Automation

Computer processing power is evolving at an extremely rapid pace (Cochrane, 2014; Moore, 1965) toward what Kurzweil (2006) coined the *point of singularity*. He envisaged a future in which technology will enable our minds and bodies to merge with machines. Assuming this to be true, there will certainly be friction along the path toward singularity as new technology is adopted inappropriately to the detriment of operator performance. Engineers develop and integrate technology into new systems at such a fast pace that its effect on human performance is often poorly understood.

An example of such unintended consequences can be observed with the introduction of the head up display (HUD) into aircraft flight decks. The HUD is a see through display positioned directly in front of the pilots forward facing field of view. The display presents flight avionics and status information at a focal length of infinity, which allows the pilot to monitor the instrument readouts while looking out of the front window of the flight deck. The displays have been shown to improve accuracy in flight path tracking and the detection of expected state changes, however, they have also been shown to degrade performance in regard to detecting unexpected state changes (Fadden, Ververs, & Wickens, 1998). This finding supports Dekker's (2005) view that unexpected strengths and weaknesses emerge when computing capability is employed to overcome human weakness.

Integrating new technology into a systems design remains a challenge for human factors engineers, particularly in regard to determining operator functions that should be automated and those that should remain under the direct control of the human. Studies have shown that humans benefit most from automation during normal conditions, but significant issues arise when the automation or feed sensors fail and the human has to intervene (Endsley, 1999). An often cited example of automation contributing to the breakdown of human performance is the Air France 447 incident, when an Airbus A330 aircraft crashed as the result of a series of events that occurred following the automatic disconnection of some systems (BEA, 2009). Incorrect speed indications brought about this change after ice crystals obstructed the pitot tubes, which are used to measure air speed. Wickens (2013) commented that the incident was assumed to have resulted from a complacent overreliance on automation, and the operators not being involved in the

initiation of actions. These conditions affected the pilots' ability to perceive and comprehend the current state of the aircraft.

In an early attempt to tackle the issue of allocating automated functionality, Fitts (1951) wrote his seminal paper of guidelines stating what 'men are better at' and what 'machines are better at', which became known as the MABA-MABA list. Subsequent attempts to evolve this list remain controversial (Dekker & Woods, 2002), however, since its inception technology has continued to outperform humans on many of the list's attributes. The onset of task automation through the integration of computers led to the term 'human-in-the-loop', which refers to a human centred design philosophy that ensures the system supports the needs of the user at all times. As computers advance and allow more complex forms of automation, researchers began suggesting that automated systems move away from traditional 'in-the-loop' systems of operator manual control toward a more supervisory role (Dekker & Woods, 2002; McCarthy, Fallon, & Bannon, 2000; Sheridan, 2000). Within most complex work environments, particularly aviation flight decks, debate continues regarding the appropriate level and application of automation. The introduction of automation to improve the performance of routine tasks has been found to be problematic when the level of automation moves beyond supporting information analysis and into supporting action selection (Onnasch, Wickens, Li, & Manzey, 2014).

With the advent of highly automated flight decks and other complex work environments, system monitoring behaviour has become one of the most important skills the operator must attain. As expected, the role of the pilot has shifted toward the monitoring and control of automation. The United States Navy conducted early research into automation in 1980, which was called the Adaptive Function Allocation for Intelligent Cockpits (AFAIC) (Parasuraman, 1987). This program looked into the different aspects of automation and its effect on human performance and led to subsequent research into what is now termed *adaptive automation*, or *adaptive aiding*. This approach considers dynamically altering the level of automation to suit each particular task or context of operational demand (Rouse, 1988) and has been found to produce improved results and decreased workload when compared with non-adaptive automation during radar and gauge monitoring task (Kaber & Riley, 1999). Kaber et al. (2001) suggest that a major hurdle toward defining strategies for implementing adaptive aiding across a broad spectrum of systems lies in the development of effective human machine interface mechanisms.

The *Swiss cheese* model for accident causation is often referred to within aviation. Slices of cheese illustrate that there are many layers of defence between hazards and a potential accident. Any line of defence has a weakness, represented by the holes in the cheese. If those weaknesses were to line up then an accident may occur. We can consider the pilot's instrument monitoring to be a line of defence occurring at multiple stages of activity between hazards and incidents. If monitoring isn't adequate to understand the current state of the system, then the effectiveness of the system degrades significantly. The accident report from Asiana Flight 214 (NTSB, 2013), in which a Boeing 777 aircraft crashed into the sea wall at San Francisco airport in 2013 reported that "insufficient monitoring of airspeed indications during the approach resulted from expectancy, increased workload, fatigue, and automation reliance" (p. xi).

We now understand that the application of modern automated technology can degrade operator performance. Operators rely heavily on the visual monitoring of systems information in order to understand the current and future state of automation. Examples have been provided where ineffective monitoring has played a significant role in the breakdown of operator performance. We must therefore continue to explore ways in which more intuitive and accommodating interfaces convey information in an appropriate and timely manner. More research is therefore required into multimodal displays, particularly into the use of spatial audio, in order to help ease an often overly burdened visual modality when performing tasks that require the gathering of complex system state information.

2.1.2 Information Processing

Establishing prior understanding regarding the way in which humans' process information is imperative in order to design an auditory interface that accommodates cognitive processing limitations while capitalising from human capabilities that exceed current technology. The human brain can be considered to contain similar structures to that of a computer in regard to how it performs functions such as input, storage, processing, and output. The experimental study of memory goes back to the 1850s with Ebbinghaus' discovery of the forgetting curve and the spacing effect (Lieberman, 2011). Miller (1955) also wrote a modern seminal paper on the retention of information being constrained to seven, plus or minus two.

Subsequent information processing models evolved based around the three components of sensory, short- and long-term memory. Baddeley's (2002) model of working memory is one of the most prominent in the field. It highlights the existence of three components, including the *central executive*, *visuospatial sketchpad*, and

phonological loop, which are fluid systems collectively termed the *executive control system*. The executive is considered to be a limited capacity attentional controller. It remains arguably ill defined, but may account for coordination of the subsystems and their interface control to long term memory. The phonological loop is an auditory store and an articulatory rehearsal system, while the visuospatial sketch pad is involved in spatial orientation and resolving visuospatial problems. The phonological store is limited to approximately two seconds without rehearsal. It is imperative that limitations in human memory be supported when designing an auditory display. Knowledge pertaining to the way in which humans retain and recall information that is presented in a particular form must be incorporated into the way in which auditory display information is structured and presented.

Situation awareness refers to one's ability to develop an internalised understanding of the environment and has been studied since the early 1990s (Sarter & Woods, 1991). Endsley (1995) defined a theory of situation awareness to account for information processing as it relates to decision making. The model can be considered an expansion on the information acquisition stage of processing as defined by Wickens & Carswell (2006). Endsley's model consists of three levels: (1) *perception*; (2) *comprehension*; and (3) *projection*. The first level relies on perceiving the state, attributes, and dynamics of elements within the environment, such as the onset appearance of a low oil pressure indicator in the flight deck. Once perceived, the information must be adequately understood to achieve level two, comprehension. This stage requires gaining understanding of the contextual meaning behind all of the interrelated elements perceived during stage one. Attaining the third and final level, termed projection, requires an understanding regarding how the current dynamic state of elements within the environment will change into the future, both with and without human intervention. This level informs the decision making stage of information processing. Endsley's model of situation awareness provided a useful framework when establishing design requirements and associated performance measures for experimental designs within this thesis.

Wickens' (2002, 2008) multiple resource theory states that our cognitive resources are comprised of capacity limited processing channels. These channels are organised by visual (foveal and ambient) and auditory modalities. A further dimension for codes of processing exist, which consider the spatial or verbal context of the stimulus. Each channel has stages of processing that relate to perception (spatial/verbal), comprehension, and responding (manual/spatial, and vocal/verbal). These channels come into conflict to a varying degree when multitasking contextually similar tasks. Multiple resource theory was

applied during the design and evaluation of auditory displays within this thesis. Doing so helped identify and mitigate potential audio/visual information processing conflicts that may have otherwise occurred through limitations in human sensory perception and cognitive processing ability.

2.1.3 Operator Workload

Automation was originally introduced into aviation to alleviate performance problems associated with the excessive workload demand being placed on pilots. As mentioned previously, a pilot's primary role now involves monitoring and managing the aircraft's automation, rather than constantly engaging in the direct manual manipulation of the aircraft's flight control interfaces. New technology has relieved the pilot from a large amount of workload associated with manual control of the aircraft, however, workload has increased in tasks associated with visually attending to the flight deck instruments in order to understand the current state of the aircraft and automation.

The aforementioned ease in which technology may provide information to operators highlights the importance of ensuring that such information is well structured and presented in an appropriate format. Sorting and prioritising available information in any high workload system can be difficult and problematic during normal operation, however, when non-normal events occur, workload can very quickly become excessive. An example of excessive data availability can be considered with the Qantas Flight 32 incident that happened on the 8th November 2010. An Airbus A380 aircraft experienced a catastrophic failure of its number two engine as it departed from Singapore Changi airport. This incident offers a case in point regarding the difficulty in attending to available information as the electronic centralised aircraft monitoring system (ECAM) generated over 100 fault check list messages in the air, and an additional 20 on the ground (de Crespigny, 2012).

Wicken's (2008) multiple resource theory tells us that capacity limitations within cognitive channels constrain the amount of information that an individual can process. He also highlights the presence of unique characteristics within each modality that constrain how effectively information can be attended to. For instance, compared with audio, visual interfaces are very effective at communicating unambiguous information, but the unidirectional nature of vision, along with limited foveal and peripheral fields of view, limit the size and location of accessible information. Therefore, such constraints can introduce excessive workload if operators are required to gather information from multiple regions. This problem has influenced the design of modern flight deck interfaces as they moved

away from individual analogue readouts toward modern 'glass cockpit' digital interfaces. These interfaces integrate common information and position data such that they support limitations in the pilot's field of view.

Because sound is omnidirectional, auditory displays help alleviate some of the constraints and excessive workload conditions associated with the visual system, thereby allowing audio signals to be attended to independently of the heads orientation or visual focus. Arguably, very little progress has been made in terms of transitioning modern auditory display design solutions into systems interfaces to alleviate visual workload. The medical industry can be observed as an exception to this, where auditory displays are being used to enable the head-up monitoring of patient vitals during theatre operations (Sanderson, 2006).

2.2 Audio Localising

2.2.1 Terminology and Spatial Reference

Figure 2-1 provides a pictorial reference for spatial audio terminology used throughout this thesis.

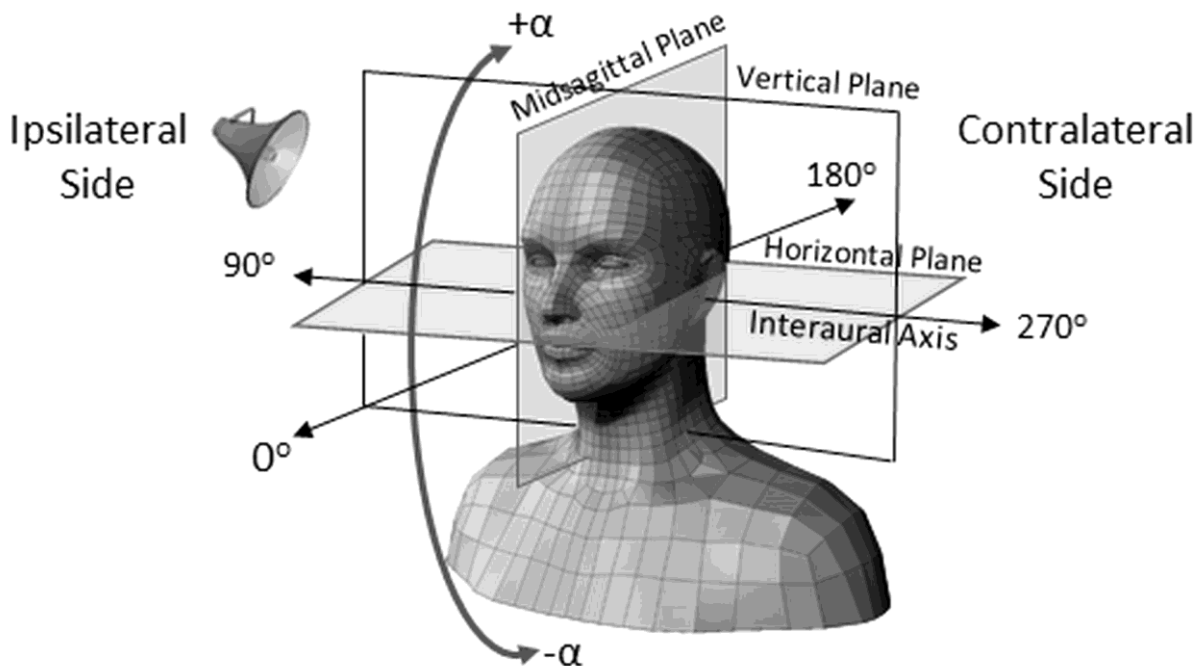


Figure 2-1. Spatial terminology used throughout the thesis. The main axis are labelled, along with the numbering conventions used for communicating spatial bearing. Elevation is shown as $\pm \alpha$, which indicates an increase or decrease in angle from 0-degrees at the horizontal plane to ± 90 -degrees up or down, respectively. The ipsilateral side refers to the side of the midsagittal plane where the stimulus is presented, while the contralateral side refers to the non-stimulus side.

Figure 2-2 illustrates regional segmentations used when grouping localisation data for analysis.

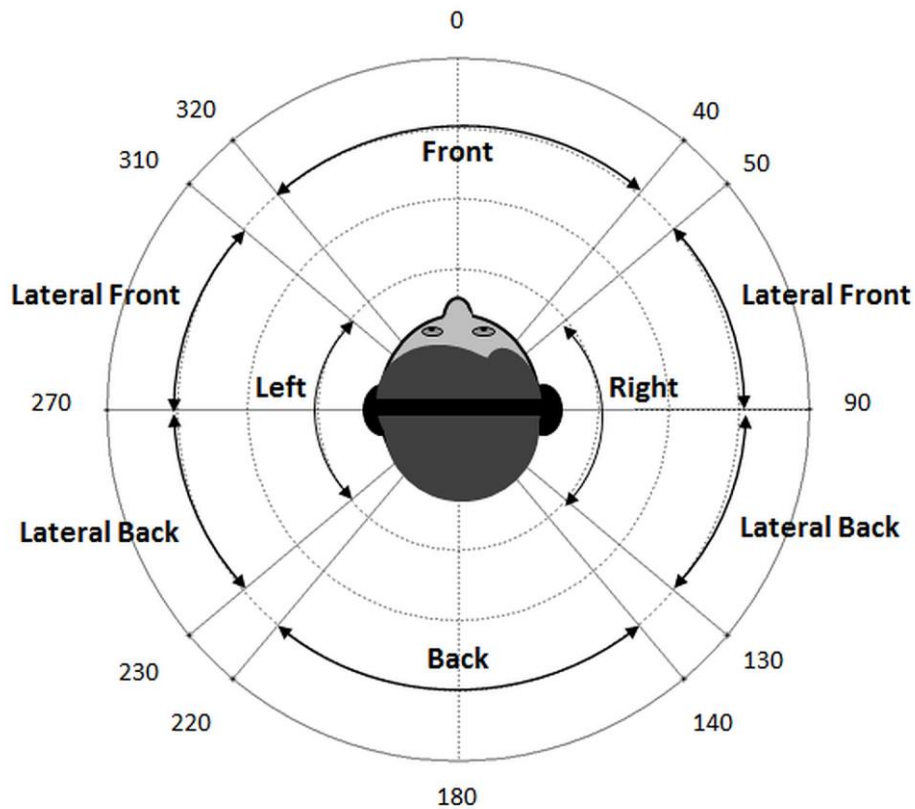


Figure 2-2. Spatial segmentation used for grouping localising estimates in azimuth.

2.2.2 Interaural Cues

The most important cues for localising a sound's azimuth position are derived through differences in the onset timing and intensity of a sound's waveform upon arrival at each ear. These cues are referred to within Rayleigh's (1907) duplex theory of localisation as *interaural time differences* (ITD) and *interaural intensity differences* (IID), which help the auditory system localise low frequency and high frequency sounds, respectively. Yost (2007) provides a good overview of these phenomena. ITDs occur when a sound source is positioned away from the midsagittal plane. As illustrated in Figure 2-3, the waveform from the source has a greater distance to travel to the right ear than the left. The subsequent relative delay in arrival of the waveform at the right ear will cause a slight phase difference between the vibrations occurring at the left and right tympanic membrane. ITDs not only occur at the leading and trailing edge of a sound, they also occur as similar portions of the continuing waveform arrive at each ear (Tobias & Schubert, 1959).

Interaural intensity differences (IID) occur due to shadowing effects. Figure 2-3 illustrates that a waveform will be attenuated at the right ear as the physical shape of the

head poses an obstacle to the direct path to the ear. An increase in frequency brought about through a decrease in wavelength will result in an increase in the magnitude of a shadowing effect (Middlebrooks & Green, 1991). IIDs become no longer effective when wavelengths are below 1 kHz as the waveform can diffract around the head. Sounds that have been localised through interaural cues alone are perceived to reside within the head along the interaural axis.

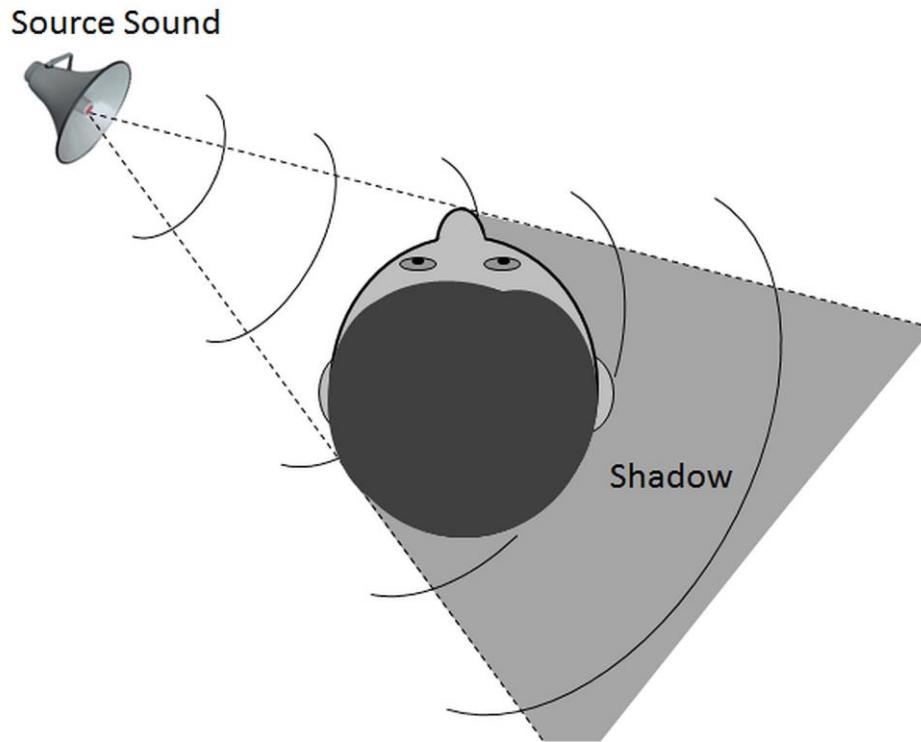


Figure 2-3. Duplex Theory of Localisation. The head shadow, or attenuation of the sound wave at one ear due to the physical shape of the head is illustrated.

2.2.3 Spectral Cues and Head Related Transfer Functions

The pinnae (outer ear), head, and upper torso, have an effect on the sound waveform prior to its arrival at the eardrum. The spectral filtering of the sound which occurs as it interacts with these body parts is termed the *head-related transfer function* (HRTF). The pinnae in particular is shown to produce localising cues, otherwise known as *spectral cues* due to the spectral changes that occur in the waveform during interaction with the asymmetric shape of the pinnae. Very small time delays, resonances, and diffractions are introduced into the sound waveform as it interacts with the pinnae, thereby providing a unique modification to the HRTF for any change in sound source position. These spectral and timing differences act as a localising cue, which aid in monaural and binaural localising when determining a sound's elevation (Alves-Pinto, Palmer, & Lopez-

Poveda, 2014) and disambiguating front-back confusions (Oldfield, S. R. & Parker, 1984b; Roffler & Butler, 1968; Yost, 2007).

2.2.4 Front-back Localising Ambiguity

In the absence of reliable spectral cues, a listener relies predominantly on interaural cues to localise a sound's front-back hemisphere of origin. In such instances, an increased potential for hemisphere ambiguity occurs (Oldfield, S. R. & Parker, 1984a). These errors occur more frequently when listening to spatial audio presented through binaural headphones rather than free field listening (Wightman & Kistler, 1989a). Such confusions occur due to the resemblance of interaural cues for sounds positioned at similar angles toward the front or back. Figure 2-4 illustrates this dilemma. Sound 'Source A', is positioned 20-degrees in azimuth forward of the interaural axis. When localising a source in this position, the listener has minimal cues to draw upon when differentiating this position from a similar angle of 20-degrees behind the interaural axis because both positions provide very similar interaural cues.

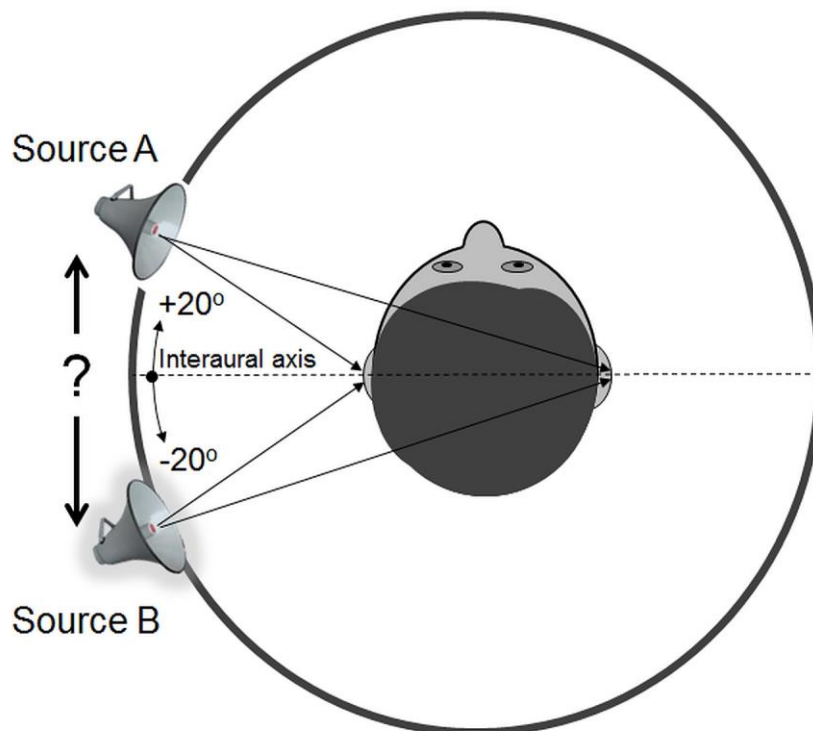


Figure 2-4. Front-back hemisphere localising ambiguity. Two potential source locations identified as Source A and Source B are shown. Ambiguous azimuth localising cues exist because of the similarity of interaural cues (ITD, IID) available at mirrored positions in the front and back hemispheres.

Front-back localising uncertainties exist at intermediate distances for source locations on points anywhere about a circle oriented perpendicular to the interaural axis.

These conic surfaces extend from the ear and are referred to as the *cone of confusion* (Yost, 2007). Figure 2-5 illustrates this phenomenon as it extends from the right ear.

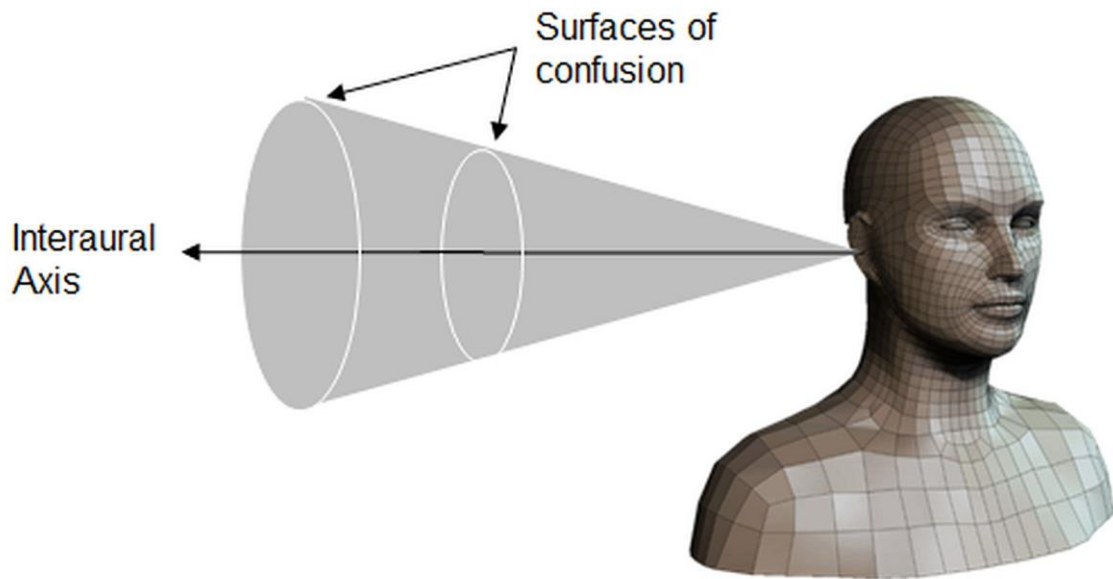


Figure 2-5. The cone of confusion. Conical surfaces are shown extending from the right ear around the interaural axis. Localising audio positioned about these surfaces are problematic.

2.2.5 Audio Localising Through Head Movement Cues

An effective method for localising sound occurs through head movements (Muller & Bovet, 1999). We turn our head in order to minimise differences between sampled interaural cues, which thereby orients the listener's head toward a sound source. Evidence suggests that infants as young as several hours old will turn their head in an attempt to localise sound and face the source (Clifton, 1992). Similarly, by turning the head we can also gather important localising cues for determining a sound's front or back hemisphere of origin, thereby overcoming front-back confusion. Figure 2-6 illustrates this method. As previously highlighted, interaural cues may not be adequate to resolve front-back hemisphere localising ambiguity for the sound labelled Source Location in the upper left side of the ambiguous localisation diagram. The source sound would produce very similar interaural cues to a possible alternate location toward the back, labelled in the diagram as the Possible Source Location. In this example, the listener turns his head to the left in order to deliberately alter the interaural cues. If the source was located in the front hemisphere, the interaural delay would decrease due to the now similar distance from the source to each ear.

The opposite is true if the source was located toward the back. In this configuration, as the head is turned to the left, the left ear moves closer to the sound

source position, thereby increasing the interaural onset timing delay and subsequent phase difference. A number of studies support the claim that head movements greatly improve one's ability to resolve front-back localising ambiguity (Iwaya, Suzuki, & Kimura, 2003; McAnally & Martin, 2014; Thurlow, Mangels, & Runge, 1967; Wightman & Kistler, 1999). Similar to the way in which head movements introduce dynamic changes in the localising cues for a sound source, a moving sound produces dynamic changes for a fixed head.

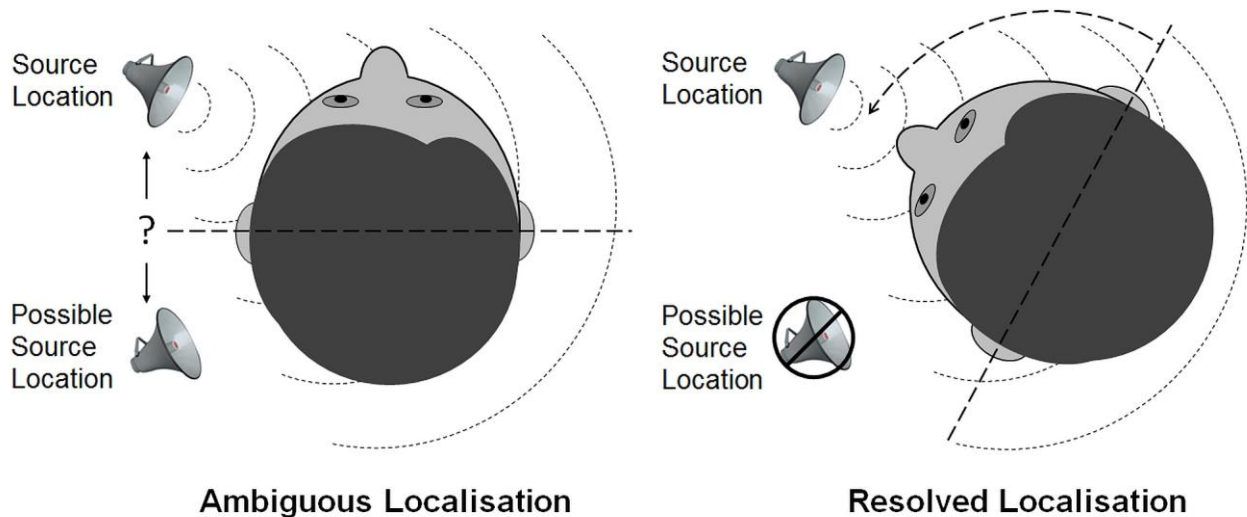


Figure 2-6. Resolving front-back localising ambiguity through head movements. A situation of ambiguous hemisphere localising for a source sound is shown to the left. The diagram to the right illustrates how turning the head to face a source minimises the interaural cues and aids in front-back hemisphere discrimination.

2.2.6 Binaural Speech

Research conducted into binaural speech remains quite focussed on speech intelligibility, particularly in regard to *binaural unmasking*. This is a common term used to describe a phenomenon whereby the masking properties found when one monaural sound interferes with another monaural sound are reduced when the sounds are presented binaurally. Pollack and Pickett (1958) wrote what is considered a seminal paper regarding the effect of stereophonic listening on speech intelligibility. They used stimuli presented through headphones to compare listening performance in dichotic and monotic configurations. In the monotic condition, normal speech was presented to one ear, along with a varying number of background talkers, who delivered a babble of monosyllabic words. The dichotic listening condition included different background talkers presented to each ear along with the normal test words, which were presented binaurally. They found that a 12 dB advantage existed in the stereophonic configuration with one background

talker, and a 5.5 dB advantage with seven background talkers. The superior performance demonstrated in the binaural condition they referred to as the *cocktail party effect*, which is a term originally coined by Cherry (1953).

Other researchers then began to demonstrate that the binaural advantage is facilitated by the auditory system's comparison of signals arriving at each ear (Levitt & Rabiner, 1967). Bronkhurst & Plomp (1988, 1992) conducted speech intelligibility tests using a KEMAR manikin to introduce HRTF cues into the speech and noise sources. They studied the effect of interaural time delay (ITD) and acoustic head shadow (interaural level differences (ILD) on speech intelligibility under conditions of competing noise and interfering talkers. They found that positioning the maskers off the midsagittal plane toward other positions about the azimuth introduced a gain in the speech-reception threshold ranging from 1.5 to 8 dB, thereby supporting the claim that binaural cues in the form of ILD and ITD significantly aid speech intelligibility. They also highlight a monaural effect that occurs when the signal to noise ratios at both ears change when a masker is moved to a new location. In asymmetrical configurations, the signal to noise ratio at the best ear can then also be used to achieve a release from masking.

Culling et al., (2004) found that ITD's alone were effective in producing a binaural unmasking effect for noise in the form of speech and non-speech shaped interferers that were spatially distributed about the listener. Their experiment showed that both ILDs and ITD's make independent contributions to spatial unmasking. They also conducted speech intelligibility studies using different interference maskers. They used maskers in the form of speech from the same talker delivering the target sentences; time-reversed sentences from the same talker; or speech spectrum shaped noise. They found that the benefits of binaural hearing are more prominent when there are multiple voiced interferers (Hawley, Litovsky, & Culling, 2004).

Informational masking refers to a degraded ability for the listener to perceive discriminating features between sounds containing similarities in context. This differs from energetic masking, where sounds overlap during physical interaction at the cochlea. Balakrishnan & Freyman (2008) found that the release from informational masking can be achieved through barely noticeable target-masker spatial differences. Release from informational masking has previously been shown to be improved through the use of audible and written forms of priming (Freyman, Balakrishnan, & Helfer, 2004). Priming in this context refers to precursory exposure to the audio stimulus. Informational masking posed a potential issue in the segregation of concurrent sonifications in studies 4 and 5.

The spatial separation of sounds provided an effective means for mitigating such interference.

Research into binaural speech has helped us understand the benefits to be gained in utilising binaural auditory displays to overcome problems with masking. Little research has been conducted to supplement this knowledge with insight into the effect of a talker's spatial location on speech comprehension. This question is explored further in Study 4 of this thesis and builds from other fields of research related to *right ear advantage*, where speech is considered more effectively attended to by the right ear (Kimura, 1961, 1967).

2.3 Auditory Stream Segregation

Our environment comprises a variety of sounds that originate from different sources and locations. Those unique sounds interact with our ears as a combined complex waveform. How an individual then undertakes the process of perceptually reconstituting the individual streams is a complex and challenging field of study. Albert Bregman (1990) coined the term *Auditory Scene Analysis (ASA)*, which is a model of auditory perception that addresses the problem of identifying which properties of sound belong together and which do not, thereby identifying audio *stream segregation*.

Bregman's ASA model proposes two processing stages that enable the perceptual organisation of sound. The first stage of the perceptual grouping process occurs when sound waves are broken down into a large number of separate chunks of audio for analysis. These chunks are then analysed during a second stage to determine which of them adhere to the attributes of grouping cues. Bregman refers to two ways in which the integration of common audio components occur: (1) *sequential integration*; and (2) *simultaneous integration*. *Schema* is also relied upon to introduce our knowledge of the environment and available sounds when interpreting the auditory scene.

2.3.1 Sequential Integration

Sequential integration is used to identify common components of sound that follow each other in time. Sound attributes, such as melody and rhythm provide temporal cues that enable grouping. Rhythm and tempo define the temporal proximity of sounds and are sometimes referred to as the beat. Within linguistics, rhythm is one of the elements of prosody, which help the listener infer the state of emotion or context of utterance from the speaker.

2.3.2 *Simultaneous Integration*

Simultaneous integration is a process involved with segregating concurrent components into groups. The process relies on identifying cues associated with the spectral qualities of a sound, such as pitch and timbre, to regroup the components into distinct streams of sound.

Pitch is the perceived frequency of sound. The *place theory* of pitch indicates that vibrations occurring at the eardrum will stimulate nerve endings along the basilar membrane such that a doubling of frequency will double the distance along the membrane (Yost, 2007). Pitch perception can be considered as absolute, or relative. Absolute perception refers to an individual's ability to identify or recreate a musical note in the absence of a reference tone. This ability is extremely rare, with only one in 10,000 people being able to do so (McDermott & Oxenham, 2008). Relative pitch perception is far more common and refers to an individual's ability to identify the distance of a note from a reference note. The playing of a musical instrument, or singing along in tune to music utilises this skill.

Timbre refers to our holistic perception of the unique spectral frequencies and shape of the waveform over time. Timbre is very difficult to define and quantify. As such it is sometimes referred to as the colour, or quality of a tone. Put simply, if one was to play the same note at the same loudness on two different musical instruments, the perceived difference in the sound is the timbre.

Loudness is the perceptual correlate of a waveform's intensity and duration. An increase in the amplitude of a sound's waveform will be perceived as an increase in loudness. The loudness of a waveform varies as a function of frequency such that a 100 Hz tone played at 60 dB would not sound as loud as a 4 kHz tone played at 60 db. While the perception of loudness is very subjective, the Fletcher-Munson graph plots the relationship between frequency and perceived loudness and is reproduced in Figure 9.2 of Annex A.

Spatial location also aids with component segregation. The different spatial locations of competing sounds produce unique localising cues that help the listener to segregate those sounds. These cues can be derived in two ways, either by the sound sources moving independently (Shestopalova et al., 2014), or by the listener moving their head (Kondo, Toshima, Pressnitzer, & Kashino, 2014).

2.3.3 Schemas

Bregman (1990) referred to schema as “a mental representation of some regularity in our experience” (pg. 43). Our existing knowledge informs schema when attempting to make sense of a particular situation, which is not unlike the process involved in determining which sounds belong together. Bar (2007) tells us that our existing knowledge contributes toward our perception of the environment as much as incoming sensory information. He believes that the brain is continuously establishing predictions of the future that guide our actions, plans and thoughts. Similarly, both Winkler et al. (2012) and Denham et al. (2014) believe that several alternate representations of stream groupings are continuously being established and vie for dominance. The established pattern of predictable components enable new incoming sounds to be detected and appropriately integrated into streams. Such predictability requires a level of established schema. Established knowledge that can guide perception during auditory stream segregation can take the form of contextual information (Snyder, Carter, Lee, Hannon, & Alain, 2008; Snyder, Holder, Weintraub, & Carter, 2009; Winkler et al., 2012), and prior learning (Snyder, Carter, Hannon, & Alain, 2009; van Zuijlen, van Zuijlen, Sussman, Winkler, & Näätänen, 2005).

2.3.4 Gestalt Principles for Auditory Perception

Many of the grouping principles that underlie *Auditory Scene Analysis* (ASA) come from Gestalt psychology. Gestalt psychology implies that we acquire meaningful perceptions of our environment when our minds consider a holistic view of the available stimulus independent of the collective sum of the individual parts themselves. Many of the Gestalt laws, such as *similarity*, *closure*, *symmetry*, *common fate*, and *continuity*, are relevant to auditory perception when considered in a temporal form, rather than spatially, which is the more commonly associated application of the laws within visual perception.

The Gestalt law of similarity states that images that are similar tend to group together. For instance, the diagram in the left of Figure 2-7 below shows how a grid of evenly spaced symbols are perceived as grouped by column or row depending on the orientation of the similar shapes. Bregman & Campbell (1971) demonstrated the auditory equivalent of the law of similarity by presenting a series of six evenly spaced tones of equal period and varying frequency. The graph in the right of Figure 2-7 illustrates that regardless of the presentation order, tones of high and low frequencies are perceived to segregate into parallel sequences as if they originate from separate sources.

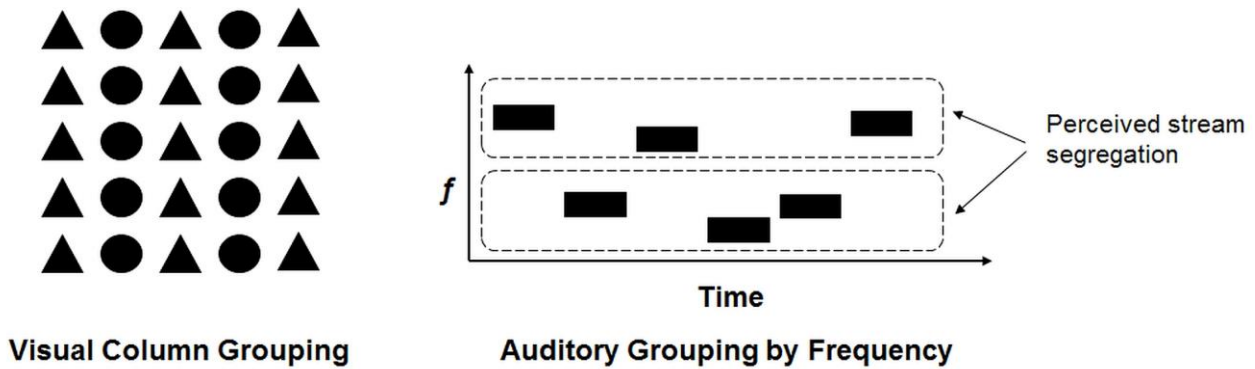


Figure 2-7. Modal similarities for the Gestalt law of similarity. The left figure illustrates a visual phenomenon where evenly spaced shapes are perceptually grouped by similarity into columns. The figure to the right illustrates a phenomenon where a sequence of audio tones of varying pitch are perceived to exist as two separate streams of high and low pitch tones from different sources.

A visual illustration of the Gestalt law of good continuity can be seen when two dots are presented moving at an angle toward a point of intersection where their paths cross. If the dots are moving at the same velocity when they intersect, they appear to bounce off each other. If they differ in velocity at the point of intersection, the viewer correctly perceives the continuation in trajectory as they cross over. McPherson, Ciocca, and Bregman (1994) have demonstrated the auditory equivalent of the law of good continuity by presenting two tones whose frequency vary such that they progress through a merging point of common frequency. The rate of change in pitch was found to be a factor that enabled segregation of the two sounds. Sounds with differing rates of change were able to be tracked through a point of cross over pitch more easily than those with the same rate of change. Figure 2-8 illustrates the visual and auditory concepts in the law of good continuity. To describe all of the Gestalt Laws as they relate to auditory perception is beyond the scope of this thesis. For a more comprehensive overview of the similarities, the reader is directed to Kubovy and Van Valkenburg (2001).

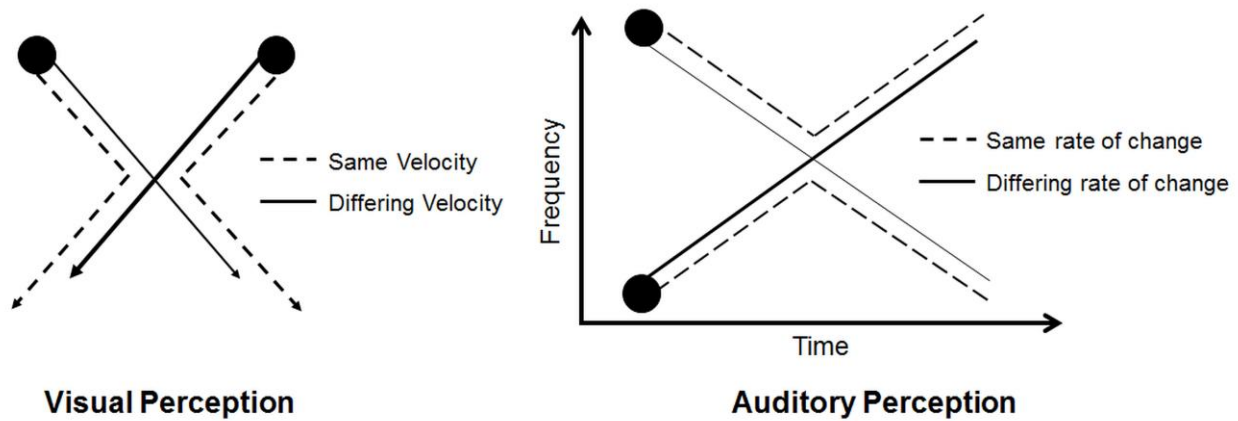


Figure 2-8. Modal similarities for the Gestalt law of good continuity. The left diagram illustrates a visual phenomenon where as two visual stimulus cross paths they will either appear to bounce off or pass through each other depending on their common or differing velocities. The figure to the right illustrates the auditory version of this phenomenon. A differing or similar rate of change in pitch between two sounds as they pass through the same pitch can produce a similar percept.

2.4 Spatial Audio Rendering

Wightman & Kistler (1989a, 1989b) conducted some of the earliest spatial audio studies utilising audio filters to render localising cues into audio streams for delivery through binaural headphones. Shortly following this work, NASA produced the Convolvotron, which was one of the first spatial audio systems employing integrated head tracking (Wenzel, Fisher, Stone, & Foster, 1990). This system demonstrated the powerful capabilities of spatial audio technology and pioneered the use of spatial audio within modern virtual reality (VR) systems. Virtual environments, such as CAVE (Kaper, Wiebel, & Tipei, 1999) and CyberStage (Eckel, 1998) have been displaying data through 3-D auditory space for some time. Modern computer processing capabilities are such that hardware solutions similar to the Convolvotron are no longer required. The average on-board processor found within most laptop computers will easily perform 3-D audio rendering functions.

Spatial audio is a key element required in the creation of an immersive virtual reality experience. Open source VR software communities, such as Blender (2015) and Rapture (Blue Ripple Sound, 2015) now support the use of spatial audio. Rapture supports the use of OpenAL, which is a cross platform audio application programming interface (API) that provides a library of software functions that enable programming of multi-channel spatial audio. Commercial VR systems, such as Sony's Project Morpheus (Sony, 2015) will

include a custom 3-D audio binaural solution in its development kit, while Facebook's new Oculus Rift will also include integrated headphones with 3-D audio support (Anthony, 2014).

SLAB3D (Miller, J. D. & Wenzel, 2002) is a real time 3-D audio rendering tool developed by NASA's Ames Research Center and was used in all of the studies conducted within this thesis. SLAB's audio processing allows for the arbitrary placement of sound sources and customisation of environmental parameters. The default SLAB3D HRTF was used, which is a particular person's HRTF measurements converted to minimum phase HRTFs. The minimum and maximum ITDs are -784 (left 90 degrees, 0 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. (1957) and as calculated for a spherical head model using a radius of 8.75 cm (Woodworth, 1938). A generalised or non-individual HRTF refers to the use of one HRTF to provide localising cues for a variety of people utilising an auditory display. Localising sound within a non-individual auditory display is often more difficult because the localising cues do not exactly match the listener's own individual cues. Spectral cues in particular are found to be the most problematic to generalise through a non-individual display because of the diverse range of anatomical shapes that exist between individuals pinnae. This greatly effects localising performance for a sound's elevation and front-back point of origin.

2.4.1 Internalisation of Sound

The use of a non-individual HRTF is known to often promote a well-documented phenomenon called *internalisation* (Yost, 2007). Non-individual HRTF spectral cues that are poorly matched to a listener's own cues make a sound appear to reside more internally toward the interaural axis, particularly in the front region as illustrated in Figure 2-9. Internalised spatial perception can be problematic as it degrades stream segregation and localising accuracy.

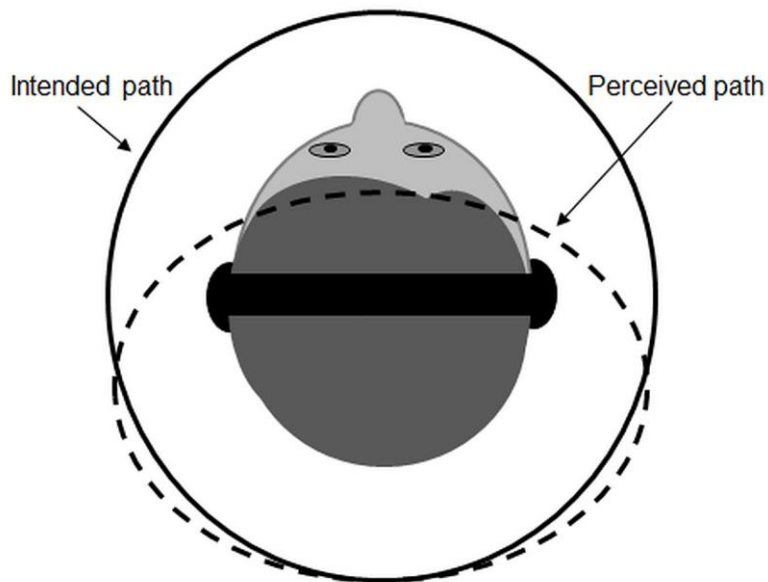


Figure 2-9. The internalised perception of sound caused by poorly matched spectral cues. The diagram shows the commonly perceived internalised path of an orbiting sound compared with the intended path.

2.4.2 Localising Skew

A larger or smaller mismatch between a non-individual HRTF interaural dimension and a listener's actual anatomy can introduce a bias in localising perception toward the interaural axis or midsagittal plane, respectively. For example, Figure 2-10 below illustrates the effect of a cue mismatch where a sound containing interaural cues created from an upper percentile head size is rendered at the location labelled HRTF database cue position. For a listener with an anatomically smaller head size, this situation would produce a perceived localisation further toward the interaural axis than intended, as shown by the position labelled perceived position. This is due to the listener's expectation of smaller interaural cues for sound positioned toward the intended location. Alternatively, a smaller mismatch in interaural distance between the HRTF and listener would skew the perceived location toward the mid-sagittal plane where interaural differences are smaller.

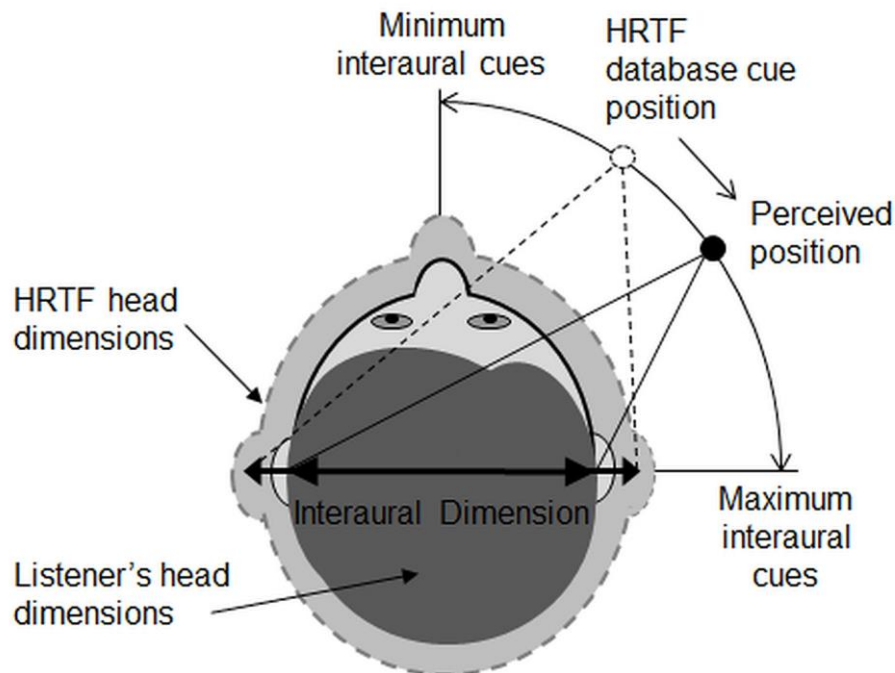


Figure 2-10. Azimuth localising error resulting from poorly matched HRTF interaural cues. A sound spatially rendered with interaural cues developed from an individual with an upper percentile head size is labelled HRTF database cue position. Due to the larger than expected interaural cue, a person with a smaller head size will localise this sound further toward the interaural axis at the position labelled perceived position.

2.5 Auditory Displays

The use of non-speech audio to convey systems information has until recently been limited to simple sounds like beeps and pings to alert the operator to certain situations. Mainly used as warnings and status alerts, these audible cues have often been poorly designed in terms of ensuring an appropriate level of salience and meaning. There are many benefits to be gained from the use of auditory displays, not the least of which relates to easing the burden on the visual system by employing otherwise underutilised auditory sensory and cognitive processing faculties. The following section introduces three different types of non-speech audio, which have laid the foundation for the design of auditory displays.

2.5.1 Multisensory Displays

Interpreting events in the real world often involves the use of at least two of our five senses, which include sight, hearing, touch, smell, and taste. Multisensory, or multimodal displays, are interfaces that utilise more than one modality to communicate information to the operator. Diversifying the presentation of information by employing different modalities has the potential to overcome some of the information processing bottlenecks mentioned

previously when outlining Wicken's (2008) Multiple Resource Theory. Access to multisensory information has been shown to improve audio-visual stimulus processing speed and accuracy, with electrophysiological recordings showing decreased neural processing for both audio and visual stimulus under multisensory conditions when compared with visual stimulus alone (Mishra & Gazzaley, 2012).

An effective example of a multisensory display can be observed in a car navigation system. Aural navigation instructions facilitate safer driving conditions through improved access to information. By supplementing the visual information presented on the display, the multisensory display allows the driver to keep her eyes on the road and attend to the visual display when conditions are favourable to do so. The automotive industry is adopting multisensory displays in a variety of ways to communicate warnings relating to driver distraction and fatigue (Spence & Ho, 2008). The medical industry is also adopting multisensory displays to facilitate improved monitoring of patient health monitoring devices during theatre operations (Sanderson, 2006). The relevance and associated benefits of multisensory displays are expanded on throughout this thesis within the context of each study.

2.5.2 Aural Alerts

High workload environments facilitate conditions where one's limited sensory and cognitive processing abilities can make it difficult to adequately acquire information and meet task objectives. The primary intent of an audible alert is to draw the operator's attention toward a high priority event or condition. The need for a system to intervene and orient an operator's attention arises when the nature of a situation makes it likely that important information may be neglected, thereby resulting in an undesirable consequence. Auditory alerts are often simply used to indicate an impending need for action by the operator, such as an aircraft's terrain awareness and warning system (TAWS), which presents an aural alert or verbal statement to the pilot in order to prevent controlled flight into the ground (Skybrary, 2014).

Peryer et al. (2005) surveyed 50 commercial pilots to obtain their thoughts on the design characteristics of flight deck audible alerts. They found that while pilots believe that most alerts are effective, approximately half of those pilots surveyed believe them to be too loud, with 74% having experienced a startled response from alerts. Forty-six percent reported having experienced impaired performance brought about through the presentation of an audible alert. As expected, such problems are certainly not isolated to aviation. Sorkin (1988) found that operators from a range of industries such as nuclear

power plants and rail systems were disabling audible alerts. He believes that an excessive number of alerts that are highly aversive in their design brought about this behaviour.

Conveying alerts with an appropriate sense of urgency is difficult. Mondor & Finley (2003) found that naive participants were unable to distinguish urgency between three levels of escalating alerts that were generated by medical equipment designed by the manufacturer. It might be argued that we have a long way to go in regard to achieving interfaces that facilitate an adequate level of situation awareness such that alerts are required less frequently. Improving operator situation awareness through the use of integrated auditory displays may assist in mitigating errors brought about through inattention to primary system information.

2.5.3 Earcons

Graphical symbols, termed *icons*, have been utilised within computer interfaces to communicate information since the very earliest days of interface design. Following along similar principles to those of icons, structured sounds referred to as *earcons* have been in use since the mid 1970's as a means to communicate information via the alternate auditory modality (Adams & Trucks, 1976). Earcons are generally non-verbal audio messages used to provide user feedback through a computer interface. They communicate information relating to messages, functions, states and labels (Blattner, Sumikawa, & Greenberg, 1989). Gaver (1986) coined the term *auditory icon* to refer to a type of earcon that has a sound with a natural mapping to a particular condition or event that it relates to. The Windows operating system provides an example of such mapping, which can be heard when placing a file in the 'trash'. An auditory icon is presented that sounds like a piece of paper being scrunched up, which has a strong association with the event of discarding the file. The mapping of basic sounds to information has proven to be a very effective feedback method within computer interface design. The design principles and associated multimodal benefits of earcons and auditory icons has served as a precursor to more ambitious mapping of information to different properties of continuously presented sound, termed *sonification*.

2.5.4 Sonification

Sonification is a relatively new field of study, therefore the associated methodologies, terms and definitions are still evolving. Hermann (2008) defines sonification as "a technique that uses data as input, and generates sound signals (eventually in response to optional additional excitation or triggering)" (p. 2). He goes on further to provide the following four attributes that must be met for a signal to be considered sonification:

1. “The sound reflects *objective* properties or relations in the data.
2. The transformation is *systematic*. This means that there is a precise definition provided of how the data (and optional interactions) cause the sound to change.
3. The sonification is *reproducible*: given the same data and identical interactions (or triggers) the resulting sound has to be structurally identical” (p. 2).

The multi-dimensional attributes of sound provide a means for encoding a variety of data elements into a continuous stream. Often used to monitor the content and relationships of datasets contained within a database, auditory displays that employ sonification map data variables to different attributes of sound, such as pitch, timbre, and loudness. These attributes are adjusted as a function of the data over time in order to convey an audible representation of the data, thereby providing more perceptive insight into the relative changing state of the data.

Hans Geiger invented the Geiger-counter in the early 1900s. This device presents audible clicks at a rate that correlate with the amount of invisible radiation in the environment. This device allows the user to maintain accurate awareness of the relative levels of radiation while not monitoring the visible readout display. This provides practical benefits in that the user is free to move about and multitask uninterrupted by an otherwise visually constrained monitoring activity. The Geiger-counter remains in use today and has been demonstrated to provide a more effective means for monitoring radiation than compared with a visual display, or more interestingly, when compared with the concurrent use of both a visual display and audible device (Tzelgov, Srebro, Henik, & Kushelevsky, 1987).

The medical industry is adopting sonification displays to enable the hands free monitoring of a patient’s vitals. One of the earliest displays was the Pulse-Oximeter, which has been in use in medical operating theatres since the mid 1980’s to present a variable pitch tone that maps to the level of oxygen in a patient’s blood. Fitch and Kramer (1993) expanded this device into a workstation that included six parameters. Similar to the Geiger-counter, this device allowed doctors to detect emergency situations more quickly than when only having access to visual displays, or a combination of both visual and audio information.

Sonification has also been used to help surgeons precisely position surgical instruments (Jovanov, Starcevic, Wegner, Karron, & Radivojevic, 1998). They referred to this type of use of sonification as *tactile audio*. Other real time biofeedback applications utilising sonification have been demonstrated through movement (Effenberg, 2005; Ghez, Rikakis, DuBois, & Cook, 2000) and self-regulation of brain activity data (EEG)

(Hinterberger & Baier, 2005). Success has been found for sonified biofeedback in use for stroke victim rehabilitation (Woodford & Price, 2007) and for amputees learning how to control the position of their prosthetics (González, Yu, & Arieta, 2010)

Throughout studies three and four of this thesis, the context of sonification is focussed on closed loop interactions, which Hermann (2008) would consider fall into the category of *human activity*. This means that the interaction extends beyond the data to have an effect on the state of a system in the real world. The current studies utilise flight simulator interfaces and require the participant to monitor and deduce navigation errors regarding the aircraft's attitude and position primarily through information derived from sonified instrument data. The participant adjusts the aircraft through physical interaction with the control yoke, which changes the aircraft state and is subsequently fed back to the participant through sonified signals. Figure 2-11 illustrates a high level overview of the sonification feedback utilised in this thesis to convey system state data.

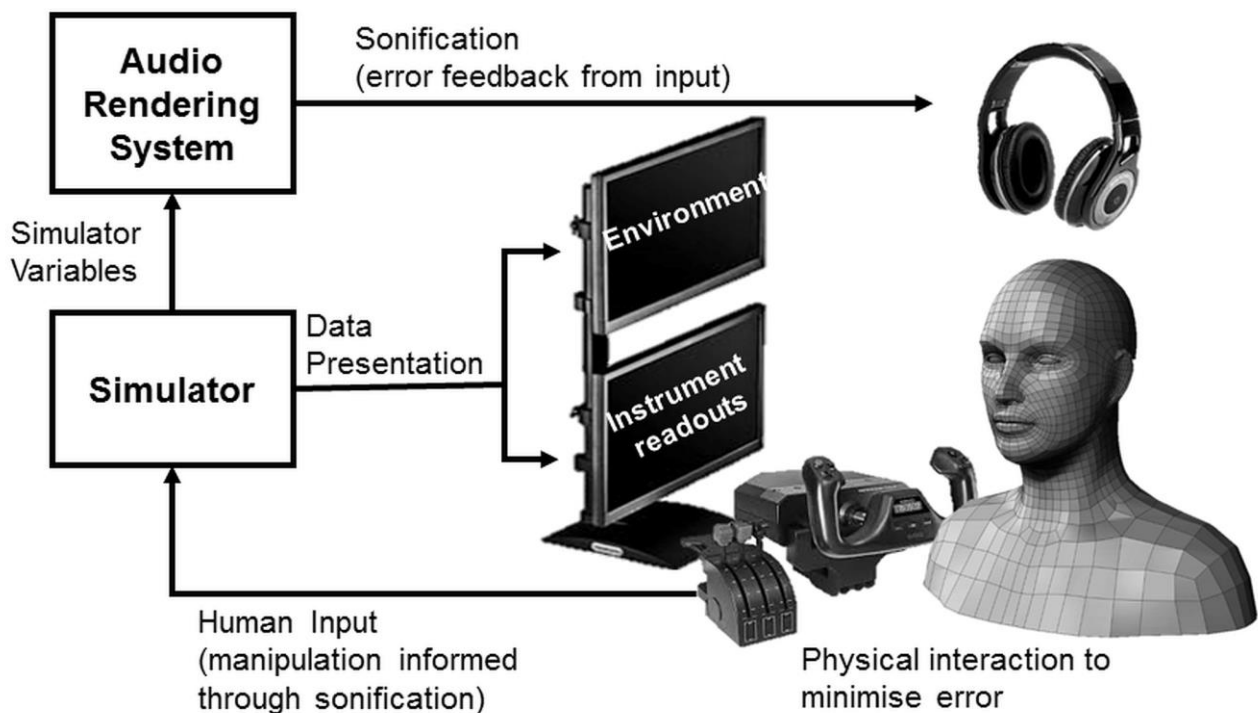


Figure 2-11. Sonified feedback of simulator data. Shown is a basic overview of how the flight simulator feeds flight simulator variables to the audio rendering tool to be sonified for presentation to the participant. Physical interaction with the flight deck controls then alter the aircraft's navigation readouts, which subsequently modifies the sonified signals.

Sonification effectively conveys a greater number of simultaneous parameters than compared with other modalities (Scaletti & Craig, 1991). As previously mentioned, modern technologies enable operators to access large amounts of data and information. Sonification is currently being utilised to monitor and extract meaning from large data sets

(Singurd, 2000). Sonified data from large EMG data sets have been found to not only be effective, but convey information in such a way as to be easily understood and utilised by non-experts (Pauletto & Hunt, 2009). A well-known example of how sonification has been utilised to diagnose a problem with a large data set came from the Voyager 2 program. A problem was detected with the spacecraft as it travelled past the rings of Saturn, however, the data could not be adequately diagnosed utilising the program's standard visual displays. The data was sonified through a synthesiser and produced a "machine gun" sound that led to a diagnosis that the spaceship was colliding with high-speed electromagnetically charged micrometeoroids (Kramer, 1993).

2.6 Handedness and Gender Effects on Lateralisation

Lateralization refers to the asymmetric distribution of functional regions between the cerebral hemispheres. The relationship between cerebral functioning and handedness is complex, however, early findings led to a belief among researchers that an individual's degree of left ear advantage correlates with the measure of handedness (Knox & Boone, 1970; Shankweiler & Studdert-Kennedy, 1975). Ear advantage will be discussed at some length in chapter six, however, for the purpose of clarifying the current statement, in most people, the right ear is found to facilitate better speech processing over the left. This is generally accepted as being because the left cerebral hemisphere has more specialised regions for processing speech. While both ears have neural pathways that connect to this region, the largest bundle connects to the right ear. This is referred to as the right ear advantage (Kimura, 1967). The association between handedness and asymmetrical functioning has been challenged over the years. Galin et al. (1982) collected EEG data from 90 participants of differing handedness as they performed a variety of activities involving writing, reading, speaking, listening, and block design construction. They found that asymmetric functioning did not differ by handedness.

In terms of gender differences, the total volume of brain has been found to differ significantly (Luders, Steinmetz, & Jancke, 2002). More recent studies support significant differences in size and structure between genders, but no relationship has been found between these differences and cognitive performance (Escorial et al., 2015). There is a body of evidence to suggest that the mature adult male has a more asymmetrically organised brain than a female or juvenile, which has been proposed as possibly accounting for the males improved visuospatial processing skills (McGlone, 1980). This notion has received further support recently through MRI studies focusing on the effects of age and gender on asymmetric functioning. Agcaoglu et al. (2015) conducted a large

study of over 600 participants ranging in age from 12 to 71. Their findings support theories that the brain functioning becomes more asymmetric with age to compensate for neural decline.

Hyde & Linn (1988) conducted a meta-analysis on 165 language studies. Overall, the studies indicated a slight female advantage for verbal ability, however, when Hyde & Linn adjusted the effect size by number of subjects, the effect was reversed. One of the studies which had found a slight male advantage had over 900,000 participants. They therefore concluded that strong evidence exists in support of there being zero difference in verbal ability between genders. A seminal paper on the topic of phonological processing between genders was published by Shaywitz et al. (1995). Their study involved collecting MRI data from 38 participants as they performed letter recognition, rhyme, and semantic tasks. They found that brain activation in male participants was far more lateralised toward the left than in females, whose data revealed a more broadly dispersed activation of regions in both the left and right cerebral hemispheres.

There remains much uncertainty and a lack of consolidated thought within the primary literature regarding the effects of handedness and gender on the perception and comprehension of spatial audio. The contributing factors are complex and multifaceted in regard to these effects. As such, the author has attempted to highlight his attempt to balance or control such variables in order to meet the aims and objectives within each study.

3. Study 1: Localising Synthesised Spatial Audio Filtered through a Generalised HRTF

Towers, J., Burgess-Limerick, R., & Riek, S. (2008). *Localising synthesised spatial audio filtered through a generalised HRTF*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society of Australia 44th Annual Conference, Adelaide, Australia.

A large number of improvements have been made to this manuscript since its first publication. Most notably, the results section has been expanded considerably to include additional data in the form of tables, figures, and descriptive analysis. The defined regions for grouping azimuth localising estimates for front-back error analysis have been changed. Front-back errors were previously grouped by front, left, right, and back regions. A more appropriate way to analyse these data is by front, lateral front, lateral back, and back regions. The reasoning is discussed further in section 3-3-4.

3.1 Abstract

The presentation of spatial audio through binaural headphones requires initial filtering of the source sound through head related transfer functions (HRTF) to provide localisation cues in the form of interaural time differences (ITD), interaural intensity differences (IID), and spectral cues. These cues occur naturally when sound interacts with an individual's physical features, such as the shape of the pinna, head, and upper shoulders. Incorporating non-individual HRTF cues into an auditory display is cost effective, however, they often degrade localising accuracy because the cues do not accurately match the individual's own cues. The purpose of the current study was to determine participants' ability to localise spatial white noise sounds developed with SLAB3D's non-individual HRTF database and compare those results with previous studies. These data also provide baseline performance information for subsequent auditory display studies that focus on improving localising performance when utilising a non-individual HRTF.

Eleven untrained participants listened to sounds randomly presented at 10-degree increments in azimuth and -20, 0, and +20 degrees in elevation. Azimuth localisation accuracy compared similarly with previous studies. It is likely that poorly matched HRTF cues caused a tendency for localisation estimates to bias considerably toward the ipsilateral side of the interaural axis. Approximately 20% of presented sounds resulted in front-back hemisphere localising confusions. SLAB3D and its non-individual HRTF database are considered adequate for use in subsequent experiments into auditory display design.

3.2 Introduction

Complex systems employing diverse input modalities such as auditory displays mitigate excessive operator workload by exploiting cognitively independent processing capabilities (Wickens, 2002). Spatial auditory displays also offer perceptual benefits such as the Cocktail Party Effect (Brungart & Simpson, 2007), where spatially isolated voice communications allow operators to monitor and attend to several concurrent conversations more effectively than through mono audio. The Duplex Theory of Localisation identifies interaural time differences (ITD) and interaural intensity differences (IID) as the primary cues for spatial audio localisation about the azimuth. The ability to differentiate between locations where ITD and IID are effectively equal, such as in elevation and across similar front-back angles, is attributed to spectral cues that appear in the form of peaks and

notches across the sound waveform. They occur when sound is distorted upon impact with the physical shape of the pinnae, upper torso, and head (Yost, 2007). This theory is described further in section two of this thesis.

Synthesising spatial cues for presentation over binaural headphones is made possible through the use of audio filters called Head Related Transfer Functions (HRTFs), which are created from binaural recordings made within an anechoic chamber (Wightman & Kistler, 1989a). Non-individual HRTFs often degrade sound localising performance because the integrated localising cues often don't closely match a listener's own cues. This is due to anatomical differences between the person used during the creation of the HRTF database and the current listener.

The objective of this study was to determine if participants could achieve a similar degree of localising accuracy to that of related studies when localising spatial white noise sounds generated through SLAB3D and its non-individual HRTF. These data are intended to provide a baseline of expected localising performance for future studies and guide design direction during the development of techniques intended to improve the localising of sound generated with a non-individualised HRTF.

3.3 Method

3.3.1 Apparatus

Recordings were made of two second white noise sounds positioned at three different elevations (+20, 0, and -20 degrees) for each 10-degree increment about the azimuth starting at 0-degrees. A total of 108 audio files were created with SLAB3D (Miller, J. D. & Wenzel, 2002) using the default non-individual HRTF, which is described further in section 2.4 of this thesis. The sounds were then randomised in order to be presented three times each over two sessions. A total of 324 sounds were combined together into two audio files comprising 162 presentations each using the audio editing software Cool Edit Pro. A five second delay was incorporated between each sound to allow time for the participant to input their localising estimates. During the localising task, recordings were played back to participants through Bose® TriPort binaural headphones on a Dell Latitude D630 laptop running Windows XP and Windows Media Player 11.

A Wacom Bamboo graphics tablet was used to record participants' localising estimates. The touch pad was overlaid with a circle comprising lines indicating the midsagittal plane, interaural axis, and each 45-degree increment about its circumference. To its right was a liner scale labelled low to high where participants input their level of

confidence when localising each sound. Figure 3-1 below shows the graphics tablet interface.

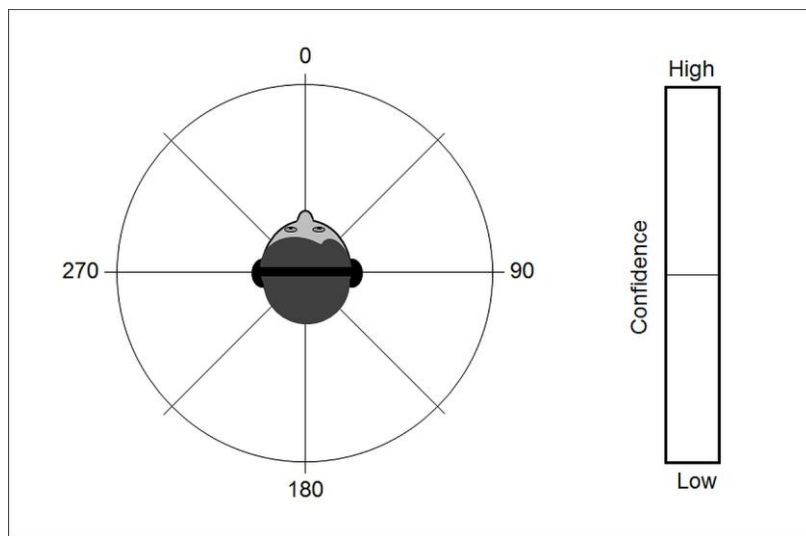


Figure 3-1. Graphics tablet overlay for recording participants' localising estimates and confidence. Participants input sound localising estimates by touching the circumference of the circle at the perceived relative bearing. Confidence was input by touching the slider bar on the right.

3.3.2 Participants

Eleven volunteers, eight male and three female, aged between 24 and 52 years of age participated in the experiment. All were Boeing Australia employees and reported no known prior or existing hearing problems. None of the participants had previous experience listening to spatial audio through binaural headphones.

3.3.3 Procedure

Prior to the experiment, participants were provided with several minutes of exposure to a white noise sound that orbited the head in a clockwise direction at a virtual distance of 1.5 meters. At this point participants adjusted the volume to a comfortable level. The front-back location of sound is quite perceptible during this type of demonstration, thereby providing participants with some familiarisation with internalisation. This occurs when a sound with constant distance cues appears closer in the front region compared with the back and is further described in section 2.4.1 of this thesis. Participants were informed that during the experiment sounds would appear at random stationary locations anywhere about the full 360-degrees in azimuth.

Participants then performed a self-paced dexterity task by touching marked intersecting points about the circumference of the localising input circle positioned on the graphics tablet on the table directly in front of them. This task was intended to test the accuracy in which participants could record predetermined azimuth bearings. Participants were asked to touch each point in a clockwise direction within one to two seconds following the previous point, but not to compromise accuracy if that seemed too fast.

As sounds were presented during the data collection phase, participants were required to remain still, facing forward with their head up and eyes closed. During a five-second pause between sound presentations, participants were required to open their eyes and input the perceived azimuth bearing of the sound. They did this by touching the azimuth location on the circle overlaid on the graphics tablet. They then input an estimate of the level of confidence they had in their localising accuracy. This was achieved by touching the linear scale located on the right side of the tablet, which recorded a numerical value to two decimal places between 0 and 10. Each session took approximately 20-minutes to complete with a short break being taken between each session. Three participants performed the experiment over two days.

3.3.4 Data Analysis

All data analysis were conducted using a significance level identified at $\alpha = .05$. The approach taken was a 2-way repeated-measures analysis of variance (ANOVA) with factors of azimuth region and elevation (-20, 0, 20 degrees). The azimuth regions used for grouping localising estimates were front, left, right, and back. These regions are illustrated in Figure 2-2 within chapter two of this thesis. Those regions were chosen because they compare similarly with previous studies and were expected to support identifying potential non-individual HRTF interaural effects that may degrade localising within the left and right regions. Post-hoc analyses were undertaken for all significant main effects using Bonferroni adjusted alpha levels. Effect sizes were expressed using Cohen's *d*. Within-subject confidence intervals were constructed by eliminating intersubject variance using the method outlined by Loftus and Masson (1994).

Front-back errors occur when a sound is localised toward the incorrect front or back hemisphere with respect to the interaural axis. These errors occur when non-individual spectral cues are inadequate for the listener to differentiate between points in the front and back where interaural cues are similar, especially about the mid-sagittal plane and cones of confusion. In line with previous studies (Oldfield, S. R. & Parker, 1984a; Wenzel et al., 1993; Wightman & Kistler, 1989b), front-back errors have been corrected prior to

undertaking azimuth error ANOVA because spectral cues are not considered to be a key discriminator in determining azimuth bearing. Front-back corrections were made by calculating the bearing from the midsagittal plane to the localised input and creating a new bearing estimate of the same magnitude in the opposite hemisphere on the same lateral side of the midsagittal plane.

The azimuth regions chosen for front-back error analysis were front, lateral front, lateral back, and back, which are also illustrated in Figure 2-2. These regions were chosen because they border the interaural axis, which is where front-back errors are relative to. This was intended to ensure that differences in front-back errors occurring between hemispheres within the lateral regions could be measured, which would not have been the case if those lateral front and lateral back hemisphere regions were consolidated as left and right regions.

3.4 Results

Azimuth localising was found to be most accurate in the left and right regions, with the front then back regions producing the least accurate localising performance. Sounds positioned in elevation at 20-degrees were localised more accurately than at 0-degrees, with -20-degrees producing the worst localising accuracy. Front-back errors were significantly more prominent in the front region compared with the back region. The back 0-degree condition produced significantly less errors than all other regions and elevations, while the front -20-degree condition produced the worst number of front-back errors of any region or elevation.

3.4.1 Input Accuracy

The dexterity task confirmed that participants could achieve a high degree of precision when inputting localising estimates, which was measured as degrees of error when selecting predetermined points about the localising interface shown in Figure 3-1 ($M = 0.24$, $SD = 0.18$).

3.4.2 Azimuth Localisation Accuracy

ANOVA was undertaken for azimuth localisation estimates grouped by region and elevation.

Table 3-1 contains a summary of the ANOVA, post-hoc and effect size data for region and elevation. Main effects were found for region ($F(3, 30)$, $p = .000$), elevation ($F(2, 20)$, $p = .000$), and an interaction was found for region by elevation ($F(6, 60)$, $p = .000$). Post-hoc analysis of localising results by region revealed a large effect size for

improved localising performance in the right region compared with the front ($d = .087$) region. A large effect size was also observed for better localising performance in both the left ($d = 1.7$) and right ($d = 1.8$) regions compared with the back region. Figures 3-2, 3-3, and 3-4 below plot the mean azimuth localising error by region, elevation, and region by elevation.

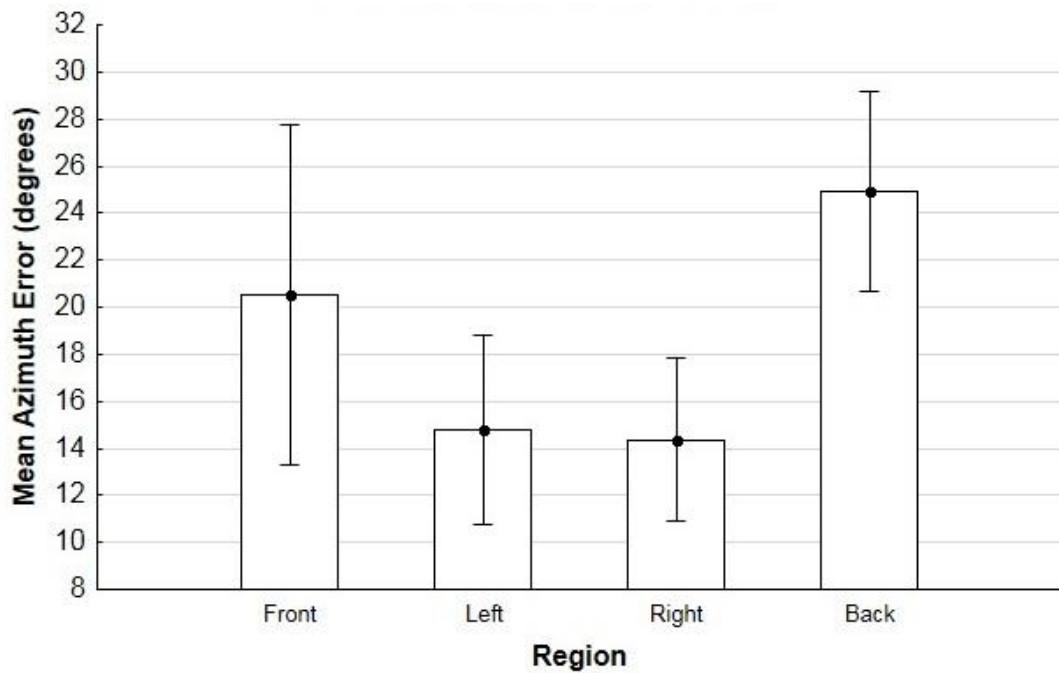


Figure 3-2. Mean azimuth localising error grouped by azimuth region. Error bars denote 0.95 confidence intervals. Significantly degraded performance was observed for the front region compared with the right region. Degraded performance was also observed in the back region compared with both left and right regions.

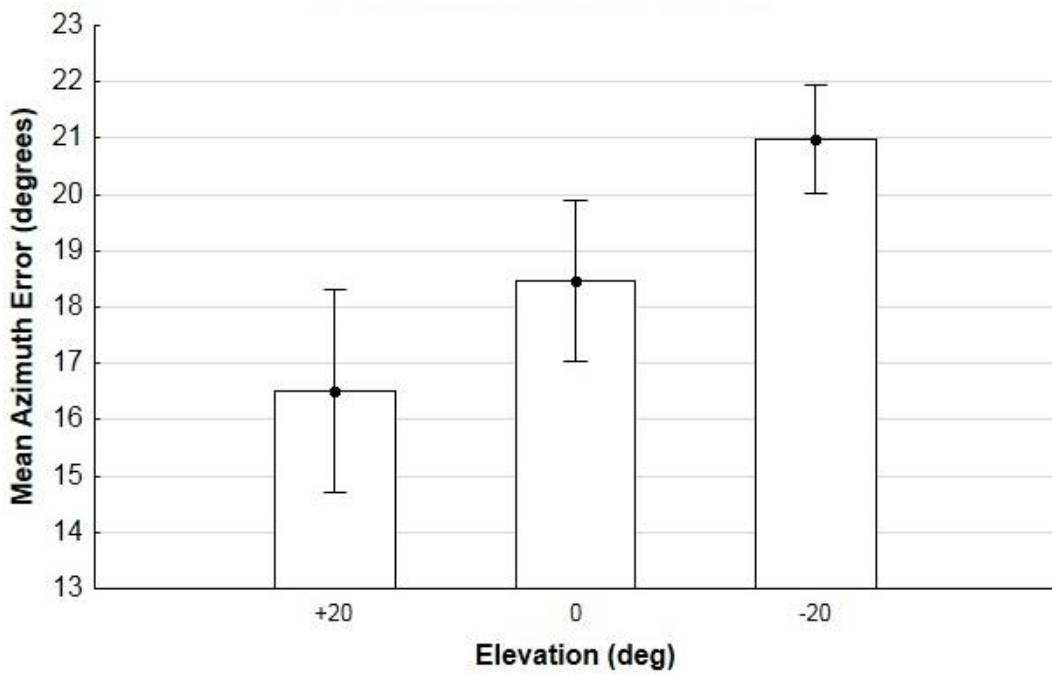


Figure 3-3. Mean azimuth localising error grouped by elevation. Error bars denote 0.95 confidence intervals. Significant effects were observed between all elevation conditions.

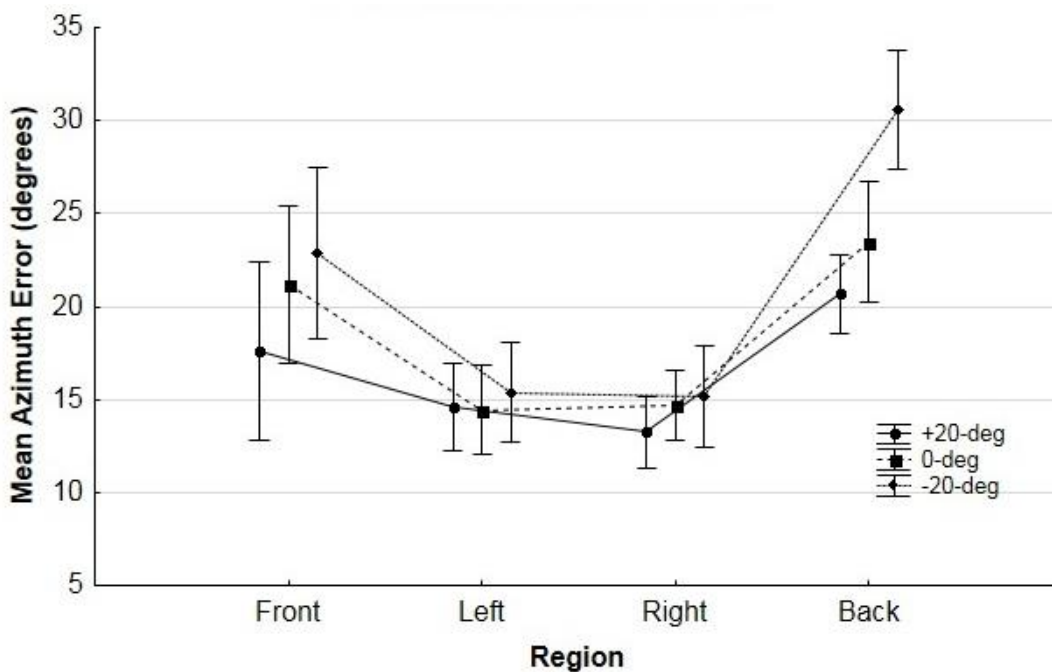


Figure 3-4. Mean azimuth localising error grouped by azimuth region and elevation. Error bars denote 0.95 confidence intervals.

Post-hoc analysis of the elevation main effect revealed that sounds positioned at 20-degrees in elevation produced a small improvement in localising performance compared

with those positioned at 0-degrees ($d = .29$), and an intermediate level of improvement compared with those positioned at -20-degrees elevation ($d = .58$). A small improvement in localising performance was observed between the 0-degree and -20-degree elevation ($p = .31$).

Table 3-1.

ANOVA and post-hoc results for azimuth localising error (degrees) occurring by region and elevation

Condition	RM-ANOVA				Bonferroni				Cohen's d					
	F	df	p	$M (SD)$	95% CI	Left	Right	Back	0	-20	Right	Back	0	-20
Region	11.483	(3, 30)	.000*											
Front				20.6 (9.8)	[13.3, 27.8]	.061	.037*	.281			0.87 _(L)			
Left				14.8 (2.5)	[10.8, 18.8]		1.0	.000*				1.7 _(L)		
Right				14.3 (2.3)	[10.9, 17.8]			.000*				1.8 _(L)		
Back				24.9 (7.8)	[20.6, 29.2]									
Elevation	32.332	(2, 20)	.000*											
20				16.5 (6.4)	[14.7, 18.3]				.007*	.000*			0.29 _(S)	0.58 _(I)
0				18.5 (7.3)	[17.0, 19.9]					.000*				0.31 _(S)
-20				21.0 (8.9)	[20.0, 21.9]									
Region/Elevation	9.164	(6, 60)	.000*											

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Left = 230 to 310, Right = 50 to 130, Back = 140 to 220; (S) = small effect size; (I) = intermediate effect size; (L) = large effect size; * $p < .05$.

Post-hoc and effect size data for the observed region by elevation interaction are provided in Table 3-2. This analysis identified that participants performed significantly better when localising sounds at all elevations within the left and right regions when compared with all other elevations by region. No differences were observed in localising between the left and right regions. Participants localised the front region 20-degree sounds better than both front 0-degree and front -20-degree sounds. The -20-degree back sounds were most problematic to localise, with participants performing significantly worse in this configuration than all other regions and elevations.

Table 3-2.

Post-hoc analysis and effect size results for azimuth localising error (degrees) occurring by region and elevation.

Condition	M (SD)	95% CI	Bonferroni												Cohen's d											
			Front 0	Front -20	Left 20	Left 0	Left -20	Right 20	Right 0	Right -20	Back 20	Back 0	Back -20	Front 0	Front -20	Left 20	Left 0	Left -20	Right 20	Right 0	Right -20	Back 20	Back 0	Back -20		
Front 20	17.6 (9.5)	[12.8, 22.4]	.054*	.000*	.322	.214	1.0	.005*	.483	1.0	.252	.000*	.000*	.38 _(S)	.54 _(I)			.64 _(I)					.70 _(I)	1.6 _(L)		
Front 0	21.2 (9.7)	[17.0, 25.4]		1.000	.000*	.000*	.000*	.000*	.000*	.000*	1.0	1.0	.000*			.93 _(L)	.96 _(L)	.82 _(L)	1.1 _(L)	.92 _(L)	.87 _(L)			1.1 _(L)		
Front -20	22.9 (10.3)	[18.2, 27.5]			.000*	.000*	.000*	.000*	.000*	.000*	1.0	1.0	.000*			1.1 _(L)	1.1 _(L)	1.0 _(L)	1.3 _(L)	1.1 _(L)	1.0 _(L)			.87 _(L)		
Left 20	14.6 (2.5)	[12.2, 16.9]				1.0	1.0	1.0	1.0	1.0	.000*	.000*	.000*										1.3 _(L)	1.7 _(L)	3.0 _(L)	
Left 0	14.4 (2.5)	[12.0, 16.8]					1.0	1.0	1.0	1.0	.000*	.000*	.000*										1.4 _(L)	1.7 _(L)	3.0 _(L)	
Left -20	15.4 (2.6)	[12.7, 18.1]						1.0	1.0	1.0	.000*	.000*	.000*										1.1 _(L)	1.5 _(L)	2.9 _(L)	
Right 20	13.2 (2.3)	[11.3, 15.2]							1.0	1.0	.000*	.000*	.000*										1.6 _(L)	2.0 _(L)	3.3 _(L)	
Right 0	14.7 (2.3)	[12.8, 16.6]								1.0	.000*	.000*	.000*										1.3 _(L)	1.7 _(L)	3.1 _(L)	
Right -20	15.1 (2.0)	[12.4, 17.9]									.000*	.000*	.000*										1.2 _(L)	1.6 _(L)	3.0 _(L)	
Back 20	20.7 (6.1)	[18.6, 22.8]										.546	.000*												1.5 _(L)	
Back 0	23.5 (7.1)	[20.2, 26.7]											.000*												1.0 _(L)	
Back -20	30.6 (7.0)	[27.4, 33.8]																								

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Left = 230 to 310, Right = 50 to 130, Back = 140 to 220; (S) = small effect size; (I) = intermediate effect size; (L) = large effect size; * $p < .05$.

Figure 3-5 compares mean azimuth localisation error for this study against previous work by Wightman & Kistler (1989b) and Wenzel et al. (1993). Similar regional comparisons about the azimuth can be made between studies, however, note that the grouping of elevation regions vary. For this reason, care needs to be taken when drawing conclusions between study comparisons. However, in the absence of relevant data published in the primary literature, this comparison does provide an indication of expected azimuth localisation accuracy when utilising non-individual HRTFs. The comparison reflects a similar trend in localisation performance by region, with the lateral regions being most accurate, then the front, and finally the back producing the worst results.

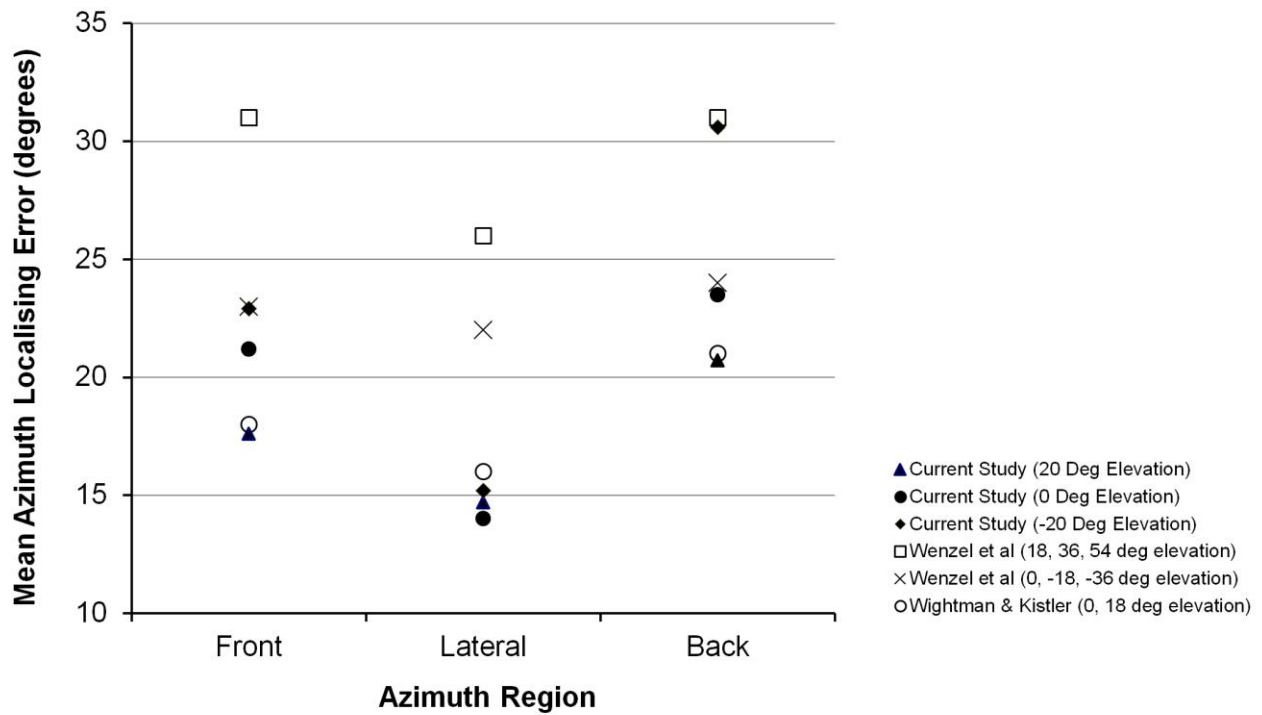


Figure 3-5. Between-study comparison of mean azimuth localisation error. Similar to better localising performance was observed by region for this study when compared with previous studies by Wightman & Kistler (1989b) and Wenzel et al. (1993).

Figure 3-6 shows a sample variability plot for localisation estimates about the azimuth at 20 degrees elevation to illustrate the general trend in data that occurred throughout the study.

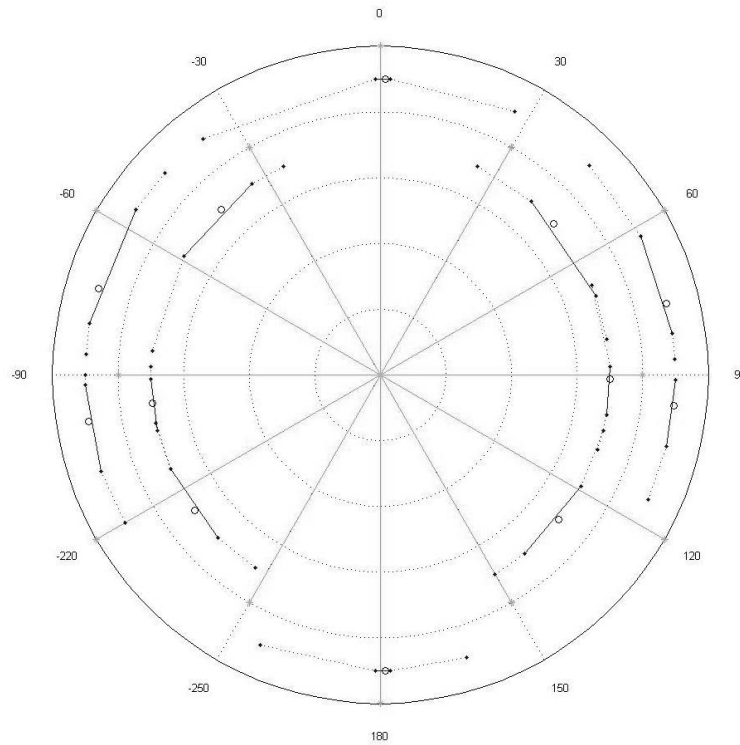


Figure 3-6. Sample localisation variability plot at 20-degree elevation. A sample of 12 plots is displayed. Each plot varies in distance from the centre simply to minimise overlap. They comprise a solid line that spans the quartile range; a circle indicates the mean; while the outer points indicate the 10% and 90% range. Azimuth localisation tends to consistently bias toward the interaural axis on the ipsilateral side of the mid-sagittal plane, which is likely due to the use of a non-individual HRTF.

3.4.3 Front-back Localising Errors

A total of 29% of front hemisphere signals resulted in reversals, which was twice the number observed in the back hemisphere. In general, regional errors were consistently worse toward lower elevations and compare similarly with previous studies. Oldfield & Parker (1984b) found that in the absence of spectral cues, reversals occurred across all azimuth bearings, with the worst being 55% of signals at 0 degrees azimuth. Wenzel et al. (1993) similarly found that 31% of signals initiated reversal errors, 25% of which occurred from the front, and 6% from the rear. Figure 3-7 shows the total front-back errors occurring about the azimuth as plotted by elevation.

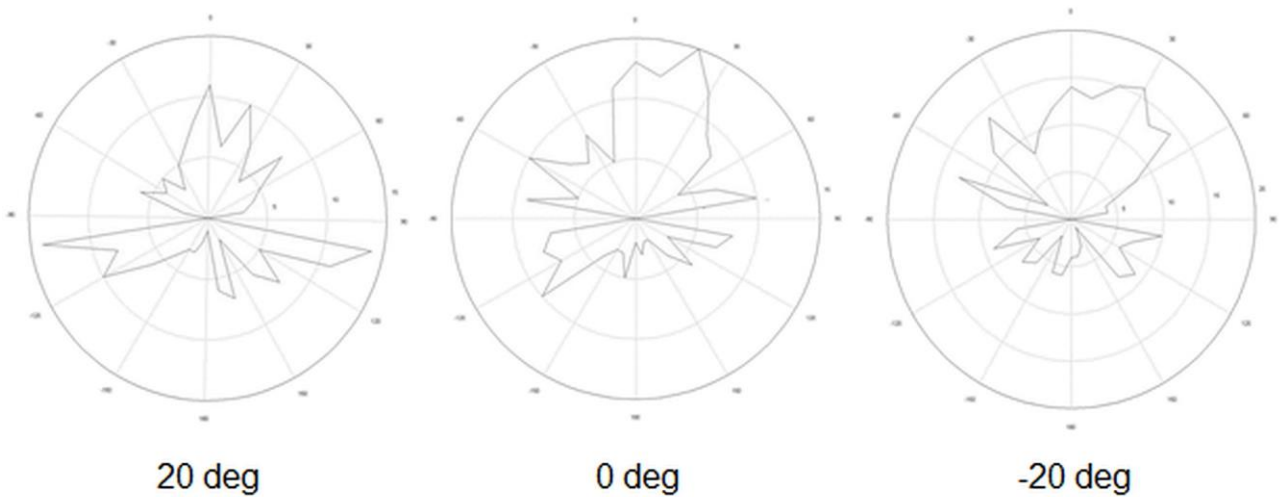


Figure 3-7. Total reversal errors plotted by azimuth bearing and grouped by 20-, 0-, and -20-degrees elevation. The plots show the total number of front-back errors occurring at each 10-degree in azimuth bearing. Each concentric circle moving out from the centre indicate increments of five.

ANOVA conducted on front-back errors revealed main effects for region ($F(3, 30) = 6.095, p = .002$), elevation ($F(2, 20) = 5.560, p = .012$), and an interaction for region by elevation ($F(6, 60) = 4.031, p = .002$). Post-hoc analysis revealed a large increase in front-back errors occurring in the front region compared to the back region. A small difference was found regarding an increased number of front-back errors occurring at -20-degrees elevation compared to 20-degrees elevation. Table 3-3 lists the ANOVA and post-hoc analysis data for front-back errors. Figure 3-8 plots the mean front-back errors by region, while Figure 3-9 plots mean front-back errors by elevation.

Table 3-3.

ANOVA and post-hoc results for mean number of front-back errors occurring by region and elevation.

Condition	RM-ANOVA					Bonferroni					Cohen's <i>d</i>	
	<i>F</i>	<i>df</i>	<i>p</i>	<i>M (SD)</i>	95% CI	Lateral Front	Lateral Back	Back	0	-20	Back	-20
Region	6.094	(3, 30)	.002*									
Front				9.6 (5.0)	[5.0, 14.2]	.116	.121	.001*			1.5 _(L)	
Lateral Front				5.8 (3.5)	[2.8, 8.8]		1.00	.516				
Lateral Back				5.8 (3.2)	[2.9, 8.8]			.496				
Back				3.1 (3.2)	[0.0, 6.7]							
Elevation	5.560	(2, 20)	.012*									
20				5.3 (4.0)	[4.3, 6.3]				.539	.010*		.35 _(S)
0				6.0 (4.4)	[4.8, 7.2]					.204		
-20				7.0 (4.8)	[5.4, 8.6]							
Region/Elevation	4.031	(6, 60)	.002*									

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Lateral Front = 270 to 310 + 50 to 90, Lateral Back = 90 to 130 + 230 to 270, Back = 140 to 220; (S) = small effect size; (L) = large effect size; * $p < .05$.

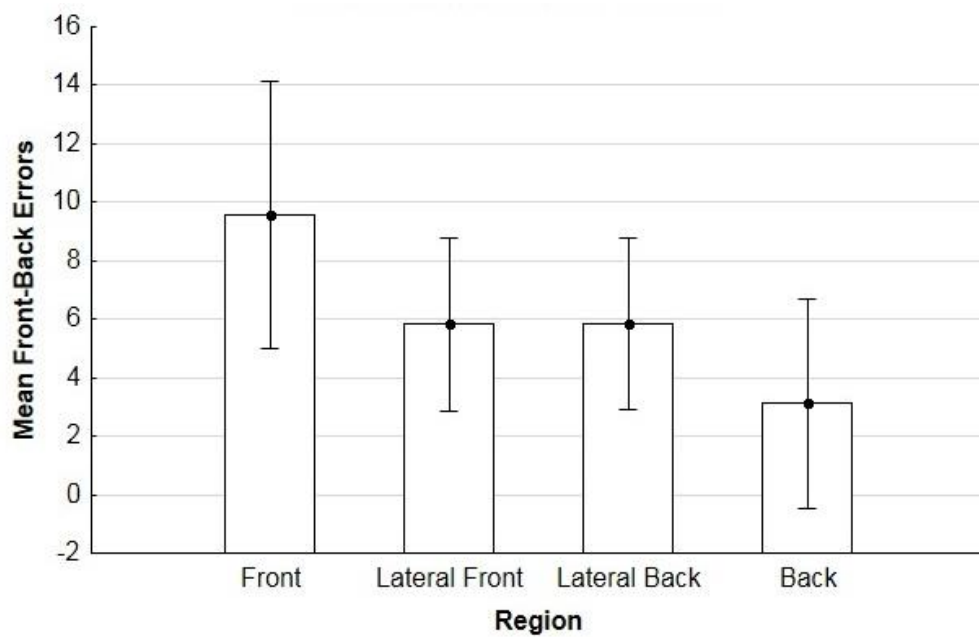


Figure 3-8. Mean front-back localising errors grouped by azimuth region. Error bars denote 0.95 confidence intervals.

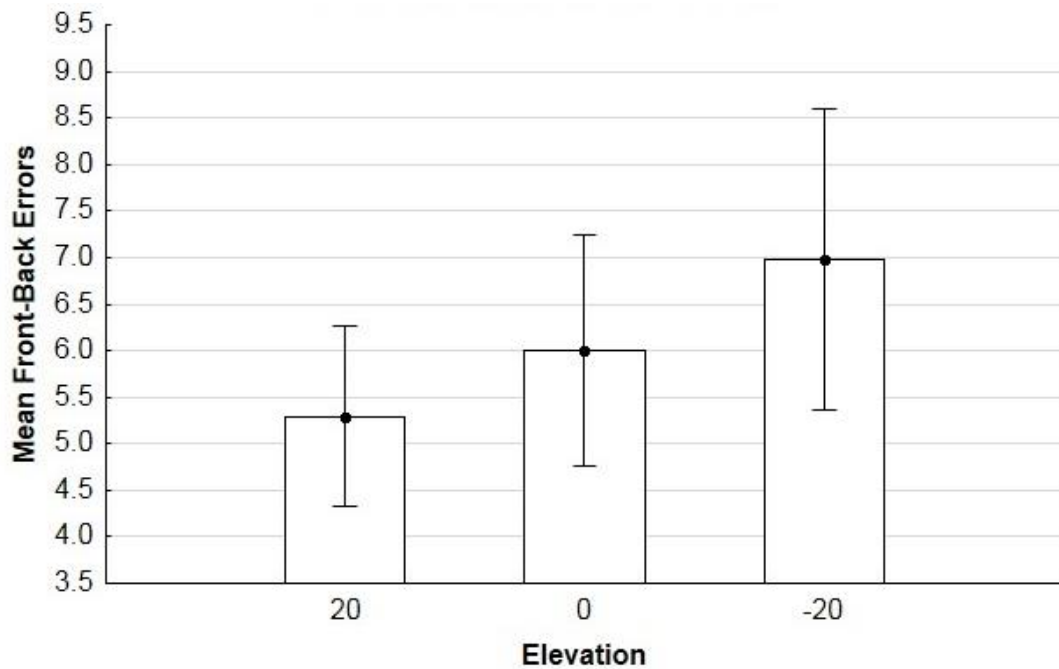


Figure 3-9. Mean front-back localising errors grouped by elevation. Error bars denote 0.95 confidence intervals.

Post-hoc analysis results for the region by elevation interaction are detailed in Table 3-4, with effect sizes listed in Table 3-5. Post-hoc analysis of the region by elevation interaction revealed that sound in the front -20-degree condition produced significantly more front-back errors than all other regions and elevations. Sounds located in the back 0-degree elevation produced significantly less front-back errors than all regions and elevations, except for the 0- and -20-degree sounds in the back region. The back -20-degree condition produced significantly fewer front-back errors than the front 0- and -20-degree configurations, while the back 20-degree configuration produced fewer front-back errors than all front elevations. Figure 3-10 plots the mean azimuth localising error grouped by region and elevation.

Table 3-4.

Post-hoc analysis results for mean number of front-back errors occurring by region and elevation.

Condition	M (SD)	95% CI	Bonferroni											
			Front 0	Fornt -20	Lateral Front 20	Lateral Front 0	Lateral Front -20	Laterral Back 20	Laterral Back 0	Laterral Back -20	Back 20	Back 0	Back -20	
Front 20	6.9 (4.7)	[3.8, 10.1]	.443	.001*	.874	1.00	1.00	1.00	1.00	1.00	1.00	.046*	.009*	.275
Front 0	9.9 (4.4)	[7.0, 12.8]		1.00	.000*	.131	.350	.558	.002*	0.007	.000*	.000*	.000*	.000*
Fornt -20	11.9 (4.9)	[8.6, 15.2]			.000*	.000*	.001*	.001*	.000*	.000*	.000*	.000*	.000*	.000*
Lateral Front 20	4.2 (2.8)	[2.3, 6.1]				1.00	1.00	.700	1.00	1.00	1.00	1.00	1.00	1.00
Lateral Front 0	6.5 (3.2)	[4.3, 8.6]					1.00	1.00	1.00	1.00	1.00	.169	.035*	.874
Lateral Front -20	6.8 (4.2)	[4.0, 9.6]						1.00	1.00	1.00	1.00	.060	.012*	.350
Lateral Back 20	7.0 (3.7)	[4.5, 9.5]							1.00	1.00	1.00	.035*	.007*	.216
Lateral Back 0	5.1 (3.3)	[2.9, 7.3]									1.00	1.00	1.00	1.00
Lateral Back -20	5.5 (2.5)	[3.8, 7.1]										1.00	.558	1.00
Back 20	3.1 (3.4)	[0.8, 5.3]											1.00	1.00
Back 0	2.5 (3.4)	[0.23, 4.9]												1.00
Back -20	3.7 (3.0)	[1.7, 5.7]												

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Lateral Front = 270 to 310 + 50 to 90, Lateral Back = 90 to 130 + 230 to 270, Back = 140 to 220; * $p < .05$.

Table 3-5.

Effect size results for mean number of front-back errors occurring by region and elevation.

Condition	M (SD)	95% CI	Cohen's d										
			Front -20	Lateral Front 20	Lateral Front 0	Lateral Front -20	Laterral Back 20	Laterral Back 0	Laterral Back -20	Back 20	Back 0	Back -20	
Front 20	6.9 (4.7)	[3.8, 10.1]	1.0 _(L)								.93 _(L)	1.1 _(L)	
Front 0	9.9 (4.4)	[7.0, 12.8]		1.5 _(L)					1.2 _(L)		1.7 _(L)	1.8 _(L)	1.6 _(L)
Formt -20	11.9 (4.9)	[8.6, 15.2]		1.9 _(L)	1.3 _(L)	1.1 _(L)	1.1 _(L)	1.6 _(L)	1.6 _(L)		2.1 _(L)	2.2 _(L)	2.0 _(L)
Lateral Front 0	6.5 (3.2)	[4.3, 8.6]										1.2 _(L)	
Lateral Front -20	6.8 (4.2)	[4.0, 9.6]										1.1 _(L)	
Lateral Back 20	7.0 (3.7)	[4.5, 9.5]									1.1 _(L)	1.3 _(L)	

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Lateral Front = 270 to 310 + 50 to 90, Lateral Back = 90 to 130 + 230 to 270, Back = 140 to 220; (L) = large effect size.

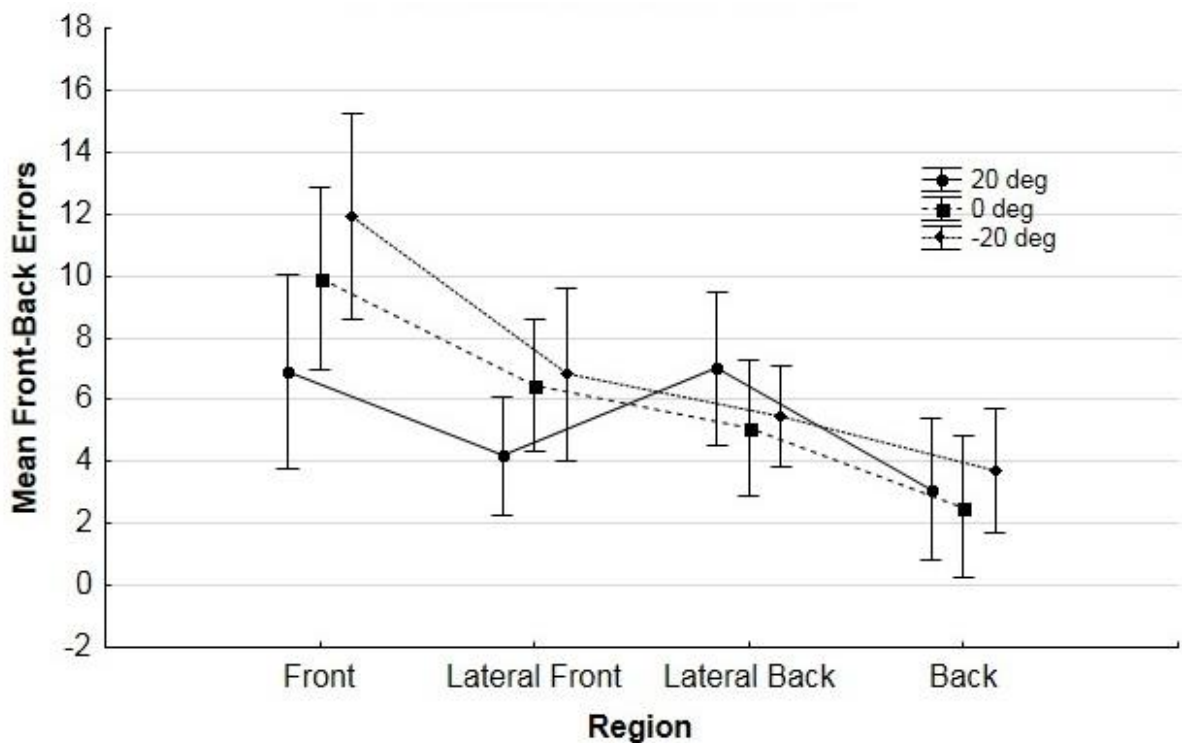


Figure 3-10. Mean front-back localising errors grouped by azimuth region and elevation. Error bars denote 0.95 confidence intervals.

3.4.4 Confidence

ANOVA conducted on confidence estimates for localising performance indicated a main effect for region ($F(3, 30) = 3.526, p = .027$). No main effects were observed for confidence estimates by elevation ($F(2, 20) = 1.167, p = .332$), or region by elevation ($F(6, 60) = .375, p = .892$). Post-hoc analysis found that confidence was significantly elevated for localising sound in the lateral back region compared with the front. No differences were found in confidence between any other region or elevation. Table 3-6 provides the data associated with the analysis undertaken for confidence estimates by region and elevation. Figure 3-11 plots the mean confidence estimates by region.

Table 3-6.

Summary analysis results for mean azimuth localising confidence estimates.

Condition	RM-ANOVA					Bonferroni			<i>d</i>
	<i>F</i>	<i>df</i>	<i>p</i>	<i>M (SD)</i>	95% CI	Lateral Front	Lateral Back	Back	Lateral Back
Region	3.526	(3, 30)	.027*						
Front				7.1 (.94)	[6.5, 7.7]	1.0	.020*	1.0	.72 _(L)
Lateral Front				7.4 (.64)	[6.9, 7.8]		.391	1.0	
Lateral Back				7.8 (1.0)	[7.4, 8.3]			.239	
Back				7.3 (1.2)	[6.6, 8.0]				
Elevation	1.167	(2, 20)	.332						
Region/Elevation	.375	(6, 60)	.892						

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Left = 230 to 310, Right = 50 to 130, Back = 140 to 220; (S) = small effect size; (I) = intermediate effect size; (L) = large effect size; Confidence scale from 0 (low) to 10 (high);

* $p < .05$.

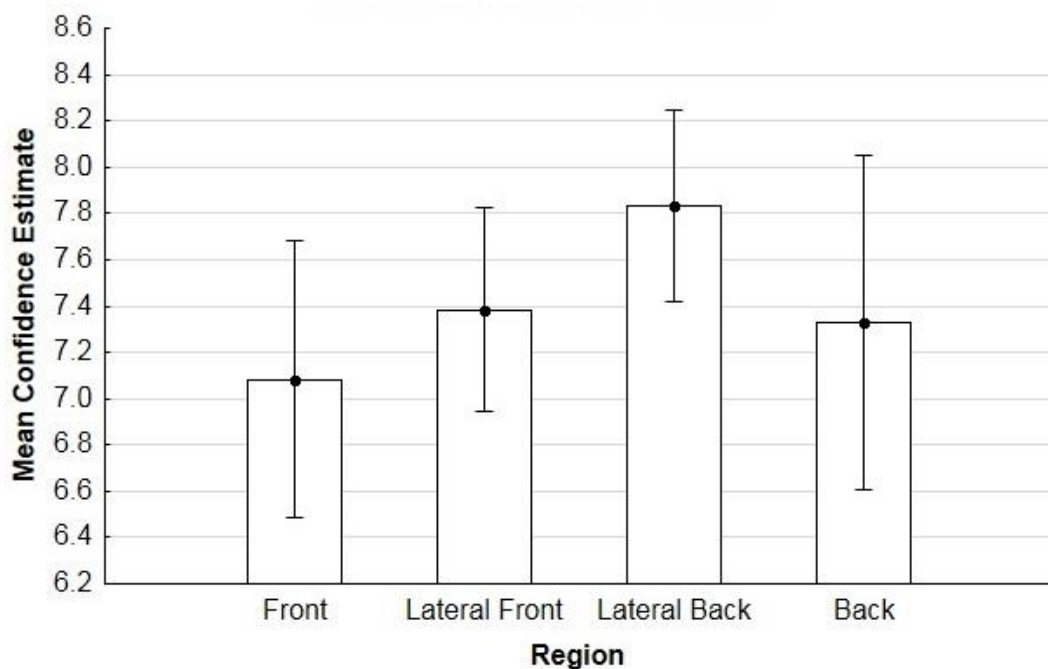


Figure 3-11. Mean confidence in localising estimates grouped by region. Error bars denote 0.95 confidence intervals. Confidence scale from 0 (low) to 10 (high). Significant effects were observed between the front and lateral back regions.

3.5 Discussion

A consistent horizontal orientation was maintained between the audio stimulus and response mode in order to minimize possible transposing errors through mismatches in polar axis. The dexterity task results are not intended to infer the absence of such errors, which should possibly be considered more objectively in subsequent studies.

The observed localising bias is likely accounted for by the use of an upper percentile interaural cue, which could skew the sound's perceived location toward the interaural axis. This phenomenon is discussed in more detail in section 2.4.2. The non-individual HRTF used in this study is an upper percentile head dimension and therefore considered likely to introduce such bias (Miller, J. D., Godfroy-Cooper, & Wenzel, 2014). When non-individual HRTF spectral cues are poorly matched to an individual, the listener is likely to have poorer spatial depth perception than normal. This is called internalisation and may also contribute toward the observed degraded localising performance. This phenomenon is also described further in section 2.4.1.

The finding that differences in confidence by region did not compare well with localising accuracy might be due to depth cues being reportedly more pronounced toward the back. This may be a significant percept that provides a false sense of familiarity with localising cues and subsequently increases perceived accuracy and confidence for the

back region compared to the front region. If this is true, it appears that participants generally have a poor awareness of the degree to which their azimuth localising accuracy has degraded due to mismatches between non-individual HRTF localising cues and their own. This finding supports a need for more research into techniques that provide the listener with performance feedback that calibrates localising confidence. Auditory displays that map system information to a sound's spatial location may require such feedback in order to mitigate the risk of applying misinterpreted information.

Figure 3-7 illustrates the slight asymmetric bias in front-back errors toward the front-right quadrant across all elevations. While this finding is most likely attributed to poorly matched localising cues within the non-individualised HRTF, the skew may also be potentially attributed to a general cognitive predisposition for superior spatial processing ability to occur within the right cerebral hemisphere, thereby producing better audio localising performance for sound originating from the left. Previous studies support this phenomenon and find it to be particularly evident when discriminating between spectral patterns, as occurs when determining the front-back origin of a sound (Abel & Paik, 2004; Butler, 1994; Sharon, Christian, Angela, & Blake, 2000). However, the finding that there was only a difference of one front-back error occurring between the lateral back regions does not readily support this inference.

Wenzel et al. (1993) found that non-individual HRTFs only slightly decreased localising performance for individuals as long as they themselves were good localisers. It is considered appropriate to use a HRTF database that has been optimised to match a broad sample of the intended user population for any future applied use. The current default HRFT is considered suitable for use in future experiments that focus on improving localising performance for non-individualised HRTF audio.

Several experiments support the idea that localisation cues can be learnt. Abel & Paik (2004) claim that learning through the encoding of spectral cues has provided increases in performance for azimuth localising. Previously, Oldfield & Parker (1984b) determined that elevation and azimuth are coded independently and that localising in elevation is independent of azimuth accuracy. Current research disputes this, with claims of up to 10% increases in azimuth performance being attributed to spectral cues (Qian & Eddins, 2008; Razavi, O'Neill, & Paige, 2005). Previous research into synthesised cues may not have observed the influence of spectral cues on improvements in azimuth localising because an inadequate amount of prior exposure to those cues may have failed to facilitate learning.

Razavi, O'Neill, & Paige (2005) support the claim that robust spectral cues, when used in conjunction with interaural cues, significantly improve localisation accuracy. Honda et al. (2007) indicate performance increases of 20% in localisation ability through perceptual-motor learning as a direct result of playing a virtual audio display game over a two week period. They found little difference in performance between individual and non-individual HRTFs after learning occurred, with the increase in ability persisting for more than one month after the practice sessions. If spectral cues can be learnt and subsequently improve azimuth localising, the current method whereby reversals are corrected prior to analysing azimuth localising performance needs to be reconsidered. This approach relies on spectral cues having negligible influence on azimuth localising performance. The related influence of training needs to be further examined and incorporated into subsequent experiments.

The current experiment constrained head movement during sound presentation. Turning the head allows a free field listener to alter the relative position of the sound so that more discernible localisation cues may be utilised. Muller & Bovet (1999) found that head movements provide a 10% increase in azimuth localisation ability. Subsequent experiments support the claim, with head movements being shown to reduce front-back errors through variation of localising cues (Iwaya et al., 2003). Kudo, Higuchi, Hokari, & Shimada (2006) optimised spatial audio localisation for systems without head tracking ability through the use of what they termed a 'swing sound image', where the horizontal positioning of a sound source was varied between five degrees either side of the intended position, thereby also considerably reducing front-back errors. They believe that establishing a dynamic localisation cue is important when using non-individualised HRTFs. There are possibly more effective dynamic localising cue techniques that are intended to be explored in future studies, such as the incorporation of an additional transient reference signal.

3.6 Conclusion

Similar to previous experiments, azimuth localising accuracy was found to be best in the lateral regions, with the back region producing worst performance. Sounds positioned at 20-degrees elevation were localised more accurately in azimuth than at 0-degrees elevation, with -20-degrees elevation producing the worst azimuth localising accuracy. Front-back errors were significantly worse in the front region compared with the back. The back 0-degree condition produced significantly less front-back errors than all other regions and elevations. The -20-degree condition produced the worst number of front-back errors

of any region or elevation. Poorly generalised interaural cues likely caused localisation estimates to skew toward the interaural axis on the ipsilateral side of the mid-sagittal plane. Poorly generalised distance cues may have inappropriately increased confidence in the back region, which requires further attention in future studies. It is considered appropriate to incorporate some form of training into future experiments as learning has been found to improve recognition of non-individual localising cues and improve localising performance. Comparisons between localising data from the current study and previous studies support the use of SLAB3D and its default HRTF in future experiments.

4. Study 2: Improving 3-D Audio Localisation through the Provision of Supplementary Spatial Audio Cues.

Towers, J., Burgess-Limerick, R., & Riek, S. (2012). Improving 3-D audio localisation through the provision of supplementary spatial audio cues. *The Ergonomics Open Journal*, 5(1-9).

A number of improvements have been made to this manuscript since its first publication, the most considerable of which is to the front-back error analysis methodology outlined in section 4-4-2. The original analysis involved grouping front-back localising errors by front, left, right, and back regions. The original paper also reported an additional analysis with front-back errors grouped by lateral front and lateral back regions, which identified a main effect. Grouping by lateral front and lateral back regions is a more appropriate way to analyse the data given that front-back errors are relative to the midsagittal plane, which is where the lateral front and lateral back regions border. The left and right regions span the interaural axis and therefore fail to differentiate front-back errors occurring within the lateral front and back hemispheres. The current revision includes ANOVA conducted on the six regions, which allows for post-hoc analysis to be performed more appropriately on main effects. Additional analysis has also been undertaken on the confidence estimates reported in section 4-4-4.

4.1 Abstract

This study examined whether azimuth localising performance for non-individualised spatial audio without integrated head tracking can be improved through the provision of supplementary reference sounds. Twenty-two participants attempted to localise spatial target sounds developed through a non-individualised HRTF while performing a visual distractor task. Target sounds were randomly presented at 0-degrees elevation for each 10-degree increment about the azimuth. Three audio conditions were tested. A control *stable* condition utilised a stationary target sound. A *swing* condition utilised a target sound that oscillated in azimuth to diversify localising cues. A final *sweep* condition introduced supplementary cues in the form of stationary and transient sounds that were intended to provide target sound distance cues toward the midsagittal plane and interaural axis. The sweep condition 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 integrated into the target sound to lessen distraction and disorientation.

4.2 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 (Middlebrooks & Green, 1991). 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° and 170°), 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 altered upon impact with the physical shape of the pinnae, head, and upper torso (Yost, 2007). 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 signals recorded at the eardrum or ear canal of an individual situated within an anechoic chamber (Wightman & Kistler, 1989a).

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 HRTF 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 (Moller, Sorensen, Jensen, & Hammershoi, 1996; Wenzel et al., 1993; Wightman & Kistler, 1989b).

The absence of integrated head tracking within 3-D auditory displays has been well documented as providing poor externalisation of sound and degraded localising performance (Iwaya et al., 2003; Wightman & Kistler, 1999). 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 into design methods that mitigate localising errors known to exist in the absence of head tracking.

The primary objective of this study was to determine whether the sound localising performance for participants utilising 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 discernible (Oldfield, S. R. & Parker, 1984a). Wightman & Kistler (1999) suggest that dynamic cues only reduce front-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 the front and back hemispheres, such dynamic cues can be used to provide an unambiguous indicator for localising a sound's front-back hemisphere. 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 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 auditory 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.

4.3 Materials and Methodology

4.3.1 Apparatus

Spatially positioned 220 Hz square wave target sounds were generated and pre-recorded as audio files by the SLAB3D audio rendering tool with its default non-individualised HRTF (Miller, J. D. & Wenzel, 2002). The sounds were virtually positioned for localising 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 recorded duration of 2-seconds for each sound. Figure 4-1 provides a pictorial representation of the three different sound conditions used in the study. The *Stable* condition provided a baseline and consisted simply of a stationary target sound. The *Swing* condition, previously designed by Kudo et al. (2006), oscillated the target sound's position about four-degrees in azimuth either side of the intended bearing.

The *sweep* condition comprised a 100 Hz square wave sweep sound that would originate from the front and transit 90 deg/s in a 0.5 m diameter arc about the participant; alternating its direction from front (0-degrees) to back (180-degrees) for the duration of each trial. The sweep sound made four 180-degree sweeps in total per trial. Each time the sweep sound passed through the midsagittal plane and interaural axis, supplementary 100-msec square wave *accent* sounds would be initiated as accents to the sweep sound in an attempt to reinforce the relative location of those azimuth bearings. The accent sounds varied by 1 Hz, with a 103 Hz sound presented at 0°; a 102 Hz sound at 90°; and a 101 Hz sound at 180°. A target sound was activated once per trial as the sweep sound passed through its location. To allow the participant adequate time to orient herself with the sweep sound prior to presentation of the target sound, different onset timings for target sounds were introduced depending on their front-back hemisphere location. Target sounds located in the front hemisphere were initiated during the second sweep, while sounds located in the back hemisphere were initiated during the third sweep. In total, each sweep condition trial lasted 8-seconds.

The aim of the sweep condition was to improve accuracy in target sound localising through the provision of supplementary sweep and accent sounds. The introduction of a sweep sound was intended to provide cues that improve relative bearing judgements between the target sound and accent sounds located at the mid-sagittal plane and interaural axis. Improvements were considered likely because of distance cues that could be derived from the time taken for the sweep sound to transit from a target sound to its adjacent accent sounds. Accent sounds were designed with frequencies limited 1 Hz to 3

Hz higher than the sweep sound to encourage them as being perceived as being grouped with the sweep sound. Potential interference between the sweep and target sounds was expected to be mitigated through sound design differences in frequency and motion. The sweep sound always originated from the front, thereby providing an established known design reference to resolve any potential front-back ambiguity that may be present due to the use of a non-individual HRTF.

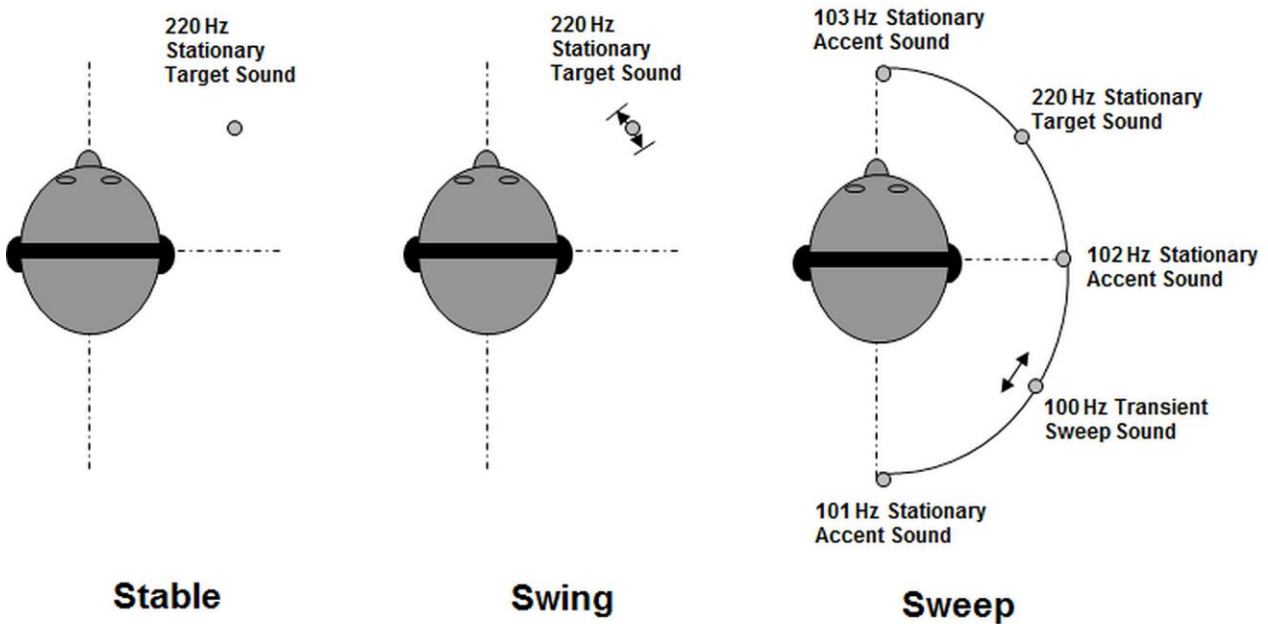


Figure 4-1. Configuration of sound conditions. The three different sound conditions used in the study are shown. The stable condition comprised a stationary target sound. The target sound in the swing condition oscillated four degrees in azimuth either side of the intended position. The swing condition included a stationary target sound and a supplementary sound that transitioned back and forward in a 180-degree arc initiating the onset of supplementary accent sounds at the interaural axis and midsagittal plane as it passed through those spatial positions.

The experiment employed a spatial distractor task in the form of a 2-D flight simulator, shown in Figure 4-2. Participants were required to position the earth within a square alignment region displayed in the centre 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. Aside from changing the starting position of the earth at the onset of each trial, there were no other factors designed to influence task difficulty. The task was considered easy, but did require allocated attention throughout each localising trial to achieve the task in the allocated time. 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 distractor activity. A Hewlett-Packard xw6200 desktop computer with Microsoft XP and Windows Media Player 11 ran the simulator and audio presentation, with audio delivered through Bose® TriPort binaural headphones.

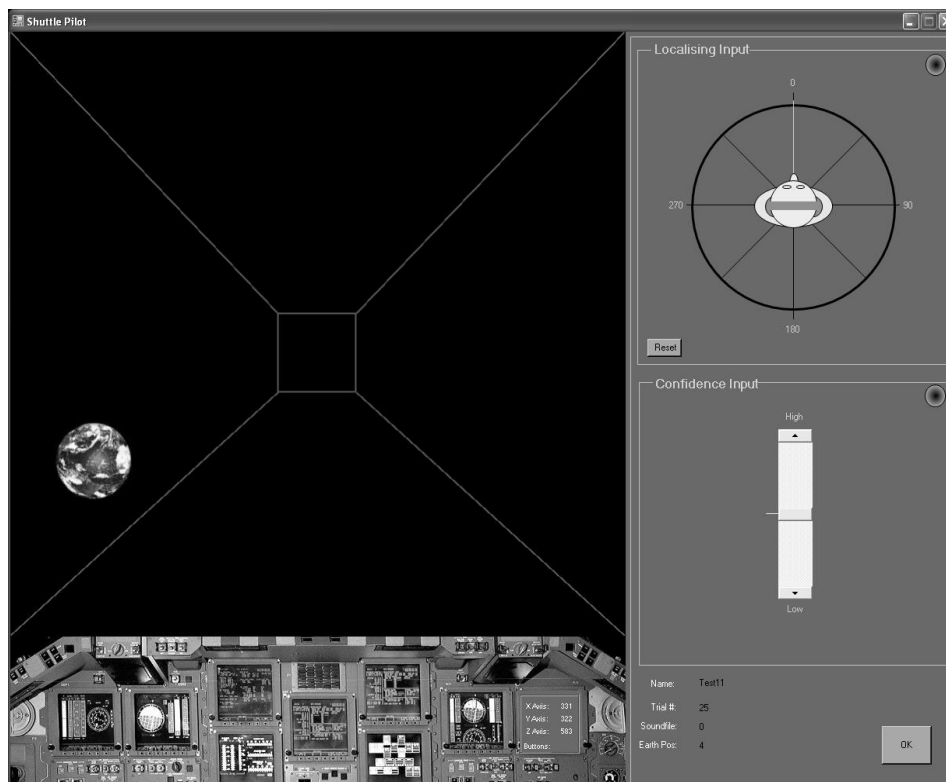


Figure 4-2. Study interface. Shown is a screenshot from the graphical user interface. The left of screen shows the secondary task interface where participants guided the spaceship toward the Earth. The upper right of the figure shows where one group input sound bearing localising estimates. Below right is a slider bar used to input confidence estimates.

The input method for registering audio localising estimates and associated confidence differed between two groups in an attempt to explore transposing effects due to differing orientations between the horizontal auditory display and vertical visual display. 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 Figure 4-2. The localising interface included a circle with consecutive lines indicating 45° increments in bearing from the midsagittal plane, with each 90° 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 displayed on the same computer monitor for both groups.

A dexterity study was previously undertaken by Towers (2008) 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-degree for participants' accuracy when touching predetermined points about the circumference of an overlaid circle.

4.3.2 Participants

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

4.3.3 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. Upon successful completion of the hearing test, participants spent a few minutes listening to a 100 Hz sound that transited about their head at a distance of 1.5 meters in a clockwise direction. This was undertaken to provide a demonstration of internalisation, which is described further in section 2-4-1. Participants were then played example presentations of each audio condition as the design features were explained using Figure 4-1. Target sounds were presented at different predetermined azimuth bearings so the participant understood the intended bearing and characterise any mismatch in the non-individual

localising cues. This demonstration took approximately 20-minutes and included 5-minutes familiarisation with the flight simulator.

Phases two and three were data collection phases, where participants attempted to localise the point of origin in 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 one 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 a message on the screen requested that the participant input localising and confidence estimates in the method determined by their group allocation.

Participants performed the two data collection phases over two non-consecutive 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. Each stimulus was presented twice, 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. Practice effects were assumed to be averaged out due to the randomised presentation of localisation trials across sound conditions and repetitions. To balance any potential spatial bias 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 within those conditions.

4.3.4 Data Analysis

All data analysis were conducted using a significance level identified at $\alpha = .05$. The approach taken was a one way RM-ANOVA (three level). Subsequent post-hoc analyses were undertaken for all significant main effects using Bonferroni adjusted alpha levels. Effect sizes were expressed using Cohen's *d*. Within-subject confidence intervals were constructed by eliminating intersubject variance using the method outlined by Loftus and

Masson (1994). ANOVA was conducted on front-back error, localising error, and confidence estimates.

4.4 Results

4.4.1 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 (1999). 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 Figure 4-3. 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. This was dismissed due to the small difference and it being the only finding.

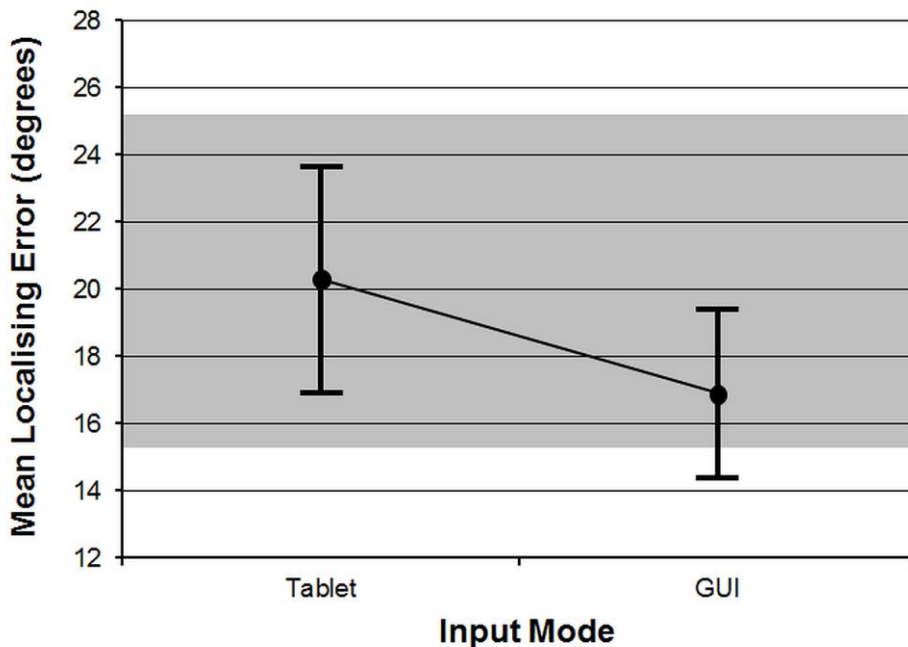


Figure 4-3. Input equivalency results for front stable condition. This figure is an example from the equivalency testing that was conducted to compare if the data from the graphics tablet and GUI input modes could be considered as practically equivalent. Error bars denote 0.90 confidence intervals.

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.

4.4.2 Front-back Errors

Front-back errors occur when a sound is localised toward the incorrect hemisphere with respect to the interaural axis. These errors often occur due to poorly matched non-individual spectral cues not providing adequate discriminating features between points where IID and ITD are similar, especially about the midsagittal plane and cones of confusion. During the debrief participants reported experiencing a varying degree of *internalisation* (Yost, 2007). 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 localise a sound toward the front or back hemisphere. 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. Section 2.4.1 within chapter two of this thesis provides further detail on internalisation.

Figure 4-4 plots total front-back errors occurring across each 10-degree 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.

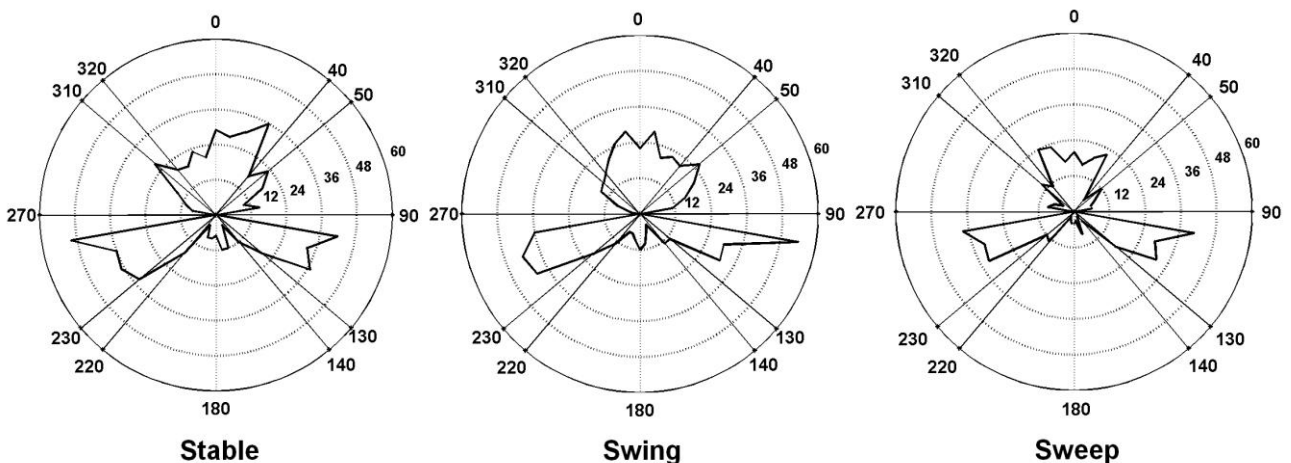


Figure 4-4. Total front-back localising errors grouped by sound condition. This figure plots the total number of front-back errors occurring at each azimuth bearing as grouped by the three sound conditions.

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, lateral front, lateral back, left, right, and back regions. Table 4-1 provides a summary of results for the statistical analysis conducted on front-back errors.

Table 4-1.

Summary statistics for mean front-back errors by sound condition.

Condition	RM-ANOVA					Sound			Region								
	F	df	p	M (SD)	95% CI	Bonferroni		d	Bonferroni			d					
						Swing	Sweep	Sweep	Lateral Front	Lateral Back	Left	Right	Back	Lateral Front	Lateral Back	Back	
Sound	17.474	(2, 42)	.000*														
Stable				9.3 (6.9)	[7.9, 10.7]	.305	.000*	0.42 _(S)									
Swing				8.4 (6.2)	[7.2, 9.7]		.000*	0.30 _(S)									
Sweep				6.5 (6.3)	[4.4, 8.2]												
Region	8.485	(5, 105)	.000*														
Front				10.6 (9.3)	[5.6, 15.7]				.000*	1.00	1.00	1.00	.001*	0.82 _(L)			0.82 _(L)
Lateral Front				4.6 (4.5)	[1.7, 7.6]					.000*	.161	.161	1.00			1.14 _(L)	
Lateral Back				11.8 (7.7)	[7.8, 15.2]						.300	.300	.000*				1.15 _(L)
Left				8.2 (4.1)	[7.0, 9.5]							1.00	.242				
Right				8.2 (4.9)	[6.9, 9.6]								.242				
Back				4.8 (3.8)	[1.6, 8.1]												
Region/Sound	0.456	(10, 210)	0.917														

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Lateral Front = 270 to 310 + 50 to 90, Lateral Back = 90 to 130 + 230 to 270, Left = 320 to 40, Right = 50 – 130, Back = 140 to 220; (S) = small effect size; (L) = large effect size; * $p < .05$.

A significant main effect was observed for front-back errors occurring by region $F(5, 105) = 8.485, p = .000$. Figure 4-5 plots mean front-back errors grouped by region.

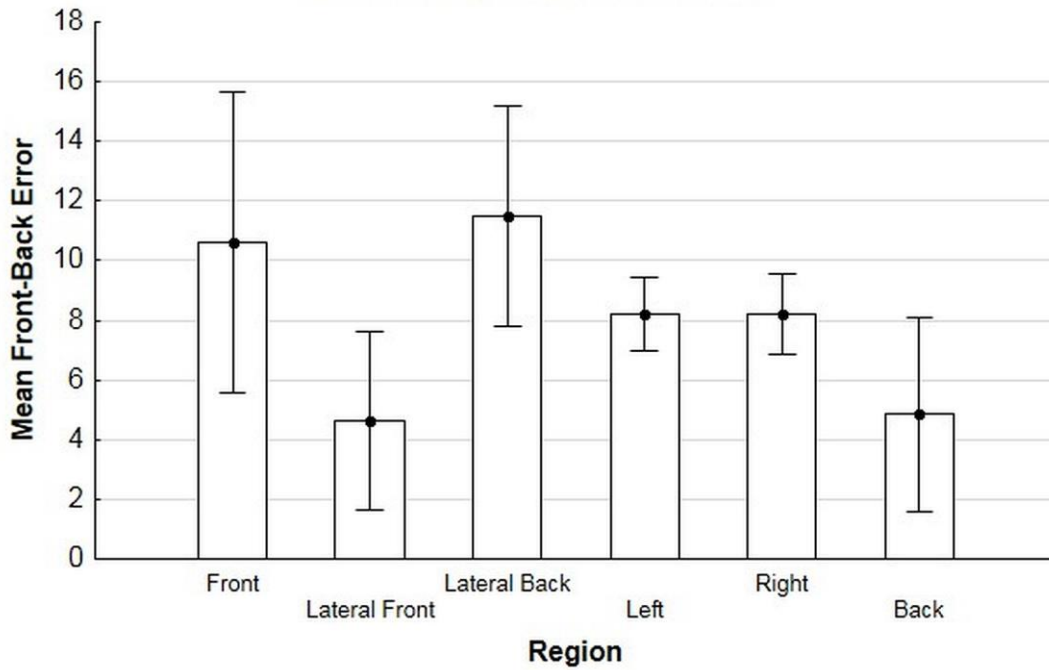


Figure 4-5. Mean front-back localising errors grouped by region. Post-hoc analysis indicated a large difference in the number of errors in the front compared with both the back ($p = .001$, $d = .82$) and lateral front ($p = .000$, $d = .82$) regions. The lateral back region was found to have a large difference in the number of errors occurring in that region when compared with both the lateral front ($p = .000$, $d = 1.14$) and back ($p = .000$, $d = 1.15$) regions.

A significant main effect was found for front-back errors occurring by sound condition $F(2, 42) = 17.474$, $p = .000$. Figure 4-6 below plots the mean front-back errors occurring by sound condition.

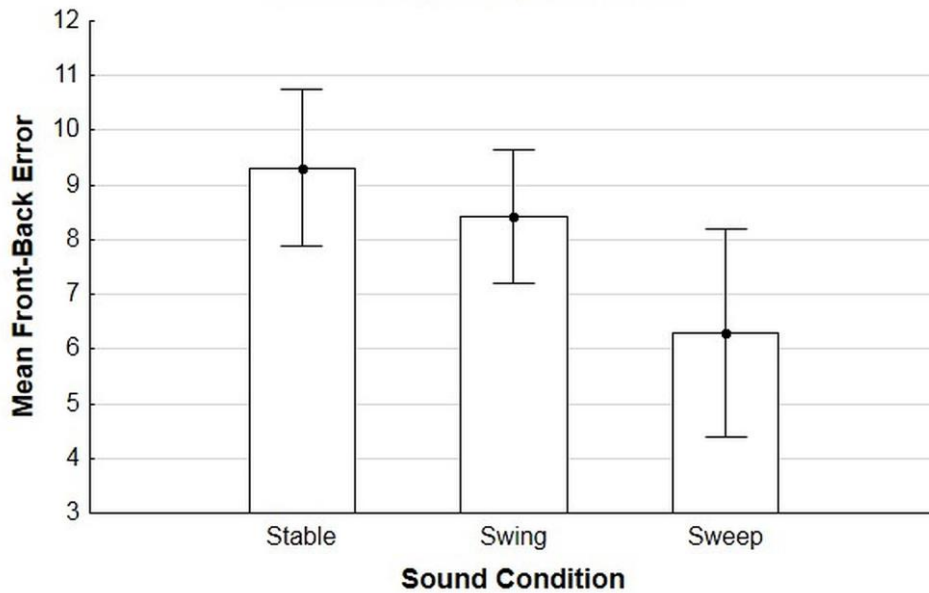


Figure 4-6. Mean front-back localising errors grouped by sound condition. Post-hoc analysis for front-back errors occurring by sound condition revealed a significantly small increase in errors for both the stable condition ($p = .0000$, $d = .42$) and swing condition ($p = .0000$, $d = .30$) when compared with the sweep condition

A front-back localising error interaction was not found for sound condition by region $F(10, 210) = .456$, $p = .916$. Figure 4-7 plots the mean front-back localising errors by region.

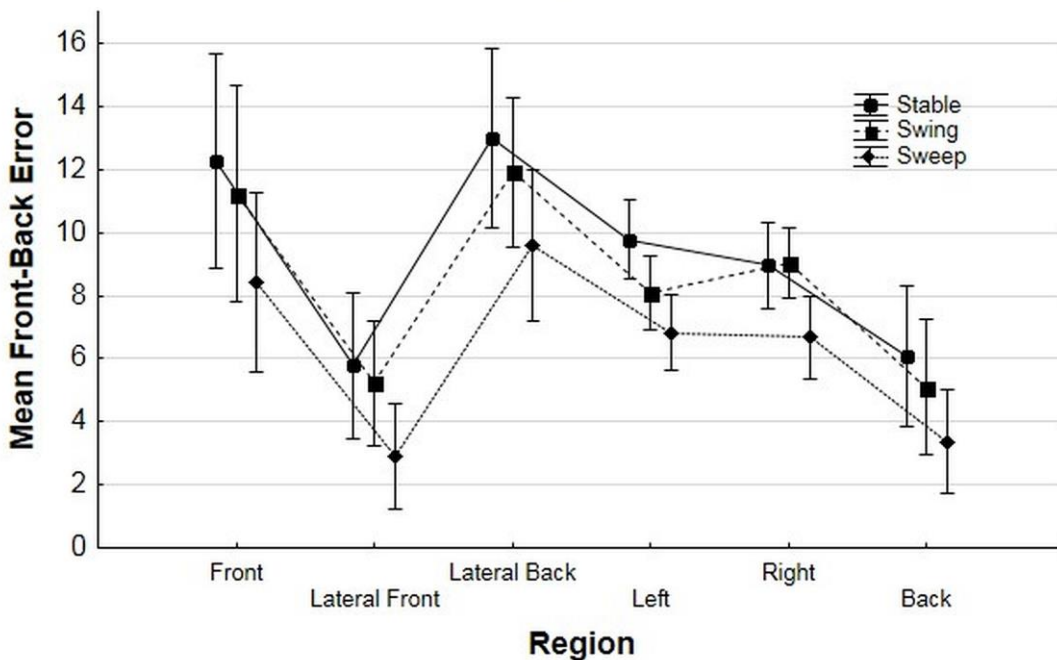


Figure 4-7. Mean front-back localising errors for sound conditions grouped by region. No significant interaction was observed. Error bars denote 0.95 confidence intervals.

4.4.3 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° (Oldfield, S. R. & Parker, 1984a; Wenzel et al., 1993; Wightman & Kistler, 1989b).

Figure 4-8 shows the mean localising error for sound condition 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.

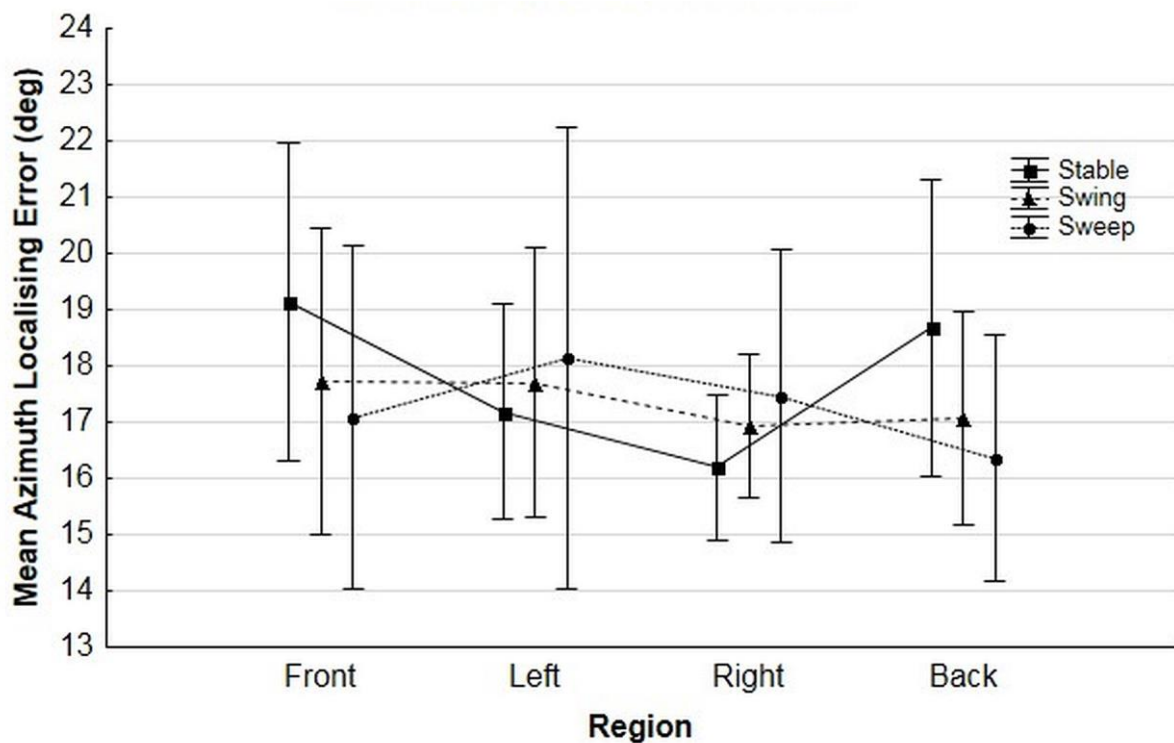


Figure 4-8. Mean localising error for sound grouped by region. This figure shows the mean azimuth localising error by region for each of the different sound conditions. Error bars denote 0.95 confidence intervals.

Figure 4-9 shows the spread in localising estimates for azimuth bearings as grouped by sound condition. A sample of 12 of the 36 possible target locations are shown in each of the three figures. Plots within each figure are staggered in distance from the centre simply to minimize overlap between adjacent data. Arrows extending toward each of the sound locations from the centre of each figure and stop to indicate the associated

plot for that bearing. Each plot comprises a solid line that spans the quartile range; a hollow circle to indicate the mean; and 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.

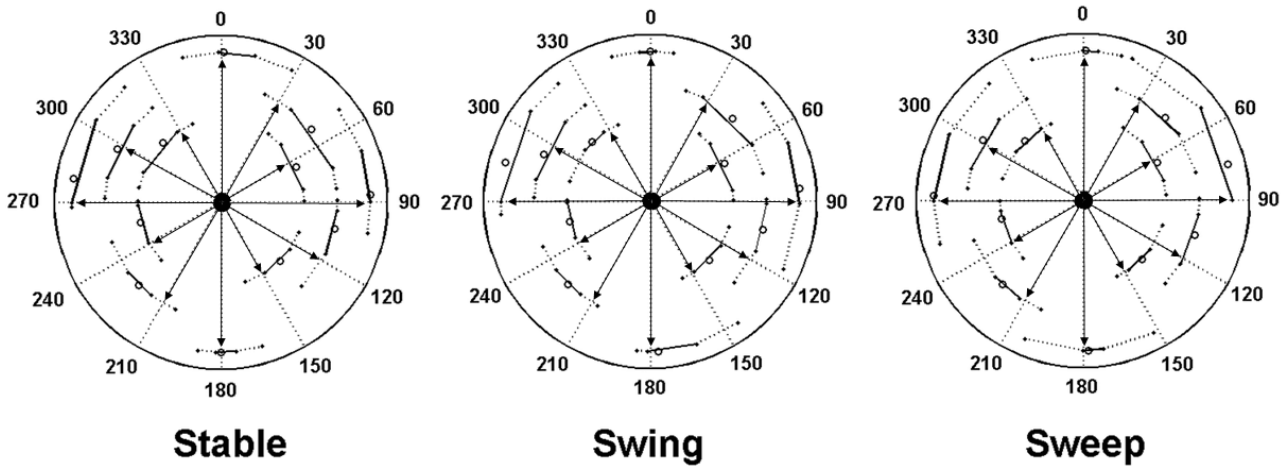


Figure 4-9. Variance in localising estimates plotted by sound condition. This figure provides an illustration of the trend in variance for localising estimates within each of the sound conditions. Each azimuth bearing has an associated variance plot which comprises a solid line representing the quartile range; a hollow circle indicates the mean; and outer points representing the 10 and 90 percentile range.

4.4.4 Confidence

Similar lateral groupings were adopted for the confidence estimates as were undertaken for azimuth localising errors. ANOVA conducted on the confidence estimates revealed a significant main effect for region ($F(3, 60) = 4.390, p = .007$), however, no main effects were observed for the sound condition ($F(2, 40) = 2.665, p = .082$). An interaction was reported for region by sound ($F(6, 120) = 4.462, p = .000$), but not for region by input ($F(3, 60) = 2.78, p = .050$), sound by input ($F(2, 40) = .715, p = .495$), or region by sound by input ($F(6, 120) = 1.371, p = .231$). Table 4-2 and Figure 4-10 below report the AVOVA main effects, interactions, and post-hoc analysis results for region alone. Post-hoc results for sound by region are detailed in Table 4-3. Post-hoc analysis revealed a large increase in participants' confidence when attending to sound in the left and right regions compared with the front region.

Table 4-2.
Summary statistics for mean confidence estimates.

Condition	RM-ANOVA					Bonferroni			Cohen's <i>d</i>	
	<i>F</i>	<i>df</i>	<i>p</i>	<i>M (SD)</i>	95% CI	Left	Right	Back	Left	Right
Sound	2.665	(2, 40)	.082							
Sound/Input	.715	(2, 40)	.495							
Region/Sound	4.462	(6, 120)	.000*							
Region/Input	2.78	(3, 60)	.050							
Region/Sound/Input	1.371	(6, 120)	.231							
Region	4.390	(3, 60)	.007*							
Front				7.95 (0.85)	[7.6, 8.4]	.012*	.026*	.103	1.1 _(L)	1.1 _(L)
Left				9.01 (1.00)	[8.2, 8.6]		1.00	1.00		
Right				8.98 (0.99)	[8.2, 8.5]			1.00		
Back				8.20 (0.86)	[8.0, 8.6]					

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Left = 320 to 40, Right = 50 – 130, Back = 140 to 220; (L) = large effect size; Confidence scale from 0 (low) to 10 (high); * *p* < .05.

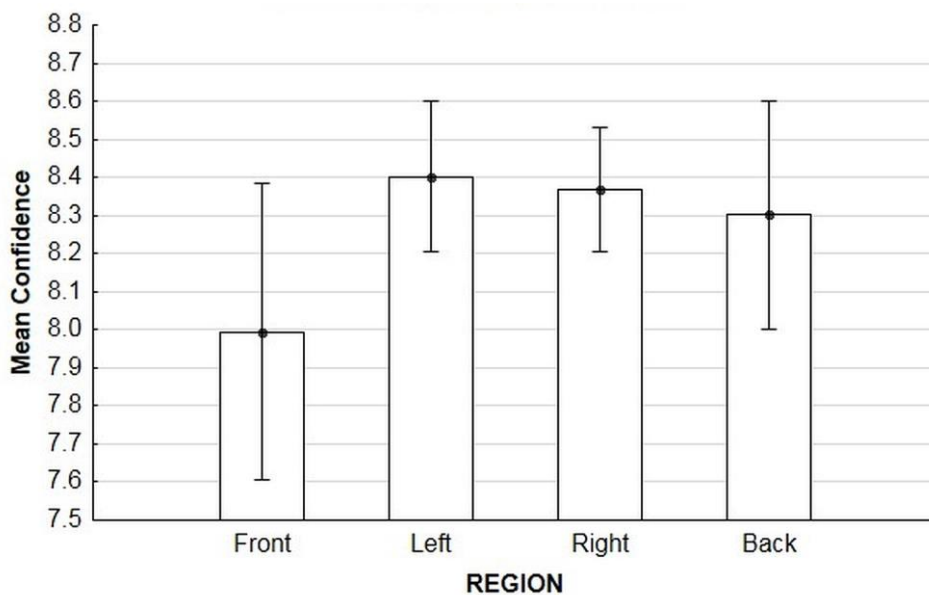


Figure 4-10. Mean confidence estimates grouped by region. Error bars denote 0.95 confidence intervals. Confidence scale from 0 (low) to 10 (high). A large increase in participants' confidence was observed when localising sound toward the left (*p* = .012, *d* =

1.1) and right ($p = .026$, $d = 1.1$) regions compared with the front region ($F(3, 60) = 4.390$, $p = .007$). No differences were observed between the back and all other regions.

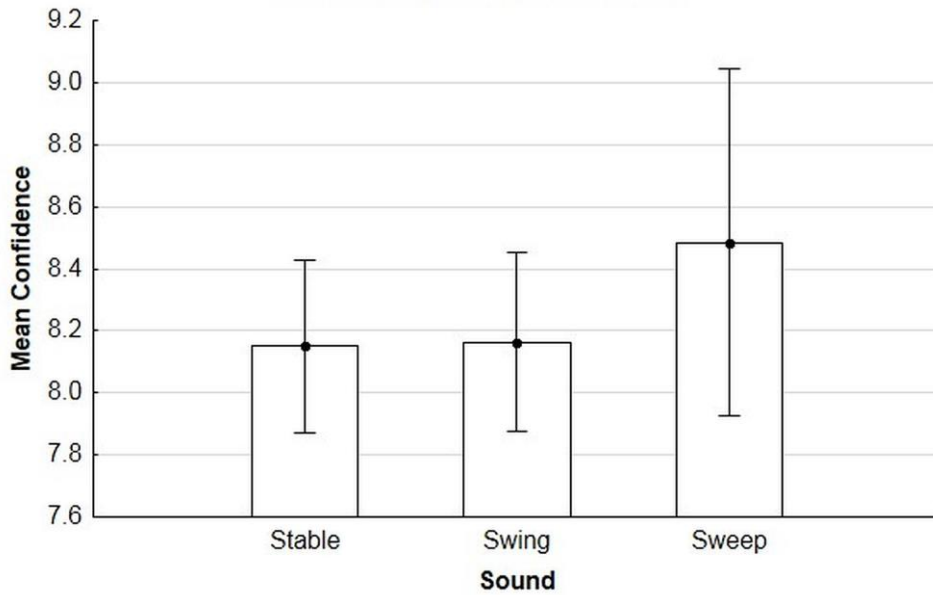


Figure 4-11. Mean confidence estimates grouped by sound condition. Error bars denote 0.95 confidence intervals. No main effect was observed for confidence grouped by sound ($p = .082$). Confidence scale from 0 (low) to 10 (high). Nine of the 22 participants remarked that the sweep condition was occasionally distracting and disorienting.

Figure 4-12 below plots the mean confidence estimates grouped by sound and region. The sweep sound condition noticeably trends higher than the other conditions in all regions.

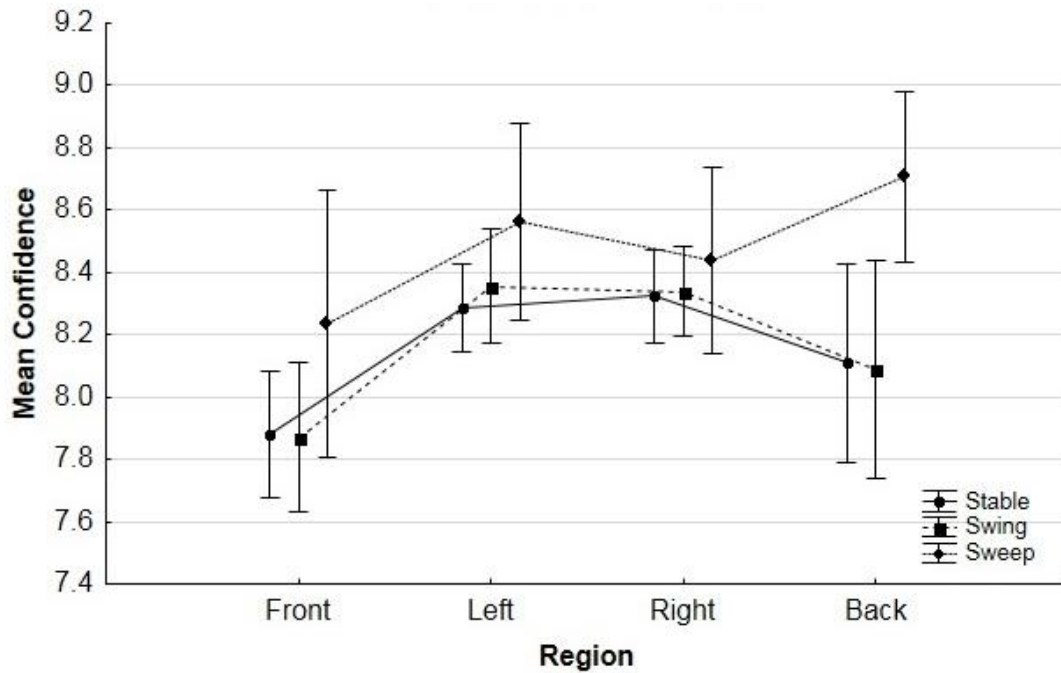


Figure 4-12. Mean confidence estimates grouped by sound and region. Error bars denote 0.95 confidence intervals. Confidence scale from 0 (low) to 10 (high).

Post-hoc analysis of the main effect for sound by region found significant effects, which are listed in Table 4-3. Table 4-4 reports the effect sizes for sound condition by region.

The front stable and front swing conditions were found to both produce a large reduction in confidence compared with all sound conditions in the left and right regions. The front sweep introduced a small increase in confidence compared with front stable and front swing, but was found to have a large decrease in confidence compared to the left sweep. The front sweep was also found to have a small decrease in confidence compared to the back sweep. Both the left and right sweep conditions were not found to differ in confidence compared to the stable and swing conditions within those regions. The left sweep produced a large increase in confidence compared with the back stable and back swing. The right sweep produced a small increase in confidence compared with the back swing.

The back sweep condition produced the highest confidence of all. It rated as having introduced a small increase in confidence compared with the left stable, left swing, and right stable conditions. It produced an intermediate level of increased confidence when compared with the right swing, back stable, and back swing. A large improvement in confidence was also found between the back sweep condition and the front stable and front swing conditions. Both front and back sweep conditions were found to produce

improved confidence compared to the stable and swing conditions within those regions. The same effect was not found in the left or right regions.

Table 4-3.

Post-hoc analysis of participants' confidence estimates for sound conditions by region.

Condition	M (SD)	95% CI	Bonferroni										
			Front Swing	Front Sweep	Left Stable	Left Swing	Left Sweep	Right Stable	Right Swing	Right Sweep	Back Stable	Back Swing	Back Sweep
Front Stable	7.9 (0.7)	[7.7, 8.1]	1.00	.003*	.000*	.000*	.000*	.000*	.000*	.000*	.423	.837	.000*
Front Swing	7.8 (0.8)	[7.6, 8.1]		.002*	.000*	.000*	.000*	.000*	.000*	.000*	.310	.625	.000*
Front Sweep	8.2 (1.1)	[7.8, 8.7]			1.00	1.00	.008*	1.00	1.00	.908	1.00	1.00	.000*
Left Stable	8.9 (0.9)	[8.1, 8.4]				1.00	.082	1.00	1.00	1.00	1.00	1.00	.000*
Left Swing	9.0 (1.0)	[8.2, 8.5]					.946	1.00	1.00	1.00	.235	.110	.003*
Left Sweep	9.2 (1.2)	[8.2, 8.9]						.310	.526	1.00	.000*	.000*	1.00
Right Stable	8.9 (0.9)	[8.2, 8.5]							1.00	1.00	.739	.371	.000*
Right Swing	9.0 (0.9)	[8.2, 8.5]								1.00	.442	.215	.001*
Right Sweep	9.1 (1.1)	[8.1, 8.7]									.007	.003*	.110
Back Stable	8.0 (0.9)	[7.8, 8.4]										1.00	.000*
Back Swing	8.0 (0.9)	[7.7, 8.4]											.000*
Back Sweep	8.6 (0.7)	[8.4, 9.0]											

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Left = 320 to 40, Right = 50 – 130, Back = 140 to 220; * p < .05.

Table 4-4.

Cohen's d effect sizes for sound conditions by region.

Condition	Cohen's d									
	Front Sweep	Left Stable	Left Swing	Left Sweep	Right Stable	Right Swing	Right Sweep	Back Stable	Back Swing	Back Sweep
Front Stable	.33 _(S)	1.2 _(L)	1.3 _(L)	1.3 _(L)	1.2 _(L)	1.3 _(L)	1.3 _(L)			1 _(L)
Front Swing	.42 _(S)	1.3 _(L)	1.3 _(L)	1.4 _(L)	1.3 _(L)	1.4 _(L)	1.3 _(L)			1.1 _(L)
Front Sweep				0.87 _(L)						0.43 _(S)
Left Stable										0.37 _(S)
Left Swing										0.43 _(S)
Left Sweep							1.12 _(L)	1.13 _(L)		
Right Stable										0.37 _(S)
Right Swing										0.5 _(I)
Right Sweep								0.09 _(S)		
Back Stable										0.74 _(I)
Back Swing										0.74 _(I)
Back Sweep										

Note. CI = confidence interval; Regions comprised the following azimuth bearings: Front = 320 to 40, Left = 320 to 40, Right = 50 – 130, Back = 140 to 220; (S) = small effect size; (I) = intermediate effect size, (L) = large effect size; * $p < .05$.

4.5 Discussion

The results indicate that supplementary spatial audio cues in the form of stationary and transient sounds do have the potential to improve sound 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.

4.5.1 Contrasting Display Orientation

Visual interfaces that convey bearing information, such as the 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 interaural cues, rather than

the more fallible spectral cues that dominate the perception of 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 non-equivalence was found in the front stable condition, the conservative equivalency range of ± 5 -degrees 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 a listener when extrapolating information presented through disparately oriented displays, which should possibly be considered in future studies.

4.5.2 *Front-back Errors and Localising Accuracy*

Oldfield & Parker (1984b) 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. (1993) similarly found that 31% of signals initiated reversal errors, 25% of which occurred from the front, and 6% from the rear. It is not surprising that we see a similar distribution of front-back errors occurring within the current study. The use of a non-individual HRTF means that the available spectral cues would have been unfamiliar and therefore difficult to utilise adequately.

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 target sound, and then focus solely on the target sound at the expense of gaining any relative azimuth positional cues that the transient sweep sound may continue to offer. They appeared to lack an ability to concurrently derive spatial cues from the sweep sound while attending to the target 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. (2006) 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 (Iwaya et al., 2003; Wightman & Kistler, 1999). 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.

4.5.3 *Internalisation and Localising Accuracy*

As mentioned previous in section 2.4.1, internalisation occurs when non-individual HRTF spectral cues poorly match a listener's own cues. The listener experiences poor spatial depth perception for sound as it appears to reside more internally toward the interaural axis. Participants in this study subjectively reported experiencing a varying degree of internalisation, which as expected was markedly more prevalent in the front region.

Muller & Bovet (1999) 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 (Honda et al., 2007).

The observed localising bias, which appeared to skew the estimates toward the interaural axis, is likely accounted for by the use of an upper percentile interaural distance within the non-individual HRTF. This phenomenon is described in more detail at section 2.4.2. Oldfield & Parker (1984a) observed the same localising bias occurring under normal hearing conditions, which has likely 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 (2008) 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.

4.5.4 Confidence

Findings that 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 may have therefore provided the listener with adequate cues to overcome front-back hemisphere uncertainties, which may have been the sole factor contributing to 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 align well with front-back errors or localising performance suggests that a less than optimal perceptual framework has been established by the supplementary cues.

4.5.5 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 (2002) 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 (Santangelo & Spence, 2007; Veltman, Oving, & Bronkhorst, 2004a, 2004b), while often facilitating more immersive situation awareness (Hopcroft et al., 2006; McCarley & Wickens, 2005). Multisensory displays may benefit from 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 navigation 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.

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.

4.6 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 non-individual 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 alignment 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.

5. Study 3: Concurrent 3-D Sonifications Enable the Head-up Monitoring of Two Interrelated Aircraft Navigation Instruments

Towers, J., Burgess-Limerick, R., & Riek, S. (2014). Concurrent 3-D sonifications enable the head-up monitoring of two interrelated aircraft navigation instruments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(8), 1414-1427. doi: 10.1177/0018720814536443

5.1 Abstract

Objective: To enable the head-up monitoring of two interrelated aircraft navigation instruments by developing a 3-D auditory display that encodes this navigation information within two spatially discrete sonifications.

Background: Head-up monitoring of aircraft navigation information utilizing 3-D audio displays, particularly involving concurrently presented sonifications, requires additional research.

Method: A flight simulator's head-down waypoint bearing and course deviation instrument readouts were conveyed to participants via a 3-D auditory display. Both readouts were separately represented by a collocated pair of continuous sounds, one fixed and the other varying in pitch, which together encoded the instrument value's deviation from the norm. Each sound pair's position in the listening space indicated the left/right parameter of its instrument's readout. Participants' accuracy in navigating a predetermined flight plan was evaluated while performing a head-up task involving the detection of visual flares in the out of cockpit scene.

Results: The auditory display significantly improved aircraft heading and course deviation accuracy, head-up time, and flare detections. Head tracking did not improve performance by providing the participant with control over interaural cues, suggesting that integrated supplementary localising cues were adequate for the task.

Conclusion: A supplementary 3-D auditory display enabled effective head-up monitoring of interrelated navigation information normally attended to through a head-down display.

Application: Pilots operating aircraft such as helicopters and unmanned aerial vehicles may benefit from a supplementary auditory display because they navigate in 2-dimensions while performing head-up, out of aircraft, visual tasks.

5.2 Introduction

The safe operation of modern aircraft relies heavily on pilots' ability to effectively monitor flight instruments. Not unlike other complex control systems, aircraft cockpits employ visual interfaces as the primary mode of data presentation. During periods of increased workload, dependencies on the visual modality may place an excessive demand on pilots' capacity limited sensory and cognitive processing channels (Wickens, 2002). This may consequently constrain the pilot's ability to establish and maintain situation awareness (Endsley, 1995). The current study aimed to enhance aircraft navigation by developing a supplementary auditory display that enables pilots to maintain awareness of the changing state of navigation instruments while performing head-up visual tasks.

By diversifying the modality in which information is presented, multisensory displays overcome constraints associated with any single sensory modality to better accommodate workload demands and facilitate improved situation awareness (McCarley & Wickens, 2005; Santangelo & Spence, 2007; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011). The use of sonification as a component within multisensory displays has been studied since the mid 1950's as a perceptually effective means for communicating information (Barrass & Kramer, 1999; Pollack & Ficks, 1954). Of particular note is the gain in head-up monitoring capabilities through the use of such displays (Parker, Smith, Stephan, Martin, & McAnally, 2004; Watson & Sanderson, 2004). While the commercial use of 3-D audio within multisensory displays remains rare, some research has found success in its use as a navigation aid by spatially positioning broadband white noise beacons or verbal commands toward the intended direction of travel (Lokki & Grohn, 2005; Marston, Loomis, Klatzky, & Golledge, 2007; Walker & Lindsay, 2006). The spatial separation of sound sources within 3-D auditory displays also facilitates the cocktail party effect, which enables the listener to focus on a single auditory source while filtering out sounds presented from other locations (Brungart & Simpson, 2007).

More advanced three-dimensional audio in the form of verbal beacons, broadband cueing signals, and sonifications have been used to supplement traffic collision avoidance and target acquisition and tracking guidance systems within cockpits (Begault & Pittman, 1996; McAnally & Martin, 2008; Tannen, Nelson, Bolia, Warm, & Dember, 2004; Veltman et al., 2004b). Veltman et al. (2004a) successfully encoded aircraft roll angle corrections into the pulse frequency and pitch of a spatially positioned tone in order to improve pilot performance during a fighter pursuit task. Simpson et al. (2005) also successfully utilized an auditory display to indicate waypoint direction through the use of a white noise beacon.

FLIGHT NAVIGATION AUDITORY DISPLAY

Aircraft attitude was encoded into music that panned horizontally to indicate aircraft roll angle, while altering in frequency to indicate aircraft pitch.

The effectiveness of absolute pitch as an informational cue is questionable as it is inaccurate and unreliable given that only one in 10,000 people are estimated to possess absolute pitch perception (McDermott & Oxenham, 2008). However, encoding information into variations in the tempo and direction of a sound's pitch can prove effective because this relies on relative perception, which most people can differentiate well and perceive from a young age (Plantinga & Trainor, 2005). Tempo and pitch were used in the current study to aurally encode changing values of navigation instruments. In an attempt to improve relative pitch perception, an original design feature in the form of a fixed reference pitch was added to each of these variable sounds.

Creating a 3-D audio signal for presentation through stereo headphones requires digital processing through a head related transfer function (HRTF). Localising cues are incorporated into the sound through spectral cue distortions and interaural onset timing and intensity differences (Yost, 2007). Due to variations in the physical features of individuals, localising cues differ to the extent that the use of a generalized, or non-individualized filter, can introduce poor localising accuracy and front-back confusions if care is not taken to design adequate supplementary localising cues. Individually calibrated HRTFs are often impractical and costly to create, requiring sound recordings be taken at the ear drum or ear canal of an individual situated within an anechoic chamber (Wightman & Kistler, 1989a).

A head tracking enabled auditory display allows the listener to control a sound's position relative to the head through deliberate changes in head orientation. When doing so, the listener has control over the ongoing interaural time differences. Changes in the delay relative to the direction in which the head is being turned can help determine a sound's apparent front or back hemisphere of origin. This is because the delay changes in opposing directions for sound located in different hemispheres on the same lateral side of the midsagittal plane (Wightman & Kistler, 1999). Poor hemisphere localising results in increased front-back errors, which may introduce overlapping interference between sounds positioned in the front and back. In the absence of head tracking, hemisphere discrimination relies solely on effective localising cues designed into the signals themselves. The objective for including conditions with and without head tracking in this study was to test if the auditory display's design effectively mitigated front-back errors to an extent that interrelated audio signals located in the front and back could be perceived as independent streams of information without head tracking information.

Utilizing current auditory display design principles based on perceptual and theoretical findings covered in Bregman's (1990) *Auditory Scene Analysis* and theoretical models that focus on integrating sound in interactive interfaces (Ahmad, Stanney, & Fouad, 2009), our aim was to develop an auditory display that improves head-up monitoring of an aircraft's position relative to a flight plan. We employed the *audio integration*, *temporal audio*, and *spatial audio* theoretical models defined by Ahmad et al. (2009) as guidelines for determining the cognitive performance objectives and subsequent acoustic wave attributes for each auditory cue. Auditory cues were used to guide psychomotor activity when adjustments were being made to the aircraft's control yoke to correct and maintain accurate flight navigation. Component information within the aircraft's visual navigation display was determined to be most appropriately encoded into an audio signal through arrhythmic, interval-based temporal and spatial cues. These design decisions embody the framework of the auditory display and have been refined through additional testing and evaluation, which is described in more detail throughout this paper.

While previous studies have focused on aircraft profile by encoding heading and attitude information, there has not previously been a focus on the aircraft's 2-dimensional deviation from the intended course. The current study required participants to manually fly an aircraft simulator about a predetermined flight plan that was displayed on the simulator's navigation display. A supplementary 3-D auditory display aurally encoded the direction toward the next waypoint and the aircraft's lateral deviation from the intended flight path through two spatially discrete sonifications. Research into supplementary 3-D auditory displays for 2-dimensional navigation is important because helicopter and unmanned aerial vehicle (UAV) pilots navigate in 2-dimensions while attending to high workload visual tasks in the out of cockpit scene.

The ability to continually hear the changing state of navigation instruments was expected to allow participants to make informed course corrections while performing head-up visual tasks. Navigation accuracy was assessed in terms of the aircraft's mean forward facing angular error relative to the next waypoint, and the mean distance in meters that the aircraft deviated from the intended flight path. Head-up time was expected to increase, resulting with improved performance on a visual scan task that required participants to detect flares that were periodically displayed in the out of cockpit scene.

5.3 Method

5.3.1 Participants

Twenty-four Boeing Defence Australia employees aged between 22 and 49 ($M = 34$, $SD = 8$) completed the study on a voluntary basis. The sample comprised 22 males and two females who each successfully completed a hearing screening test. Equal loudness tests were conducted on both ears of each participant across a frequency range of 30 Hz to 16 kHz, with results compared to a standardized dBA weighted curve. None of the participants had previous experience with spatial audio localising or flight simulation. An additional five volunteers were withdrawn prior to beginning the study. Two were considered to have abnormal hearing and three failed to reach an adequate level of performance when flying the simulator. A fifty-dollar gift voucher was offered as an incentive reward to the participant who obtained the most accurate overall flight performance. 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 (HMS11/2704).

5.3.2 Flight Simulator

Figure 5-1 shows the Cessna 172 simulator used in the study, which employed Microsoft Flight Simulator X (FSX) and a simulated Garmin G1000 glass cockpit. The out of cockpit scene was projected at a resolution of 1920 x 1080 pixels using a Dell 2400MP projector. The image was 180 cm x 110 cm in size and positioned at a distance of 180 cm directly in front of the participant. The G1000 instrumentation was displayed on a 76 cm LCD computer monitor approximately 80 cm in front of the participant at a resolution of 1920 x 1080 pixels. The monitor was positioned offset to the right of center so that the primary flight display was directly in front of the participant. A Saitek Pro Flight yoke was used to control the simulator and the throttle was configured to remain fixed at the maximum setting. The flight simulator realism setting was configured to novice, which minimized introduced handling effects on the aircraft to an indiscernible level.



Figure 5-1. Flight simulator.

Figure 5-2 shows the G1000 flight instrumentation, which includes the primary flight display (PFD) on the left, and the multifunction flight display (MFD) on the right.

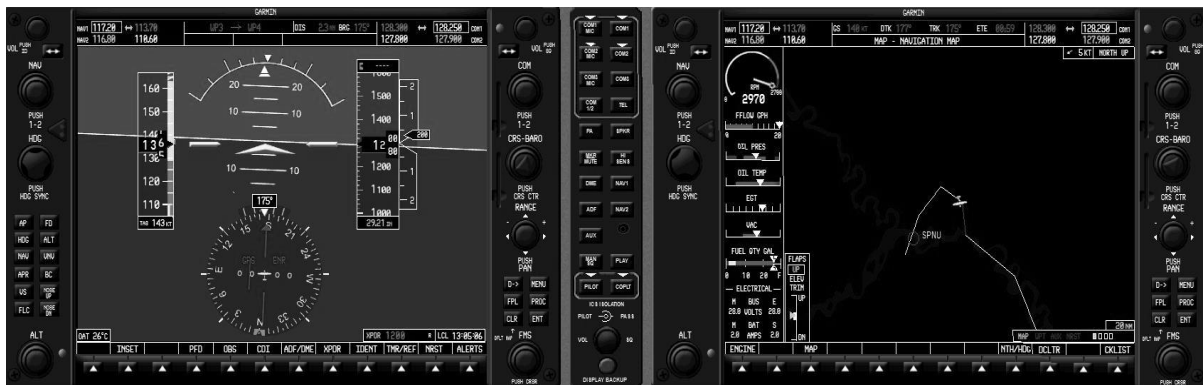


Figure 5-2. G1000 glass cockpit.

Figure 5-3 illustrates the flight plan used within the study, which was divided into three equal parts. Each part corresponded to a different experimental condition and data collection phase. Each phase comprised two legs and three associated waypoints that identified geographical coordinates where changes in aircraft heading were planned to occur. Phases were separated by an entry leg where no data were collected. The entry leg was used to verbally introduce the altered auditory display configuration and allowed the participant time to stabilize the aircraft for entry into the next data collection phase. The order of conditions was counter balanced between participant groups to address any potential ordering effects that may have otherwise occurred. The total time to complete the course was approximately 16 min, with each phase taking approximately 4.5 min.

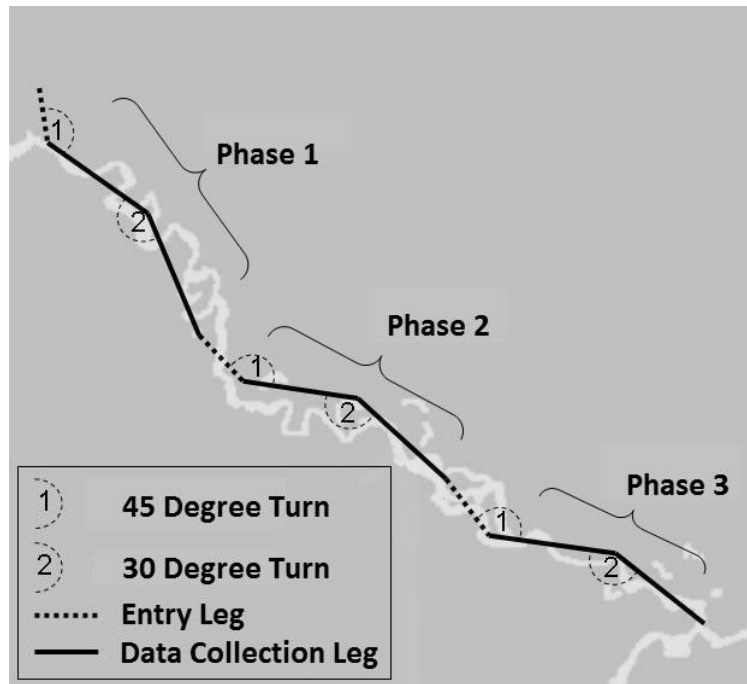


Figure 5-3. Flight plan. The three data collection phases of the flight plan are shown. Each phase comprised a 45-degree entry turn into each phase and two legs separated by a 30-degree turn. The dotted lines indicate entry legs, where no performance data was collected and participants prepared for the next audio display condition.

Figure 5-4 identifies the location of the waypoint bearing indicator (WP), course deviation indicator (CDI), and altimeter within the primary flight display. These were the primary navigation instruments utilized throughout the study.

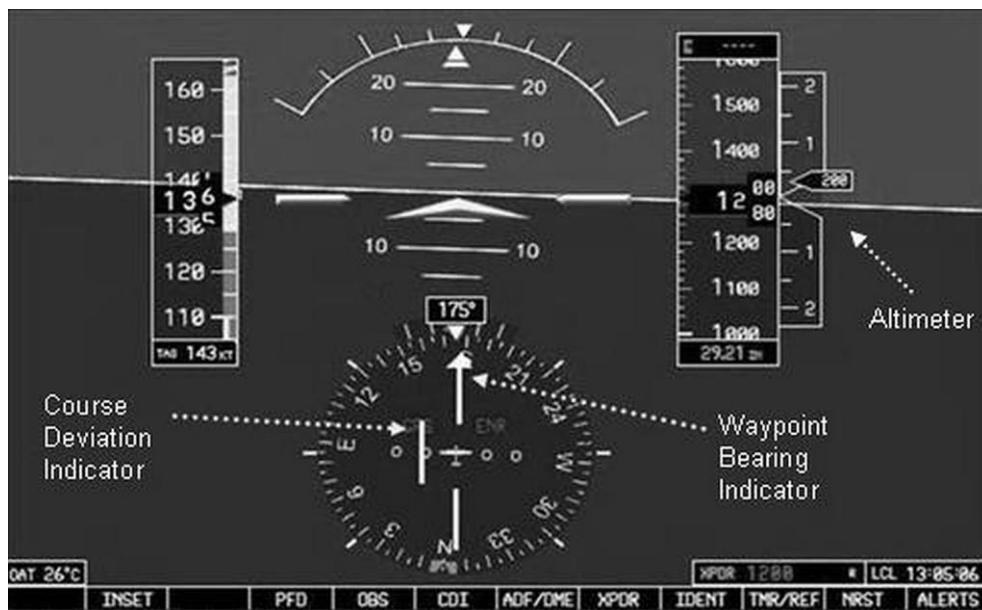


Figure 5-4. Primary flight display (PFD). The course deviation (CDI), waypoint (WP), and altimeter readouts are labelled.

The waypoint bearing indicator maintains its orientation toward the next waypoint in a similar manor to that of a compass needle pointing north. The CDI is a line that moves laterally off centre of the waypoint bearing needle to indicate the direction and distance in which the aircraft has deviated from the current leg of the flight plan. The scale of the CDI was fixed at a full scale deflection of 555 m. The circles either side of the aircraft image on the CDI display indicate distances of 275 m and 555 m in lateral deviation from the flight path. The altimeter presents aircraft altitude as a scrolling numerical tape.

5.3.3 3-D Auditory Display

A 3-D auditory display was developed to present redundant navigation information for waypoint bearing and course deviation through two spatially discrete sonifications that are referred to from this point forward as the *audio signals*. SLAB3D (Miller, J. D. & Wenzel, 2002) was installed on the simulator computer to provide 3-D rendering of the audio signals for delivery through Bose® TriPort binaural headphones. The default SLAB3D HRTF was used, which is a particular person's HRTF measurements converted to minimum phase HRTFs. Custom software was developed to monitor the flight simulator navigation variables and manipulate their encoded values within the signals.

Figure 5-5 illustrates the spatial positioning of the two audio signals. Changes in aircraft heading caused the waypoint signal to transit about the azimuth in order to maintain its alignment toward the fixed geographical position of the next waypoint. The course deviation signal (CDI) is shown in both of its alternate stationary positions, which were fixed at 140-degrees on either side of forward as defined by the type of auditory display (simple binaural vs. head tracking enabled). The CDI spatial position encodes bearing information from the CDI visual display for the lateral direction in which the aircraft is deviating from the current leg of the flight plan. A Polhemus Fastrak head tracking system was utilized during one experimental condition to measure the participant's head position. This enabled the auditory display to maintain spatial alignment with the visual instruments and out of cockpit scene regardless of changes in the participant's head orientation.

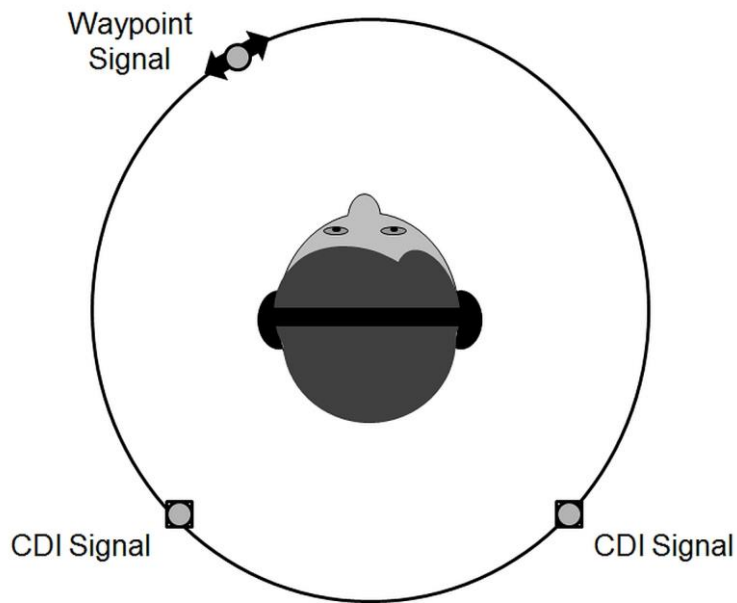


Figure 5-5. Auditory display spatial configuration. The display comprised a waypoint sound that maintained its orientation toward the next waypoint, and a CDI sound that was presented on the side in which the aircraft deviated away from the intended flight path.

Both audio signals were composed of two sounds, a pitch invariant *carrier note* and a variable *supplementary sound*. Audio signal component sounds were always collocated and played concurrently to ensure that their combination was perceived as a unified point of auditory information. The characteristics of the audio signal component sounds are detailed in Table 5-1.

Table 5-1.

Auditory display signals

Signal	Component Sounds		
	Carrier Note	Supplementary Sound (Square Wave)	
		Range	Rate of Change
Waypoint	C#3 Muted Trumpet, 139 Hz*	Varying 140.5 Hz to 229 Hz	0.5 Hz/deg
CDI	B ^b 2 Baritone Sax, 117 Hz*	Varying 123 Hz to 450 Hz	0.6 Hz/m

Note. * Fundamental *f*; Sounds were sampled at the same dB level.

The carrier note's purpose was twofold. Firstly, it provided unique timbre properties that assisted in differentiating between the two concurrently presented audio signal streams. Secondly, Fourier transforms confirmed that the notes contained high amplitude harmonic frequencies above 4 kHz, which provide spectral cues that aid in front-back localization and externalization of sound (Yost, 2007).

The supplementary sound component of each audio signal was a square wave sound generated by SLAB3D, which did not contain spectral cue frequencies. The value of the waypoint bearing and lateral course deviation were encoded into changes in the fundamental frequency of each signal's supplementary sound. For example, if the aircraft altered heading and the waypoint direction proceeded about the azimuth toward the rear of the aircraft, the fundamental frequency of the waypoint signal's supplementary sound would increase at a rate of 0.5 Hz/deg. Similarly, as the aircraft deviated from the flight path, the fundamental frequency of the CDI signal's supplementary sound would increase at a rate of 0.6 Hz/m. Conversely, the frequency would decrease at the same rate as the navigation values approached zero. The direction and rate of change (tempo) of the navigation instrument readouts were encoded into the supplementary sound pitch direction and tempo. The fundamental frequencies of the waypoint signal component sounds were the same when the aircraft was facing the waypoint, while the CDI signal component frequencies were the same when the aircraft was directly over the intended flight path. Each audio signal stream thereby communicated a continuous relative pitch differential between the fixed frequency of the carrier note and the varying frequency of the supplementary sound, which was intended to make the absolute value of the encoded navigation information aurally salient.

When altering the aircraft's heading to correct course deviations, the pilot must gain awareness of the angle of approach and distance from the current leg of the flight plan by concurrently monitoring the WP and CDI signals. These course corrections change the aircraft's rate of closure toward the leg and subsequently the tempo of pitch change in the CDI supplementary sound. The CDI pitch tempo therefore provided an additional redundant cue for changes in WP bearing, and in the design phase of the study, it was noted that this might offer a more effective cue than the WP signal itself. However, when this was tested by flying the simulator without the WP signal, it was found that the CDI signal was not informative enough to replace the waypoint signal. Most notably, without the WP signal, hunting effects were observed as the aircraft continually overshot the leg during course corrections. This conclusion was confirmed in the data analysis phase of

the study by ensuring that hunting effects were not present to the extent observed in the design phase.

Figure 5-6 plots the frequency of the two waypoint signal component sounds by azimuth bearing. The WP signal was muted when the aircraft heading was within +/- 3-degrees of the waypoint bearing and the frequency of the WP signal was 140.5 Hz.

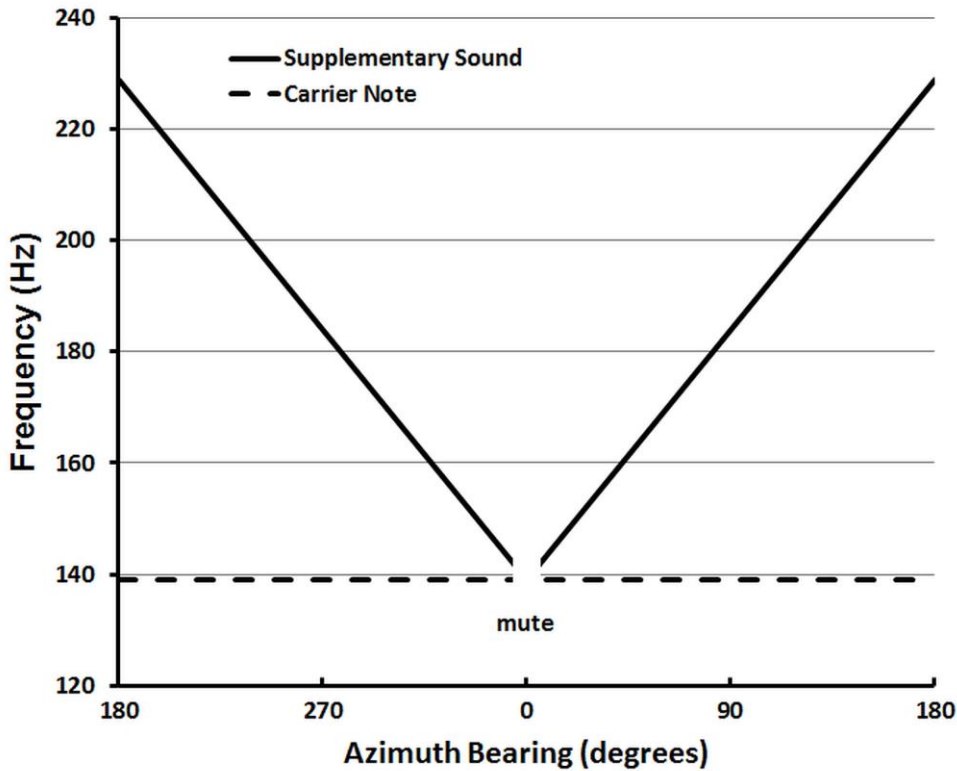


Figure 5-6. Waypoint audio signal component sound configuration. The supplementary sound frequency varied as a function of the waypoint’s relative azimuth bearing. The carrier sound maintained a consistent frequency. Both sounds were spatially collocated and muted when the waypoint bearing was within +/- 3-degrees of the aircraft’s heading.

Figure 5-7 illustrates the mapping of audibly encoded information between the CDI visual display and CDI audio signal. The pitch of the supplementary sound varied as a function of changes in the course deviation visual display. In this example the CDI indicator has moved to the right, thereby causing the CDI signal to be positioned on the right with a proportionate increase in the frequency of the supplementary sound. The CDI audio signal was muted when the aircraft was within +/- 10 m of the intended leg and the CDI frequency was 123 Hz.

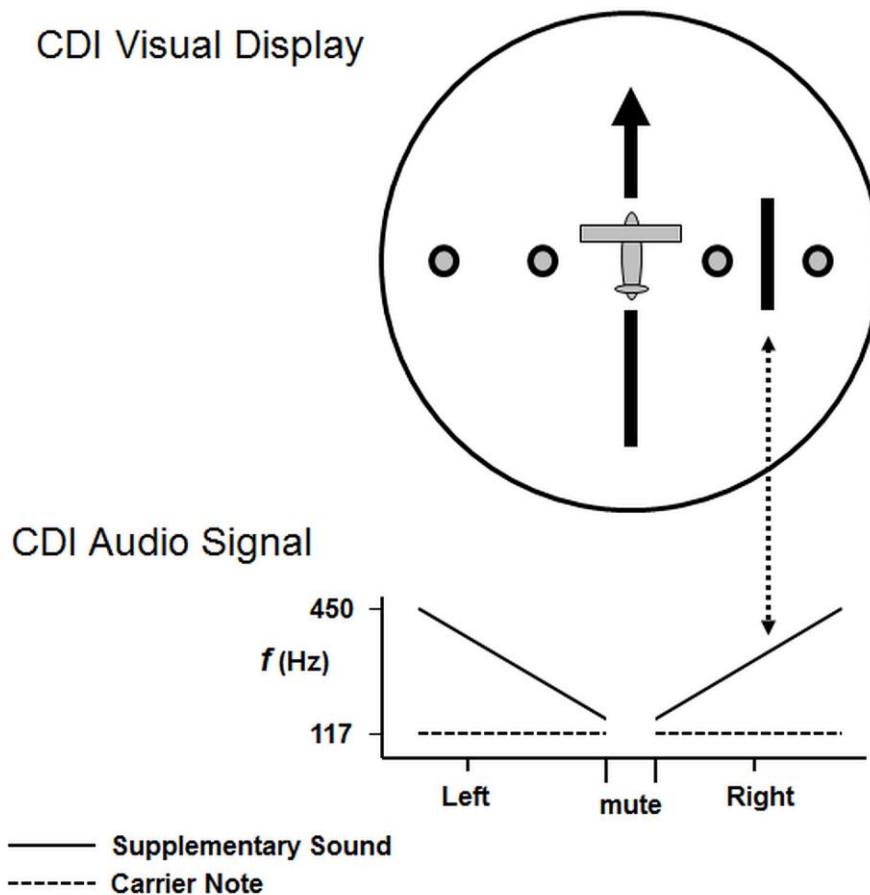


Figure 5-7. CDI visual interface to auditory display configuration mapping. The CDI supplementary sound varied as a function of the aircraft's lateral deviation from the intended flight path. The distance that the CDI indicator deviates from the center line of the visual display is mapped to a change in the supplementary sound's pitch. The carrier sound maintains a consistent frequency. The CDI signal was muted if the deviation error was less than +/- 10m from the intended path.

Two-months prior to completing the study, three participants assisted the authors with evaluating the auditory display's candidate component sounds. Learning effects were considered unlikely to have occurred during this activity because the component sounds were evaluated in a different context to their final application within the auditory display. Carrier notes were chosen from 35 candidate sounds that were sampled from eight different instruments on a Yamaha DGX-630 keyboard using Cool Edit Pro. They evaluated the 3-D characteristics of each candidate note in turn as SLAB3D orbited the sound about their head at a virtual distance of one meter and an elevation of zero degrees. Additional subjective ratings were evaluated, the first of which required listeners to draw a line about the top-down view of a head to indicate the perceived path and distance in which each note appeared to transit. The consonance of each audio signal was rated as a

number between one and 10. The evaluation resulted in a B^b2 note from a baritone saxophone being chosen to serve as a carrier note for the CDI, while a C[#]3 note from a muted trumpet was selected for the waypoint bearing. A square wave was chosen rather than a sine wave to serve as the supplementary sound because it rated most highly in the subjective evaluation for consonance and spatial perception when paired with the chosen carrier note.

Bregman's *Auditory Scene Analysis* identifies perceptual cues that enable the auditory system to separate individual sounds from the combined sound wave arriving at a listener's ear (1990). In addition to timbre and spatial location cues providing between-signal differentiators, the current design incorporated other cues to ensure that each signal's component sounds were perceived as a unified stream of information. A transition cue aided grouping by the supplementary sound being perceived to turn into the carrier note as their primary frequencies approached the same value. The synchronized muting and onset of each signal's component sounds also aided in their perceived grouping, while providing an additional cue to indicate the point where navigation readouts were close to zero. The frequency range of the supplementary sounds did overlap; however, operationally this did not appear to introduce any issues in regard to perceptual interference between signals. This was attributed to sufficient differences in the magnitude, rate, and direction of change between the two signals at any point in time while flying.

5.3.4 Experimental Design

The study was a within-subjects, repeated measures design, with the auditory display configuration manipulated as the independent variable. Three auditory display conditions were identified as *mute*, *audio*, and *audio HT*. As the name suggests, no audio was presented in the mute condition and participants navigated utilizing the head-down instrument readouts. The audio condition included the supplementary 3-D auditory display without head tracking. In this condition, the auditory display's zero-degree azimuth point, which corresponds with the forward facing direction of the simulator, was constantly aligned toward the participant's midsagittal plane. The final condition was audio HT, which introduced the supplementary auditory display with head tracking adjustments to continuously align the listening space with the forward-facing direction of the flight simulator, regardless of the orientation of the participant's head.

5.3.5 Procedure

Training was conducted in 30-minute sessions and focused on teaching the participants to fly the simulator and interpret the auditory display signals. The amount of total training required to reach a predetermined level of proficiency varied between participants ($M = 151$ min, $SD = 37$ min). The auditory display was only included in each participant's final 2 sessions of training, with head tracking only enabled in the final session. Students were encouraged to adjust the orientation of their head periodically to make best use of the head tracker cues during that condition. Flight performance measures were evaluated during training to ensure participants had attained proficiency prior to undertaking the experiment. The criteria included maintaining aircraft altitude within a window of 60 m; correctly interpreting navigation instruments; and achieving a mean course deviation of less than 100 m.

Gaze region dwell times, waypoint bearing error, course deviation error, and flare detections were recorded throughout the study. A Facelab (Seeing Machines, Canberra) gaze tracker was used to record the participants' gaze regions, which were categorized as either *out of cockpit*, *PFD* (primary flight display), or *MFD* (multifunction display). Waypoint bearing error was recorded as a streaming log of absolute angular error in degrees for the aircraft's forward facing direction relative to the next waypoint. Course deviation was recorded as a streaming log of absolute error in meters for the aircraft's distance from the intended leg of the flight plan.

Prior to beginning the experiment, the gaze tracker was calibrated and the auditory display volume was adjusted to a comfortable level. Participants were then instructed not to be complacent with their navigation performance and to constantly correct for the slightest course deviation. The trial began in a paused state with the aircraft positioned at 400 m in altitude facing the first entry leg of the flight plan. Participants activated the simulator by pressing the "P" button on a keyboard positioned behind the control yoke. A pre-recorded voice file automatically played when transiting through each entry leg, indicating which auditory display condition was being applied to the next phase. If the altitude envelope set at 60 m about the flight altitude of 400 m was exceeded, a voice recording would be automatically initiated to instruct the participant to either "climb", or "descend" as appropriate.

Participants were asked to look for flares in the out-of-cockpit scene as they navigated each leg of the experiment's flight plan. This visual search task was communicated as being more important than the navigation task and provided motivation for participants to minimize time spent visually attending to the navigation instruments.

Flares were generated by FSX and displayed at predetermined locations throughout the course on river banks and clearings between trees. Each flare was automatically activated as the aircraft approached its position and remained in a stationary position for ten-seconds at a height of two-meters off the ground. A total of 30 flares were evenly distributed throughout the data collection phases of the flight plan and were visible for a total of 5 min and absent for 8 min 30 sec. Correct-positive flare detections were automatically recorded when the participant pressed the control yoke trigger button while a flare was being displayed. False-positives were unable to be identified when flares were active, but were recorded during periods when flares were not being presented. False-positive errors were expected to be minimal because flares were carefully positioned to be easily distinguishable from the surrounding terrain during fixated inspection.

On completion of the study, participants were asked to provide verbal feedback during an interview that focused on three areas of auditory display design. Of particular interest was the degree to which audio signals were found to interfere with each other when presented on the same lateral side. Opinion was sought to determine how easily the supplementary reference cue was perceived, along with its effectiveness in aiding participants to determine the absolute value of encoded navigation information. Finally, we asked if the participants considered themselves able to attend to the two navigation signals concurrently while performing the head-up visual scanning activity.

5.3.6 Data Analysis

Unless otherwise specified, all data analyses were conducted using a significance level identified at $\alpha = .05$. The general approach taken for data analysis was a one-way RM-ANOVA (three-level) performed on the statistical software package Statistica (version 10). Subsequent post-hoc analyses were undertaken for all significant main effects using Bonferroni adjusted alpha levels. Effect sizes were expressed using Cohen's *d*. Within-subject confidence intervals were constructed by eliminating intersubject variance using the method outlined by Loftus and Masson (1994).

An inferential test of equivalence in the form of two simultaneous one-sided hypothesis tests (Rogers, Howard, & Vessey, 1993) was used to determine the extent to which auditory display conditions were identical. This test was used to supplement traditional hypothesis testing when audio conditions were not found to differ significantly. The equivalence test null hypothesis asserts that the difference between two groups is at least as large as a predetermined equivalence interval. As in traditional hypothesis testing, the goal is to reject the null hypothesis and accept the alternative hypothesis. An

equivalence interval of $\pm 10\%$ was defined around a difference of zero. Any difference between the two groups small enough to fall within this interval was considered practically unimportant. Therefore, if the audio condition mean fell within the $\pm 10\%$ interval set around the audio HT condition mean, the null hypothesis was rejected and the alternative hypothesis was accepted.

5.4 Results

Main effects were observed for all aircraft navigation and gaze performance measures, with subsequent post-hoc analysis consistently indicating degraded performance within the mute condition when compared with both the audio and audio HT conditions. Pairwise comparisons revealed no significant differences between the two audio conditions across all performance measures, which were subsequently tested for statistical equivalence. The audio conditions produced statistically equivalent results for all navigation measures and visual dwell times, while flare detections were found not to be statistically equivalent between audio conditions.

Table 5-2 provides a summary of results for the statistical analysis conducted on all performance measures.

Table 5-2.

Summary statistics for audio conditions grouped by performance measure

Condition	RM-ANOVA				Bonferroni		Cohen's <i>d</i>		Equivalence ^{a, b}		
	<i>F</i> (2, 46)	<i>p</i>	<i>M</i> (<i>SD</i>)	95% CI	Audio	Audio HT	Audio	Audio HT	SE	<i>z</i>	<i>p</i>
Course Deviation (m)	9.827	.000*									
Mute			84.02 (6.87)	[81.18, 86.91]	.001*	.001*	1.29 _(L)	1.38 _(L)			
Audio			75.23 (6.72)	[72.31, 78.10]		1.0		NA			
Audio HT			75.35 (5.63)	[73.08, 77.76]					1.79	4.14	.000*
Waypoint Error (deg)	5.439	.008*									
Mute			6.68 (0.44)	[6.50, 6.87]	.010*	.041*	1.10 _(L)	0.86 _(L)			
Audio			6.26 (0.31)	[6.13, 6.40]		1.0		NA			
Audio HT			6.33 (0.37)	[6.18, 6.49]					.098	5.71	.000*
Total Dwell Time (sec)	5.95	.005*									
Mute			145.3 (19.27)	[137.2, 153.5]	.020*	.009*	1.01 _(L)	1.03 _(L)			
Audio			162.5 (14.48)	[156.4, 168.6]		1.0		NA			
Audio HT			164.2 (17.26)	[156.9, 171.5]					4.60	3.2	.000*
Mean Dwell Time (sec)	7.92	.001*									
Mute			2.10 (.54)	[1.87, 2.33]	.001*	.002*	1.02 _(L)	1.23 _(L)			
Audio			2.57 (.37)	[2.42, 2.73]		1.0		NA			
Audio HT			2.65 (.33)	[2.51, 2.79]					.101	1.83	.034*
Flare Detection	5.07	.010*									
Mute			3.96 (1.26)	[3.43, 4.50]	.037*	.018*	0.95 _(L)	0.96 _(L)			
Audio			5.08 (1.09)	[4.62, 5.55]		1.0		NA			
Audio HT			5.20 (1.31)	[4.65, 5.77]					.348	1.15	.125

Note. CI = confidence interval; (L) = large effect size; SE = standard error.

^a Equivalence test conducted on the audio and audio HT conditions with only the highest *p* value of the two one-sided tests reported.

^b Criterion is +/- 10 % of the audio HT condition mean.

* *p* < .05.

5.4.1 Course Deviation

Figure 5-8 plots the mean course deviation by audio condition. A main effect was observed, with pairwise comparisons revealing a large significant difference between the mute condition and both audio conditions. No significant differences were observed between the audio conditions, which were subsequently found to be statistically equivalent.

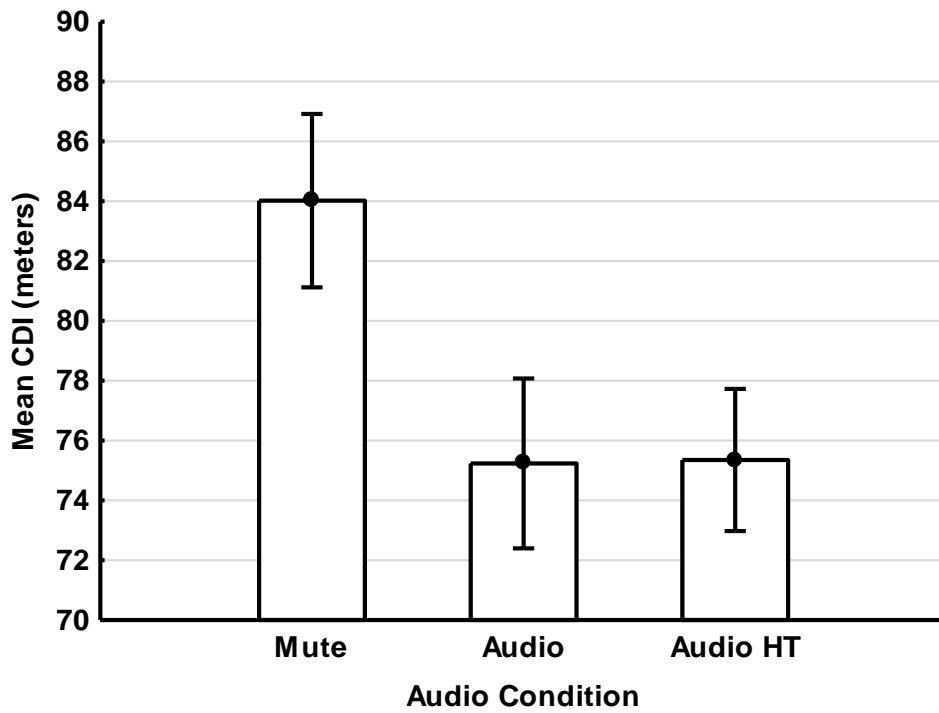


Figure 5-8. Mean course deviation by audio condition. Error bars denote within-subjects 95% confidence intervals.

5.4.2 Waypoint Angle (WP) Error

Figure 5-9 plots the mean waypoint angular error by audio condition. A main effect was observed, with pairwise comparisons revealing a large significant difference between the mute condition and both audio conditions. No significant differences were observed between the audio conditions, which were subsequently found to be statistically equivalent.

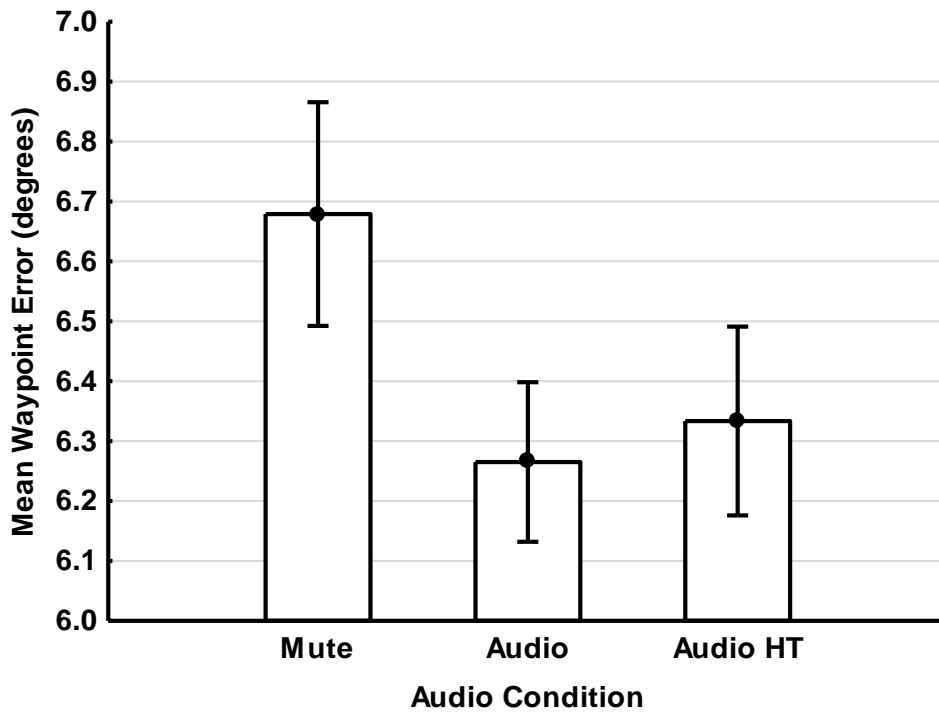


Figure 5-9. Mean waypoint angle by audio condition. Error bars denote within-subjects 95% confidence intervals.

To determine if the WP signal was being neglected in favor of the CDI signal, the data were further analyzed for hunting effects similar to those observed during the design of the auditory display. The presence of such effects was not detected, even during periods of elevated course corrections, such as occurred for 20-sec following the initiation of new waypoints in the flight plan.

5.4.3 Head-up Dwell Time

Figure 5-10 plots the total gaze dwell time out of the cockpit by audio condition. A main effect was observed, with pairwise comparisons revealing a large significant difference between the mute and both audio and audio HT conditions. No significant differences were observed between the audio conditions, which were subsequently found to be statistically equivalent.

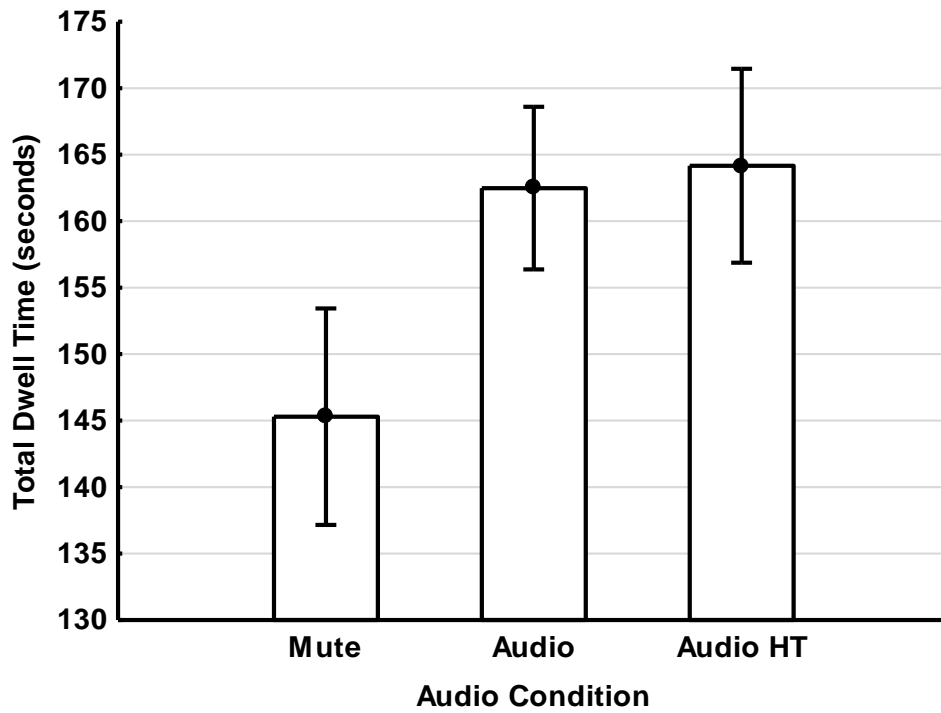


Figure 5-10. Total out of window gaze dwell time by audio condition. Error bars denote within-subjects 95% confidence intervals.

Figure 5-11 plots the mean out of cockpit gaze dwell time by audio condition. These data represent the average amount of time that visual attention was directed away from the head-down visual display to scan the out of flight deck scene. A main effect was observed, with pairwise comparisons revealing a large significant difference between the mute and both audio and audio HT conditions. No significant differences were observed between the audio conditions, which were subsequently found to be statistically equivalent.

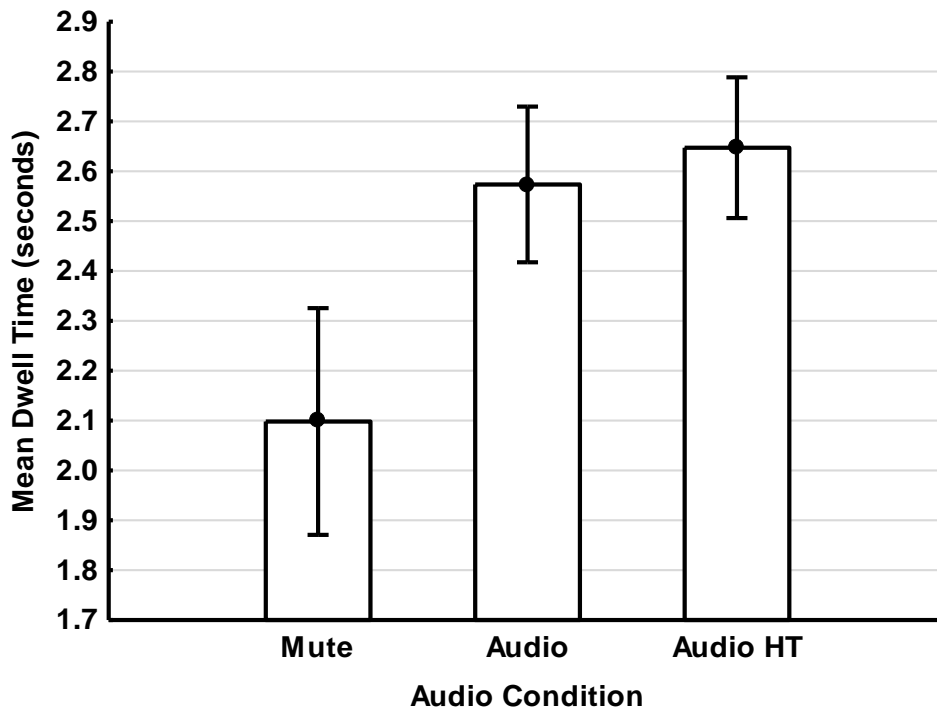


Figure 5-11. Mean out of cockpit gaze dwell time by audio condition. Error bars denote within-subjects 95% confidence intervals.

5.4.4 Flare Detection

Figure 5-12 plots the mean number of correct flare detections occurring by audio condition. A main effect was observed, with pairwise comparisons revealing a large significant difference between the mute and both audio and audio HT conditions. No significant differences were observed between the audio conditions, which were subsequently not found to be statistically equivalent at the $\pm 10\%$ criterion used in this analysis. However, an equivalence criterion set at 14% for the audio and audio HT conditions resulted in statistical equivalence ($SE = 0.35$, $z = 1.75$, $p = 0.04$). As expected, there were very few false-positive flare detections recorded ($M = 0.92$, $SD = 1.02$), suggesting that flares were adequately distinguishable from the surrounding terrain. The experimental design limited the identification of false-positives to periods when flares were not being presented. It is considered unlikely that a greater amount of false-positive responses occurred during flare presentations to the extent that they biased the correct-positive response data.

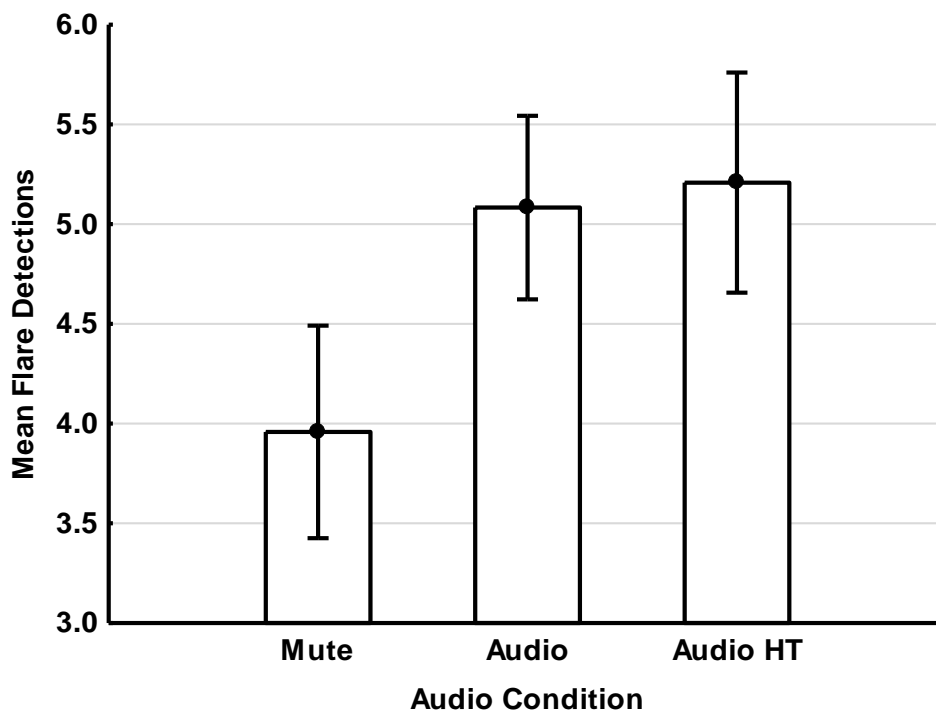


Figure 5-12. Mean number of detected flares by audio condition. Error bars denote within-subjects 95% confidence intervals.

5.5 Discussion

We explored the effect on aircraft navigation accuracy and head-up time for participants flying an aircraft simulator with the aid of a supplementary 3-D auditory display. Our findings indicate that two discrete 3-D audio sonifications can be used concurrently to facilitate the head-up monitoring of two interrelated aircraft navigation instruments.

5.5.1 Concurrent Monitoring of Interrelated Audio Signals

When correcting for deviations in flight path navigation accuracy, participants had to consider the aircraft's angle of approach, closing rate, and distance from the active leg of the flight plan. Only when situation awareness had been attained for these three attributes could a participant improve navigation accuracy by correctly adjusting the aircraft's attitude. All three attributes can be determined from the two audio navigation signals, which need to be considered concurrently as they are interrelated aspects of aircraft profile. Participants were able to effectively maintain awareness of the changing state of both CDI and WP readouts within the navigation display during head-up flight. This claim is further supported through the finding that no hunting effects were observed in the participants' flight path, as was otherwise expected if the CDI signal was being favored.

Given that both flight accuracy and head-up time improved when using the auditory display, we may conclude that the auditory display design worked effectively.

5.5.2 Head Tracking

Participants' head orientation remained forward facing throughout the study, which meant that even without the head tracker, the auditory display remained quite stable in its spatial alignment with the simulator. This design limited the usefulness of the head tracker to situations where the participant deliberately performed head movements for the sole purpose of improving front-back localising. Participants were reportedly confident in their ability to localize the front-back hemisphere of audio signals without the head tracker, while also claiming to have had little difficulty in discriminating between the two signals when they were located on the same lateral side. While the head tracker failed to improve task performance, it reportedly did provide noticeable improvement in hemisphere localising when participants deliberately turned their head. Audio and audio HT conditions were found to be statistically equivalent in all performance data with the exception of flare detections. These findings suggest that the auditory display adequately facilitated the localising of each signal's hemisphere of origin without the need for a head tracker, even when audio signals were located such that front-back errors were expected to introduce overlapping interference between signals.

5.5.3 Auditory Display Design

We adopted an original auditory display design feature into this study by collocating carrier notes with supplementary sounds. This was intended to assist the participant in determining the absolute value of information aurally encoded within each supplementary sound. All participants reportedly found this design to be helpful and indicated that the auditory display facilitated perceptive monitoring of the navigation information. The ability to perceive audio signal component sounds as a unified stream of information was reportedly very strong. Participants reported being confident in their interpretation of the audio cues when correcting course deviations.

The audio sounds were designed to accommodate a pleasant listening experience. As expected, feedback during post study interviews suggested that extended use of the display beyond the length of the study would be annoying to the listener. Given the relatively long period of time that participants listened to the display, this feedback suggested an adequate level of comfort was achieved.

5.5.4 Future Research

To build on findings outlined within this study, we recommend that:

- Further research be undertaken to quantify the utility of the reference pitch cue utilized in the current study. A localising study might be designed that requires participants to determine the final resting position of the signal after it has traversed from the midsagittal plane to different positions about the azimuth. A control condition might include the waypoint carrier note without the supplementary sound.
- The ability for a listener to maintain accurate perception of the current spatial location and state of sonifications should be explored under conditions when the auditory display is being periodically muted. These conditions would possibly be more representative of an applied use of the current design.
- Future studies should be conducted in more representative operational environments to explore the effect of increased head movements on the spatial perception of the auditory display. Such environments may require visual attention to be allocated toward more disparately located regions than the current study required.
- The degree to which voice communications and other operational audio alerts interfere with current sonifications should be explored further.

5.5.5 Application

Helicopter aircrews operating in active war zones often need to deviate from predetermined flight routes to avoid incoming enemy fire or respond to urgent operational requests. In these situations it is imperative that the crew maintain situation awareness for changes in the aircraft's 2-dimensional geographic position relative to the flight plan. This knowledge helps the crew avoid other known enemy positions and limits deviation from the intended flight path. The current research may find use as a supplementary cue to assist the crew in navigating while conducting high workload out of cockpit visual tasks associated with detecting and avoiding enemy threats.

The research may also find useful application as a supplementary navigation cue for unmanned aerial vehicle pilots during manual recovery. For some platforms, this task requires the pilot to fly by joystick while viewing video from the vehicle's forward facing camera. The pilot must also monitor navigation and system state information located on adjacent computer monitors. This high workload activity has the potential to degrade the pilot's awareness of the vehicle's spatial orientation when viewing the camera video. The use of an auditory display similar to the current design may provide the pilot with improved

head-up awareness of the vehicle's heading and deviation from the intended path toward the recovery site.

5.6 Conclusion

We demonstrated that two spatially discrete sonifications can be used to facilitate the head-up monitoring of two interrelated aircraft navigation instruments. Invariable reference pitch cues reportedly provided benefit in determining the absolute value of collocated variable pitch cues that were used to aurally encode navigation information. Even in the absence of head tracking, an effective use of localising cues prevented concurrent sounds located in the front and back hemispheres from interfering with each other. Our results support further research into the use of supplementary 3-D auditory displays within helicopter and unmanned aerial vehicles to assist pilots with 2-dimensional navigation while performing out of aircraft visual activities during emergency procedures.

5.7 Key Points

- Head-up time and aircraft navigation accuracy were improved through the introduction of an auditory display that aurally encoded waypoint and course deviation information.
- Interrelated aircraft navigation instruments were effectively presented for concurrent monitoring within a 3-D auditory display by means of two spatially discrete sonifications.
- An invariable reference pitch reportedly facilitated relative pitch perception for a collocated variable pitch, which encoded navigation data.
- Participants reported greater confidence when navigating the aircraft simulator with the auditory display enabled. They effectively made aircraft attitude adjustments to correct for course deviations utilizing information gleaned from the auditory display.

6. Study 4: Spatial Positioning of Verbal Audio within a Multi-Source 3-D Auditory Display

Journal review pending.

6.1 Abstract

Objective: Determine if the spatial positioning of verbal instructions associated with a working memory task effect a listener's ability to process that information while attending to a secondary task involving a flight simulator auditory display that competes for right cerebral hemisphere processing resources.

Background: Additional research is required to determine the existence of ear and lateralization effects on the processing of spatial audio information best suited for processing in the right (spatial) cerebral hemisphere.

Method: The spatial location of a verbal navigation task was manipulated within six different azimuth positions as the independent variable. Performance was measured on a secondary task that required participants to navigate a flight simulator utilizing encoded information delivered through two spatially discrete sonifications.

Results: Performance was better when the verbal instructions were delivered to the left ear compared with the right. Positions on the midsagittal plane produced results similar to the left ear. The forward left and forward right positions produced somewhat worse performance, while presentation toward the right ear produced the worst results over all.

Conclusion: Priming of the right cerebral hemisphere was considered likely to have aided the processing of information best suited for that hemisphere and supported a left ear advantage. Directional cue bias was considered a possible factor that contributed towards poorer performance in forward locations positioned away from the midsagittal plane.

Application: Pilots and air traffic controllers are often exposed to high workload conditions where visual and verbal tasks compete for right cerebral hemisphere resources. This study attempts to further knowledge into the applied effects of ear advantage and hemispheric lateralization in order to develop spatial auditory display designs that support improvements in the processing of competing information.

6.2 Introduction

Pilots and air traffic controllers regularly perform high workload tasks that involve attending to competing audio information. In an attempt to improve inter-stream segregation, 3-D auditory displays have been successfully used to virtually position disparate audio information toward different spatial locations (MacDonald, Balakrishnan, Orosz, & Karplus, 2002; Towers, Burgess-Limerick, & Riek, 2014). An enabling phenomenon being exploited in this endeavour is the cocktail party effect (Cherry, 1953; Pollack & Pickett, 1958), where compared with alternate dichotic or mono configurations, spatial cues are utilised to improve discrimination between streams and thereby enhance one's ability to monitor and attend to any one stream under competing conditions.

Spatial auditory displays require the synthesis of localising cues into an audio stream prior to its delivery through binaural headphones. Interaural onset timing and intensity differences between signals arriving at each ear provide the most robust spatial localising cues about the azimuth, while spectral distortions in the frequencies above 4 kHz introduce cues that aid in localising sound in elevation and toward the front or back (Yost, 2007). Spatial displays have been shown to improve speech perception in conditions of competing speech and ambient noise (Abouchacra, Breitenbach, Mermagen, & Letowski, 2001; MacDonald et al., 2002; McAnally & Martin, 2007). However, when the monaural quality of speech is already quite intelligible, such displays have been found to offer little benefit (Hawley, Litovsky, & Colburn, 1999; Yost, 1997).

Altering the spatial positioning of a sound within an auditory display has the potential to aid localising perception, however, the interface designer must be mindful that for any particular auditory display configuration, degraded performance may occur in a variety of unforeseen ways. For instance, when attending to a certain contextualised type of information, one cerebral hemisphere will often show dominant processing capability over the other. Doreen Kimura (1961, 1967) demonstrated this phenomenon when she discovered that a right-ear advantage exists when attending to speech. Speech delivered in a dichotic configuration was most effectively attended to when presented to the right ear. This finding contributed significantly toward an understanding that asymmetric, or lateralized functioning exists within the cerebral structure. Approximately 85% of right-handed people are known to exhibit a right ear advantage for verbal material (Wexler & Halwes, 1983). Kinsbourne (1973, 1974, 1980) defined an attentional model that accounts for lateralisation of cerebral processing, suggesting that a dynamic imbalance exists in the activation of the cerebral hemispheres. Kinsbourne's model suggests that anticipation of

verbal stimuli activates specialised regions for speech processing located in the left cerebral hemisphere and subsequently facilitates an involuntary bias of attention contralaterally toward the right ear.

While the left cerebral hemisphere is considered to dominate linguistic processing, the right hemisphere is generally accepted as being better suited for processing prosodic verbal stimuli (Iaccino, 1993). Altering the context of verbal stimuli or response method has been shown to reduce the magnitude of a previously present right ear advantage, or even produce a left ear advantage (Bryden & MacRae, 1988; Grimshaw, 1998; Techentin & Voyer, 2007; Techentin, Voyer, & Klein, 2009). Hiscock & Kinsbourne (1996) attributed the reduction in magnitude of an observed right ear advantage to right hemisphere activation when they asked participants to relate coded vowel-varied consonant-vowel-consonant syllables to bird sounds, rather than coded words. The spatialized processing capabilities of the right hemisphere in regard to language have been found to extend to written communication, where the ability to process hand written cursive script, rather than typed print has been found to be best attended to within the right hemisphere (Bryden, 1982; Iaccino, 1993; Springer & Deutsch, 1989).

Schwartz & Tallal suggest that the left hemisphere is possibly more dominant in speech processing due to its ability to process rapidly changing acoustic patterns (1980). More recent studies challenge the temporal reasoning for left hemisphere dominance when tested with temporal gap experiments (Carmichael, Hall, & Phillips, 2008), and even claim that the right hemisphere shows greater preference for pitch and temporal properties of sound within language (Scott & McGettigan, 2013). Uncertainties pertaining to lateralisation remain, however, there is evidence to suggest that both hemispheres perform complementary roles in both speech processing and comprehension (Holtgraves, 2012; Obleser, Eisner, & Kotz, 2008; Segalowitz & Cohen, 1989).

Priming of a particular cerebral hemisphere has been shown to influence an ear advantage within dichotic studies (Sætrevik & Hugdahl, 2007). Morris & Landercy (1977) primed participants' right hemisphere by asking them to retain musical melodies within memory. When asked to identify a target syllable pair from other pairs with differing consonants or vowels, a small non-significant left ear advantage emerged to replace the previously significant REA. Gadea, Espert, & Chirivella (1997) enhanced left ear recall for verbal dichotic stimulation by priming participants' right hemisphere with a left hand manipulospatial secondary task, which eliminated a previously observable right ear advantage. Kinsbourne (1982) believes that a priming effect can only be found when the secondary task is easy and does not impose interference with the primary task, while

Hiscock & Chipuer (1993) state that priming is dependent on interaural competition. There still remains uncertainty regarding the ability to replicate lateralization effects for ear advantages in situations other than dichotic testing.

Springer & Deutsch (1989) suggest the possibility for both a *direct access model* and an *indirect relay model* to account for lateralization. The first suggests that the hemisphere receiving the information will process it directly, otherwise, if the non-specialised hemisphere receives the information first, it will pass the information to the specialised hemisphere via the corpus callosum. In addition to conveying audio information, more recent studies suggest that the corpus callosum carries attentional-control signals (Pollmann, Maertens, von Cramon, Lepsien, & Hugdahl, 2002) and is involved in the excitation and inhibition of the cerebral hemispheres (2005). Imaging studies support the notion that the corpus callosum is involved in the allocation of audio attention toward the right ear, and routing of left ear signals to the left hemisphere (Westerhausen et al., 2009; Westerhausen et al., 2006).

Studies conducted into the factors that influence ear advantage have generally employed dichotic testing methods. More research is required to explore the external validity of these findings under conditions where spatial auditory displays present information of different form and context. The current study was conducted with the aim of determining if an ear advantage exists under dual task conditions that facilitate high workload demand on the right cerebral hemisphere. Our objective was to determine if altering the spatial location of right cerebral hemisphere contextualised speech would affect a listener's ability to attend to spatialized sonifications that were in competition for right cerebral hemisphere processing resources. The primary task required participants to attend to a working memory navigation activity where a series of verbal instructions in the form of "forward one, turn right, forward one", etc., guided the participant away from the centre of an imaginary grid. The goal was to determine which grid quadrant the task finished in, such as "forward right" in the current example.

Six different conditions were examined, which differed based on the spatial location from which the verbal navigation commands were presented. A single condition was examined in each block of trials. A secondary task required that participants navigate an aircraft simulator about a flight plan, relying heavily on two spatially discrete sonifications that provided encoded information regarding the aircraft's lateral deviation from the intended course and direction toward the next waypoint. Variability in the processing demand imposed by the positioning of the verbal instructions were measured as

performance on the secondary task. Measures included lateral deviation error from the intended flight path and mean angular error toward each waypoint.

It was expected that the secondary task sonifications would maintain the right cerebral hemisphere in a primed state. Given the familiar and predictable language structure utilised in the working memory task, a left ear, right cerebral hemisphere processing advantage may be observed. The centre position was expected to produce a similar level of performance or greater than that of the left ear. This was expected because the centre position was unlikely to introduce masking interference with the sonifications and would better facilitate potential signal routing optimisation from either ear to the appropriate cerebral hemisphere. The forward-left, forward-right, and rear positions were all expected to perform worst due to degraded spatial localising capability for those regions and possible proximity induced masking interference with sonifications.

6.3 Method

6.3.1 Participants

Twenty-four Boeing Defence Australia employees aged between 22 and 49 years ($M = 34$, $SD = 8$) participated in the study on a voluntary basis. The sample comprised 20 males and four females who each successfully completed a hearing screening test 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 standardized curve and dBA weighted curve. Participants were assessed for left- and right-handedness using a modified version of the Edinburgh Handedness Inventory (Oldfield, R. C., 1971). The assessment produced an inventory score, called a laterality quotient (LQ), which ranged from -100 (extreme left-handed) to +100 (extreme right-handed). Twenty-two participants were determined to be right-handed ($LQ > 75$), one participant mixed-handed (-19), and another left-handed (-75). Participants were asked to identify their preferred listening ear when talking on the telephone using the same scale as the Edinburgh Handedness Inventory. Twelve participants indicated an always right side preference, 10 indicated always left, and two did not have a strong preference for either side.

6.3.2 Experimental Design

The study employed a within-subjects, repeated measures design and a dual-task paradigm. The primary task was a working memory navigation activity that was guided by verbal instruction. The spatial location of verbal instructions was manipulated as the independent variable. A secondary task involved navigating a flight simulator about a

predetermined flight plan with the aid of an auditory display. Dependent variables included measures associated with accuracy in flight path navigation on the secondary task. Participants were randomly assigned into six groups of four participants each, which determined the ordering of presentation for the different conditions that were organised utilising a balanced Latin square methodology.

6.3.3 Working Memory Navigation Task

Participants were required to imagine themselves standing at the centre of a grid. Each trial comprised four navigation instructions that asked the participant to imagine themselves turning left or right, or moving forward or backward up to 4 positions. An example instruction is as follows: “Forward two, turn right, forward one, back three, Answer?” Participants were informed during an initial brief that both left and right turns were always 90-degrees. The forward and back commands were considered relative to the current direction in which the participant imagined himself facing, rather than the seated forward facing direction. When instructed to answer, participants were required to verbally identify the grid quadrant in which they had navigated to. Quadrants were identified as *a*, *b*, *x*, and *y*, and labelled on the wall over the simulator’s out of flight deck scene in a clock-wise, top-down direction starting in the upper left, as shown in Figure 6-1. In the previous example, the correct answer was the forward left quadrant, or ‘a’. Each new working memory trial began with the participant imagining themselves positioned facing forward in the centre of the grid.

6.3.4 Flight Simulator

Figure 6-1 shows the flight simulator used in the study, which comprised Microsoft Flight Simulator X controlled by a Saitek Pro Flight yoke control. The out of flight deck visuals were projected at a resolution of 1920 x 1080 pixels using a Dell 2400MP projector. The image was 180 cm x 110 cm in size and positioned at a distance of 180 cm directly in front of the participant. The simulator realism setting was configured to novice and the aircraft throttle was fixed at its maximum setting throughout the trials.

To the right of Figure 6-1 can be seen a Garmin 500 global positioning system (GPS) navigation display, which was centred on a 76 cm LCD monitor at a resolution of 1920 x 1080 pixels and positioned approximately 80 cm directly in front of the participant. The GPS displayed a moving map of the flight plan, which comprised 12-legs that were configured in a zigzag pattern of 25-degree turns. The scale of the flight plan remained zoomed out such that the participant could not accurately determine the aircraft’s distance and bearing from the intended flight path, thereby forcing attention toward sonifications

within the auditory display as a primary means of navigation. The visual display was utilised simply as a supplementary aid that enabled participants to spatially reorient themselves during more difficult trials that may expose the participant to excessive spatial processing workload demands. This safeguard was intended to mitigate against biasing between trials, where poor performance from a previous trial may otherwise continue on into the following trial.

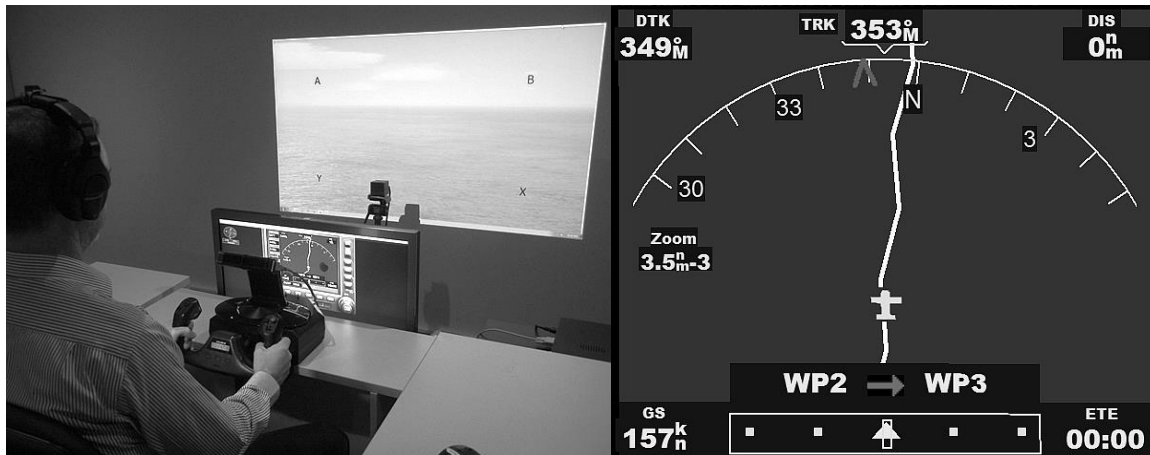


Figure 6-1. Flight simulator with GPS display. The GPS display was zoomed out to a distance such that fine navigation cues were not made available, thereby forcing the participant to utilise the auditory display. The GPS was only intended as a backup display to reorient the participant should he become disoriented.

6.3.5 Auditory Display Design

The auditory display utilized in this study was previously designed by Towers, Burgess-Limerick, & Riek (2014) as a navigation aid to facilitate head-up monitoring of an aircraft's waypoint bearing (WP) and course deviation indicator (CDI) readouts. The auditory display employed two concurrent, yet spatially discrete sounds to represent the WP and CDI. A waypoint refers to the geographical junction between two consecutive legs of a flight plan, or simply the intended direction in which the aircraft should be heading in accordance with the set flight plan. The WP sound was unconstrained in its ability to transit 360 degrees about the participant as it maintained spatial alignment toward the waypoint as the aircraft deviated from its direction.

The CDI readout indicates the lateral direction and distance in which the aircraft deviates from the current leg of the flight plan. The CDI auditory signal was presented stationary toward the left or right at +/- 140 degrees, and muted if the error was less than 10 m from the intended leg. The auditory navigation display design is illustrated in Figure 6-2.

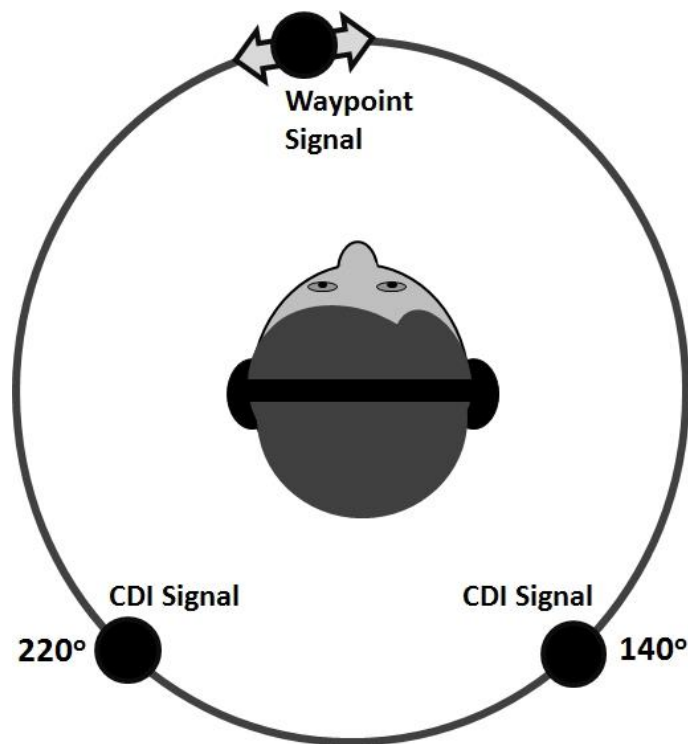


Figure 6-2. Auditory display navigation signal layout. The waypoint and CDI signals are shown. The WP was always presented and could rotate 360-degrees about the participant at a virtual distance of 1.5 meters. The CDI signal was only presented if the lateral deviation error was more than 10-meters from the intended leg of the flight plan. The CDI signal remained stationary at a distance of 1.5 meters and was only presented in one of the two positions shown, which indicated the lateral direction of deviation away from the intended flight path.

Both audio signals comprised two sounds, a pitch invariant *carrier note* and a variable *supplementary sound*. The carrier notes were intended to help segregate the auditory streams for the two signals by providing unique timbre properties and rich frequencies above 4 kHz that aid with front-back localising. The WP carrier sound included a C#3 note that was sampled from a muted trumpet, while the CDI carrier note utilized a B^b2 note sampled from a baritone sax.

Both signals' supplementary sounds were variable pitch square wave tones, which were used to encode navigation information through variations in pitch and spatial position as the aircraft deviated from the flight path. The CDI direction information was mapped to the spatial position (140-deg left or right) of the CDI sound, while distance and rate of change were mapped to changes in pitch relative to a co-located pitch invariant reference note. A more comprehensive explanation of the design rationale and testing of this navigation display is detailed in section 5-3-3 of this thesis.

A head tracker was utilised throughout the study to maintain spatial alignment between the auditory display and the forward facing direction of the simulator. In doing so, the auditory display adjusted the position of the auditory signals to compensate for movements in the participants head position away from a forward facing orientation. Figure 6-3 shows the auditory display layout for the six spatial locations where verbal instructions for the primary task were positioned.

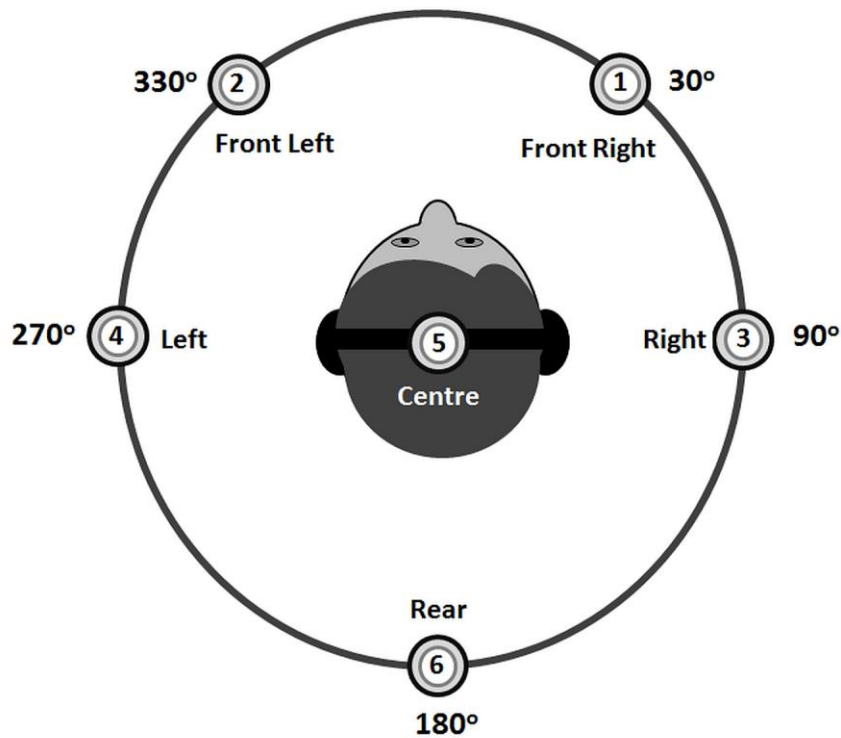


Figure 6-3. Spatial positioning of working memory verbal instructions.

During the design phase of this study, five people assisted with evaluating the primary task. They were each provided with an overview of the primary task methodology and a diagram of Figure 6-3 to illustrate the intended spatial position of verbal instructions. They each sat at a desk in a quiet room with headphones on and were presented with an automated presentation of the 12 different trials of working memory task. They did not perform the secondary flight navigation task and recorded their responses on a piece of paper, rather than verbally responding as per the final methodology. Each person reportedly found the task moderately challenging. Two participants made one error, one participant made two errors, and the final two participants made no errors. When asked during the debrief interview, none of the participants reported any difficulty in attending to speech at any location.

6.3.6 Procedure

Prior to commencing the experiment, a briefing was conducted which included informing the participant that the working memory navigation task was the highest priority and that they should neglect flying the simulator as needed to ensure the successful completion of that task. Participants were also instructed not to adopt a strategy whereby they immediately neglect the secondary task in favour of the primary task, but rather to attempt to perform both tasks concurrently.

Participants sat at the simulator and adjusted the seat height and the volume of the auditory display to a comfortable level. A Fovio eye tracking system was used to track each participant's gaze throughout the experiment and record how often the GPS display was attended to. The eye tracker was calibrated at this point prior to commencement of the trial.

In their own time, participants then initiated the first trial by pressing the p-key on a computer keyboard positioned behind the yoke controller. Guided by the auditory display, the participant then proceeded to fly the aircraft by manipulating the yoke controller in order to continuously minimize any deviation from the flight plan. The time taken to transit through each leg of the flight plan totalled approximately 24-seconds. The audio files delivering the working memory navigation instructions were automatically initiated one-second into each leg of the flight plan and took approximately six-seconds in total to deliver the verbal instructions. Participants could take their time in responding to the working memory task, but generally responded within a few seconds following the instructions. The remainder of the leg would take approximately 17-seconds and was devoted to correcting any course deviation resulting from the working memory task.

At the end of the flight plan, which comprised 12-legs, the simulator would automatically pause, reposition the aircraft back to the start of the flight plan, and reconfigure the auditory display to reposition the verbal instructions toward the next condition location. The participant then turned to face a laptop computer positioned on a desk beside the simulator to complete a software version of the NASA Task Load Index (TLX) (Cao, Chintamani, Pandya, & Ellis, 2009). Once complete, the participant turned back toward the simulator and in their own time initiated the next trial by pressing the p-key on the keyboard. Once the six conditions were complete, a post experiment interview was conducted to record participants' thoughts on the auditory display design and indicate which positions for the verbal instructions seemed to offer the best working memory task performance.

6.3.7 Data Analysis

Data analysis were undertaken as a one-way RM-ANOVA (six-level) performed on the statistical software package Statistica (version 12) using a significance level identified at $\alpha = .05$. Subsequent post-hoc analyses were undertaken for all significant main effects using Bonferroni adjusted alpha levels. Effect sizes were expressed using Cohen's *d*. Within-subject confidence intervals were constructed by eliminating intersubject variance using the method outlined by Loftus and Masson (1994).

An inferential test of equivalence in the form of two simultaneous one-sided hypothesis tests (Rogers et al., 1993) was used to determine the extent to which auditory display conditions could be considered practically equivalent. This test was used to supplement traditional hypothesis testing when audio conditions were not found to differ significantly. The equivalence test null hypothesis asserts that the difference between two groups is at least as large as a predetermined equivalence interval. As in traditional hypothesis testing, the goal is to reject the null hypothesis and accept the alternative hypothesis. An equivalence interval of $\pm 10\%$ was defined around a difference of zero. Any difference between the two groups small enough to fall within this interval was considered practically unimportant. Therefore, if the audio condition mean fell within the $\pm 10\%$ interval set around the second condition's mean, the null hypothesis was rejected and the alternative hypothesis was accepted. Workload was self-assessed utilising a software version of the NASA Task Load Index (version 2.1.2).

Data analysis was limited to the first 15-seconds following the onset of each leg of the flight plan. This time included the presentation of the working memory instructions and a further 8-seconds following when flight navigation performance was found to be most disrupted by the primary task. Over the remaining 9-seconds of each leg, participants demonstrated an ability to correct their flight performance and settle in preparation for the next leg.

6.4 Results

Participants were proficient in performing the primary working memory task. Table 6-1 provides a summary of the mean results achieved on the task for each audio spatial condition. ANOVA conducted on the results found no main effects ($F(5, 110) = 1.613, p = .163$).

Table 6-1.

Mean correct working memory task responses by verbal audio spatial condition.

	Condition					
	1	2	3	4	5	6
<i>M</i>	10.9	10.9	11.5	11.4	11.3	11.5
<i>SD</i>	1.3	1.0	0.9	0.9	1.4	0.6

Note. Results show the mean number of correct working memory task responses out of a possible 12.

Figure 6-4 plots the total head-down duration in seconds by verbal audio spatial condition. This measure recorded the time spent visually attending to the GPS navigation display. A main effect was not observed and no conditions were found to be practically equivalent ($F(5, 115) = 2.287, p = .051$).

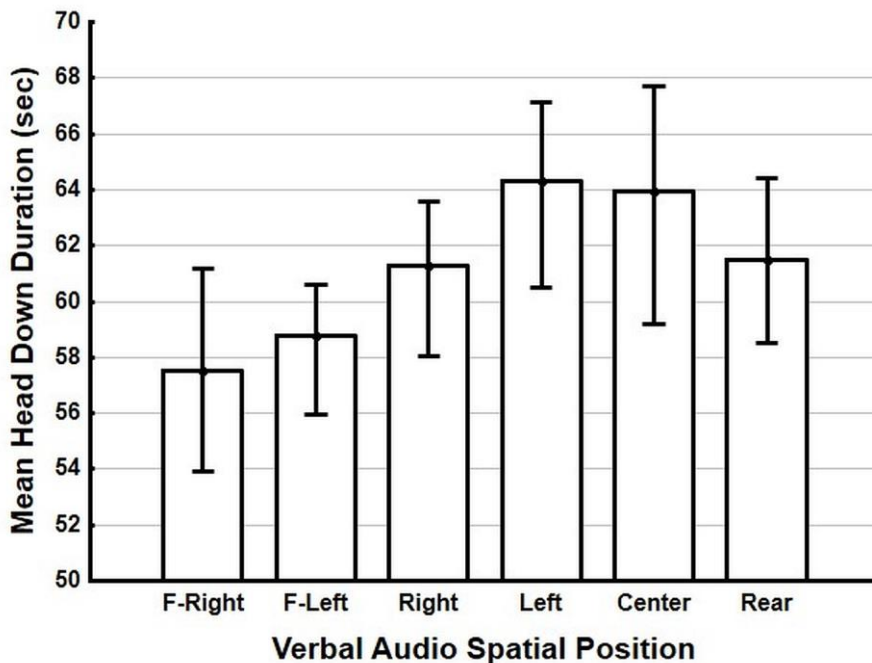


Figure 6-4. Mean time attending to GPS by verbal audio spatial position. No main effects were observed. Error bars denote within-subjects 95% confidence intervals.

Figure 6-5 plots the mean course deviation error (meters) by verbal audio spatial condition. A main effect was observed, with pairwise comparisons revealing a large improvement in navigation performance for the centre condition when compared with each of the front right, front left, and right conditions. Performance was substantially worse for the right condition when compared with each of the left, centre, and rear conditions.

Practical equivalence was observed between the front-right and front left; front-left and right; and left and rear conditions.

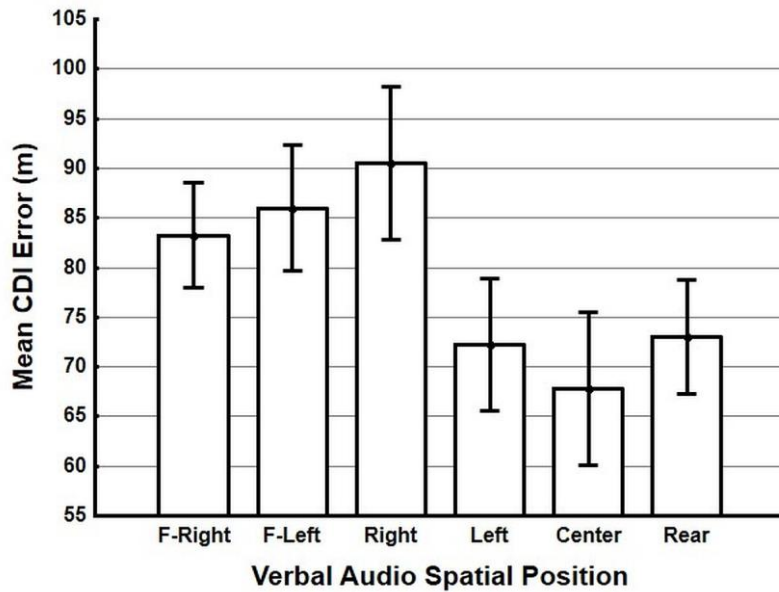


Figure 6-5. Mean CDI error by verbal audio spatial position. Error bars denote within-subjects 95% confidence intervals.

Figure 6-6 plots the mean waypoint error by verbal audio spatial condition. A main effect was observed, with pairwise comparisons for the right condition revealing significant differences when compared with each of the left, centre, and rear conditions. The front-right condition was found to be practically equivalent to the front-left. The left, centre, and rear conditions were also found to be practically equivalent.

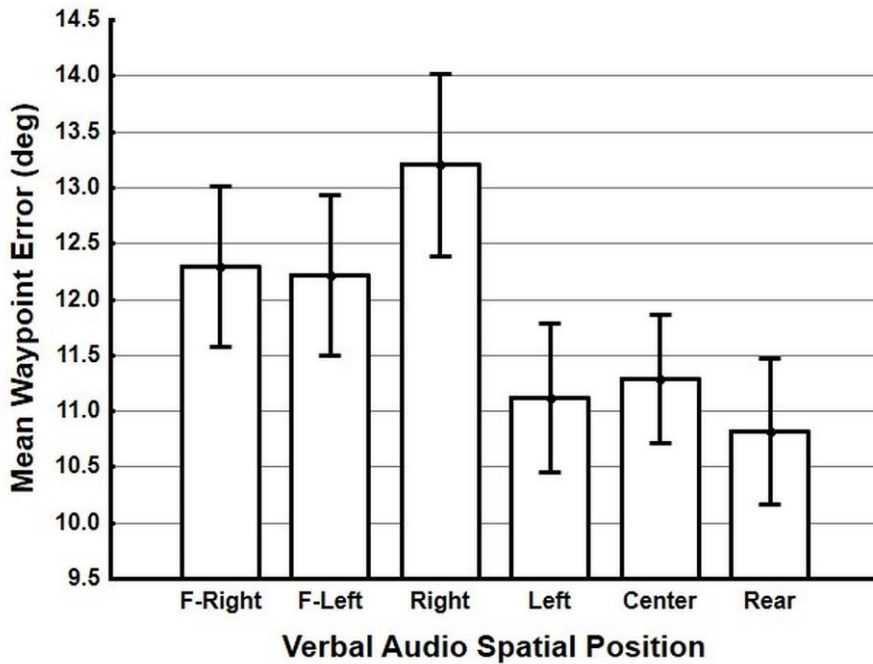


Figure 6-6. Mean waypoint error by verbal audio spatial position. Error bars denote within-subjects 95% confidence intervals.

No significant main effects were observed for any of the workload measures reported in the NASA TLX. TLX Score ($F(5, 115) = 0.712, p = .616$), Mental ($F(5, 115) = 0.60, p = .703$), Performance ($F(5, 115) = 1.76, p = .126$), Temporal ($F(5, 115) = 0.456, p = .808$). Figure 6-7 plots the TLX Mental score by condition.

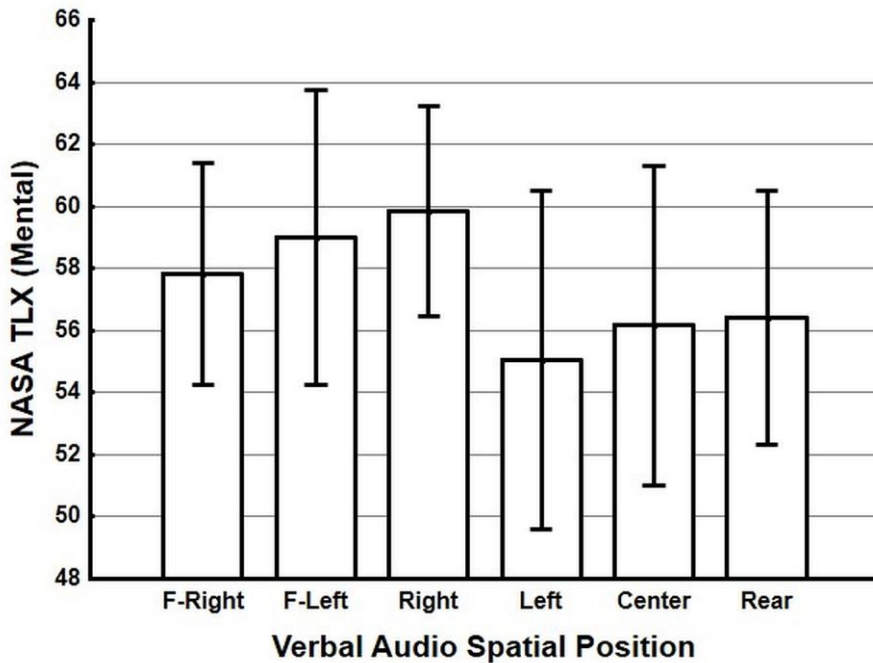


Figure 6-7. TLX scores by verbal audio spatial position. Error bars denote within-subjects 95% confidence intervals.

Tables 6-2 and 6-3 provide a summary of results for the statistical analysis conducted on all performance measures.

Table 6-2.

Summary statistics for verbal audio conditions grouped by performance measure.

Condition	RM-ANOVA				Bonferroni					Cohen's <i>d</i>		
	<i>F</i> (5,115)	<i>p</i>	<i>M</i> (<i>SD</i>)	95% CI	F-Left	Right	Left	Center	Rear	Left	Center	Rear
Course Deviation (meters)	6.599	.000										
F-Right			83.23 (12.55)	[88.25, 78.21]	1.0	1.0	.433	.035*	.622			1.4(L)
F-Left			85.99 (14.92)	[91.96, 80.02]		1.0	.099	.006*	.151			1.6(L)
Right			90.50 (18.29)	[97.82, 83.18]			.005*	.00*	.001*	1.5(L)	1.8(L)	1.6(L)
Left			72.26 (15.83)	[78.60, 65.92]				1.0	1.0			
Center			67.79 (18.10)	[75.03, 60.55]					1.0			
Rear			73.01 (13.49)	[78.41, 67.61]								
Waypoint Error (degrees)	6.006	.000										
F-Right			12.29 (1.70)	[12.97, 11.61]	1.0	1.0	.399	.842	.084			
F-Left			12.21 (1.70)	[12.89, 11.53]		.872	.581	1.0	.130			
Right			13.21 (1.93)	[13.98, 12.44]			.002*	.005*	.000*	1.7(L)	1.7(L)	2(L)
Left			11.12 (1.58)	[11.75, 10.49]				1.0	1.0			
Center			11.29 (1.35)	[11.83, 10.75]					1.0			
Rear			10.82 (1.55)	[11.44, 10.20]								

Note. CI = Confidence interval; (L) = large effect size;

* *p* < .05

Table 6-3.

Summary statistics for equivalence between verbal audio conditions grouped by performance measure.

Condition	F-Left			Right			Centre			Rear		
	SE	z	p	SE	z	p	SE	z	p	SE	z	p
Course Deviation (meters)												
F-Right	3.98	3.62	.013*	4.53	.97	.166						
F-Left				4.82	1.69	.045*						
Left							4.91	2.97	.124	4.25	2.21	.014*
Centre										4.61	.928	.177
Waypoint Error (degrees)												
F-Right	.490	3.34	.000*	.525	1.53	.062						
F-Left				.525	1.36	.087						
Left							.425	3.28	.001*	.453	4.01	.003*
Centre										.420	2.65	.004*

Note. * $p < .05$

6.5 Discussion

The spatial position of verbal navigation instructions was manipulated in order to explore its effect on a secondary task that involved navigating an aircraft using spatially positioned sonifications encoded with navigation information. We found that the position of the verbal communications did have an effect on secondary task performance. Improved performance was observed when the verbal communications were presented at the centre of the interaural axis on the midsagittal plane, or toward either the left ear or the back on the midsagittal plane. The front left, front right and right conditions produced the worst results. The following discussion considers the results in terms of ear advantage and lateralization effects.

We begin our discussion by considering the significant difference in performance observed between the left and right ear conditions, which indicate the presence of a left ear advantage. We had anticipated that this result may occur as it was encouraged by the experimental design, which facilitated an established priming of the right hemisphere, and included a simple and limited range of language within the primary task. Approximately 13 participants made similar comments inferring that the right ear condition was more difficult

to attend to than any other condition. Most participants believed that during the right ear condition they had to occasionally ignore the secondary task completely in order to allocate enough resources to ensure that they achieved a correct response on the primary task. The high workload and spatial context associated with both primary and secondary tasks has not been conducive for optimal performance to occur in the right ear, left hemisphere processing condition. We can therefore cautiously reason that we have observed an instance of left ear advantage facilitated through conditions similar to other previously described examples of right hemisphere priming.

Further consideration is required before attributing right hemisphere priming as a contributing factor to the observed left ear advantage. As mentioned previously, the presence of a competing interaural stimulus (Hiscock & Chipuer, 1993), and the secondary task being easy and not imposing interference with the primary task (Kinsbourne, 1982), are criteria believed to be required for priming to occur. The secondary task within this study was considered to be very difficult, which possibly challenges this claim. Perhaps the nature of the secondary task meets the aforementioned criteria because participants were encouraged to neglect the secondary task as required to effectively complete the primary task. Any resultant deviations from the flight plan could be corrected in a timely manner after completing the primary task. So, while the task was difficult, there perhaps existed an ability to allocate attention adequately between tasks to maintain a degree of ease such that the secondary task was not relatively difficult. Competing interaural stimulus would only have been present periodically, depending on the configuration and state of the display at any particular time. These criteria were originally defined in the context of dichotic testing, and therefore require further consideration for their relevance within dynamic spatial displays.

The magnitude of difference observed between the left and right conditions was unexpected, as was the right ear consistently having the worst results compared with all other conditions. If a left ear advantage exists for reasons of hemisphere priming and context, the improved performance observed within the centre and back positions may be attributed to the audio being available at both ears, thereby possibly enabling the dynamic routing of signals via the most effective path to cerebral processing faculties. Given the current findings regarding the role the corpus callosum plays in this function, it is possible that these conditions offer such an advantage.

The improved performance found for conditions located on the midsagittal plane (centre and back) may be attributed to a directional cue bias that may degrade performance in conditions where verbal instructions were located toward the left or right.

Lee (2010) found that directional cues presented to the corresponding ear result in fewer errors on a navigation task. Given the counterbalanced design of the current study, this effect alone should have degraded both left and right conditions equally, but not have biased poorer performance for the right ear over the left. Perhaps positioning the verbal instructions anywhere on the midsagittal plane offers benefits in counteracting such effects. We can only speculate regarding the impact of directional cue bias on this study, however, a possibility is that all positions not located on the midsagittal plane suffer some degree of directional cue bias. Lateralisation and hemispheric priming may be far more influential in regard to the overall performance observed with the left and right ear conditions.

The front left and front right conditions were expected to facilitate the least effective performance partially due to the potential for proximity interference to occur with the secondary task sonifications. Studies have previously acknowledged the benefit in positioning verbal communications about the interaural axis, particularly toward the side rather than the front to mitigate such issues (MacDonald et al., 2002). It is difficult to reason the practical equivalence found between the two front conditions when there exists such a significantly large effect occurring between the left and right conditions.

Head-down data indicate total time spent observing the GPS navigation display. As mentioned previously, the GPS display was scaled such that it only offered benefit to participants in circumstances where they found themselves experiencing considerable disorientation. While no significant effects were observed for head-down behaviour, the general trend in the data seems counter intuitive given the results and feedback from participants. Disorientation was only observed twice during the study and subsequent analysis of the data suggest that both participants overcame the situation adequately without impacting subsequent trials. The data suggest a trend whereby attention was being allocated to the GPS more often in conditions of best performance. Given the scaling of the display, best performance required adequate attention be allocated to the sonifications, not the visual display. We suggest that the optimal configurations accommodated performance to an extent that spare resources were available. Participants may have implicitly scanned the visual display more frequently in an attempt to attain additional information regarding distance to run on the current leg, and direction of the next turn.

Several people commented that the rear condition sounded lower in volume than the other conditions. This is possibly due to the increased externalisation experienced for sounds positioned toward the rear of the listener. Given the already favourable results

obtained in the rear condition, any future studies should enable independent volume control for each condition. Perhaps if the volume was adjusted by condition, we may have observed even greater improvements in the rear condition. During initial testing of the experiment design, differences in perceived volume between conditions was reviewed and determined not to be noticeable. The additional workload experienced when performing the study activity may have made participants more sensitive to perceived volume.

6.5.1 Reliability of Participant Reporting Methodology

While no significant effects were observed in the NASA-TLX data, the general trend in workload matches expectations based on the secondary task performance. It is considered likely that the TLX self-reporting method was not sensitive enough to detect changes in workload between conditions. Most participants commented on the noticeable difference in workload between conditions, particularly for the left and right. However, the ability for participants to reliably recall their performance during the post study interview and during the TLX rating should be considered further. When asked if there were certain locations where the verbal audio was best attended to, one participant commented that he didn't realize that the position of verbal audio had changed throughout the entire session. The automated configuration of the spatial rendering was subsequently checked and found to be working appropriately. Six participants commented that they detected when the location of verbal instructions changed position, but could not recall which positions facilitated the best performance. Hiscock & Kinsbourne (2011) similarly observed a consistent dissociation between subjects ability to detect signals at a specified ear and their ability to identify the ear of entry for signals that have been detected.

6.5.2 Future Research

Previous studies focussing on factors that influence ear advantage and lateralization have employed dichotic testing methodologies. More research is required to determine how a potential ear advantage may be supported or degraded through the spatial positioning of audio. A left ear advantage was not observed between the front left and front right conditions, which could be an artefact of the display design where sonifications degraded performance for those conditions; or possibly because audio from the forward conditions was received at both ears to an extent that compromised an ear advantage.

The current experiment utilised a non-individual HRTF that is known to facilitate worse sound localising performance than would an individual's own cues. This HRTF may have therefore effected speech localising performance at the right ear to an extent that

introduced a left ear advantage. Follow up studies should either employ individualised HRTFs or mirror left-right HRTF cues so that any localising error is counterbalanced. As such, any ear advantage that may have been introduced as an artefact of the HRTF will disappear. As mentioned previously, in the absence of the secondary task, an initial evaluation of the primary task failed to expose any difficulty for participants to localise or attend to speech at each intended location compared with other locations.

Limited understanding exists regarding the optimal spatial layout for auditory displays that contain multiple instances of competing speech. If speech can be determined as being more effectively processed in the left or right cerebral hemisphere, factors that facilitate ear effects need to be better understood such that they guide the spatial positioning of speech. Further studies should be undertaken to explore this spatial effect.

More understanding is required regarding the potential for directional bias effects to occur in spatial auditory navigation displays. Such knowledge is crucial for designing future displays where competing speech may necessitate the spatial repositioning of speech toward locations that introduce potential directional bias effects on performance.

Current thought that defines hemisphere priming criteria within the primary literature, such as the need for a competing interaural stimulus (Hiscock & Chipuer, 1993), and an easy secondary task that does not impose interference with the primary task (Kinsbourne, 1982), need to be further verified within studies that employ different audio configurations other than the dichotic presentations utilised in previous studies.

6.5.3 Application

Most modern computing platforms have adequate processing capability to accommodate the local rendering of a 3-D auditory display. This capability now facilitates cost effective utilisation of spatial audio to declutter concurrent audio and alleviate an often overburdened visual modality by presenting more information through auditory displays. The current study may help focus research toward factors that affect auditory display design in applied configurations. Our finding that an ear advantage was present in a binaural configuration should focus future research toward identifying spatial auditory display configurations that improve processing of information that differs in form and context. Future findings may help guide the configuration of auditory display layouts to suit the processing of contextually specific information. Environments such as command and control consoles, air traffic control, and aircraft flight decks may utilise such knowledge to accommodate more effective application of concurrent sonifications and speech audio.

6.6 Conclusion

Changing the spatial position of verbal navigation instructions effects the ability to attend to that information while processing sonifications that communicate encoded navigation error on a secondary flight simulator navigation task. A left ear advantage for processing verbal instructions was observed over the right, while similarly effective performance was also observed for other positions located on the centre and rear of the midsagittal plane. Front left and front right positions were equally poor, with the right ear condition facilitating the worst performance over all conditions. These findings were attributed to having established priming of the right cerebral hemisphere, which aided the processing of information best suited for that hemisphere. Our results support the notion that ear advantage can be present in audio configurations other than dichotic. Further research is required to determine the factors that impact the processing of cerebral hemisphere contextualized information under conditions of competing audio streams within spatial audio display configurations.

7. Summary and Discussion

7.1 Aims and Scope of the Thesis

This thesis set out to develop spatial auditory display design solutions that improve a listener's ability to attend to simultaneous streams of sonifications and verbal dialogue. A primary aim was focussed on understanding how to design encoding methods for sonifications that also improve sound localising and stream segregation. Another aim was to identify if cognitive processing effects regarding ear advantages are observed when altering the location of speech delivered concurrently with similarly contextualised sonifications. These designs target applications within flight decks where visual monitoring tasks often expose pilots to excessive workload demands and constrain their attendance to concurrent head-up activities.

The first study is detailed in chapter three and was undertaken to determine if listeners could localise sound rendered through SLAD3D with its non-individual HRTF to a similar degree of accuracy that was obtained in previous binaural localising studies. Non-individual HRTF cues are known to promote front-back hemisphere localising ambiguity and degraded localising performance. It was therefore important to determine the degree to which these issues were prevalent at different points of azimuth and elevation with the current setup. The study was intended to confirm that the spatial audio rendering solution was suitably capable of supporting subsequent studies that focussed on developing design solutions that address these problems. Resultant data from this study was expected to provide a baseline of performance expectations and inform design direction for future display designs.

Chapter four provides an overview of study two, which explored the utility of introducing supplementary reference sounds to improve the localising of a target sound's azimuth position. Cues were designed to provide the listener with relative reference from the target sound to the midsagittal plane and interaural axis. The aim was to reduce front-back hemisphere ambiguity and improve azimuth localising for target sounds. This focus was considered particularly relevant for auditory displays that are not head tracking enabled. Head tracking is known to provide the listener with cues that aid with improving or resolving these localising uncertainties. Supplementary cues were expected to provide azimuth reference cues that substitute head tracking cues.

Chapter five details the design and evaluation of an applied auditory display that presented real time aircraft navigation error to the pilot. The primary aim of the display

was to facilitate the head-up monitoring of aircraft visual navigation readouts. Information regarding the aircraft's relative heading and lateral deviation from a set flight path were encoded into spatially discrete sonifications. Additional sounds were collocated with the sonifications in an attempt to improve stream segregation between concurrent sounds. The sonifications themselves were designed such that the methods for encoding the navigation information were also expected to contribute toward improved localising and stream segregation.

The fourth study is detailed in chapter six and focusses on the final objective of this thesis. That was to determine the most effective spatial position for verbal audio within the auditory navigation display designed in the previous study. The context of verbal information was spatial and therefore expected to compete with existing sonifications for processing resources located within the right cerebral hemisphere. The primary objective of this study was to improve knowledge regarding how to effectively position verbal audio to support stream segregation and the concurrent cognitive processing of those streams. Known effects regarding ear advantage, cerebral hemisphere priming, and directional cue bias are considered in the design and subsequent findings.

7.2 Summary of Key Findings

The following sections summarise specific findings from each of the studies undertaken throughout this thesis.

7.2.1 Key Findings from Study One

The aims of study one regarding the assessment of SLAB3D and its non-individual HRTF were achieved with sound localising results comparing well with previous studies. Azimuth localising accuracy was found to be best in the lateral regions, with the front then back regions producing the worst performance. Sounds were localised in azimuth more accurately when positioned at higher elevations. Front-back hemisphere localising errors were significantly worse in the front region compared with the back. The back 0-degree elevation condition produced significantly less front-back errors than all other regions and elevations. The -20-degree elevation condition produced the worst number of front-back errors of any region or elevation. The use of a non-individual HRTF likely caused localisation estimates to skew toward the interaural axis on the ipsilateral side of the mid-sagittal plane. Confidence was elevated in the back region, which was attributed to poor localising cues causing internalisation of the sound. The localising performance achieved with the use of SLAB3D and its non-individual HRTF was considered adequate to support future studies. The knowledge gained into localising performance occurring in azimuth

regions provided a baseline of expectations that informed design decisions in subsequent studies.

7.2.2 Key Findings from Study Two

The second study found a significant reduction in front-back hemisphere localising errors occurred due to cues made available through a supplementary sweep sound that transited back and forth about the listener in a 180-degree arc. Additional momentary accent tones that were activated when the transiting sound passed through the midsagittal plane and interaural axis reportedly helped the listener localise those bearings. Confidence was found to increase with the provision of these supplementary cues, particularly when the target sound was located toward the back region. Azimuth localising accuracy was not found to improve with the introduction of supplementary cues.

The study found that participants were able to accurately transpose auditory display azimuth bearing information to a vertically oriented visual display. This means that a sound located at 45-degrees in the more perceptive azimuth plane would be represented on a computer monitor in the upper right of the display at a 45-degree angle from its centre. This finding is important because localising accuracy for a sound's elevation is unreliable to an extent that it is impractical to use elevation bearing to encode information such as aircraft navigation error. It is likely that future azimuth oriented auditory display designs may be considered less constrained by perception problems associated with differing axis orientations between visual and auditory displays than perhaps previously thought.

The supplementary sounds were found to be distracting and disorienting to some listeners. In order to reduce the workload associated with attending to supplementary cues, future designs should include less salient cues that are more closely integrated into the target sound. It also became apparent at the time that very accurate and low cost head tracking technology was becoming more accessible, which therefore meant that future spatial auditory displays would not have to compromise their complexity in order to mitigate potential hemisphere localising ambiguity. The significant reduction in front-back hemisphere localising confusions alone were not considered large enough for the design in its current form to have adequate applied value. This conclusion was supported by the finding that a reportedly excessive level of additional workload was being introduced by the supplementary cues.

7.2.3 Key Findings from Study Three

The third study successfully employed the use of two concurrent sonifications to enable the head-up monitoring of two interrelated aircraft navigation readouts. The aircraft's heading and lateral deviation from the intended flight path were successfully encoded into two spatially discrete sonifications. The variant pitch supplementary cues successfully communicated direction and rate of change for navigation information. The use of a collocated invariant carrier sound aided with stream segregation and resolving front-back confusions. It was considered very likely that in the absence of such cues, energy and informational masking effects would have been problematic. In general, the design was found to reportedly facilitate inter-stream segregation very distinctly, which contributed to the successful mitigation of front-back confusions and masking effects. This claim is considered to be supported through the finding that participant's performed equally well with absent head tracking cues, which aid in resolving front-back hemisphere localising confusions and stream segregation.

Informed primarily by the auditory display signals alone, participants could accurately alter the aircraft's profile to correct navigation errors, thereby allowing significantly more time to be devoted to an out of flight deck visual search task. These findings are important as they demonstrate that concurrent sonifications can be used to convey high fidelity spatial information and facilitate the head up monitoring of interrelated visual displays. This strategy helped reduce reliance on an already overburdened visual modality during multitasking conditions. Multiple sonifications have been used concurrently in previous auditory displays, however, limited research to date has focussed on the use of concurrent spatially discrete sonifications.

7.2.4 Key Findings from Study Four

The final study examined the ability for a listener to process speech delivered from different locations while secondary task sonifications of similar contextualised information compete for right cerebral hemisphere processing resources. The spatial location of the talker was found to have an effect on the listener's ability to process that information. A left ear advantage for processing speech was observed when compared with the right ear. A similar level of improved processing ability was observed for other locations situated along the centre and back regions of the midsagittal plane when compared with both front left and front right positions.

The secondary task in this study was thought to have facilitated a condition whereby the right cerebral hemisphere was constantly primed. Hemisphere priming suggests that

an active cerebral hemisphere provides improved performance toward enabling the processing of subsequent information of similar context (Freyman et al., 2004; Hiscock & Chipuer, 1993; Kinsbourne, 1982). This likely improved the processing of information best suited for that cerebral hemisphere, which includes speech information of a spatial context delivered to the left ear.

This study differed from previous studies in that it did not use a dichotic display to explore the presence of an ear advantage. To the best of the authors' knowledge, this is the first time that an ear advantage has been observed in a binaural listening configuration. This finding has significant implications regarding the way in which future auditory displays might be configured. The study contributes toward enabling the development of more complex auditory displays by improving our knowledge regarding how the spatial positioning of speech optimises cerebral hemisphere processing for information best suited for a particular cerebral hemisphere.

7.3 Limitations of Current Research

7.3.1 *Localising in Elevation*

Problems associated with accurately localising a sound's elevation are well known and mentioned previously. The scope of this thesis has been deliberately constrained toward developing design improvements for spatial auditory displays that are limited to presentation about the horizontal plane. This is considered an important first step toward developing applied auditory display design solutions. Findings from this thesis support the claim that such displays are suitable for a broad range of ambitious applications without requiring elevation as a design attribute or encoding technique.

7.3.2 *HRTF Individualisation*

During the development of studies contained within this thesis, no attempt was made to modify the non-individual HRTF database in order to achieve a closer match between integrated localising cues and those of each participant. This was considered appropriate because a general aim of the thesis was to develop auditory display design solutions that improve localising accuracy for sound containing poorly generalised HRTF cues. It is considered likely that techniques for individualising HRTF cues will continue to evolve rapidly, however, a need will remain for knowledge into auditory display design methods that support localising, regardless of HRTF quality.

A variety of HRTF cue optimisation techniques are currently being proposed and reported on within the primary literature. These range from simply providing the listener

with several HRTFs to subjectively evaluate; to allowing the listener to actively adjust different properties of the HRTF in order to resolve a solution that provides the best spatial perception (Runkle, Yendiki, & Wakefield, 2000). Current approaches toward individually tailoring HRTFs through the use of anatomical measures of the pinnae are providing encouraging improvements in localising performance (Zotkin, Duraiswami, & Davis, 2004). New automated HRTF selection methods are being developed that aim to select an appropriate HRTF based on contour measures taken from a 2D image of the listener's ear. Such approaches are claimed to improve elevation performance by 17% (Geronazzo, Spagnol, Bedin, & Avanzini, 2014).

7.3.3 Context of Information

As discussed in the final study detailed in chapter six, the spatial location of verbal audio can affect the listener's ability to process that information. This is considered partly influenced by the informational context of the audio and which cerebral hemisphere is best suited to process that type of information. For example, a right ear advantage for speech is considered to be facilitated partly because of the larger neural connection from the right ear to the left cerebral hemisphere, where speech is most effectively processed. All speech and non-speech audio utilised in studies throughout this thesis contained information of a spatial context, which in most people is processed more effectively in the right cerebral hemisphere. The external validity of findings from this thesis are therefore constrained to information of a spatial context and not immediately applicable when considering audio information of differing context. Speech that requires abstract or qualitative interpretation, such as a word reordering task, might fall into this category and perhaps be better processed by the left cerebral hemisphere through a different spatial position of the audio than results from this thesis suggest.

7.4 Future Research and Practical Implications

Applied auditory interfaces largely remain limited to simple alerts that often appear as a cacophonous chorus of beeps and buzzers. The overarching goal of this thesis was to develop perceptive and versatile spatial auditory display design solutions that facilitate system state monitoring. As explained previously, this research is important because auditory displays improve operator situation awareness and accommodate workload by releasing the operator from constraints associated with overreliance on the visual system. Results obtained from studies detailed throughout this thesis support more ambitious applied research into the utilisation of spatial sonifications to convey system state

information concurrently with voice communications. The following section summarises some suggestions for future research.

7.4.1 *Misplaced Confidence in Localising with Non-Individual Cues*

Individuals vary in their ability to localise audio containing spatial cues generated through a non-individual HRTF. Sounds containing poorly matched HRTF cues have been shown to cause problems regarding the listener misplacing confidence in their ability to accurately localise sounds in certain regions. It was reasoned that internalisation caused listeners to have misplaced confidence when localising sounds in the back region during study two. Misplaced confidence may increase trust for incorrectly perceived information encoded into a sound's location, which could be problematic operationally when making decisions regarding future interaction with the system. If the listener receives no feedback regarding their localising accuracy, confidence can't be regulated or improvements achieved.

More research is required to determine new methods for providing localising performance feedback to the listener. This is important for the reasons mentioned, but also possibly because a body of evidence suggests that feedback during training improves the listener's ability to learn non-individual cues and improve localising. This is discussed further in section 3-5. More research into the effects of real time feedback as a learning aid for non-individual HRTF cues should be undertaken to determine if supplementary cues in a form similar to those used in studies three and four can over time facilitate such learning.

A future study might involve participants firstly localising white noise sounds with non-individual HRTF cues prior to spending some time flying the flight simulator utilised in studies three and four. After a period of time when cues are determined to have been learnt, the localising activity would be conducted again to determine if improvements had been achieved in the participants' ability to correctly identify localising cues and therefore more accurately localise sound.

7.4.2 *Supplementary Cues*

The carrier sounds introduced in study three (chapter five) of this thesis were found to be an effective means for providing a reference pitch for a collocated variable pitch supplementary sound. They also provided additional cues that aided in front-back localising and stream segregation. Further objective research is required for this design feature to better determine the extent which these inferences are true. An additional study

similar to study three is required to compare navigation performance when utilising the carrier sound to that of a control condition when the cue is not available.

Additional research should be undertaken to determine how effectively the relative cues gained through the supplementary and carrier pitch sounds facilitated improved azimuth localising accuracy. This design evolved from knowledge gained in the design of the sweep sound in study two (see section 4.3.1). The cues contribution to azimuth localising performance could be more accurately verified through a study that perhaps requires participants to localise the resting position of a sound that transits in azimuth about the listener before stopping. Control conditions might include a stationary sound and a transient carrier note without the variant pitch of the supplementary sound.

Studies throughout this thesis presented a relatively constant presentation of auditory sounds. For a variety of reasons, an applied use of this type of design would require the sound to be muted intermittently. Further research is required to determine the effects on localising performance and disorientation when sounds are periodically muted. Perhaps new methods for reorienting the listener toward the sound's new position might be required.

7.4.3 Workload

Focus on workload throughout this thesis has generally related to consequences associated with introducing excessive demand on the operator. Performance has also been found to degrade during underloaded conditions, particularly during the failure of automation (Desmond, Hancock, & Monette, 1998). Young & Stanton (2002) measured attention through eye movements recorded during a driving simulator study and found that capacity for attention degrades proportionately with workload demand. If an auditory display were designed such that it provided a level of misguided reliance and complacency, underloaded conditions may occur where operators neglect attending to more detailed primary visual displays. Further research into underloaded conditions is required to ensure an appropriate approach is taken toward integrating auditory displays into highly automated environments.

Supplementary cues utilised in this thesis were designed to enhance the quality of information presented to the listener, however, more research is required into the potential for excessive workload to be introduced through the applied use of such cues. The degree to which supplementary cues effect workload should be studied further by assessing the performance of real pilots' as they conduct more representative activities within high fidelity flight simulators. Further research is required to ensure that auditory displays

supplement visual information in such a way that they enhance situation awareness while maintaining an adequate level of workload.

7.4.4 *Speech Processing*

The fourth study found that a talker's location can affect a listener's ability to process the information. This finding occurred under conditions where the talker was delivering right cerebral hemisphere contextualised speech while concurrent sonifications compete for similar processing resources. The findings from this study are considered relevant as pilots frequently experience high workload situations while attending to multiple streams of audio information. Certain channels of speech may be consistently more contextually spatial than others and therefore better positioned to optimise processing. For instance, flight path vectoring instructions, or traffic collision avoidance system (TCAS) instructions might be best positioned toward the left ear to benefit from findings identified in the fourth study. However, naively positioning spatially contextualised audio to optimise cerebral processing may introduce problematic side effects, such as directional bias. For instance, perhaps directional commands such as those presented through TCAS are more effectively presented in the direction of the instruction to capitalise from directional bias effects; or along the midsagittal plane to mitigate any such bias. More research is required into design solutions that optimise cerebral hemisphere processing performance while mitigating directional bias for spatially contextualised information.

Additional studies are required to replicate the left ear advantage observed in study four. This finding challenges some of the basic criteria expected for an ear advantage to occur, such as the need for a competing interaural stimulus (Hiscock & Chipuer, 1993), and an easy secondary task that does not impose interference with the primary task (Kinsbourne, 1982). More evidence is required to support the claim that an ear advantage can occur in a high workload spatial audio environment. If further support were to be found, a subsequent body of research may have significant influence on improving future auditory display designs for high workload environments such as those experienced by fighter pilots.

Within this thesis, limited focus has been applied toward determining the possible ways in which the design of sonifications can come into conflict with speech in a spatial auditory display. During the design phase of study four, considerable thought was applied toward mitigating energetic interference between audio streams and optimising the possibility for informational processing advantages to occur through spatial position of speech. More research is required to determine if informational masking effects are

introduced by encoding information into particular sound attributes that are problematic when delivered concurrently with a certain context of speech.

7.4.5 Utilising Spatial Auditory Displays in Pilot Upset Recovery

Findings from study three indicate that an effective approach toward communicating deviations from a flight plan can be achieved by encoding aircraft navigation information into the spatial position and movement of sound. These findings may contribute toward developing more effective alerts for correcting pilot spatial disorientation. Gibb et al. (2011) claim that spatial disorientation “contributes to nearly 33% of all mishaps with a fatality rate of almost 100%” (p. 717). Research is currently being undertaken into the development of new flight simulators with an intent to improve training in upset recovery (Kvrgic, Kvrgic, Visnjic, Cvijanovic, & Divnic, 2015). More research is required to determine if improved spatial awareness might be gained through spatial auditory cues that alert pilots to an impending upset condition, along with the best course of action to recover.

7.5 The Future of Spatial Auditory Displays

The past decade or so has seen computer processing power evolve to the point where even the smallest of mobile computing devices are now capable of supporting the spatial rendering of multiple streams of concurrent audio (Vazquez Alvarez & Brewster, 2010). It is now perhaps a deficiency in design know-how that constrains the application of spatial auditory displays. The medical and computer gaming industry can be considered an exception to this. The medical industry has been mentioned several times throughout this thesis, however, the gaming industry is fast emerging as a likely incubation industry for future spatial audio enabling technologies. Game engineers are mindful of the need to incorporate effective spatial audio technology into their products in order to enhance a player’s gaming experience.

Given the spatial audio expertise and momentum established within the gaming industry, it is considered likely that emerging virtual reality (VR) and augmented reality (AR) technology in gaming will be the main drivers behind future research and development in the field of spatial auditory displays. These systems require effectively integrated spatial audio to ensure an immersive experience. With the introduction of Microsoft’s (2015a) HoloLens, it appears that core AR technology has evolved to a level of maturity that enables its use across a broad range of practical applications within entertainment, business, military, and everyday life. As expected, Microsoft (2015b) has acknowledged that the HoloLens will include built in spatial audio capability. The market

size and diverse range of intended applications for this technology will likely nurture more research into the field.

8. References

- Abel, S. M., & Paik, J. E. (2004). The benefit of practice for sound localization without sight. *Applied Acoustics*, *65*(3), 229-241.
- Abouchacra, K. S., Breitenbach, J., Mermagen, T., & Letowski, T. (2001). Binaural helmet: Improving speech recognition in noise with spatialized sound. *Human factors*, *43*(4), 584-594.
- Adams, K. S., & Trucks, L. B. (1976). A Procedure for Evaluating Auditory Warning Signals. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *20*(8), 166-172. doi: 10.1177/154193127602000803
- Agcaoglu, O., Miller, R., Mayer, A. R., Hugdahl, K., & Calhoun, V. D. (2015). Lateralization of resting state networks and relationship to age and gender. *NeuroImage*, *104*, 310-325. doi: 10.1016/j.neuroimage.2014.09.001
- Ahmad, A. M., Stanney, K. M., & Fouad, H. (2009). Theoretical foundations for integrating sound in interactive interfaces: identifying temporal and spatial information conveyance principles. *Theoretical Issues in Ergonomics Science*, *10*(2), 161-186.
- Alves-Pinto, A., Palmer, A. R., & Lopez-Poveda, E. A. (2014). Perception and coding of high-frequency spectral notches: potential implications for sound localization. *Frontiers in neuroscience*, *8*, 112. doi: 10.3389/fnins.2014.00112
- Anthony, S. (2014). Oculus reveals Crescent Bay VR headset with 360-degree head tracking and headphones. from <http://www.extremetech.com/computing/190619-oculus-reveals-crescent-bay-vr-headset-with-360-degree-head-tracking-and-headphones>
- Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85-97. doi: 10.1027//1016-9040.7.2.85
- Balakrishnan, U., & Freyman, R. L. (2008). Speech detection in spatial and nonspatial speech maskers. *The Journal of the Acoustical Society of America*, *123*(5), 2680-2691. doi: <http://dx.doi.org/10.1121/1.2902176>
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280-289. doi: 10.1016/j.tics.2007.05.005
- Barrass, S., & Kramer, G. (1999). Using sonification. *Multimedia Systems*, *7*(1), 23-31. doi: 10.1007/s005300050108

- BEA. (2009). Air France 447 Incident Final Report. 2009, from <http://www.bea.aero/en/enquetes/flight.af.447/rapport.final.en.php>
- Begault, D. R., & Pittman, M. T. (1996). Three-dimensional audio versus head-down Traffic Alert and Collision Avoidance System displays. *The International Journal of Aviation Psychology*, 6(1), 79-93. doi: 10.1207/s15327108ijap0601_5
- Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-computer interaction*, 4(1), 11-44. doi: 10.1207/s15327051hci0401_1
- Blender. (2015). Blender website. from www.Blender.org
- Bloom, J. S., & Hynd, G. W. (2005). The role of the corpus callosum in interhemispheric transfer of information: Excitation or inhibition? *Neuropsychology Review*, 15(2), 59-71. doi: 10.1007/s11065-005-6252-y
- Blue Ripple Sound. (2015). Rapture3D. from <http://www.blueripplesound.com/products/rapture-3d-user>
- Bregman, A. S. (1990). *Auditory scene analysis : the perceptual organization of sound*. MIT Press: Cambridge, Mass.
- Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 89(2), 244-249. doi: 10.1037/h0031163
- Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83(4), 1508.
- Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America*, 92(6), 3132-3139. doi: 10.1121/1.404209
- Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception & psychophysics*, 69(1), 79-91.
- Bryden, M. P. (1982). *Laterality: Funtional asymetry in the intact brain*. New York: Academic Press.
- Bryden, M. P., & MacRae, L. (1988). Dichotic Laterality Effects Obtained with Emotional Words. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology Fall*, 1(3), 171-176.
- Butler, R. A. (1994). Asymetric performances in monaural localization of sound in-space. *Neuropsychologia*, 32(2), 221-229.

- Cao, A., Chintamani, K., Pandya, A., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods*, 41(1), 113-117. doi: 10.3758/BRM.41.1.113
- Carmichael, M. E., Hall, S. E., & Phillips, D. P. (2008). Ear and contralateral masker effects on auditory temporal gap detection thresholds. *Hearing Research*, 245(1-2), 18-23. doi: 10.1016/j.heares.2008.08.002
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 25(5), 975-979. doi: [doi:http://dx.doi.org/10.1121/1.1907229](http://dx.doi.org/10.1121/1.1907229)
- Clifton, R. K. (1992). The development of spatial hearing in human infants. In L. A. W. E. W. Rubel (Ed.), *Developmental psychoacoustics* (pp. 135-157). Washington, DC, US: American Psychological Association.
- Cochrane, P. (2014). Exponential Technology and The Singularity: The technological singularity (Ubiquity symposium). *Ubiquity*, 2014(November), 1-9. doi: 10.1145/2667644
- Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America*, 116(2), 1057.
- de Crespigny, R. (2012). *Inside QF32*. Sydney, NSW: Macmillan Australia.
- Dekker, S. W. A. (2005). *Ten questions about human error: a new view of human factors and system safety*. Mahwah, N.J: Lawrence Erlbaum Associates.
- Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work*, 4(4), 240-244. doi: <http://dx.doi.org/10.1007/s101110200022>
- Denham, S., Böhm, T. M., Bendixen, A., Szalárdy, O., Kocsis, Z., Mill, R., & Winkler, I. (2014). Stable individual characteristics in the perception of multiple embedded patterns in multistable auditory stimuli. *Frontiers in neuroscience*, 8, 25. doi: 10.3389/fnins.2014.00025
- Desmond, P., Hancock, P., & Monette, J. (1998). Fatigue and Automation-Induced Impairments in Simulated Driving Performance. *Transportation Research Record: Journal of the Transportation Research Board*, 1628, 8-14. doi: 10.3141/1628-02
- Eckel, G. (1998). *A Spatial Auditory Display for the CyberStage*. Paper presented at the International Conference on Auditory Display, Glasgow, UK.
- Effenberg, A. O. (2005). Movement sonification: Effects on perception and action. *MultiMedia, IEEE*, 12(2), 53-59. doi: 10.1109/MMUL.2005.31

- Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32-64. doi: 10.1518/001872095779049543
- Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492. doi: 10.1080/001401399185595
- Escorial, S., Román, F. J., Martínez, K., Burgaleta, M., Karama, S., & Colom, R. (2015). Sex differences in neocortical structure and cognitive performance: A surface-based morphometry study. *NeuroImage*, 104(0), 355-365. doi: <http://dx.doi.org/10.1016/j.neuroimage.2014.09.035>
- Fadden, S., Ververs, P. M., & Wickens, C. D. (1998). Costs and benefits of head-up display use: A meta-analytic approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(1), 16-20. doi: 10.1177/154193129804200105
- Feddersen, W. E., Sandel, T. T., Teas, D. C., & Jeffress, L. A. (1957). Localization of high-frequency tones. *Journal of the Acoustical Society of America*, 29, 988-991.
- Fitts, P. M. (1951). *Human Engineering for an Effective Air-navigation and Traffic-control System*: National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115(5), 2246.
- Gadea, M., Espert, R., & Chirivella, J. (1997). Dichotic Listening: Elimination of the Right Ear Advantage Under a Dual Task Procedure. *Applied Neuropsychology*, 4(3), 171-175. doi: 10.1207/s15324826an0403_5
- Galín, D., Ornstein, R., Herron, J., & Johnstone, J. (1982). Sex and handedness differences in EEG measures of hemispheric specialization. *Brain and Language*, 16(1), 19-55. doi: 10.1016/0093-934X(82)90070-0
- Gaver, W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-computer interaction*, 2(2), 167-177. doi: 10.1207/s15327051hci0202_3
- Geronazzo, M., Spagnol, S., Bedin, A., & Avanzini, F. (2014, 4-9 May 2014). *Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions*. Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference.

- Ghez, C., Rikakis, T., DuBois, R. L., & Cook, P. R. (2000). *An Auditory Display System for Aiding Interjoint Coordination*. Paper presented at the International Conference on Auditory Display.
- Gibb, R., Ercoline, B., & Scharff, L. (2011). Spatial disorientation: decades of pilot fatalities. *Aviation, Space, and Environmental Medicine*, *82*(7), 717-724.
- González, J., Yu, W., & Arieta, A. H. (2010). Multichannel audio biofeedback for dynamical coupling between prosthetic hands and their users. *Industrial Robot-an International Journal*, *37*(2), 148-156. doi: 10.1108/01439911011018920
- Grimshaw, G. M. (1998). Integration and Interference in the Cerebral Hemispheres: Relations with Hemispheric Specialization. *Brain and Cognition*, *36*(2), 108-127. doi: <http://dx.doi.org/10.1006/brcg.1997.0949>
- Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction research: past, present, and future. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(2), 9-14. doi: 10.1177/1064804613477099
- Hawley, M. L., Litovsky, R. Y., & Colburn, H. S. (1999). Speech intelligibility and localization in a multi-source environment. *Journal of the Acoustical Society of America*, *105*(6), 3436-3448. doi: 10.1121/1.424670
- Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *The Journal of the Acoustical Society of America*, *115*(2), 833.
- Hermann, T. (2008). *Taxonomy and definitions for sonification and auditory display*. Paper presented at the Proceedings of the 14th international conference on auditory display, Paris, France.
- Hinterberger, T., & Baier, G. (2005). Parametric orchestral sonification of EEG in real time. *MultiMedia, IEEE*, *12*(2), 70-79. doi: 10.1109/MMUL.2005.36
- Hiscock, M., & Chipuer, H. (1993). Children's ability to shift attention from one ear to the other: Divergent results for dichotic and monaural stimuli. *Neuropsychologia*, *31*(12), 1339-1350. doi: [http://dx.doi.org/10.1016/0028-3932\(93\)90102-6](http://dx.doi.org/10.1016/0028-3932(93)90102-6)
- Hiscock, M., & Kinsbourne, M. (2011). Attention and the right-ear advantage: What is the connection? *Brain and Cognition*, *76*(2), 263-275. doi: 10.1016/j.bandc.2011.03.016
- Hiscock, M., Lin, J., & Kinsbourne, M. (1996). Shifts in children's ear asymmetry during verbal and nonverbal auditory-visual association tasks: A "virtual stimulus" effect. *Cortex*, *32*(2), 367-374. doi: [http://dx.doi.org/10.1016/S0010-9452\(96\)80058-X](http://dx.doi.org/10.1016/S0010-9452(96)80058-X)

- Holtgraves, T. (2012). The role of the right hemisphere in speech act comprehension. *Brain and Language*, 121(1), 58-64. doi: <http://dx.doi.org/10.1016/j.bandl.2012.01.003>
- Honda, A., Shibata, H., Gyoba, J., Saitou, K., Iwaya, Y., & Suzuki, Y. (2007). Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. *Applied Acoustics*, 68(8), 885-896.
- Hopcroft, R., Burchat, E., & Vince, J. (2006). Unmanned aerial vehicles for maritime patrol: Human factors Issues. Melbourne: Defence Science and Technology Organisation.
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104(1), 53-69. doi: 10.1037/0033-2909.104.1.53
- Iaccino, J. F. (1993). *Left brain-right brain differences*. Hillsdale, NJ: Erlbaum.
- Iwaya, Y., Suzuki, Y., & Kimura, D. (2003). Effects of head movement on front-back error in sound localization. *Acoustical Science and Technology*, 24(5), 322-324.
- Jovanov, E., Starcevic, D., Wegner, K., Karron, D., & Radivojevic, V. (1998). *Acoustic rendering as support for sustained attention during biomedical processing*. Paper presented at the International Conference on Auditory Display, Glasgow, UK.
- Kaber, D. B., & Riley, J. M. (1999). Adaptive Automation of a Dynamic Control Task Based on Secondary Task Workload Measurement. *International Journal of Cognitive Ergonomics*, 3(3), 169.
- Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics*, 5(1), 37-57.
- Kaper, H. G., Wiebel, E., & Tipei, S. (1999). Data sonification and sound visualization. *Computing in Science & Engineering*, 1(4), 48-58. doi: 10.1109/5992.774840
- Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 15(3), 166-171. doi: 10.1037/h0083219
- Kimura, D. (1967). Functional Asymmetry of the Brain in Dichotic Listening. *Cortex*, 3(2), 163-178. doi: [http://dx.doi.org/10.1016/S0010-9452\(67\)80010-8](http://dx.doi.org/10.1016/S0010-9452(67)80010-8)
- Kinsbourne, M. (1973). The control of attention by interaction between the cerebral hemispheres. In S. Kornblum (Ed.), *Attention and Performance* (pp. 4--276): , Vol 4.
- Kinsbourne, M. (1974). The mechanisms of hemisphere asymmetry in man. In M. Kinsbourne & W. Smith (Eds.), *Hemispheric Disconnection and Cerebral Function*: Charles C.

- Kinsbourne, M. (1980). Dichotic imbalance due to isolated hemisphere occlusion or directional rivalry? *Brain and Language*, 11(1), 221-224. doi: [http://dx.doi.org/10.1016/0093-934X\(80\)90123-6](http://dx.doi.org/10.1016/0093-934X(80)90123-6)
- Kinsbourne, M. (1982). Hemispheric specialization and the growth of human understanding. *American Psychologist*, 37(4), 411-420. doi: 10.1037/0003-066X.37.4.411
- Knox, A. W., & Boone, D. R. (1970). Auditory Laterality and Tested Handedness. *Cortex*, 6(2), 164-173. doi: [http://dx.doi.org/10.1016/S0010-9452\(70\)80025-9](http://dx.doi.org/10.1016/S0010-9452(70)80025-9)
- Kondo, H. M., Toshima, I., Pressnitzer, D., & Kashino, M. (2014). Probing the time course of head-motion cues integration during auditory scene analysis. *Frontiers in neuroscience*, 8, 170. doi: 10.3389/fnins.2014.00170
- Kramer, G. (1993). *Auditory Display: Sonification, Audification, and Auditory Interfaces*: Perseus Publishing.
- Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, 80(1–2), 97-126. doi: [http://dx.doi.org/10.1016/S0010-0277\(00\)00155-4](http://dx.doi.org/10.1016/S0010-0277(00)00155-4)
- Kudo, A., Higuchi, H., Hokari, H., & Shimada, S. (2006). Improved method for accurate sound localization. *Acoustical Science and Technology*, 27(3), 134-146.
- Kurzweil, R. (2006). *The Singularity is Near: When Humans Transcend Biology*: Penguin Books.
- Kvrgic, V. M., Kvrgic, V. M., Visnjic, Z. M., Cvijanovic, V. B., & Divnic, D. S. (2015). Dynamics and control of a spatial disorientation trainer. *Robotics and computer-integrated manufacturing*, 35, 104-125. doi: 10.1016/j.rcim.2015.03.003
- Lee, Y. (2010). The effect of congruency between sound-source location and verbal message semantics of in-vehicle navigation systems. *Safety Science*, 48(6), 708-713. doi: <http://dx.doi.org/10.1016/j.ssci.2010.02.006>
- Levitt, H., & Rabiner, L. R. (1967). Binaural release from masking for speech and gain in intelligibility. *The Journal of the Acoustical Society of America*, 42(3), 601.
- Lieberman, D. A. (2011). *Human Learning and Memory* Retrieved from <http://UQL.ebib.com.au/patron/FullRecord.aspx?p=824445>
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476-490. doi: 10.3758/bf03210951
- Lokki, T., & Grohn, M. (2005). Navigation with auditory cues in a virtual environment. *MultiMedia, IEEE*, 12(2), 80-86. doi: 10.1109/mmul.2005.33
- Luders, E., Steinmetz, H., & Jancke, L. C. A. (2002). Brain size and grey matter volume in the healthy human brain. *Neuroreport*, 13(17), 2371-2374.

- MacDonald, J. A., Balakrishnan, J. D., Orosz, M. D., & Karplus, W. J. (2002). Intelligibility of Speech in a Virtual 3-D Environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(2), 272-286. doi: 10.1518/0018720024497934
- Marston, J. R., Loomis, J. M., Klatzky, R. L., & Golledge, R. G. (2007). Nonvisual Route Following with Guidance from a Simple Haptic or Auditory Display. *Journal of Visual Impairment & Blindness*, *101*(4), 203-211.
- McAnally, K. I., & Martin, R. L. (2007). Spatial Audio Displays Improve the Detection of Target Messages in a Continuous Monitoring Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(4), 688-695. doi: 10.1518/001872007x215764
- McAnally, K. I., & Martin, R. L. (2008). Localization of Sound Presented Via a Spatial Audio Display during Visually Induced Vection in Pitch, Roll, and Yaw. *Aviation, Space, and Environmental Medicine*, *79*(6), 611-615. doi: 10.3357/asem.2248.2008
- McAnally, K. I., & Martin, R. L. (2014). Sound localization with head movement: implications for 3-D audio displays. *Frontiers in neuroscience*, *8*. doi: 10.3389/fnins.2014.00210
- McCarley, J. S., & Wickens, C. D. (2005). Human factors implications of UAVs in the national airspace (Report No. AHFD-05-05/FAA-05-01). Savoy: University of Illinois at Urbana-Champaign.
- McCarthy, J. C., Fallon, E., & Bannon, L. (2000). Dialogues on function allocation. *International Journal of Human-Computer Studies*, *52*(2), 191-201. doi: <http://dx.doi.org/10.1006/ijhc.1999.0284>
- McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452-463. doi: 10.1016/j.conb.2008.09.005
- McGlone, J. (1980). Sex differences in human brain asymmetry: a critical survey. *Behavioral and Brain Sciences*, *3*(2), 215-227. doi: 10.1017/S0140525X00004398
- McPherson, L. M. P., Ciocca, V., & Bregman, A. S. (1994). Organization in audition by similarity in rate of change: Evidence from tracking individual frequency glides in mixtures. *Perception & psychophysics*, *55*(3), 269-278. doi: 10.3758/BF03207598
- Microsoft. (2015a). HoloLens. from <https://www.microsoft.com/microsoft-hololens/en-us>
- Microsoft. (2015b). HoloLens spatial audio. from <http://lumiaconversations.microsoft.com/2015/04/30/build-2015-a-closer-look-at-the-microsoft-hololens-hardware/>

- Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual review of psychology*, 42(1), 135.
- Miller, G. A. (1955). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2), 81-97. doi: 10.1037/h0043158
- Miller, J. D., Godfroy-Cooper, M., & Wenzel, E. M. (2014, June 22-25). *Using published HRTFs with SLAB3D: Metric-based database selection and phenomena observed*. Paper presented at the 20th International Conference on Auditory Display (ICAD2014), New York, NY.
- Miller, J. D., & Wenzel, E. M. (2002). *Recent Developments in SLAB: A Software-Based System for Interactive Spatial Sound Synthesis*. Paper presented at the Proceedings of the International Conference on Auditory Display, Kyoto, Japan.
- Mishra, J., & Gazzaley, A. (2012). Attention distributed across sensory modalities enhances perceptual performance. *The Journal of Neuroscience*, 32(35), 12294-12302. doi: 10.1523/JNEUROSCI.0867-12.2012
- Moller, H., Sorensen, M. F., Jensen, C. B., & Hammershoi, D. (1996). Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, 44(6), 451-469.
- Mondor, T., & Finley, G. A. (2003). The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anaesthesia*, 50(3), 221-228. doi: 10.1007/BF03017788
- Moore, G. E. (1965). *Cramming More Components Onto Integrated Circuits*: McGraw-Hill.
- Morais, J., & Landercy, M. (1977). Listening to speech while retaining music: What happens to the right-ear advantage? *Brain and Language*, 4(2), 295-308. doi: [http://dx.doi.org/10.1016/0093-934X\(77\)90024-4](http://dx.doi.org/10.1016/0093-934X(77)90024-4)
- Muller, B. S., & Bovet, P. (1999). Role of pinnae and head movements in localizing pure tones. *Swiss Journal of Psychology*, 58(3), 170-179.
- NTSB. (2013). Descent below visual glidepath and impact with seawall. Asiana Airlines Flight 214. Washington, D.C.: National Transportation Safety Board.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, 28(32), 8116-8123. doi: 10.1523/jneurosci.1290-08.2008
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

- Oldfield, S. R., & Parker, S. P. (1984a). Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions. *Perception*, *13*(5), 581-600.
- Oldfield, S. R., & Parker, S. P. (1984b). Acuity of sound localisation: A topography of auditory space. II. Pinna cues absent. *Perception*, *13*(5), 601-617.
- Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(3), 476-488. doi: 10.1177/0018720813501549
- Parasuraman, R. (1987). Human-computer monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(6), 695-706. doi: 10.1177/001872088702900609
- Parker, S. P. A., Smith, S. E., Stephan, K. L., Martin, R. L., & McAnally, K. I. (2004). Effects of supplementing head-down displays with 3-D audio during visual target acquisition. *International Journal of Aviation Psychology*, *14*(3), 277 - 295.
- Pauletto, S., & Hunt, A. (2009). Interactive sonification of complex data. *International Journal of Human-Computer Studies*, *67*(11), 923-933. doi: <http://dx.doi.org/10.1016/j.ijhcs.2009.05.006>
- Peryer, G., Noyes, J., Pleydell-Pearce, K., & Lieven, N. (2005). Auditory alert characteristics: A survey of pilot views. *The International Journal of Aviation Psychology*, *15*(3), 233-250. doi: 10.1207/s15327108ijap1503_2
- Plantinga, J., & Trainor, L. J. (2005). Memory for melody: infants use a relative pitch code. *Cognition*, *98*(1), 1-11. doi: 10.1016/j.cognition.2004.09.008
- Pollack, I., & Ficks, L. (1954). Information of elementary multidimensional auditory displays. *The Journal of the Acoustical Society of America*, *26*(2), 155-158. doi: 10.1121/1.1907300
- Pollack, I., & Pickett, J. M. (1958). Stereophonic listening and speech intelligibility against voice babble. *The Journal of the Acoustical Society of America*, *30*(2), 131-133. doi: <http://dx.doi.org/10.1121/1.1909505>
- Pollmann, S., Maertens, M., von Cramon, D. Y., Lepsien, J., & Hugdahl, K. (2002). Dichotic listening in patients with splenial and nonsplenial callosal lesions. *Neuropsychology*, *16*(1), 56-64. doi: 10.1037/0894-4105.16.1.56
- Qian, J., & Eddins, D. A. (2008). The role of spectral modulation cues in virtual sound localization. *Journal of the Acoustical Society of America*, *123*(1), 302-314.

- Rayleigh, L. (1907). On the dynamical theory of gratings. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 79(532), 399-416.
- Razavi, B., O'Neill, W. E., & Paige, G. D. (2005). *Both interaural and spectral cues impact sound localization in azimuth*. Paper presented at the Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on.
- Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, 43(6), 1255-1259. doi: [doi:http://dx.doi.org/10.1121/1.1910976](http://dx.doi.org/10.1121/1.1910976)
- Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, 113(3), 553-565. doi: 10.1037/0033-2909.113.3.553
- Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 30(4), 431-443. doi: 10.1177/001872088803000405
- Runkle, P., Yendiki, A., & Wakefield, G. (2000). *Active sensory tuning for immersive spatialized audio*. Paper presented at the ICAD, Atlanta, GA.
- Sætrevik, B., & Hugdahl, K. (2007). Priming inhibits the right ear advantage in dichotic listening: Implications for auditory laterality. *Neuropsychologia*, 45(2), 282-287. doi: <http://dx.doi.org/10.1016/j.neuropsychologia.2006.07.005>
- Sanderson, P. M. (2006). The multimodal world of medical monitoring displays. *Applied Ergonomics*, 37(4), 501-512. doi: <http://dx.doi.org/10.1016/j.apergo.2006.04.022>
- Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology-Human Perception and Performance*, 33(6), 1311-1321.
- Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology*, 1(1), 45-57. doi: 10.1207/s15327108ijap0101_4
- Scaletti, C., & Craig, A. B. (1991). *Using sound to extract meaning from complex data*.
- Schwartz, J., & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, 207(4437), 1380-1381. doi: 10.2307/1683464
- Scott, S. K., & McGettigan, C. (2013). Do temporal processes underlie left hemisphere dominance in speech perception? *Brain and Language*, 127(1), 36-45. doi: <http://dx.doi.org/10.1016/j.bandl.2013.07.006>

- Segalowitz, S. J., & Cohen, H. (1989). Right hemisphere EEG sensitivity to speech. *Brain and Language*, 37(2), 220-231. doi: [http://dx.doi.org/10.1016/0093-934X\(89\)90016-3](http://dx.doi.org/10.1016/0093-934X(89)90016-3)
- Shankweiler, D., & Studdert-Kennedy, M. (1975). A continuum of lateralization for speech perception? *Brain and Language*, 2(2), 212-225. doi: 10.1016/S0093-934X(75)80065-4
- Sharon, M. A., Christian, G., Angela, C., & Blake, C. P. (2000). The effect of aging on horizontal plane sound localization. *The Journal of the Acoustical Society of America*, 108(2), 743-752.
- Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., & et al. (1995). Sex differences in the functional organization of the brain for language. *Nature*, 373(6515), 607-609.
- Sheridan, T. B. (2000). Function allocation: algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies*, 52(2), 203-216. doi: <http://dx.doi.org/10.1006/ijhc.1999.0285>
- Shestopalova, L., Böhmer, T. M., Bendixen, A., Andreou, A. G., Georgiou, J., Garreau, G., . . . Winkler, I. (2014). Do audio-visual motion cues promote segregation of auditory streams? *Frontiers in neuroscience*, 8, 64. doi: 10.3389/fnins.2014.00064
- Simpson, B. D., Brungart, D. S., Dallman, R. C., Joffrion, J., Presnar, M. D., & Gilkey, R. H. (2005). Spatial audio as a navigation aid and attitude indicator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(17), 1602-1606. doi: 10.1177/154193120504901722
- Singurd, S. (2000). *A model for interaction in exploratory sonification displays*. Paper presented at the International conference on auditory display, Atlanta, US.
- Skybrary. (2014). Terrain Avoidance and Warning System (TAWS). from [http://www.skybrary.aero/index.php/Terrain_Avoidance_and_Warning_System_\(TAWS\)](http://www.skybrary.aero/index.php/Terrain_Avoidance_and_Warning_System_(TAWS))
- Snow, M. P., Reising, J. M., Barry, T. P., & Hartsock, D. C. (1999). Comparing new designs with baselines. *Ergonomics In Design*, 7(4), 28-33(26).
- Snyder, J. S., Carter, O. L., Hannon, E. E., & Alain, C. (2009). Adaptation reveals multiple levels of representation in auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1232-1244. doi: 10.1037/a0012741
- Snyder, J. S., Carter, O. L., Lee, S.-K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology:*

- Human Perception and Performance*, 34(4), 1007-1016. doi: 10.1037/0096-1523.34.4.1007
- Snyder, J. S., Holder, W. T., Weintraub, D. M., & Carter, O. L. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, 46(6), 1208.
- Sony. (2015). Project Morpheus. from <https://www.playstation.com/en-au/explore/ps4/features/project-morpheus/>
- Sorkin, R. D. (1988). FORUM: Why are people turning off our alarms? *The Journal of the Acoustical Society of America*, 84(3), 1107-1108. doi: doi:<http://dx.doi.org/10.1121/1.397232>
- Spence, C., & Ho, C. (2008). Tactile and multisensory spatial warning signals for drivers. *Haptics, IEEE Transactions on*, 1(2), 121-129. doi: 10.1109/TOH.2008.14
- Springer, S. P., & Deutsch, G. (1989). *Left brain, right brain* (3 ed.). New York: Freeman.
- Tannen, R. S., Nelson, T., Bolia, R. S., Warm, J. S., & Dember, W. N. (2004). Evaluating adaptive multisensory displays for target localization in a flight task. *The International Journal of Aviation Psychology*, 14(3), 297 - 312. doi: 10.1207/s15327108ijap1403 5
- Techentin, C., & Voyer, D. (2007). Congruency, attentional set, and laterality effects with emotional words. *Neuropsychology*, 21(5), 646-655. doi: 10.1037/0894-4105.21.5.646
- Techentin, C., Voyer, D., & Klein, R. M. (2009). Between- and within-ear congruency and laterality effects in an auditory semantic/emotional prosody conflict task. *Brain and Cognition*, 70(2), 201-208. doi: <http://dx.doi.org/10.1016/j.bandc.2009.02.003>
- Thurlow, W. R., Mangels, J. W., & Runge, P. S. (1967). Head movements during sound localization. *The Journal of the Acoustical Society of America*, 42(2), 489-493. doi: doi:<http://dx.doi.org/10.1121/1.1910605>
- Tobias, J. V., & Schubert, E. D. (1959). Effective onset duration of auditory stimuli. *The Journal of the Acoustical Society of America*, 31(12), 1595-1605. doi: doi:<http://dx.doi.org/10.1121/1.1907665>
- Towers, J., Burgess-Limerick, R., & Riek, S. (2008). *Localising synthesised spatial audio filtered through a generalised HRTF*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society of Australia 44th Annual Conference, Adelaide, Australia.

- Towers, J., Burgess-Limerick, R., & Riek, S. (2012). Improving 3-D audio localisation through the provision of supplementary spatial audio cues. *The Ergonomics Open Journal*, 5(1-9).
- Towers, J., Burgess-Limerick, R., & Riek, S. (2014). Concurrent 3-D sonifications enable the head-up monitoring of two interrelated aircraft navigation instruments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 56(8), 1414-1427. doi: 10.1177/0018720814536443
- Tzelgov, J., Srebro, R., Henik, A., & Kushelevsky, A. (1987). Radiation search and detection by ear and by eye. *Human factors*, 29(1), 87.
- Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, 55(3), 1208-1218. doi: 10.1016/j.neuroimage.2010.12.068
- van Zuijlen, T. L., van Zuijlen, T. L., Sussman, E., Winkler, I., & Näätänen, R. (2005). Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Brain research. Cognitive brain research*, 23(2), 270-276. doi: 10.1016/j.cogbrainres.2004.10.007
- Vazquez Alvarez, Y., & Brewster, S. (2010, 2010). *Designing spatial audio interfaces to support multiple audio streams*. Paper presented at the Mobile HCI '10, New York, NY.
- Veltman, J. A., Oving, A. B., & Bronkhorst, A. W. (2004a). 3-D audio in the fighter cockpit improves task performance. *International Journal of Aviation Psychology*, 14(3), 239-256. doi: 10.1207/s15327108ijap1403 2
- Veltman, J. A., Oving, A. B., & Bronkhorst, A. W. (2004b). Effectiveness of 3-D audio for warnings in the cockpit. *International Journal of Aviation Psychology*, 14(3), 257-276. doi: 10.1207/s15327108ijap1403 3
- Walker, B. N., & Lindsay, J. (2006). Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(2), 265-278. doi: 10.1518/001872006777724507
- Watson, M., & Sanderson, P. M. (2004). Sonification supports eyes-free respiratory monitoring and task time-sharing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(3), 497-517. doi: 10.1518/hfes.46.3.497.50401
- Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1), 111-123.

- Wenzel, E. M., Fisher, S. S., Stone, P. K., & Foster, S. H. (1990). *A system for three-dimensional acoustic "visualization" in a virtual environment workstation*. Paper presented at the Proceedings of the 1st conference on Visualization '90, San Francisco, California.
- Westerhausen, R., Moosmann, M., Alho, K., Medvedev, S., Hamalainen, H., & Hugdahl, K. (2009). Top-down and bottom-up interaction: manipulating the dichotic listening ear advantage. *Brain Res, 1250*, 183-189. doi: 10.1016/j.brainres.2008.10.070
- Westerhausen, R., Woerner, W., Kreuder, F., Schweiger, E., Hugdahl, K., & Wittling, W. (2006). The role of the corpus callosum in dichotic listening: A combined morphological and diffusion tensor imaging study. *Neuropsychology, 20*(3), 272-279. doi: 10.1037/0894-4105.20.3.272
- Wexler, B. E., & Halwes, T. (1983). Increasing the power of dichotic methods: The fused rhymed words test. *Neuropsychologia, 21*(1), 59-66. doi: [http://dx.doi.org/10.1016/0028-3932\(83\)90100-8](http://dx.doi.org/10.1016/0028-3932(83)90100-8)
- Wickens, C. D. (2002). Multiple Resources and Performance Prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177. doi: 10.1080/14639220210123806
- Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 449-455. doi: 10.1518/001872008x288394
- Wickens, C. D., & Carswell, C. M. (2006). Information Processing. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3 ed., pp. 111-149). Hoboken, NJ: John Wiley & Sons.
- Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America, 85*(2), 858-867.
- Wightman, F. L., & Kistler, D. J. (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America, 85*(2), 868-878.
- Wightman, F. L., & Kistler, D. J. (1999). Resolution of front-back ambiguity in spatial hearing by listener and source movement. *Journal of the Acoustical Society of America, 105*(5), 28-41. doi: 10.1121/1.426899
- Winkler, I., Denham, S., Mill, R., Bohm, T. M., & Bendixen, A. (2012). Multistability in auditory stream segregation: a predictive coding view. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 367*(1591), 1001-1012. doi: 10.1098/rstb.2011.0359

- Woodford, H., & Price, C. (2007). EMG biofeedback for the recovery of motor function after stroke. *Cochrane database of systematic reviews*(2), CD004585. doi: 10.1002/14651858.CD004585.pub2
- Woodworth, R. S. (1938). *Experimental psychology*. Holt.
- Yost, W. A. (1997). The cocktail party problem: Forty years later. In R. H. G. T. R. Anderson (Ed.), *Binaural and spatial hearing in real and virtual environments* (pp. 329-347). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Yost, W. A. (2007). *Fundamentals of hearing: An introduction* (5th ed.). Burlington, MA: Academic Press.
- Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human factors*, 44(3), 365.
- Zotkin, D. N., Duraiswami, R., & Davis, L. S. (2004). Rendering localized spatial audio in a virtual auditory space. *Multimedia, IEEE Transactions on*, 6(4), 553-564. doi: 10.1109/TMM.2004.827516

9. Appendices

9.1 Appendix A: Participant Hearing Test

Source: www.phys.unsw.edu.au/~jw/hearing.html

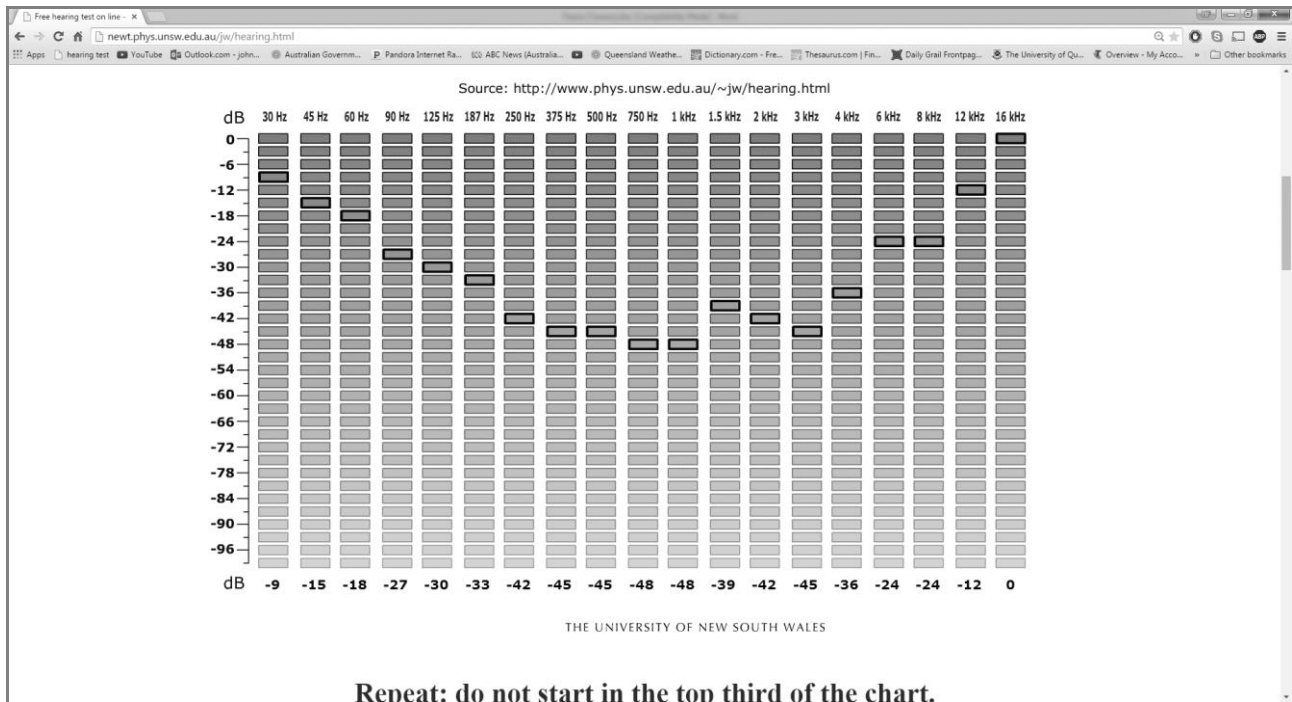


Figure 9-1. Online equal loudness hearing test.

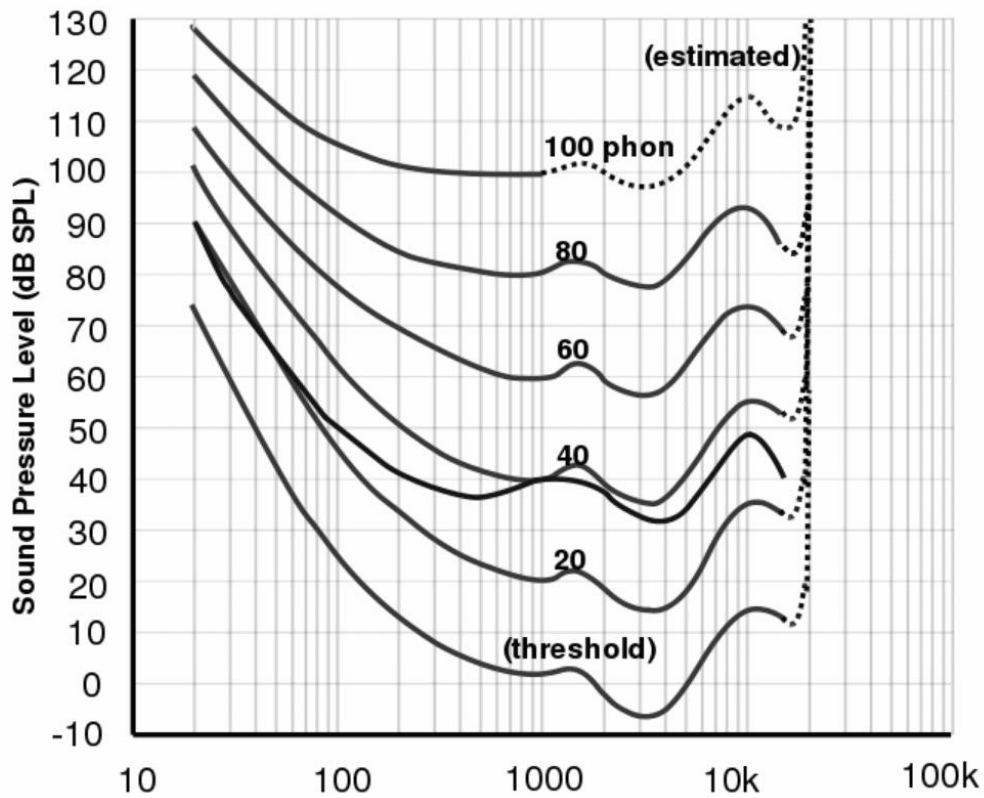


Figure 9-2. The Fletcher-Munson graph of equal loudness contours

9.2 Appendix B: Study 1 Participant Briefing

Participant Name:

Session #:

Demonstration

1. Ensure volume and balance are set appropriately
2. Test graphics tablet log
3. Provide demonstration of orbiting noise. (file: "Orbiting white noise demo.wav")
4. Ask participant to adjust volume to a comfortable level

Conduct dexterity task.

1. Ask the participant to touch the 45-degree intersecting points of the circle in a clockwise direction in their own time.
2. Save log

Conduct Study:

1. During the presentation of each sound, remain still, facing forward with your head up and eyes closed.
2. When complete, there will be a 5-sec pause where you can indicate the location of the sound source on the graphics tablet.
3. Also indicate the confidence you have in your ability to localise the sound by touching the slider bar on the right of the tablet.
4. Load the appropriate session file
5. Tell participant that if they need to stop to raise their hand and I will pause the session.
6. Ask participant if they are ready.
7. Start the session

Shut down

1. Archive logs

9.3 Appendix B: Study 1 Tablet Overlay

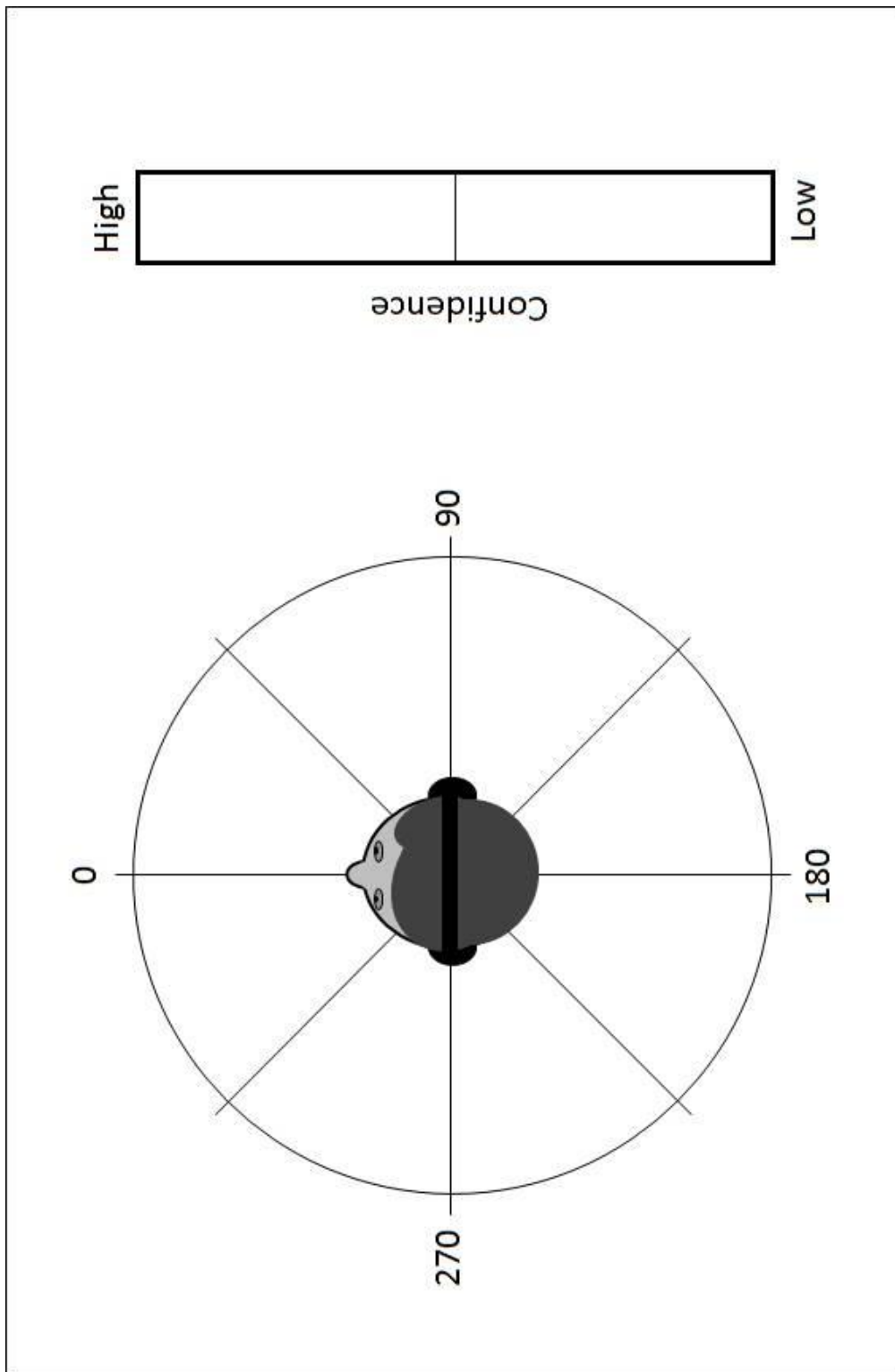


Figure 9-3. Tablet overlay for input of sound localising and confidence estimates in experiment 1.

9.4 Appendix B: Study 2 Participant Information Sheet



School of Human Movement Studies

The University of Queensland
 Brisbane Qld 4072 Australia
 Telephone (07) 3365 6240
 International +61 7 3365 6240
 Facsimile (07) 3365 6877
 Email enquiries@hms.uq.edu.au
 Internet www.hms.uq.edu.au
 CRICOS PROVIDER NUMBER 00025B

Participant Information Sheet

Study Title: Optimising Localisation Cues for Spatial Audio Filtered through a Generalised HRTF.

Investigator: John Towers (Advanced Human Factors Technologist, Boeing Research & Technology Australia.)

Purpose: The purpose of this study is to validate techniques that have been designed to optimise a listeners localising performance for 3-D audio displays.

Duration: Over the course of three non-consecutive days, each participant is requested to participate in 1 x 15 minute session for a hearing test and 2 x 30 minute localising sessions.

Procedure: Each participant will use a joystick to fly a computer simulation of a space ship while listening to 3-D tones presented randomly about the azimuth. Depending on the assigned group, the participant will indicate the perceived origin of the sound either via graphics tablet or through a graphical user interface. A confidence estimate will also be recorded to indicate the accuracy with which the participant believes the sounds origin could be localised. A more comprehensive overview of the study, along with some information relating to 3-D audio display fundamentals will be provided to participants prior to the study.

Risk: This study is not considered to pose cognitive or physical risk to participants beyond that which would be expected in normal every day life.

Benefits: It is expected that participants will gain insight into the fundamentals associated with 3-D audio, along with an appreciation of the constraints and considerations associated with integrating spatial audio displays into C3 systems. This will be a valuable experience for C3 systems design engineers wishing to expand their knowledge base in this field.

Data: Performance data will be obtained and recorded for sound localising estimates, along with self reported confidence levels associated with each estimate. Data obtained from this study will be treated in confidence and stored in a manner that prevents results being associated with an individual.

Emergency: The study will be conducted within the usability laboratory at Boeing House, Brisbane. As such, any emergency will be treated as per the guidelines within the Boeing Defence Australia employee handbook (HB-BDA-001).

Withdrawal: Participation in this study is voluntary. It is the right of any participant to withdraw from the study at any time without prejudice. Data collected prior to withdrawal will be erased from all working and/or archived files.

Contact: Should you have any questions regarding the study, please do not hesitate to contact John Towers on 3306 3527, or email john.towers@boeing.com.

Debrief: Upon completion of the study, results will be provided to all participants through a report paper. A debrief presentation will be held for participants to explain the findings and future application of the research.

Complaints: This study has been cleared in accordance with the ethical review guidelines and processes of the University of Queensland. These guidelines are endorsed by the University's principal human ethics committee, the Human Experimentation Ethical Review Committee, and registered with the Australian Health Ethics Committee as complying with the National Statement. You are free to discuss your participation in this study with project staff (contactable on 3306 3527). If you would like to speak to an officer of the University not involved in the study, you may contact the School Ethics Officer on 3346 7904.

9.5 Appendix B: Study 2 Participant Informed Consent Form



School of Human Movement Studies

The University of Queensland
Brisbane Qld 4072 Australia
Telephone (07) 3365 6240
International +61 7 3365 6240
Facsimile (07) 3365 6877
Email enquiries@hms.uq.edu.au
Internet www.hms.uq.edu.au
CRICOS PROVIDER NUMBER 00025B

Participant Informed Consent Form

Study Title: Optimising Localisation Cues for Spatial Audio Filtered through a Generalised HRTF

Researcher:

John Towers
Advanced Human Factors Technologist
Boeing Research & Technology Australia
Brisbane, Qld, 4001
Ph 3306 3527
Email: john.towers@boeing.com

Participant Declaration:

I _____ acknowledge that I agree to participate in the above mentioned study, which has been appropriately explained to my satisfaction.

I have been informed that at any stage throughout the study I may withdraw my participation without prejudice.

I understand that all information/data collected throughout the course of the study will be treated in confidence and therefore not presented or stored in a manner that identifies my individual performance.

I understand that participation in the study is voluntary and does not involve financial remuneration.

Signature: _____

Date: ____/____/____

Witness

Name: _____

Date: ____/____/____

Signature: _____

9.6 Appendix B: Study 2 Participant Briefing

Participant Name:

Group #:

Demonstration

1. Ensure volume and balance are set appropriately
2. Test graphics tablet log or GUI input log. Ensure the correct configuration.
3. Ask participant to adjust volume to a comfortable level while playing demo sound

Conduct Study:

1. Load the appropriate session file
2. Ask the participant to “sit still with your head upright facing forward for the duration of each trial.”
3. “Pressing the joystick trigger will initiate the earth-tracking task and the sound will be presented two-seconds later.”
4. “When the task starts, use the joystick to guide the spaceship to the Earth as accurately as you can”
5. “The session will then pause and you can enter the localisation and confidence estimates on the tablet/screen.”
6. Tell participant that if they need to stop to raise their hand and I will pause the session.
7. Ask participant if they are ready.
8. Start the session.

Shut down

1. Archive logs

9.7 Appendix B: Study 2 Interview Protocol

Participant ID: _____

Date: ____/____/____

Q1. Did you find anything particularly difficult or easy when localising sounds?

Q2. How did you find the Stable condition?

Q3. How did you find the Swing condition?

Q3. How did you find the Sweep condition?

Q4. Do you think you were able to accurately input your localising estimates via the interface/graphics tablet?

Q5. Do you have any other comments about the activity?

9.8 Appendix C: Study 3 Participant Information Sheet



School of Human Movement Studies

The University of Queensland
Brisbane Qld 4072 Australia
Telephone (07) 3365 6240
International +61 7 3365 6240
Facsimile (07) 3365 6877
Email enquiries@hms.uq.edu.au
Internet www.hms.uq.edu.au
CRICOS PROVIDER NUMBER 00025B

Participant Information Sheet

Study Title: Utilising Spatial Sonification to Aid Head-Up Monitoring of Aircraft Navigation Displays.

Investigator: John Towers (Advanced Human Factors Technologist, Boeing Research & Technology Australia.)

Purpose: The purpose of this study is to validate techniques that have been designed to optimise a listeners localising performance for 3-D audio displays.

Duration: Over the course of three non-consecutive days, each participant is requested to participate in 1 x 15 minute session for a hearing test and 2 x 30 minute sessions flying the simulator.

Procedure: Each participant will use a control yoke to fly a flight simulator about a predetermined path. Participants will be required to attempt to fly as accurate a course as possible while visually searching for flares that will be periodically presented in the forward cockpit view. When a flare is observed, participants will acknowledge the detection by pressing a button on the control yoke. Spatially positioned sonification will be presented to supplement visually presented information within flight gauges for waypoint bearing and course deviation. The following three audio conditions will be experienced by participants over the course of the experiment.

1. A mute condition where no audio is presented.
2. Spatial sonification. Without head tracking so that 0 degree alignment for the audio display is consistently maintained in the forward facing direction of the participants head.
3. Spatial sonification with head tracking. This condition will maintain a 0 degree alignment for the audio display toward the front of the aircraft and compensate for variations in head movement.

Risk: This study is not considered to pose cognitive or physical risk to participants beyond that which would be expected in normal every day life. There is a slight potential for participants to experience motion sickness due to the flight simulator. If this does occur, participants are requested to stop the experiment immediately and withdraw from the study without prejudice.

Benefits: It is expected that participants will gain insight into the fundamentals associated with 3-D audio, along with an appreciation of the constraints and considerations associated with integrating spatial audio displays into systems. This will be a valuable experience system design engineers wishing to expand their knowledge within this field.

Data: Performance data will be obtained and recorded for flight path accuracy, flare detection, and gaze fixation. Data obtained from this study will be treated in confidence and stored in a manner that prevents results being associated with an individual.

Emergency: The study will be conducted within the usability laboratory at Boeing House, Brisbane. As such, any emergency will be treated as per the guidelines within the Boeing Defence Australia employee handbook (HB-BDA-001).

Withdrawal: Participation in this study is voluntary. It is the right of any participant to withdraw from the study at any time without prejudice. Data collected prior to withdrawal will be erased from all working and/or archived files.

Contact: Should you have any questions regarding the study, please do not hesitate to contact John Towers on 3306 3527, or email john.towers@boeing.com.



Debrief: Upon completion of the study, results will be provided to all participants through a report paper. A debrief presentation will be held for participants to explain the findings and future application of the research.

Complaints: This study has been cleared in accordance with the ethical review guidelines and processes of the University of Queensland. These guidelines are endorsed by the University's principal human ethics committee, the Human Experimentation Ethical Review Committee, and registered with the Australian Health Ethics Committee as complying with the National Statement. You are free to discuss your participation in this study with project staff (contactable on 3306 3527). If you would like to speak to an officer of the University not involved in the study, you may contact the School Ethics Officer on 3346 7768.

9.9 Appendix C: Study 3 Participant Informed Consent Form



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

School of Human Movement Studies

The University of Queensland
Brisbane Qld 4072 Australia
Telephone (07) 3365 6240
International +61 7 3365 6240
Facsimile (07) 3365 6877
Email enquiries@hms.uq.edu.au
Internet www.hms.uq.edu.au
CRICOS PROVIDER NUMBER 000258

Participant Informed Consent Form

Study Title: Utilising Spatial Sonification to Aid Head-Up Monitoring of Aircraft Navigation Displays

Researcher:

John Towers
Advanced Human Factors Technologist
Boeing Research & Technology Australia
Brisbane, Qld, 4001
Ph 3306 3527
Email: john.towers@boeing.com

Participant Declaration:

I _____ acknowledge that I agree to participate in the above mentioned study, which has been appropriately explained to my satisfaction.

I have been informed that at any stage throughout the study I may withdraw my participation without prejudice.

I understand that all information/data collected throughout the course of the study will be treated in confidence and therefore not presented or stored in a manner that identifies my individual performance.

I understand that participation in the study is voluntary and does not involve financial remuneration.

Signature: _____

Date: ____/____/____

Witness

Name: _____

Date: ____/____/____

Signature: _____

9.10 Appendix C: Study 3 Sound Evaluation Sheet

Draw a line about each head to indicate the perceived path and relative distance from your head for each sound.

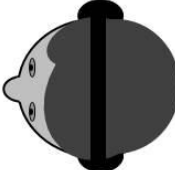
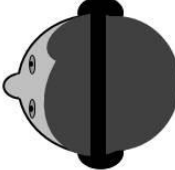
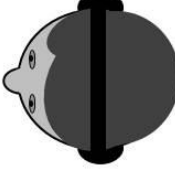
1)		Consonance ____/10 Comment: _____ _____ _____
2)		Consonance ____/10 Comment: _____ _____ _____
3)		Consonance ____/10 Comment: _____ _____ _____

Figure 9-4. Sound evaluation sheet used for sound selection task in experiment 3.

9.11 Appendix C: Study 3 Participant Briefing

Participant Name: _____

Ordering #: _____ **& Sequence:** _____

Set up prior to participant arrival

1. Archive / delete previous logs
2. Ensure head tracker is connected and started
3. Ensure NTP is on
4. Ensure SIM sound is on 'Q' to toggle
5. Make sure flight is smooth and instruments positioned correctly

When Participant arrives

6. Set eye height
7. Use A9000 world model in faceLAB
8. Ask to turn off phone
9. Calibrate eye tracker

Tell student:

1. Not to make control movements too large – lead into WP turn.
2. Correct CDI accuracy at the smallest deviation – don't be complacent.
3. Keep elevation between 1300 and 1400 feet.
4. Search aggressively for flares and press button to acknowledge detection, but if in doubt about a flare, don't press the button. The flares can be anywhere in and around trees etc.
5. Point out the flare button!!
6. Remind motivation of \$50 reward
7. Remind that data collection starts on entry into 3rd leg of flight plan
8. Try not to move head around too much as there is a limited field of view for the gaze tracker
9. Show them the map indicating when the trial will end
10. Use the audio to keep eyes up searching for flares
11. Ensure faceLAB is logging
12. Tell participant to un-pause the sim and begin

Shut down

Archive logs!!!

9.12 Appendix C: Study 3 Interview Protocol

Participant ID: _____

Date: ____/____/____

Q1. Were you comfortable flying the simulator?

Q2. Did you have any difficulty attending to the auditory display?

Q3. Were you able to concurrently attend to both the CDI and WP auditory display signals?

Q4. How did you find the supplementary sound that was collocated with the variable sound?

Q5. How did you find the Mute condition?

Q6. How did you find the Audio display without the head tracker?

Q7. How did you find the Audio display with the head tracker enabled?

Q8. How did you find the Audio display when both signals were on the same side?

9.13 Appendix D: Study 4 Participant Information Sheet



School of Human Movement Studies

The University of Queensland
Brisbane Qld 4072 Australia
Telephone (07) 3365 6240
International +61 7 3365 6240
Facsimile (07) 3365 6877
Email enquiries@hms.uq.edu.au
Internet www.hms.uq.edu.au
CRICOS PROVIDER NUMBER 00025B

Participant Information Document

Study Title: Modifying the Spatial Position of Verbal Communication Improves Comprehension amongst Competing Sonified Signals within 3-D Audio Displays.

Researcher: John Towers (Advanced Human Factors Technologist, Boeing Research & Technology Australia.)

Purpose: The purpose of this study is to validate design techniques utilised in determining the optimal spatial position of verbal audio within a 3-D audio display.

Duration: Over the course of two non-consecutive days, each participant is requested to participate in 1 x 40 minute session for a hearing test and simulator training, and 1 x 40 minute session flying the simulator.

Procedure: The experiment will take place on an aircraft simulator at Boeing Defence Australia head office, Brisbane. Each participant will navigate the simulator about a predetermined flight plan utilising an audio display. Participants will be required to attempt to fly as accurate a course as possible while undertaking a verbal working memory task. The task requires the participant to imagine being directed about an imaginary 2-D grid through verbal instructions, such as 'forward 2, left 3', etc. After the last instruction, the participant will be asked to record their final resting quadrant on the grid.

Spatially positioned sounds will replace visual navigation instruments for waypoint bearing and course deviation. The spatial position of the spoken voice commands will vary between 6 positions over the course of the flight.

Risk: This study is not considered to pose any cognitive or physical risk to participants beyond that which would be expected in normal every day life. There is a slight chance that participants may experience motion sickness while viewing the flight simulator visuals. If this does occur, participants are requested to stop the experiment immediately and consider withdrawing from the study without prejudice.

Benefits: It is expected that participants will gain insight into the fundamentals of 3-D audio, along with an appreciation of the constraints and considerations associated with integrating spatial audio displays into systems interfaces. This will be a valuable experience for system design engineers wishing to expand their knowledge within this field.

Data: Performance data will be obtained and recorded for the following:

- accuracy in flying the simulator across a flight path
- answers for the verbal navigation task
- subjective workload ratings for each phase of the flight plan
- gaze fixation will be recorded throughout the experiment

Data obtained from this study will be treated in confidence and stored in a manner that prevents results from being associated with an individual. You will be provided full access to your results upon request, as well as the overall outcomes of the experiment once the study has been completed.

Reimbursement: Participation in this study is voluntary and requires no expense on the part of the participant. There will be no reimbursement to participants for their time.

Emergency: The study will be conducted within the usability laboratory at Boeing House, Brisbane. As such, any emergency will be treated as per the guidelines within the Boeing Defence Australia employee handbook (HB-BDA-001).



Withdrawal: Participation in this study is voluntary. It is the right of any participant to withdraw from the study at any time without prejudice. Data collected prior to withdrawal will be erased from all working and/or archived files.

Contact: Should you have any questions regarding the study, please do not hesitate to contact John Towers on 3306 3527, or email john.towers@boeing.com.

Debrief: Upon completion of the study, results will be provided to all participants through a report paper. A debrief presentation will be held for participants to explain the findings and future application of the research.

Complaints: This study has been cleared in accordance with the ethical review guidelines and processes of the University of Queensland. These guidelines are endorsed by the University's principal human ethics committee, the Human Experimentation Ethical Review Committee, and registered with the Australian Health Ethics Committee as complying with the National Statement. You are free to discuss your participation in this study with project staff (contactable on 3306 3527). If you would like to speak to an officer of the University not involved in the study, you may contact the School of Human Movement Studies Ethics Officer on 3365 6380 (Dr Tim Carroll).

9.14 Appendix D: Study 4 Participant Informed Consent Form



School of Human Movement Studies

The University of Queensland
Brisbane Qld 4072 Australia
Telephone (07) 3365 6240
International +61 7 3365 6240
Facsimile (07) 3365 6877
Email enquiries@hms.uq.edu.au
Internet www.hms.uq.edu.au
CRICOS PROVIDER NUMBER 00025B

Participant Informed Consent Form

Project Title: Modifying the Spatial Position of Verbal Communication Improves Comprehension amongst Competing Sonified Signals within 3-D Audio Displays.

Researcher: John Towers (Boeing Research & Technology Australia).

This study has been cleared in accordance with the ethical review guidelines and processes of the University of Queensland. These guidelines are endorsed by the University's principal human ethics committee, the Human Experimentation Ethical Review Committee, and registered with the Australian Health Ethics Committee as complying with the National Statement. You are free to discuss your participation in this study with project staff (contactable on 3306 3527; John Towers). If you would like to speak to an officer of the University not involved in the study, you may contact the School of Human Movement Studies Ethics Officer on 3365 6380 (Dr Tim Carroll).

1. I, the undersigned..... hereby acknowledge that I have read the information document, and that the specific sections of the document that are relevant to the present experiment have been drawn to my attention. I have been provided with a description of the experiment, including the purposes, methods, demands, and possible risks and inconveniences involved.
2. I am aware that I may withdraw from this research project at any time without penalty (even after I have signed this statement of participation), and that I am entitled to a thorough explanation of any procedure employed in the study. I understand that any information I provide will be treated confidentially, and that it I will not obtain any direct benefits from my participation other than what has been outlined in the participant information sheet.
3. I hereby consent to being a research participant in this study.

(Signed)

Date:

(Witnessed by).....

Date:

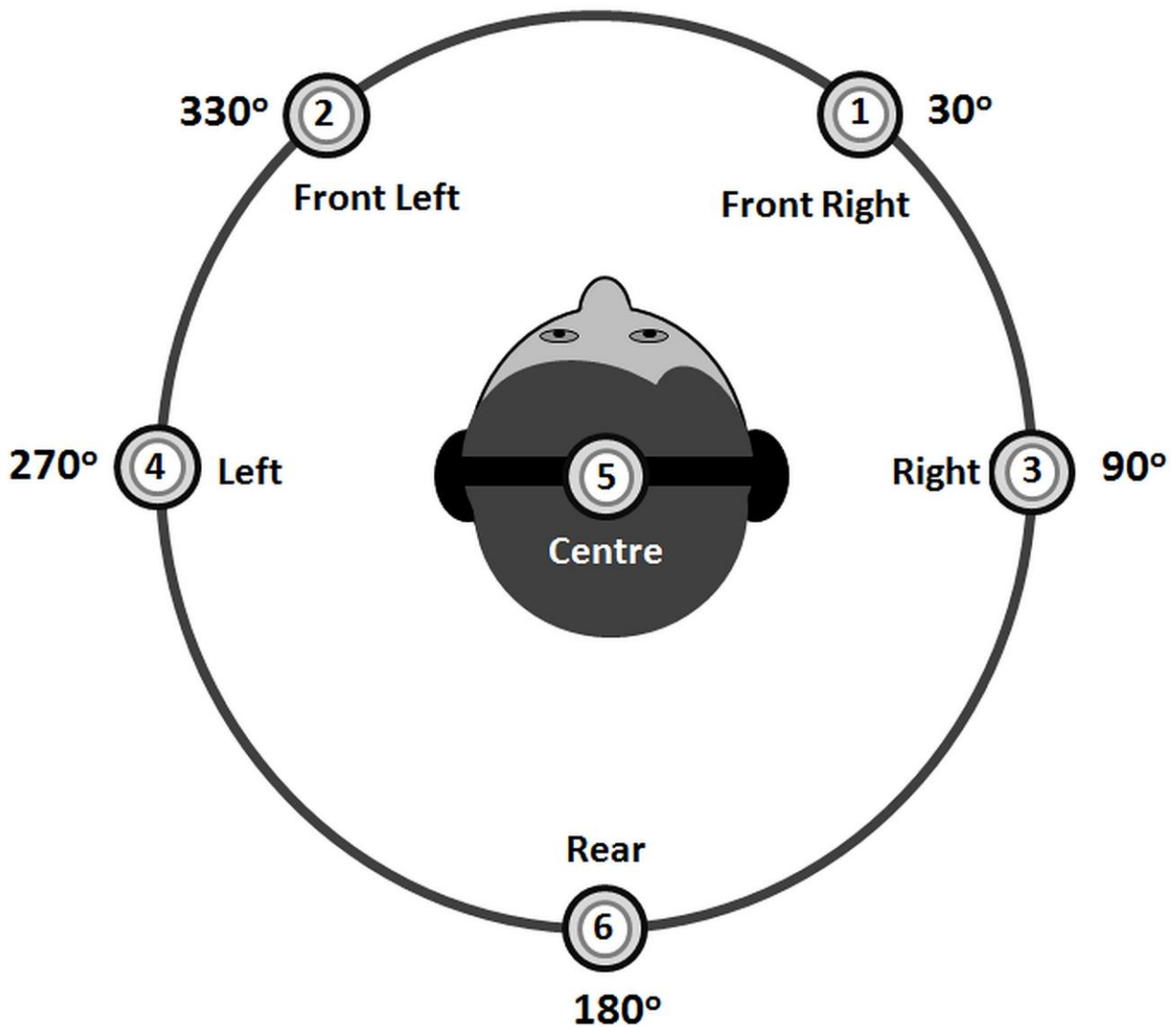
9.15 Appendix D: Study 4 Subjective Evaluation of Primary Task

Figure 9-5. Verbal instruction spatial positions for experiment 4. This diagram was provided to five listeners who, as part of a subjective evaluation of the primary task, evaluated the spatial locations of verbal instructions.

9.16 Appendix D: Study 4 Briefing Protocol

Participant Name: _____

Group: _____

Set up prior to participant arrival

1. Archive / delete previous logs
2. Ensure head tracker is connected and started
3. Ensure NTP is on
4. Ensure SIM sound is on 'Q' to toggle
5. Make sure flight is smooth and instruments positioned correctly

When Participant arrives

1. Set seat/eye height
2. Use A9000 world model in faceLAB
3. Ask to turn off phone
4. Calibrate eye tracker
5. Load program and set correct group

Tell student:

1. Not to make control movements too large – lead into WP turn.
2. Correct navigation at the smallest deviation – don't be complacent.
3. Remind them that the working memory navigation task is the highest priority and that they should neglect flying the simulator if needed to ensure that they successfully complete that task.
4. Don't adopt a strategy where you immediately neglect the secondary task in favour of the primary task.
5. Try to attempt to perform both tasks concurrently to the best of your ability.
6. Move head around to improve better localising
7. Use the audio to keep eyes up ensuring you maintain elevation. The GPS display won't provide enough information for accurate flight, so don't rely on it. It will only be valuable if you get really disoriented.
8. Upon completion of working memory instructions you should indicate in a loud and clear voice the quadrant that the navigation instructions ended in. No need to rush this answer.
9. The simulator will pause at the end of the flight plan and reposition back at the start. At this point, turn and complete the NASA-TLX ratings.
10. Ensure faceLAB is logging
11. Tell participant to un-pause the sim and begin

Shut down

Archive logs!!!

9.17 Appendix D: Study 4 NASA-TLX Interface

Screenshot of NASA-TLX Interface Utilised in Study 4



Figure 9-6. Computer Based NASA-TLX Interface

9.18 Appendix D: Study 4 Edinburgh Handedness Inventory Form

Participant ID: _____

Date: ____/____/____

Edinburgh Handedness Inventory (revised)					
<i>Please mark the box that best describes which hand you use for the activity in question</i>					
	<i>Always Left</i>	<i>Usually Left</i>	<i>No Preference</i>	<i>Usually Right</i>	<i>Always Right</i>
Writing					
Throwing					
Scissors					
Toothbrush					
Knife (without fork)					
Spoon					
Match (when striking)					
Computer mouse					

Indicate which ear you normally use when talking on the telephone? LEFT RIGHT

9.19 Appendix D: Study 4 Interview Protocol

Participant ID: _____

Date: ____/____/____

Q1. What are your thoughts on the different conditions?

Q2. Did you find any particular positioning of the verbal instructions easier to listen to than others?

Q3. Did you adopt any strategies when conducting the activity?

Q4. Did you notice that any positions were more difficult to concurrently attend to the sonifications and voice communications?
