

# **Spectral Analysis and Resolving Spatial Ambiguities in Human Sound Localization**

A dissertation presented to  
the Faculty of Electrical and Information Engineering  
of the University of Sydney  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy

by  
Craig T. Jin

© Craig T. Jin 2001

ALL RIGHTS RESERVED

To **Eugenia**

# Acknowledgements

If most Serene Prince I wished to set forth in this place all the praises due to your Highness's own merits and those of your distinguished family, I should be committed to such a lengthy discourse that this preface would far outrun the rest of the text, whence I shall refrain from even attempting the task, uncertain that I could finish half of it, let alone all (Galileo Galilei in *Operations of the Geometric and Military Compass*, see Sobel, 1999).

Having lived and grown up in the United States, I am grateful to Australia for harbouring me these last 5 years and providing me with a place at the University of Sydney to pursue a meaningful course of research. As a student, I have found only generosity on the part of Australia to support my research. The Dora Lush Postgraduate Research Scholarship provided me with the funds to pursue my Ph.D. and the Australian Academy of Science provided travel funds for a 3 week workshop in the U.S.A. (Scientific Visits to the U.S.A. for Young Australian Researchers).

Of course, even when the resources on a larger scale are in place, it still requires the determined efforts of those closer to home to enable those resources to become available. In this sense, it gives me great pleasure to thank Simon Carlile for his unfailing support. Despite the fact that I have been enrolled in the Department of Electrical and Information Engineering, it has really been Simon who has supported my research from start to finish. A phrase that I have come to associate with Simon is “good taste and judgement” – a way of life, really. His guidance has been most crucial in the development of this dissertation at *every* step and his friendship most helpful in finishing it.

In the early stages of my research, Philip Leong and Marcus Schenkel provided much needed support and guidance. If you just had to get something working, Philip could always help you. Now that I've finished the easier task, Philip can teach me the harder one: fly fishing. In the early modeling work, Marcus helped me to see which issues were important and showed me how time-delay neural networks worked. I would like to thank both Philip and Marcus for their friendship and support.

In the later stages of my research, André van Schaik played a greater and greater role in the unfolding of ideas. At times, it seemed like that of all people, I needed the fewest words to describe an idea to him. Why this should be so is puzzling, for his native language and background differ from mine. André's critical eye has in numerous ways been invaluable and his concern always helped carry things along. Thanks André.

Frequently, I have felt that it would have been impossible to finish if it were not for Johahn Leung, Anna Corderoy, and Anandhi Anandan. They provided ideas, laughs, and day-to-day support in the lab. At one time, Jogi and I were the only postgraduate students in the lab and we could basically do anything we wanted to and so we did. We did the first HRTF morphing experiments together and, with Philip's help, got the water balloon into the anechoic chamber. Jogi, of course, made Microsoft's frustrating way

of computing almost bearable. Anna helped tremendously with the experiments and Anandhi helped in countless ways with so many details. Jogi, Anna and Anandhi – you have such class and insight, it's a heart-felt thanks that I give you for enriching my time here immeasurably.

Then there are those discussions that changed a stale environment into an invigorating one full of humour and warmth. Oliver Behrend has shown me his understanding of those final stages. Virginia Best and Ruben Kurilowich, now in the midst of their own Ph.D.'s, have shared their research, laughs, and sympathies with me. Stacey Harris shared her honours year with me and Stephanie Hyams my first auditory localization experiment. Thanks to all of those who participated in my auditory localization experiments. Thanks to those around the physiology department and SEDAL for their friendship and more: Sam Solomon, Andrew White, Paul Martin, Tom FitzGibbon, Bogden Dreher, Paddy Fitzgerald, John Dodson, Richard Coggins.

Thanks to John Ferris and Louise Berben for the wonderful Wednesday evening outings, swims and discussions. Thanks to Andrew, Alison and little Kathleen Garvie for the relaxing break at Lake Jindabyne.

Thanks to Dennis Moore and Sanjoy Mahajan whose friendships over the years have given me so much strength and pleasure.

Thanks to my family and Eugenia's family for their unconditional support and love and for giving me a home.

Thanks to my beloved wife for her inspiration, support and love. To me, as it has been said before,

She walks in beauty, like the night  
Of cloudless climes and starry skies;  
And all that's best of dark and bright  
Meet in her aspect and her eyes . . .

Finally, if research can be likened to a search in the forest in which nobody, not even yourself, knows or has seen exactly where you want to go, then it is by working with others that you can often find the paths that might lead in an interesting, if not always correct direction. To those who have shared the journey, I would like to say that perhaps your most valuable gift to me is something that I may not easily be able to see myself. Time passes and we change, perhaps develop. In those changes are the results of much time and effort and yet I suspect that the changes in oneself are actually the most difficult to see. But if you see them, you can reflect upon them and the gifts you've given me.

# **Spectral Analysis and Resolving Spatial Ambiguities in Human Sound Localization**

Craig T. Jin

University of Sydney 2001

## **Abstract**

This dissertation provides an overview of my research over the last five years into the spectral analysis involved in human sound localization. The work involved conducting psychophysical tests of human auditory localization performance and then applying analytical techniques to analyze and explain the data. It is a fundamental thesis of this work that human auditory localization response directions are primarily driven by the auditory localization cues associated with the acoustic filtering properties of the external auditory periphery, i.e., the head, torso, shoulder, neck, and external ears. This work can be considered as composed of three parts.

In the first part of this work, I compared the auditory localization performance of a human subject and a time-delay neural network model under three sound conditions: broadband, high-pass, and low-pass. A “black-box” modeling paradigm was applied. The modeling results indicated that training the network to localize sounds of varying center-frequency and bandwidth could degrade localization performance results in a manner demonstrating some similarity to human auditory localization performance.

As the data collected during the network modeling showed that humans demonstrate striking localization errors when tested using bandlimited sound stimuli, the second part of this work focused on human sound localization of bandpass filtered noise stimuli. Localization data was collected from 5 subjects and for 7 sound conditions: 300 Hz to 5 kHz, 300 Hz to 7 kHz, 300 Hz to 10 kHz, 300 Hz to 14 kHz, 3 to 8 kHz, 4 to 9 kHz, and 7 to 14 kHz. The localization results were analyzed using the method of cue similarity indices developed by Middlebrooks (1992). The data indicated that the energy level in relatively wide frequency bands could be driving the localization response directions, just as in Butler’s covert peak area model (see Butler and Musicant, 1993).

The question was then raised as to whether the energy levels in the various frequency bands, as described above, are most likely analyzed by the human auditory localization

system on a monaural or an interaural basis. In the third part of this work, an experiment was conducted using virtual auditory space sound stimuli in which the monaural spectral cues for auditory localization were disrupted, but the interaural spectral difference cue was preserved. The results from this work showed that the human auditory localization system relies primarily on a monaural analysis of spectral shape information for its discrimination of directions on the cone of confusion.

The work described in the three parts lead to the suggestion that a spectral contrast model based on overlapping frequency bands of varying bandwidth and perhaps multiple frequency scales can provide a reasonable algorithm for explaining much of the current psychophysical and neurophysiological data related to human auditory localization.

# Table of Contents

Abstract . . . . .	iii
List of Tables . . . . .	x
List of Figures . . . . .	xi
Glossary . . . . .	xvi
Preface . . . . .	xxi
0.1 Auditory Perception . . . . .	xxi
0.2 A Personal Statement . . . . .	xxiii
0.3 Reading the Dissertation . . . . .	xxvi
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Outline . . . . .	5
1.3 Question in Perspective . . . . .	7
1.4 Hearing . . . . .	10
Notes to Chapter 1 . . . . .	17
<b>2 Background – Interaural Cues</b>	<b>18</b>
2.1 Overview . . . . .	19
2.2 Interaural Time Difference . . . . .	20
2.2.1 ITDs and Spatial Information . . . . .	20
2.2.2 ITDs and Psychophysics . . . . .	22
2.2.3 ITDs, Models, and the Auditory System . . . . .	26
2.3 Interaural Intensity Difference . . . . .	27
2.3.1 IIDs and Spatial Information . . . . .	27
2.3.2 IIDs and Psychophysics . . . . .	29
2.3.3 IIDs, Models, and the Auditory System . . . . .	32
2.4 Implications . . . . .	34
Notes to Chapter 2 . . . . .	35
<b>3 Background – Spectral Shape Cues</b>	<b>37</b>
3.1 Spectral Shape and Spatial Information . . . . .	37
3.1.1 Acoustics of the External Auditory Periphery . . . . .	38
3.1.2 Spectral Variations Across Space . . . . .	40
3.2 Spectral Shape and Psychophysics . . . . .	54
3.2.1 Primary Spectral Notch . . . . .	56
3.2.2 Localization of Narrow-Band Sounds . . . . .	59



3.2.3	Low Frequency Cues and Across-Frequency Interactions . . . . .	69
3.2.4	Spectral Profile Analysis . . . . .	72
3.2.5	Discrimination of Spectral Peaks and Notches . . . . .	74
3.2.6	Deciphering Source Spectrum . . . . .	76
3.2.7	Temporal Factors and Spectral Shape . . . . .	81
3.2.8	Spectral Scrambling . . . . .	83
3.3	Computational Models and Spectral Shape . . . . .	90
3.3.1	Template-Matching and Covert Peaks . . . . .	90
3.3.2	Three Sound Localization Models . . . . .	92
3.4	Spectral Shape and the Auditory System . . . . .	98
3.4.1	Intensity Discriminations . . . . .	98
3.4.2	Spectral Shape and the Auditory Nerve . . . . .	99
3.4.3	Spectral Shape and the Brainstem . . . . .	100
3.4.4	Spectral Contrast in the MGB . . . . .	102
3.4.5	Spectral Shape in Auditory Cortex . . . . .	105
3.5	Implications . . . . .	111
	Notes to Chapter 3 . . . . .	114
<b>4</b>	<b>Methods for Sound Localization</b>	<b>123</b>
4.1	Overview . . . . .	123
4.2	Sound Localization Environment and Paradigm . . . . .	124
4.2.1	Head Pointing Training . . . . .	127
4.2.2	Free-Field Sound Presentation . . . . .	128
4.2.3	Measurement of the Directional Transfer Functions . . . . .	130
4.2.4	VAS Sound Presentation . . . . .	134
4.3	Localization Analysis . . . . .	139
4.3.1	Quadrature Plots . . . . .	139
4.3.2	Spherical Statistics . . . . .	140
4.3.3	Spherical Interpolation . . . . .	143
4.3.4	Circular Statistics . . . . .	144
4.3.5	Relating Spectral Cues to Localization Responses . . . . .	150
	Notes to Chapter 4 . . . . .	159
<b>5</b>	<b>Frequency Division and Integration</b>	<b>165</b>
5.1	Overview . . . . .	165
5.2	Previous Auditory Localization Models . . . . .	167
5.2.1	Difficulties in Modeling Auditory Localization . . . . .	168
5.2.2	Localization Models using HRTFs . . . . .	169
5.2.3	Neural System Modeling . . . . .	170
5.3	Methods . . . . .	171
5.3.1	Measuring Sound Localization Performance . . . . .	171
5.3.2	Measuring DTFs . . . . .	173
5.3.3	A Network Model of Sound Localization . . . . .	173
5.4	Network Architectures . . . . .	181
5.4.1	MLP Architecture . . . . .	182
5.4.2	TDNN Architecture A . . . . .	183

5.4.3	TDNN Architecture B . . . . .	184
5.4.4	TDNN Architecture C . . . . .	185
5.5	Training Networks to Localize Sounds . . . . .	186
5.5.1	Learning Algorithm . . . . .	186
5.5.2	Structure of the Training Data . . . . .	187
5.5.3	Frequency Selective Training . . . . .	188
5.6	Results . . . . .	189
5.6.1	Localization Performance of the Subject . . . . .	189
5.6.2	Localization Performance of the MLP . . . . .	192
5.6.3	Localization Performance of the TDNN with Architecture A . . . . .	195
5.6.4	Localization Performance of the TDNN with Architecture B . . . . .	196
5.6.5	Localization Performance of the TDNN with Architecture C . . . . .	196
5.6.6	Restricted High Frequency Sound Localization . . . . .	199
5.6.7	Sound Localization at Varying Sound Levels . . . . .	203
5.6.8	TDNN Encoding of Temporal Information . . . . .	206
5.7	Discussion . . . . .	208
5.7.1	Tonotopic Processing from a Computational Viewpoint . . . . .	208
5.7.2	Matched Filtering and Sound Localization . . . . .	210
5.8	Conclusions and Implications . . . . .	212
	Notes to Chapter 5 . . . . .	214
<b>6</b>	<b>Localization of Bandpass Sounds</b>	<b>215</b>
6.1	Overview . . . . .	215
6.2	Previous Localization Experiments Using Bandlimited Sounds . . . . .	217
6.2.1	Effect of Bandwidth on Auditory Localization . . . . .	218
6.2.2	Auditory Localization of Bandpass Peaks and Bandstop Notches . . . . .	224
6.3	Methods . . . . .	229
6.3.1	Listeners and Localization Paradigm . . . . .	229
6.3.2	Stimuli and Procedure . . . . .	230
6.3.3	Directional Transfer Functions and Directional Excitation Pat- terns . . . . .	232
6.3.4	Statistical Methods for Localization Performance Data . . . . .	233
6.4	Results . . . . .	234
6.4.1	Control Localization Performance . . . . .	235
6.4.2	Localization Performance for the Block A Tests . . . . .	235
6.4.3	Localization Performance for the Block B Tests . . . . .	238
6.5	Analysis of the Lateral and Polar Angles of the Localization Data . . . . .	244
6.5.1	Lateral Angle Analysis . . . . .	246
6.5.2	Polar Angle Analysis . . . . .	248
6.6	Analysis of the Spectral Cues with Respect to the Localization Data . . . . .	265

6.6.1	Mislocalizations and Spectral Cues . . . . .	265
6.6.2	Computing the Correlation between Localization Responses and Spectral Cues . . . . .	268
6.6.3	Spatial Distribution of the Cue Similarity Indices . . . . .	274
6.6.4	Spatial Distribution of the Cue Similarity Indices for the Covert Spectral Contrast Cue . . . . .	298
6.6.5	Statistical Comparison of the Spectral Cues with the Response Directions . . . . .	298
6.6.6	Monaural Spectral Shape Cue . . . . .	314
6.6.7	Explaining Auditory Localization of Low-Pass Filtered Noise . . . . .	316
6.7	Spectral Shape and the Covert Peak Area . . . . .	318
6.8	Conclusion . . . . .	320
	Notes to Chapter 6 . . . . .	321
<b>7</b>	<b>Contrasting the Spectral Cues</b> . . . . .	<b>322</b>
7.1	Overview . . . . .	322
7.2	Introduction . . . . .	324
7.3	VAS and the ISD Cue . . . . .	330
7.4	Methods . . . . .	332
7.4.1	Listeners and Localization Paradigm . . . . .	332
7.4.2	Measuring Directional Transfer Functions . . . . .	333
7.4.3	Estimation of the Interaural Spectrum . . . . .	333
7.4.4	The VAS Sound Stimuli . . . . .	334
7.4.5	Testing Procedure . . . . .	339
7.5	Graphical Presentation of Localization Data . . . . .	339
7.5.1	Spherical Statistics . . . . .	339
7.5.2	Scatter Plots of Localization Data . . . . .	340
7.5.3	Circular Statistics . . . . .	341
7.6	Results . . . . .	342
7.6.1	Psychophysical Validation of DTFs in VAS . . . . .	342
7.6.2	Control Localization Performance . . . . .	344
7.6.3	Localization Performance in the Two Test Sound Conditions . . . . .	345
7.7	Analysis of the Lateral and Polar Angles of the Localization Data . . . . .	351
7.7.1	Lateral Angle Analysis . . . . .	353
7.7.2	Polar Angle Analysis . . . . .	355
7.8	Analysis of the Spectral Cues with respect to the Localization Data . . . . .	357
7.8.1	Extracting the Spectral Cues . . . . .	362
7.8.2	Cue Similarity Indices . . . . .	363
7.9	Contrasting the Monaural and Interaural Spectral Cues . . . . .	364
7.10	Statistical Analysis of the Cue Similarity Indices . . . . .	372
7.11	Explaining the Localization Data for the Left Hemisphere of Space . . . . .	375

7.12 Conclusion . . . . .	381
Notes to Chapter 7 . . . . .	384
<b>8 Conclusions</b>	<b>385</b>
8.1 A Few Questions and Answers . . . . .	386
8.2 A Broader Context . . . . .	387
8.2.1 Computational Perspective . . . . .	387
8.2.2 Spectral Contrast and Wide Frequency Bands . . . . .	390
8.3 Psychophysical Perspective . . . . .	399
8.4 Neurophysiological Perspective . . . . .	401
8.5 Concluding Remarks . . . . .	404
Notes to Chapter 8 . . . . .	405
<b>A Spherical Directionality Plots</b>	<b>406</b>
A.1 Directionality Plots . . . . .	406
<b>B Multiscale Spectral Analysis</b>	<b>417</b>
<b>C Spherical Thin-Plate Spline</b>	<b>421</b>
<b>D Individualized Virtual Auditory Space</b>	<b>427</b>
<b>Bibliography</b>	<b>433</b>

# List of Tables

5.1	The command line options used for the auditory image model. . . . .	176
5.2	Comparison of different models and their spectral resolution. . . . .	177
5.3	The MLP network. . . . .	183
5.4	The TDNN with architecture A. . . . .	183
5.5	The TDNN with architecture B. . . . .	184
5.6	The TDNN with architecture C. . . . .	185
5.7	A comparison of weights and patterns . . . . .	188
5.8	Comparison of the localization performance results for the subject and the competing models. Fixed and random refer to the type of training. .	202
5.9	Performance statistics for the restricted high-frequency sound condition	203
6.1	Summary localization statistics for the two control conditions. . . . .	238
6.2	Summary localization statistics for the bandpass sound conditions. . . .	245
6.3	Linear fit between target and response lateral angles . . . . .	248
7.1	A comparison of free-field and VAS localization performance . . . . .	343
7.2	Summary localization statistics for the two control conditions. . . . .	347
7.3	Summary localization statistics for the two test conditions. V. I. and V. R. refer to Veridical Interaural and Veridical Right, respectively. . .	355
7.4	Linear fit between target and response lateral angles . . . . .	355
7.5	Comparison of the localization performance between the V. I. and V. R. sound conditions in the left and right hemispheres of space . . . . .	357

# List of Figures

1	RoboCraig – a life-like acoustical mannequin . . . . .	xxv
1.1	Physical structure of the auditory sensory apparatus . . . . .	11
1.2	Cross-section of the cochlea and the organ of Corti . . . . .	13
1.3	Schematic of the basilar membrane . . . . .	14
1.4	The auditory sensory system . . . . .	16
2.1	Path length difference between the ears . . . . .	21
2.2	ITD curves . . . . .	22
2.3	Colored ITD surface plot . . . . .	23
2.4	Cone of confusion . . . . .	24
2.5	Iso-ITD and Iso-ILD contours as a function of spatial location in the proximal region of space for an acoustically transparent head . . . . .	30
2.6	Colored ILD surface plot . . . . .	31
3.1	Morphology of the outer ear . . . . .	40
3.2	Measured resonance modes of the pinna . . . . .	40
3.3	Computed resonance modes of the pinna . . . . .	41
3.4	Planar directionality plots . . . . .	44
3.5	Spherical directionality plots average across subjects . . . . .	46
3.6	Directional excitation patterns . . . . .	48
3.7	Primary spectral notch . . . . .	52
3.8	More spectral features . . . . .	54
3.9	Front-back spectral characteristics . . . . .	55
3.10	DTF alignment by frequency scaling . . . . .	57
3.11	Iso-notch-frequency contours . . . . .	58
3.12	Directional bands . . . . .	60
3.13	Narrow-band polar angle plots . . . . .	67
3.14	Low-frequency cues and the horizontal plane . . . . .	71
3.15	Low-frequency cues and the median plane . . . . .	72
3.16	Size of spectral peaks and notches . . . . .	75
3.17	Notch and peak detection . . . . .	77
3.18	Frequency DLs . . . . .	78
3.19	Localization of temporally-different broadband noise . . . . .	84
3.20	Explanation of temporal factors . . . . .	85
3.21	Spectrally-scrambled noise . . . . .	86
3.22	Human localization of spectrally-scrambled noise . . . . .	87
3.23	Monaural Localization of Spectrally Scrambled Noise . . . . .	89

3.24	Localization performance of the spectral gradient model . . . . .	93
3.25	Four spectral-feature operators . . . . .	94
3.26	Localization performance of the feature detection model . . . . .	95
3.27	Localization performance of a multi-scaled gradient model . . . . .	96
3.28	Spectral contrast mechanism . . . . .	103
3.29	Location-level response areas in MGB . . . . .	104
3.30	Frequency response areas in MGB . . . . .	106
3.31	Frequency response areas in A1 . . . . .	108
3.32	Spatial distribution of receptive field parameters in A1 . . . . .	109
3.33	Spatial distribution of relative firing rate and onset latency in A1 . . . . .	110
3.34	Auditory cortex . . . . .	112
4.1	Sound localization environment and tools . . . . .	125
4.2	Localization coordinate system . . . . .	126
4.3	Frequency response of the loudspeaker and HRTF measurement system . . . . .	129
4.4	Noise floor of the HRTF recordings . . . . .	132
4.5	VAS sound localization . . . . .	135
4.6	ER-2 earphone calibration curves . . . . .	136
4.7	Attenuation of the primary spectral notch . . . . .	138
4.8	Circular raw data plot of a response polar angle distribution . . . . .	145
4.9	Mean response direction calculated as a vector summation . . . . .	147
4.10	Five circular data distributions with increasing values for the concentration parameter . . . . .	148
4.11	Circular hair plots show the mapping between the response and the target polar angles . . . . .	150
4.12	Calculation of the Directional Excitation Pattern . . . . .	152
4.13	Cue visualization toolbox . . . . .	156
4.14	Four examples of cue directionality plots . . . . .	157
5.1	Neural system modeling paradigm . . . . .	171
5.2	Sound localization model . . . . .	175
5.3	Summing node of a time-delay neuron . . . . .	178
5.4	Time-delay neural network . . . . .	179
5.5	Target neural activity pattern . . . . .	182
5.6	Tonotopic neural architecture . . . . .	186
5.7	Free-field and VAS localization responses . . . . .	191
5.8	MLP localization responses . . . . .	194
5.9	Localization responses of TDNN with architecture A . . . . .	197
5.10	Localization responses of TDNN with architecture B . . . . .	198
5.11	Localization responses of TDNN with architecture C . . . . .	200
5.12	Comparison of the human and the TDNN with architecture C . . . . .	201
5.13	Localization for sounds bandpass filtered 7 637 to 13 264 Hz . . . . .	204
5.14	Matched filtering for sounds bandpassed 7 637 to 13 264 Hz . . . . .	205
5.15	Sound level and localization performance accuracy . . . . .	206
5.16	TDNN with architecture C and sound level . . . . .	207
5.17	TDNN neural network weights . . . . .	208
6.1	Bandwidth and Localization Accuracy . . . . .	220

6.2	Elevation discrimination and low-pass filtering . . . . .	222
6.3	Elevation discrimination and high-pass filtering . . . . .	223
6.4	DEPs and bandpass peaks and bandstop notches . . . . .	226
6.5	Localization performance data for Control A . . . . .	236
6.6	Localization performance data for Control B . . . . .	237
6.7	Localization performance data for the 300 Hz to 5 kHz sound condition	239
6.8	Localization performance data for the 300 Hz to 7 kHz sound condition	240
6.9	Localization performance data for the 300 Hz to 10 kHz sound condition	241
6.10	Localization performance data for the 3 to 8 kHz sound condition . . .	242
6.11	Localization performance data for the 4 to 9 kHz sound condition . . .	243
6.12	Localization performance data for the 7 to 14 kHz sound condition . .	244
6.13	Scatter plot of the lateral angle data for Control A . . . . .	247
6.14	Scatter plot of the pooled lateral angle data . . . . .	249
6.15	Scatter plot of the pooled lateral angle data . . . . .	250
6.16	Scatter plot of the pooled lateral angle data . . . . .	251
6.17	Scatter plot of the pooled lateral angle data . . . . .	252
6.18	Scatter plot of the pooled lateral angle data . . . . .	253
6.19	Scatter plot of the pooled lateral angle data . . . . .	254
6.20	Circular hair plot of the polar angle data for Control A . . . . .	256
6.21	Circular hair plot of the polar angle data for the 300 Hz to 5 kHz sound condition . . . . .	259
6.22	Circular hair plot of the polar angle data for the 300 Hz to 7 kHz sound condition . . . . .	260
6.23	Circular hair plot of the polar angle data for the 300 Hz to 10 kHz sound condition . . . . .	261
6.24	Circular hair plot of the polar angle data for the 3 to 8 kHz sound condition . . . . .	262
6.25	Circular hair plot of the polar angle data for the 4 to 9 kHz sound condition . . . . .	263
6.26	Circular hair plot of the polar angle data for the 7 to 14 kHz sound condition . . . . .	264
6.27	Circular hair plot of the polar angle data for Subject C in 4 sound con- ditions . . . . .	267
6.28	Directionality plots for the overt and covert spectral cues for the 300 Hz to 5 kHz sound condition for Subject A . . . . .	277
6.29	Directionality plots for the overt and covert spectral cues for the 300 Hz to 5 kHz sound condition for Subject B . . . . .	278
6.30	Directionality plots for the overt and covert spectral cues for the 3 to 8 kHz sound condition for Subject C . . . . .	280
6.31	Directionality plots for the overt and covert spectral cues for the 3 to 8 kHz sound condition for Subject B . . . . .	282
6.32	Directionality plots for the overt and covert spectral cues for the 4 to 9 kHz sound condition for Subject D . . . . .	284
6.33	Directionality plots for the overt and covert spectral cues for the 4 to 9 kHz sound condition for Subject C . . . . .	286
6.34	Directionality plots for the overt and covert spectral cues for the 7 to 14 kHz sound condition for Subject B . . . . .	288



6.35	Directionality plots for the overt and covert spectral cues for the 7 to 14 kHz sound condition for Subject D . . . . .	289
6.36	Directionality plots for the overt and covert spectral cues for the 300 Hz to 5 kHz sound condition for Subject B . . . . .	292
6.37	Directionality plots for the overt and covert spectral cues for the 3 to 8 kHz sound condition for Subject C . . . . .	293
6.38	Directionality plots for the overt and covert spectral cues for the 4 to 9 kHz sound condition for Subject B . . . . .	295
6.39	Directionality plots for the overt and covert spectral cues for the 7 to 14 kHz sound condition for Subject C . . . . .	297
6.40	Directionality plots of the covert spectral contrast area . . . . .	299
6.41	Bar plots show the percentage of responses accounted for by the overt and covert spectral cues in the 300 Hz to 5 kHz sound condition . . . . .	302
6.42	Bar plots show the percentage of responses accounted for by the overt and covert spectral cues in the 3 to 8 kHz sound condition . . . . .	306
6.43	Bar plots show the percentage of responses accounted for by the overt and covert spectral cues in the 4 to 9 kHz sound condition . . . . .	308
6.44	Bar plots show the percentage of responses accounted for by the overt and covert spectral cues in the 7 to 14 kHz sound condition . . . . .	310
6.45	Directionality plots of the ipsilateral covert notch area . . . . .	312
6.46	Bar plots show the percentage of responses accounted for by the monaural and interaural spectral shape cues in four sound conditions: 300 Hz to 5 kHz, 3 to 8 kHz, 4 to 9 kHz, 7 to 14 kHz . . . . .	315
6.47	Spatial distribution of the covert spectral cues for a series of low-passed noise stimuli . . . . .	318
7.1	Comparison of interaural spectrum for the Control and V. I. sound condition . . . . .	338
7.2	Free-field versus VAS performance data . . . . .	344
7.3	Control localization performance data for Subject D . . . . .	346
7.4	Control B localization performance data . . . . .	348
7.5	Localization performance data for the Veridical Interaural condition . . . . .	349
7.6	Localization performance data for the Veridical Right condition . . . . .	350
7.7	Scatter plot of front-back and elevation angles (Subject A) . . . . .	351
7.8	Scatter plot of front-back and elevation angles (Subject B) . . . . .	352
7.9	Scatter plot of front-back and elevation angles (Subject C) . . . . .	353
7.10	Scatter plot of front-back and elevation angles (Subject D) . . . . .	354
7.11	Scatter plot of the pooled lateral angle data . . . . .	356
7.12	Circular hair plot of the polar angle data for the Control sound condition . . . . .	358
7.13	Circular hair plot of the polar angle data for the V. I. sound condition . . . . .	360
7.14	Circular hair plot of the polar angle data for the V. R. sound condition . . . . .	361
7.15	Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $-46^\circ, 20^\circ$ ) . . . . .	365
7.16	Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $-64^\circ, 40^\circ$ ) . . . . .	366
7.17	Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $-46^\circ, -20^\circ$ ) . . . . .	367

7.18 Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $-134^\circ, 20^\circ$ ) . . . . 368

7.19 Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $-134^\circ, -20^\circ$ ) . . . 369

7.20 Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $0^\circ, 0^\circ$ ) . . . . . 370

7.21 Cue directionality plots of the monaural spectral cues and the ISD cue for the Control sound condition and target location ( $180^\circ, 0^\circ$ ) . . . . . 371

7.22 Bar plots show the percentage of responses accounted for by the monaural and interaural spectral shape cues in 3 sound conditions: Control sound condition, V. I. sound condition, V. R. sound condition . . . . . 373

7.23 Bar plots show the percentage of responses accounted for by the monaural and interaural spectral shape cues in the free-field sound condition . 374

7.24 Cue directionality plots of the monaural spectral cue for the left ear for the V. I. and V. R. sound condition . . . . . 376

7.25 Cue directionality plots of the monaural spectral cue for the left ear combined with the ISD cue for the V. I. and V. R. sound condition . . . 377

7.26 Average DEP for the response directions in the left hemisphere of space 379

7.27 Cue directionality plots of the covert peak area for the left ear for the V. I. and V. R. sound condition . . . . . 381

7.28 Cue directionality plots of the covert peak area for the left ear combined with the ISD cue for the V. I. and V. R. sound condition . . . . . 382

8.1 Directional excitation patterns for flat-spectrum and spectrally-scrambled broadband noise . . . . . 392

8.2 Cue directionality plots of the spectral contrast areas for several frequency bands for the flat-spectrum broadband noise . . . . . 393

8.3 Cue directionality plots of the spectral contrast areas for several frequency bands for the spectrally-scrambled broadband noise . . . . . 394

8.4 Directional excitation patterns for a broadband and bandpass filtered broadband noise . . . . . 396

8.5 Cue directionality plots of the spectral contrast areas for several frequency bands for the bandpass filtered broadband noise . . . . . 397

8.6 Distribution of center frequency in the lateral belt area . . . . . 398

A.1 Spherical directionality plots . . . . . 408

# Glossary

**A1:** A1 refers to primary auditory cortex.

**AIM:** AIM refers to the Auditory Image Model (Patterson and Allerhand, 1995; Giguère and Woodland, 1994) which simulates the spectro-temporal characteristics of peripheral auditory processing.

**ANTERIOR:** Anterior refers to the region in front.

**AUDIO-VISUAL HORIZON:** The audio-visual horizon refers to the horizontal plane containing the interaural axis between the two ears.

**CONTRALATERAL:** Contralateral refers to the opposite side.

**CF:** Characteristic frequency refers to the best response frequency of a neuron.

**CIRCULAR HAIR PLOT:** A circular hair plot is a graphical plot used for showing the mapping between two circular variables. In this plot, a circle is drawn with “hair lines”. One end of the hair line segment touches the circle and its position on the circle indicates the value of one of the circular variables. The other end of the hair line segment points in the direction that maps or corresponds to the other circular variable.

**CM:** CM refers to the caudalmedial area adjacent to the primary auditory cortex.

**CPA:** The covert peak area for a given frequency refers to the location in space which has maximum gain for that frequency relative to all other locations. Importantly, this location does not have to be the same as the location(s) with a local peak in the sound spectrum or excitation pattern at that frequency.

**CRITICAL BAND:** The critical band is a frequency band that is approximately 15% of a frequency band’s center frequency. Psychophysical data indicate that the auditory system seems to analyze spectral information within a critical band differently from that outside of a critical band. For a further discussion see endnote 17 for Chapter 3.

**CUE CORRELATION VALUE:** A cue correlation value refers to numerical measure of the similarity between two auditory localization cues.

**CUE SIMILARITY INDEX:** A cue similarity index refers to a cue correlation value that has been normalized by subtracting the mean cue correlation value across space and dividing by the standard deviation.

**DIOTIC:** Diotic refers to a listening condition in which the same sound stimuli are presented to each ear.

**DICHOTIC:** Dichotic refers to a listening condition in which the different sound stimuli are presented to each ear.

**CUE DIRECTIONALITY PLOTS:** Cue directionality plots refer to a graphical presentation that indicates which directions in space best correlate with a given acoustic cue. Generally, the brighter or lighter the color, the better the direction matches the given acoustic cue.

**DEP:** DEP refers to the directional excitation pattern which is the excitation pattern for a flat-spectrum broadband sound originating from a specific direction in space. It is generally computed by filtering a Gaussian white noise with a DTF filter and then passing the directional sound through a cochlear model.

**DTF:** A DTF refers to an HRTF that has had the RMS of the HRTFs across all locations deconvolved from it.

**ER-2:** The ER-2 is an earphone manufactured by Etymotic Research which is designed to have a flat frequency transfer function to the human eardrum. This earphone was used for all VAS experiments and it is shown in Figure 4.5.

**EXCITATION PATTERN:** An excitation pattern refers to a pattern of neural excitation within the auditory nerve. It is generally computed using a cochlear model.

**EXTRAPERSONAL SPACE:** Extrapersonal space relates to the image of objects outside of the body.

**FREQUENCY DIVISION:** Frequency division refers to the progressive integration of acoustic information across frequency as computational processing proceeds from the input layers to the output layers of a neural network model.

**HRTF:** The head-related transfer function refers to the acoustic frequency response of the external auditory periphery. It is a complex-valued frequency spectrum composed of the magnitude and phase spectrums that mathematically describe the acoustical filtering properties of the external auditory periphery. The magnitude spectrum describes the the acoustic gain or attenuation of the external auditory periphery as a function of frequency and varies with spatial location.

**IPSILATERAL:** Ipsilateral refers to the same side.

**ITD:** The interaural time difference cue refers to the time delay between the signal at the two ears.

**IID:** The interaural intensity difference cue refers to the intensity differences between the signal at the two ears.

**ILD:** The interaural level difference cue refers to the overall difference in signal intensity at the two ears averaged across frequency.

**ISD:** The interaural spectral difference cue refers to the difference in the intensity patterns across frequency for the two ears.

**JND:** The just-noticeable difference refers to the smallest quantum of change in a psychophysical variable that is perceptually detectable.

**LATERAL ANGLE:** The lateral angle indicates the laterality of a spatial position in the double-pole or lateral-polar angle coordinate system. Consider the common spherical coordinate system with the Z-axis pointing up. The lateral-polar angle coordinate system is then formed by simply rotating the spherical coordinate system so that the Z-axis is now aligned with the old Y-axis. In the auditory localization literature, the Z-axis of the lateral-polar angle coordinate system is usually aligned with the listener's interaural axis (see Figure 4.2). Let  $\theta$  be the angle with respect to the Z-axis and let  $\phi$  be the polar angle in the XY-plane. The angle  $\theta$  is then the lateral angle in the double-pole or lateral-polar angle coordinate system.

**MEDIAN PLANE:** The median plane is the same as the midsagittal plane.

**MIDSAGITTAL PLANE:** The midsagittal plane or median plane is the vertical plane perpendicular to the interaural axis that divides the body into two approximately bilaterally symmetric halves.

**MSO:** The Medial Superior Olive is part of the Superior Olivary Complex (SOC). The nucleus receives binaural input and is generally associated with interaural phase differences.

**NORMALIZED ENERGY LEVEL:** A normalized energy level refers to the energy in a given frequency band that has been normalized with respect to the energy level in its adjacent side bands.

**OCTAVE:** An octave refers to a frequency interval corresponding to a doubling of frequency.

**PINNA:** The pinna refers to the external ear.

**POLAR ANGLE:** The polar angle refers to a component angle of the double-pole or lateral-polar angle coordinate system that is frequently used to indicate a direction in space. Consider the common spherical coordinate system with the Z-axis pointing up. The lateral-polar angle coordinate system is then formed by simply rotating the spherical coordinate system so that the Z-axis is now aligned with the

old Y-axis. In the auditory localization literature, the Z-axis of the lateral-polar angle coordinate system is usually aligned with the listener's interaural axis (see Figure 4.2). The polar angle indicates the angle around the interaural axis. Let  $\theta$  be the angle with respect to the Z-axis and let  $\phi$  be the angle in the XY-plane. The angle  $\phi$  is then the polar angle in the double-pole or lateral-polar angle coordinate system.

**POSTERIOR:** Posterior refers to the region in back.

**QUADRATURE LOCALIZATION PLOT:** A quadrature localization plot refers to a set of 4 spherical localization plots showing auditory localization data from 4 points of view: front, back, left and right.

**R:** R refers to the rostral area adjacent to the primary auditory cortex.

**SCD:** The deep layers of the Superior Colliculus refers to a nucleus in the auditory midbrain that has been shown to contain a topographic map of space.

**SPECTRAL CONTRAST AREA:** The spectral contrast area refers to the the region of space which has a maximum normalized energy level for a given frequency band relative to all other locations in space.

**TDNN:** TDNN refers to a time-delay neural network.

**TONOTOPIC:** Tonotopic refers to an ordered arrangement by frequency value.

**VAS:** Virtual auditory space refers to the electronic synthesis of spatial hearing using head-related transfer functions.



# Preface

We cannot be absolutely sure, since one cannot ever explain inductive reasoning – one cannot ever explain how to proceed, when one knows only a little, in order to learn even more (Feynman, 1995).

## 0.1 Auditory Perception

Things change and the way they do so are governed by physical laws. Physical laws, it is thought, do not change. Science is the study of the physical world and its laws. Within the last five years, I have been studying the human perception of auditory space. Why perception? It is clear that in generating perceptions the brain is solving computational problems related to the physical world. However, the study of computation itself is relatively new. In the past, man did not make things to perform complex calculations. Now, everyday, microprocessors are getting clocked faster and faster, but do man-made things really perform complex calculations? Complexity involves patterns: the formation of dunes in the sand, the distribution of petals on a flower, acoustic patterns in fluent speech. You and I perceive these patterns and from them can make complex predictions about the changes in the environment around us. We do not understand how this happens, nor for that matter the mathematical rules that govern pattern formation. With just two ears and some pattern analysis, our auditory system can often determine the direction of a transient sound. However, it is not locating a sound that directly interests us, for we can do that easily enough with four microphones (i.e., six pairs of “ears”) using a method of mathematical triangulation. Rather, it is the auditory system’s ability to detect the spectral shape information related to the acoustic filtering properties of



our external ear *despite* the fact that the sound may be spectrally-scrambled  $\pm 20$  dB in level in 1/3-octave bands and presented concurrently with competing sound sources in a reverberant acoustic environment that teaches us respect for the biology.

We desire so much to make things that perform complex computations that in this last decade, the “Decade of the Brain,” brain scientists have been writing books that tell us “How the Mind Works” and engineers have been claiming they are trying to be inspired by biology. It seems that scientists have been swept away by their ability to explain almost anything at all about the brain. I do not mean this disrespectfully and modern medicine can definitely perturb the homeodynamic regulation of the human brain in helpful ways. But I do not even know what a perception is or what it means to feel and experience it. People say we are “conscious”, but I do not know what that means. It is not clear to me whether “consciousness” is a key ingredient for complex computations. Perhaps it is only required for complex computations about the self, perhaps not. When studying auditory sensory perception, I have mostly aimed my questions at issues that remain close to the sensory input when trying to learn something about the complex signal processing occurring within the human brain.

In addition to complex brain processing, there is another rather obvious aspect to the study of the human auditory perception of space: 3-D audio. We are not yet technically able to record 3-D sound and reproduce it for *any* listener with a reasonable degree of fidelity. Reproducing 3-D audio for only one person, however, is simple because we only have to put a microphone in each of his/her ears and record the sound. If you believe that surround sound or ambisonic sound has accomplished the recording and playback of high-fidelity 3-D audio, our standards probably differ. The study of the human auditory perception of space will provide us with a better understanding of the technical requirements for efficiently reproducing 3-D audio for human listeners.

## 0.2 A Personal Statement

In the course of doing my degree, I have become convinced that I am suffering from a “neglect syndrome.” In doing my research, I have found that no matter how hard I work, I cannot really see past the end of my nose. Most of us go around comfortable that we are intelligent creatures, but put it to the test and try to understand something fundamentally new and we often flounder. To show this, I will summarize my entire dissertation in a single sentence: A spectral contrast model based on overlapping frequency bands of varying bandwidth and perhaps multiple frequency scales can provide a reasonable algorithm for explaining much of the current psychophysical and neurophysiological data related to human auditory localization. If I could have made this observation in the beginning, I could have saved five years’ time. It is curious what sometimes gives people depth of insight. I am reminded of Ramachandran and Blakeslee’s (1998) comment that for most of us it is difficult to come up with several metaphors for “overdoing things,” and that Shakespeare came up with: “To gild refined gold, to paint the lily, to throw perfume on the violet, to smooth the ice, or add another hue to the rainbow . . . is wasteful and ridiculous excess.” I could, of course, add my thesis to that list. I believe that as we peer out at the world we rarely, if ever, see what is actually before us.

Some say that obtaining a Ph.D. is like getting a driver’s license. For me, admittedly an American in my educational upbringing, this makes it all the more remarkable that at the University of Sydney there is no viva voce. In other words, there is no direct examination of the individual behind the work. It is like awarding a driver’s license on the basis of watching a video tape. It is ironic that this is the attitude of an institute of higher learning that teaches *first-hand*, empirical observation is the basis for all scientific understanding and knowledge.

I would also like to say that I have recently come to the opinion that within the last five years I have successfully completed *two* projects, not one, but that this dissertation

is only concerned with only one of them. Nonetheless, in this preface I would like to describe my “other” project because it has played a significant role in my trials and tribulations over the last five years and I also believe it is a significant accomplishment.

What should really be considered as my first project was concerned with the problem of customizing acoustic transfer functions, known as head-related transfer functions (HRTFs), for individual listeners. That is to say, the human external auditory periphery is a directional acoustic filter whose filtering properties vary from one listener to another. It turns out that human auditory localization performance is sensitive to the individual differences in these filtering properties (I later quantified the extent to which this is true). Therefore each listener requires his/her own set of acoustic transfer functions in order to be able to synthesize a realistic virtual auditory space, i.e., high-fidelity 3-D audio over earphones, for that listener. Acoustically recording these acoustic transfer functions is expensive both in time and equipment. Therefore, the goal of my first project was to create a simple method for generating HRTFs that does not require acoustical measurements to be made in the laboratory.

In the midst of my first project, my thesis advisor, Philip Leong, moved to the Chinese University in Hong Kong. For better or worse, however, I was committed to remaining at the University of Sydney and was mostly supervised by Simon Carlile on a second and new project which involved trying to understand, as best as possible, the spectral analysis involved in human auditory localization. Nonetheless, I believe that to a large degree, I have actually completed both projects.

As my first step in the project with Philip, I developed numerical solutions to the acoustic wave equation for circular and elliptical disks and then for a prolate spheroid. It became clear that solving the acoustic wave equation in a realistic manner for the human external auditory periphery would require a sophisticated mathematical software package, such as a Boundary Element Method package, and an imaging technique for recording the shape of the human external auditory periphery. As these resources were

not available, I moved away from a direct simulation of the acoustic wave equation in search of a more practical solution.

As a second step in this line of research, I built a life-like acoustical mannequin of myself (see Figure 1) and recorded the differences in the acoustical transfer functions with and without the torso. With the help of an honour student, a lighter and more durable cast of the head was then created in which the ears could be rotated to change the angle of the external ear with respect to the side of the head. The acoustic transfer functions of the mannequin were then recorded for 7 different angles of the ear with respect to the head. Using a directional averaging technique for implementing principle component analysis, I numerically modeled the functional dependence of the HRTFs on the “ear angle.”



Figure 1: A life-like acoustical mannequin was made.

The reasonable degree of success with the numerical modeling (Carlile, Jin and Harvey, 1998) described above led to the creation of a database of 11 sets of HRTFs for 11 different individuals. Using this database, an HRTF morphing model was created in which 7 parameters (PCA weights) could be tuned to produce a complete set of HRTFs. The 7 parameters were tuned manually in response to how well the listener

could localize a set of test sounds. Tuning these parameters was difficult and resulted in auditory localization performance that was still significantly worse than control performance levels.

The above model was then improved in two ways: (1) a better PCA approximation method was used, and (2) a new database of HRTFs was created using an *identical* recording technique for 36 different human subjects. The new HRTF morphing model provided a generative statistical technique to compress or smooth (in a lossy fashion) the HRTFs for each of the 36 human subjects. An auditory localization experiment with 5 human subjects was then carried out to determine how many PCA weights were required for high-fidelity auditory localization. Following this, a measurement process was developed for physically measuring the Cartesian coordinates of 20 morphological landmarks defining the shape of the listener's external auditory periphery. A bite bar was made and a 3-D stylus pen was set up for recording the coordinates of the morphological landmarks. Multivariable linear regression analysis was then successfully used to develop a functional mapping between the morphology of the external auditory periphery and the HRTFs (see Jin, Leong, Leung, Corderoy and Carlile, 2000). It turns out that approximately 68% of the morphological differences in individual ear shape are significant for high-fidelity VAS. This work has resulted in 2 refereed conference papers (those cited above) and a provisional patent application for the University of Sydney.

### **0.3 Reading the Dissertation**

Chapters 1, 2 and 3 provide a fairly extensive background review related to human auditory localization. These chapters can be read on their own. Chapter 4 describes the experimental methods that have been used. Chapters 5, 6, and 7 describe the three phases of my research and each chapter can be read on its own. Chapter 5 describes a time-delay neural network model of human auditory localization; Chapter 6 describes

a psychoacoustical experiment investigating human sound localization of bandpass filtered noise stimuli; Chapter 7 describes a psychoacoustical experiment that employs the techniques of virtual auditory space to probe the relative contribution of the monaural and interaural spectral cues. Chapter 8 provides a summary of the research and the conclusions that can be derived from it. It can be read both first and last as it provides a quick overview of the focus of the work described in this dissertation.