

# Active Audition for Robots using Parameter-Less Self-Organising Maps

A thesis submitted for the degree of Doctor of Philosophy at  
The University of Queensland in October, 2005.

Erik Johan Berglund

berglund@itee.uq.edu.au

School of Information Technology and Electrical Engineering  
University of Queensland.



THE UNIVERSITY  
OF QUEENSLAND  
AUSTRALIA

July 2, 2006

## *Acknowledgements*

The candidate would like to thank the following:

Mamma and Pappa.

Dr. Gordon Wyeth for help, support and keen insights.

My associate supervisor A. Prof. Joaquin Sitte for guidance, ideas, motivation and help in providing equipment.

The Government of Norway for a stipend of 16,000 AUD received in 2002.

The Queensland University of Technology for a top-up scholarship of 2,700 AUD received in 2003.

The University of Queensland for a 9,418 AUD research scholarship received in 2005.

Dr. Frederic Maire for help with my many mathematical troubles.

Dr. Steffen Log for believing in me at a time of my academic career when I needed it.

Dr. Kazuhiro Nakadai for input and feedback.

Dr. Thomas Voegtlin for background information and source code for the Recursive SOM.

Mr. Fabrice Miras for kindly donating the Wacom Digitizer II graphics tablet used for many of the illustrations.

A. Prof. Janet Wiles for help in editing my thesis.

My brother and sisters.

All my colleagues and friends.

Various anonymous reviewers for helpful input.

*Publications by the Candidate Relevant to the Thesis:*

- E. Berglund and J. Sitte: “The Parameter-Less SOM algorithm”, Proceedings of the Eighth Australian and New Zealand Intelligent Information Systems Conference, 2003, Sydney, Australia. Pages 159-164.
- E. Berglund and J. Sitte: “The Parameter-Less Self-Organising Map algorithm”, IEEE Transactions on Neural Networks, to appear in March, 2006.
- E. Berglund and J. Sitte: “Sound source localisation through active audition”, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton Canada, 2-6 August 2005. Pages 509-514.

Awaiting editorial decision:

- E. Berglund, J. Sitte and G. Wyeth: “Active audition and directional filtering using the Parameter-Less Self-Organising Map”, submitted to Autonomous Robots, Springer.

*Additional Publications by the Candidate Relevant to the Thesis but not Forming Part of it:*

- A. Campbell, E. Berglund and A. Streit: “Graphics Hardware Implementation of the Parameter-Less Self-Organising Map”, Proceedings of the Sixth International Conference on Intelligent Data Engineering and Automated Learning, Brisbane Australia, 6-8 July, 2005. Pages 343-350. Lecture Notes in Computer Science 3578, Springer.

*Candidate's Statement of Originality*

I, Erik Johan Berglund, declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text, and that the material has not been submitted, either in whole or in part, for a degree at this or any other university.

Candidate's signature:

Date:

## ABSTRACT

How can a robot become aware of its surroundings? How does it create its own subjective, inner representation of the real world, so that relationships in the one are reflected in the other? It is well known that structures analogous to Self-Organising Maps (SOM) are involved with this task in animals, and this thesis undertakes to explore if and how a similar approach can be successfully applied in robotics. In order to study the environment-to-abstraction mapping with a minimum of guidance from directed learning and built-in design assumptions, this thesis examines the active audition task in which a system must determine the direction of a sound source and orient towards it, both in horizontal and vertical direction.

Previous explanations of directional hearing in animals, and the implementation of directional hearing algorithms in robots have tended to focus on the two best known directional clues; the intensity and time differences.

This thesis hypothesises that it is advantageous to use a synergy of a wider range of metrics, namely the phase and relative intensity difference. A solution to the active audition problem is proposed based on the Parameter-Less Self-Organising Map (PLSOM), a new algorithm also introduced in this thesis. The PLSOM is used to extract patterns from a high-dimensional input space to a low-dimensional output space. In this application the output space is mapped to the correct motor command for turning towards the source and focusing attention on the selected source by filtering unwanted noise. The dimension-reducing capability of the PLSOM enables the use of more than just two directional clues for computation of the direction.

This thesis presents the new PLSOM algorithm for SOM training and quantifies its performance relative to the ordinary SOM algorithm. The mathematical correctness of the PLSOM is demonstrated and the properties

and some applications of this new algorithm are examined, notably in automatically modelling a robot's surroundings in a functional form: Inverse Kinematics (IK). The IK problem is related in principle to the active audition problem - functional rather than abstract representation of reality - but raises some new questions of how to use this internal representation in planning and execution of movements. The PLSOM is also applied to classification of high-dimensional data and model-free chaotic time series prediction.

A variant of Reinforcement Learning based on Q-Learning is devised and tested. This variant solves some problems related to stochastic reward functions. A mathematical proof of correct state-action pairing is devised.

## CONTENTS

1. <i>Introduction</i> . . . . .	1
1.1 Audition and robots . . . . .	1
1.2 Thesis overview . . . . .	4
1.3 Thesis contribution . . . . .	4
2. <i>Physical considerations regarding sound and binaural audition</i> . . . . .	6
2.1 Distance considerations . . . . .	6
2.2 Phase . . . . .	7
2.3 Frequency, bandwidth and subband . . . . .	11
2.4 Reflection or echo . . . . .	12
2.5 Refraction . . . . .	12
2.6 Speed of sound . . . . .	13
2.7 Sound recording . . . . .	15
2.8 Ultrasound and infrasound . . . . .	15
3. <i>Binaural audition in biological systems</i> . . . . .	17
3.1 The Barn Owl . . . . .	17
3.2 Hearing in bats . . . . .	18
3.3 Ears of humans and other mammals . . . . .	20
3.4 Sound processing in the mammalian brain . . . . .	22
3.5 Human hearing . . . . .	26
3.5.1 Precedence effect . . . . .	27
3.5.2 Acuity . . . . .	27
3.6 Summary . . . . .	30

---

4. <i>Earlier works</i> . . . . .	32
4.1 Head-Related Transfer Function . . . . .	34
4.2 Barn owl simulation . . . . .	34
4.3 SIG Humanoid . . . . .	35
4.4 Cocktail party effect . . . . .	37
4.5 Elevation detection . . . . .	39
4.6 Summary . . . . .	39
5. <i>Overview of proposed system</i> . . . . .	42
5.1 Feature selection rationale . . . . .	43
5.2 Data association strategy and possible obstacles to implemen- tation . . . . .	43
5.3 How this work differs from previous works . . . . .	45
6. <i>The Parameter-Less Self-Organising Map</i> . . . . .	46
6.1 Self-Organizing Maps . . . . .	46
6.1.1 Overview . . . . .	48
6.1.2 Details . . . . .	49
6.2 Map calibration and labelling . . . . .	51
6.3 Parameter selection schemes for SOM . . . . .	51
6.3.1 Previous works . . . . .	52
6.4 Parameter-Less Self-Organising Map . . . . .	53
6.4.1 Algorithm . . . . .	54
6.5 Experiments with the PLSOM . . . . .	56
6.5.1 Comparison of SOM variants to the PLSOM . . . . .	56
6.5.2 PLSOM analysis . . . . .	66
6.5.3 Application to Inverse Kinematics . . . . .	71
6.5.4 Application to classification of the ISOLET data set . . . . .	75
6.5.5 Recursive PLSOM . . . . .	76
6.6 Discussion . . . . .	83
6.7 Conclusion . . . . .	85
7. <i>Active Audition With PLSOM</i> . . . . .	86
7.1 Brief system overview . . . . .	86



---

7.2	Preprocessing . . . . .	86
7.3	PLSOM . . . . .	89
7.4	Reinforcement Learning . . . . .	90
7.4.1	Q-Learning . . . . .	91
7.4.2	Reinforcement Learning Algorithm Variation . . . . .	92
7.5	Directional filtering . . . . .	95
7.5.1	Description of directional filtering system . . . . .	96
7.6	Conclusion . . . . .	97
8.	<i>Experiments and Results</i> . . . . .	98
8.1	Experimental setup . . . . .	101
8.1.1	List of data sets used throughout this thesis . . . . .	103
8.2	Horizontal direction detection . . . . .	104
8.2.1	Simple reference system . . . . .	105
8.2.2	SIG system . . . . .	105
8.2.3	PLSOM system . . . . .	106
8.2.4	Comparison of SIG and PLSOM systems . . . . .	106
8.2.5	Further tests with the PLSOM system, and discussion .	106
8.3	Horizontal source localisation . . . . .	112
8.4	The cone of confusion . . . . .	114
8.5	Front-back resolution . . . . .	115
8.6	Distance resolution . . . . .	117
8.7	Limits to response time . . . . .	121
8.8	Effect of source elevation on accuracy . . . . .	121
8.9	Taking advantage of the cone of confusion for elevation estimation . . . . .	122
8.10	Robustness . . . . .	128
8.11	Directional filtering . . . . .	130
8.12	Conclusion . . . . .	135
9.	<i>Discussion</i> . . . . .	137
9.1	Implementation design considerations . . . . .	139
9.2	System software architecture . . . . .	140

9.3 Responsiveness . . . . .	143
9.4 Implicit versus explicit data representation . . . . .	144
9.5 Summary . . . . .	146
10. <i>Conclusion</i> . . . . .	147
10.1 Future directions . . . . .	148
 <i>Appendix</i>	 150
<i>A. Proof of guaranteed ordering of a PLSOM with 3 nodes and 1-dimensional input and output.</i> . . . . .	151
<i>B. Proof of convergence of learning neighbourhood-algorithm.</i> . . . . .	159
<i>C. Overview of accompanying compact disc</i> . . . . .	161

## LIST OF FIGURES

2.1	Schematic head seen from above, with incidence angle ( $\theta$ ) and sound paths from the sound source to the two microphones $M_l$ and $M_r$ drawn in. The path the sound would follow if it were not for the damping of the robot head is indicated by the dotted line from the sound source to the left microphone $M_l$ .	9
2.2	Effect of frequency on the phase difference. 12 cm between the microphones and actual angle = $\frac{3\pi}{4}$ . Observe that while the phase difference increases with increased frequency, so does the time of non-overlapping phase.	11
2.3	Angle calculated according to Equation (2.12) vs. time for a 440 Hz sound with 12 cm between the microphones. Actual angle: $\frac{3\pi}{4}$ . Note that the estimation gives a complex value; the plot represents the real part.	12
2.4	Speed of sound in air vs. air pressure at 20°C and 50% relative humidity.	13
2.5	Speed of sound in air vs. relative humidity at 20°C and 101.325 kPa.	14
2.6	Speed of sound in air vs. temperature at 101.325 kPa and 50% relative humidity.	14
3.1	The ear canal inlets on the face of a Barn Owl (grey), showing the asymmetry. (Original drawing.)	18
3.2	Frontal view of an insectivore bat (Big eared townsend bat, <i>Corynorhinus townsendii</i> ). Note the large ear size, especially compared with the small eyes. (Original drawing.)	19

3.3	The major sound processing centres in a bat brain. Observe the widening of the frequency-sensitive area around 61 kHz, the frequency this bat species ( <i>Pteronotus parnellii</i> ) uses for echolocation. Adapted from [1]. . . . .	20
3.4	The human pinnae, right side. Showing helix (A), scaphoid fossa (B), triangular fossa (C), tragus (D), concha (E), anti-tragus (F) and the opening of the ear canal (G). (Original drawing.) . . . . .	21
3.5	A stylised human ear. Shown are the eardrum (A), the ossicles (hammer (B), anvil (C) and stirrup (D)), the labyrinth (E) and the cochlea (F). (Original drawing.) . . . . .	22
3.6	Primary auditory cortex (1) of the human brain. The enlarged area shows the frequency-dependent structure. (Original drawing.) . . . . .	23
3.7	Schematic view of the main auditory pathways in human brain. Shown are the Cochlear Nucleus (CN), the Superior Olivary Complex (SOC), the Nuclei of the Lateral Lemniscus (NLL), the Inferior Colliculus (IC), the Superior Colliculus (SC), the Medial Geniculate Body (MGB) and the auditory cortex. (Original drawing.) . . . . .	24
3.8	Schematic view of neurons responsible for detecting ITD. The excitation of the neurons is inversely proportional to the time difference between the delayed signals from the left and right ear. . . . .	25
3.9	Average error of localisation vs. sound frequency (from [2]). . . . .	28
3.10	Minimum audible angle vs. sound frequency for various initial sound source positions (from [3]). . . . .	28
3.11	Sound source direction estimation error as a result of a phase measurement error of 0.1 radians vs. angle between the sagittal plane and the incidence vector, according to Equation (2.12). . . . .	29
4.1	The SIG Humanoid auditory direction detection system, from [4]. . . . .	36

---

4.2	The SIG Humanoid auditory direction pass filter system, from [5]. . . . .	36
5.1	A simplified diagram of the proposed system. . . . .	42
6.1	SOM training sequence. Each line intersection represents the position of a weight vector. . . . .	47
6.2	Plot showing the effect of different $\epsilon$ values on the neighbourhood function. . . . .	55
6.3	Three different 20-by-20 node SOM algorithms mapping the same uniformly distributed input space. . . . .	57
6.4	Graph of the decrease of uncovered space as training progresses for the PLSOM, the SOM and the Matlab SOM implementation. Note the quick expansion of the PLSOM and that it consistently covers a larger area than the SOM variants. . . . .	58
6.5	Graph of the average skew for the PLSOM, the SOM and the Matlab SOM implementation. For the first 24000 iterations the PLSOM is more ordered, before the SOM variants narrowly overtake it. . . . .	59
6.6	Graph of the average skew for the PLSOM, detail. . . . .	59
6.7	Graph of the absolute mean deviation of cell size for the PLSOM, the SOM and the Matlab SOM. The PLSOM is more regular up until around iteration 10000. . . . .	60
6.8	Graph of the absolute mean deviation of cell size for the PLSOM, the SOM and the Matlab SOM, excluding the edge cells. Compare Figure 6.7. The PLSOM outperforms the Matlab SOM in both adaptation time and accuracy, and the SOM needs until iteration 30000 to reach the same level of ordering. . . . .	60
6.9	SOM first trained with inputs ranging from 0 to 0.5 for 50000 iterations shown after 20000 further training iterations with inputs ranging from 0 to 1.0. . . . .	61

---

6.10	PLSOM first trained with inputs ranging from 0 to 0.5 for 50000 iterations shown after 20000 further training iterations with inputs ranging from 0 to 1.0. Note the difference between this and Figure 6.9. . . . .	62
6.11	PLSOM first trained with inputs ranging from 0 to 1 for 50000 iterations shown after 20000 further training iterations with inputs ranging from 0 to 0.5. Note that while the weights have a higher density in the new input space, the same area as before is still covered, which means none of the old input space has been left uncovered. . . . .	63
6.12	Response of the SOM and PLSOM to 150000 normal distributed inputs with a mean of 0 and a standard deviation of 0.3. The inputs include 1068 mild outliers and no extreme outliers. The top left-hand corner of the frame is at [-1,-1] and the bottom right-hand corner is at [1,1]. Inputs may fall outside the frame. . . . .	64
6.13	The same PLSOM as in Figure 6.12(b) before and after a mild outlier at [0.809488, 0]. . . . .	64
6.14	The same PLSOM as in Figure 6.12(b) before and after an extreme outlier at [1.416529, 0]. . . . .	65
6.15	PLSOM after 5000000 normal distributed inputs including 11 extreme outliers. . . . .	65
6.16	Maps after 100000 iterations of normally distributed input with mean 0.5, standard deviation 0.2, clipped to the [0, 1] interval. See also Figure 6.23 . . . . .	66
6.17	Update size x likelihood for a corner node $v$ of a 20x20 node ordinary SOM algorithm. The position of $v$ in the input space is marked by a vertical white line. The position of $v$ in the map is (1,1). . . . .	68

---

6.18	Update size $x$ likelihood for a corner node $v$ of a 20x20 node Matlab SOM algorithm. The position of $v$ in the input space is marked by a vertical white line. The position of $v$ in the map is (1,1). Note that the shape of this update function is very similar to the one pictured in Figure 6.17, indicating that neighbourhood function shape has limited influence on the update function. . . . .	68
6.19	Update size $x$ likelihood for a corner node $v$ of a 20x20 node PLSOM. The position of $v$ in the input space is marked by a vertical white line. The position of $v$ in the map is (1,1). . . .	69
6.20	The expected displacement vectors for the edge nodes along one edge of an ordinary SOM. Note that the vectors are changing direction abruptly from node to node, causing the warping.	69
6.21	The expected displacement vectors for the edge nodes along one edge of a Matlab SOM. The effect that was noted in Figure 6.21 is even more pronounced here. . . . .	70
6.22	The expected displacement vectors for the edge nodes along one edge of a PLSOM. . . . .	70
6.23	Weight density vs. distance from centre for the SOM and the PLSOM. The 2-dimensional input was normal distributed with a 0 mean and 0.2 standard deviation. Observe that although the PLSOM has less correlation between input density and weight density, it has far less variance and covers a larger area. See also Figures 6.16(a) and 6.16(b) . . . . .	71
6.24	Convergence of some IK algorithms. Please note that the Jacobian methods can be improved by choosing different scaling variables based on the distance from the target. Here a constant scaling variable which is large without causing too much oscillation has been used. . . . .	73
6.25	Error after 500 iterations of some IK algorithms. Please note that the transpose Jacobian does not converge to one final value, but oscillates close to it. . . . .	74

---

6.26	Average execution speed of one iteration of some IK algorithms. Note that adding nodes does not induce delays in the PLSOM method. . . . .	75
6.27	Part of the Mackey-Glass chaotic time series used in this experiment. . . . .	79
6.28	Comparing the Recursive SOM and the Recursive PLSOM. . .	79
6.29	The prediction error $E$ vs. the function value $x(\tau)$ . . . . .	81
6.30	The 3-beacon navigation mapping by the SOM and the PLSOM in the unit square. Both maps have a neighbourhood size of 17. Beacons were positioned at (0.3,-0.3),(1.3,0.5) and (-0.5,0.8). The origin is in the upper left-hand corner of the square that circumscribes the input space. . . . .	84
7.1	Active audition system layout . . . . .	87
7.2	Directional filtering system layout. The subband selection algorithm picks the subbands that match closely to the weight vectors of nodes near the middle of the PLSOM. All other subbands are muted. Compare Figure 7.1 . . . . .	96
8.1	The Sony Aibo ERS-210. Note that the artificial auricles are mounted separately from the microphones. . . . .	99
8.2	Directional response of the Aibo microphones. Test conducted with speaker playing 79 dB white noise at 1 metre in a room with a background noise of 50 dB and $T60 = 0.4$ seconds. Note the increased sensitivity around $\pm 60-70^\circ$ . . . . .	99
8.3	Aibo sensitivity response at 100 cm to 800 Hz sine wave. Note that the sensitivity is switched along the central axis compared with that for white noise, see Figure 8.2. This is caused by the sound source interfering with itself in the near field. . . . .	100
8.4	Plot showing the input position each node is most sensitive to on average, relative to the head. The semi-circle represents 300 cm, the larger circle represents 50 cm. . . . .	102



---

8.5	A typical recording setup. The tape measure is extended to 1 metre for scale. This particular picture shows the recording of data set H, see Section 8.1.1. . . . .	102
8.6	The grey area is one standard deviation, scaled to match the axis units. For an explanation of the different units along the vertical axes, please refer to Section 9.4 . . . . .	107
8.7	The estimated angle of a labelled PLSOM vs. the actual angle. An idealised straight line has been inserted for reference. The grey area represents one standard deviation. . . . .	108
8.8	The grey area is one standard deviation, scaled to match the axis units. For an explanation of the different units along the vertical axes, please refer to Section 9.4 . . . . .	110
8.9	The direction detection results for the three different sounds presented above overlaid in one graph. . . . .	111
8.10	Average winning node vs. actual angle using the PLSOM method with 36x10 nodes. The grey area is one standard deviation. . . . .	111
8.11	The absolute difference between the robot head direction and the direction from the robot to the sound source using Q-Learning and Reinforcement Learning variant algorithms. . . .	113
8.12	The cone of confusion can be seen as a cone with its vertex between the microphones obtained by rotating the vector to the sound source around the tilt axis. The actual position of the sound source may lie anywhere on the cone. (Original drawing.) . . . . .	114
8.13	Aibo dealing with front/rear ambiguity. Seen from above. Adapted from [6]. . . . .	116

- 
- 8.14 Average winning node vs. source distance with one standard deviation for various angle ranges. The stronger the correlation between distance and winning node is, the better. Note that this map has only been trained with inputs from a distance of 50-300 cm, which partly explains the poor correspondence over 300 cm. Note that the nodes are not explicitly labelled with the distance they represent, see Section 9.4 for discussion. . . . . 118
- 8.15 Estimated source-robot distance vs. actual source-robot distance with one standard deviation. These points represent cases where the source is located 60-70° to the left or right of the median plane of the robot head. An idealised line has been inserted for reference. . . . . 119
- 8.16 The effect of a relative measurement error of 1% on the calculated intensity difference vs. distance in arbitrary units. . . . 120
- 8.17 Average winning node vs. source angle for different speaker and head elevations. Note that when the head is also elevated, the deviation from the non-elevated case is small, indicating that internal harmonics in the robot and echoes from the ground play a limited role. For an explanation of the units on the vertical axis, see Section 9.4. . . . . 122
- 8.18  $\mu$  is the elevation of the sound source above the horizontal plane expressed as an angle around the tilt axis. . . . . 123
- 8.19 The geometric relations of  $\nu$ ,  $\nu'$ ,  $\psi$  and  $\psi'$ . . . . . 124
- 8.20 Elevation each node is sensitive to vs. node index with one standard deviation. The map is 5x2 nodes and the neighbourhood size is 4.27. . . . . 125
- 8.21 Estimated elevation vs. actual elevation with one standard deviation. An idealised line has been added for reference. The map is 5x2 nodes and the neighbourhood size is 4.27. . . . . 126
- 8.22 Elevation each node is sensitive to vs. node index with one standard deviation. The map is 20x8 nodes and the neighbourhood size is 9. . . . . 127

---

8.23	Estimated elevation vs. actual elevation with one standard deviation. An idealised line has been added for reference. The map is 20x8 nodes and the neighbourhood size is 9. . . . .	127
8.24	The effect of reversing the output of the PLSOM on the RL weights. Each square represents the weight of a state-action pair in the RL grid, where black equals 0 and white equals 1. States are along the horizontal axis and actions along the vertical axis. Note the strong but opposite correspondence both before and after the switch. . . . .	129
8.25	The directional error of the Q-Learning and Reinforcement Learning variant algorithms during direction reversal of the PLSOM. The reversal of the PLSOM occurs at iteration 5000.	130
8.26	Signal spectrogram. A female voice utters the phrase “she had your dark suit in greasy wash-water all year” from a speaker 120 cm straight in front of the robot. The phrase is repeated continuously. . . . .	132
8.27	Noise spectrogram. White noise from the left and a male voice that utters the phrase “the bungalow was pleasantly situated near the shore” from the right. Observe that the white noise covers the entire spectrum, as is the nature of white noise. . .	132
8.28	Signal and noise spectrogram. White noise, male speaker and female speaker all recorded at once. . . . .	133
8.29	Recovered signal spectrogram. . . . .	133
9.1	Execution time for 100, 1000, 10000 and 100000 iterations of pseudo-random input for the default Matlab SOM and a Java implementation of the SOM. Note that while Matlab pre-computes the input data outside of the timing loop, Java computes the input on the fly, resulting in a number of calls to the random number generator for each iteration. This explains why the Java implementation seems to slow down relative to the Matlab implementation as input dimensions increase. . . .	141

---

9.2	Time required for 100 000 iterations of the FFT transform vs. input size on different language platforms. . . . .	141
9.3	UML class diagram of the main components of the audio processing pipeline. Shown are the <code>PipelineElement</code> and <code>SoundObject</code> interfaces, and the convenient basic implementation of the <code>PipelineElement</code> interface; <code>PipelineElementImpl</code> .	143
9.4	The data is transferred between subsystems in explicit, human-readable form. . . . .	145
9.5	The data is transferred between subsystems in implicit, non-human-readable form. No conversion is necessary. . . . .	145
A.1	The unordered subspace $U$ . All other unordered states are mirrors or inversions of states in this subspace. . . . .	155
A.2	All unordered states in the volume of all possible states. . . . .	155
A.3	All ordered states in the volume of all possible states. . . . .	155
A.4	Evolution of the weight positions of a 64-node 2D PLSOM initialised to a difficult position. Neighbourhood size is 11, minimum neighbourhood size is 0. To simulate what will happen if this configuration appears late in training, an $r$ value of 0.65 is enforced. . . . .	158

## LIST OF TABLES

6.1	Recursive PLSOM prediction compared to random guess. . . .	81
6.2	The mean percentage of winning node movements in the pre- dominant direction and standard deviation for some input se- ries to the Recursive PLSOM. . . . .	83

## LIST OF ABBREVIATIONS AND GLOSSARY

*audition* - the act of hearing. Audition is to hearing as vision is to seeing.

*auricle* - see *pinna*.

*dB* - deci Bel, a unit of relative intensity. In this thesis dB always refers to dB Sound Pressure Level (SPL), C weighting. This is defined by the formula  $10 \log_{10}(\frac{I}{I_0})$  where  $I$  is the intensity and  $I_0$  is the threshold of hearing, defined as  $2.0 \times 10^{-5}$  Pa.

*FFT* - Fast Fourier Transform, transforms a signal from the time domain to the frequency domain, allowing analysis of the frequency components of the signal.

*foveation* - the act of moving one's body in such a way as to focus the image of objects of interest on the *fovea*, the area of the retina with the highest resolution. Auditory foveation is the act of bringing the front of the head towards a sound source using sound information.

*HRTF* - Head Related Transfer Function, describes how a sound with a given direction of incidence and frequency will be distorted by the head. This is usually computed from data gathered from a large number of measurements in an anechoic chamber.

*IID* - Interaural Intensity Difference is the difference of sound intensity from one ear/microphone to the other. This is caused by two things; the difference in distance to the sound source and the damping of any material in between the ears/microphones. Compare *ILD*.

*ILD* - Interaural Level Difference is the difference of sound level (a logarithmic scale relative to a reference level) from one ear/microphone

to the other. This is caused by two things; the difference in distance to the sound source and the damping of any material in between the ears/microphones. Compare *IID*.

*IPD* - Interaural Phase Difference, the difference in phase angle between the two ears or microphones of an audition system. Similar to the ITD, but limited to one cycle, for example  $[-\pi, \pi]$  (in radians). The relation between the IPD and the incidence angle is frequency dependent.

*ITD* - Interaural Time Difference, the time that elapses from a sound event is detected at one ear/microphone till it is detected at the other ear/microphone. Related to the IPD, but is not limited to one cycle and is not frequency-dependent.

*pinna* - the visible outer part of the ear, see Figure 3.4.

*RL* - Reinforcement Learning, see Section 7.4.

*SOM* - Self-Organising Map, see Section 6.1.

*subband* - a range of frequencies that are part of the total frequency range being studied. The *FFT* extracts information about phase and amplitude for each of a given number of *subbands*.

*PLSOM* - Parameter-Less Self-Organising Map, see Chapter 6.

*T60* - reverberation time, a property of the environment that indicates the time it takes for the reverberations of a sound signal to decrease by 60 dB.