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

LIST OF FIGURES

LIST OF TABLES

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 2.0×10^{-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.