

Copyright is owned by the author of the thesis. Permission is only given for a copy to be downloaded by an individual for the purpose of research and private study. Any part of this thesis may not be reproduced elsewhere without the permission of the author.

**Multi-microphone Speech  
Enhancement Technique using a  
Novel Neural Network Beamformer**

A thesis presented in partial fulfilment of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY  
in  
Engineering

at Massey University, Albany,  
New Zealand.

Vaitheki Yoganathan

2014

# Abstract

This thesis presents a novel speech enhancement algorithm to reduce the background noise from the acquired speech signal. It introduces an innovative idea for the speech beamformer using an input delay neural network based adaptive filter for noise reduction.

Speech communication is considered as the most popular and natural way for humans to communicate with computers. In the past few decades, there has been an increased demand for speech-based applications; examples include personal dictation devices, hands-free telephony, voice recognition for robotics, speech-controlled equipment, automated phone systems, etc. However, these applications require a high signal-to-noise ratio to function effectively. The background noise sources such as factory machine noises, television, radio, computer or another competing speaker, often degrade the performance of the acquired signals. The problem of removing these unwanted signals from the acquired speech signal has been investigated by various authors. However, there is still room for improvement to the existing methods.

A multi-microphone neural network based switched Griffiths-Jim beamformer structure was implemented using the Labview software. The conventional noise reduction section of the Griffiths and Jim beamformer structure was improved with a non-linear neural network approach. A partially connected three-layer neural network structure was implemented for rapid real-time processing. The error back-propagation algorithm was used here to train the neural network structure. Although it is a slow gradient learning algorithm, it can be easily replaced with other algorithms such as the fast back-propagation algorithm.

The proposed algorithms show promising noise reduction improvement over the previous adaptive algorithms like the normalised least mean squares adaptive filter. However, the performance of the neural network depends on its chosen parameters such as learning rate, amount of training given, and the size of the neural network structure. Tests with a speech-controlled system demonstrate that the neural network based beamformer significantly improves the recognition rate of the system.

# **Acknowledgements**

First of all, I am truly grateful to my former supervisor Associate Professor Tom Moir for his guidance throughout my studies. I have learned a lot from him, and this research wouldn't have been possible without his knowledge and advice. I would like to thank my current supervisors, Dr. Fakhrul Alam and Dr. M. A. Rashid, for their guidance in writing my thesis and their support during my time at Massey University.

This research was supported by the Technology for Industry Fellowships New Zealand scholarship for the first three years of my studies. This scholarship was awarded to work with the Vibration Consultants Company to develop a noise canceller for the speech-controlled instrument. Also, I would like to thank the company owners, Dennis Page and Elisabeth Page, for allowing me to work with them as part of this scholarship programme. I specially want to thank the company mentor, Darryl Ovens, for his guidance and support during this time.

I would also like to thank my family and friends for being there for me and encouraging me throughout my studies. I am also extremely grateful to my fellow colleagues at the university for their on-going support.

# Table of Contents

ABSTRACT.....	II
ACKNOWLEDGEMENTS.....	III
TABLE OF CONTENTS .....	IV
LIST OF PUBLICATIONS .....	VI
LIST OF FIGURES .....	VII
LIST OF TABLES.....	X
LIST OF ABBREVIATIONS.....	XI
LIST OF SYMBOLS.....	XII
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 Introduction.....	2
1.2 Statement of the Problem.....	5
1.3 Research Objectives.....	6
1.4 Contribution of the Thesis to Knowledge.....	6
1.5 Outline of the Thesis.....	7
<b>CHAPTER 2 LITERATURE REVIEW .....</b>	<b>9</b>
2.1 Survey of Speech Enhancement Area.....	10
2.1.1 Beamforming Techniques .....	13
2.2 Adaptive Filters.....	18
2.2.1 Adaptive Algorithms.....	20
2.3 Neural Networks .....	24
2.3.1 Speech Enhancement Using Neural Networks.....	25
2.4 Voice Activity Detectors.....	26
2.5 Speech Recognition in Noisy Environments .....	31
<b>CHAPTER 3 DESIGN AND DEVELOP A SPEECH ENHANCEMENT ALGORITHM.....</b>	<b>32</b>
3.1 Problem Formulation .....	33
3.2 Proposed Speech Beamformer Algorithm .....	34
3.2.1 Beam-steering Filter.....	38
3.2.1 A New Neural Network Based Noise Reduction Filter.....	41
3.2.2 Voice Activity Detector.....	52
3.3 Summary.....	55
<b>CHAPTER 4 EXPERIMENTAL RESULTS FROM THE SPEECH ENHANCEMENT ALGORITHM.....</b>	<b>56</b>
4.1 Experimental Results from the Variance Based VAD Algorithm .....	57
4.1.1 Experiment 1 .....	58
4.1.2 Experiment 2 .....	59
4.1.3 Experiment 3 .....	60
4.1.4 Experiment 4 .....	61
4.2 Testing the Neural Network based Adaptive Filter .....	63
4.2.1 Network Size .....	64
4.2.2 Effect of Learning Rate .....	64
4.2.3 Selection of Initial Weights.....	66
4.3 Evaluation of Various Adaptive Filters .....	67
4.3.1 Experiment 1 .....	67
4.3.2 Experiment 2 .....	70

4.4	Real-Time Experiments of the SGJBF .....	74
4.4.1	Experiments with Dual-Microphone SGJBF.....	75
4.4.2	Experiment Using More Than Two Microphones.....	79
<b>CHAPTER 5 EVALUATION OF A SPEECH-CONTROLLED INSTRUMENT WITH THE NEW SPEECH ENHANCEMENT ALGORITHM</b>		
<b>87</b>		
5.1	Speech-Controlled Vibration Monitoring System .....	88
5.2	Proposed Design of the System .....	89
5.2.1	Hardware Configuration.....	90
5.2.2	Software Implementation .....	91
5.3	Evaluation of this System in a Noisy Environment.....	96
<b>CHAPTER 6 CONCLUSIONS AND FUTURE WORK.....</b>		<b>101</b>
6.1	Conclusions.....	102
6.2	Suggestions for Future Work.....	104
<b>REFERENCES.....</b>		<b>106</b>
<b>APPENDIX A .....</b>		<b>117</b>
<b>PUBLISHED PAPERS .....</b>		<b>117</b>
<b>A - I.....</b>		<b>118</b>
<b>APA BASED SGJBF PAPER PRESENTED AT THE M2VIP08 CONFERENCE.....</b>		<b>118</b>
<b>A - II.....</b>		<b>125</b>
<b>PAPER PRESENTED AT THE ISSPA 2010 CONFERENCE .....</b>		<b>125</b>
<b>A - III.....</b>		<b>130</b>
<b>PAPER PRESENTED AT THE ICSP 2010 CONFERENCE .....</b>		<b>130</b>
<b>A - IV .....</b>		<b>135</b>
<b>PAPER PRESENTED AT THE WCSP 2010 CONFERENCE .....</b>		<b>135</b>
<b>APPENDIX B .....</b>		<b>141</b>
<b>LABVIEW IMPLEMENTATION OF OTHER ADAPTIVE FILTER ALGORITHMS.....</b>		<b>141</b>

# List of Publications

Published conference papers:

1. Yoganathan, V., & Moir, T. J. (2010a, 24-28 Oct). Multi-microphone adaptive neural switched Griffiths-Jim beamformer for noise reduction. Paper presented at the 10th IEEE International Conference on Signal Processing (ICSP), Beijing, China.
2. Yoganathan, V., & Moir, T. J. (2010b, 21-23 Oct). Speech enhancement using microphone array neural switched Griffiths-Jim beamformer. Paper presented at the International Conference on Wireless Communications and Signal Processing (WCSP), Suzhou, China.
3. Yoganathan, V., & Moir, T. J. (2010c, 10-13 May). Speech enhancement using a nonlinear neural switched Griffiths-Jim beamformer. Paper presented at the 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA), Kuala Lumpur, Malaysia.
4. Yoganathan, V., & Moir, T. J. (2008, 2-4 Dec). Switched Griffiths-Jim beamformer using the affine projection algorithm. Paper presented at the 15th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Auckland, New Zealand.

# List of Figures

Figure 2.1.	Delay-and-Sum Beamformer .....	14
Figure 2.2.	Frost Beamformer .....	14
Figure 2.3.	Adaptive Noise Canceller Structure.....	15
Figure 2.4.	Two-channel Griffiths-Jim Beamformer .....	16
Figure 2.5.	Switched Griffiths-Jim Beamformer.....	18
Figure 2.6.	Adaptive Filter in Adaptive Noise Cancelling Configuration .....	19
Figure 2.7.	McCulloch-Pitts Model.....	24
Figure 3.1.	Acoustical Model of the Microphone Signals .....	33
Figure 3.2.	Proposed Non-linear Neural SGJBF Structure .....	35
Figure 3.3	Normalised Least Mean Square Algorithm Structure.....	39
Figure 3.4.	Implementation of the NLMS Adaptive Filter in the Labview Software .....	41
Figure 3.5.	The Direction of the Data Flow in the NN Structure (a) Showing the FF Network and (b) Showing the FB Network.....	43
Figure 3.6.	Overall Structure of the MLP with IDNN .....	44
Figure 3.7.	Fully and Partially Connected Network Structures .....	45
Figure 3.8.	Activation Function (a) Hyperbolic Tangent Graph, (b) Its Derivative .. ..	46
Figure 3.9.	Non-linear Partially Connected Three-layer FF NN ANC Structure...47	
Figure 3.10.	Single Neuron Model Structure .....	48
Figure 3.11	Applying the Learning Algorithm in the NN Adaptive Filter Structure.. ..	49
Figure 3.12.	IDNN Implementation in the Labview Software.....	51
Figure 3.13.	Automatic Variance Estimator.....	53
Figure 3.14.	Overall Process of the Voice Activity Detector Algorithm .....	54
Figure 3.15.	VAD Based on Automatic Variance Estimation .....	55
Figure 4.1.	Experiment 1 Output Results from the VAD Function .....	58
Figure 4.2.	Experiment 2 Output Results from the VAD Function .....	59
Figure 4.3.	Experiment 3 Results from the VAD Function (Threshold Value=0.2).. ..	60



Figure 4.4.	Experiment 3 Results from the VAD Function (Threshold Value=0.3).. .....	60
Figure 4.5.	Experiment 4 Output Results from the VAD Function .....	61
Figure 4.6.	Experiment 4 Improved Results (Threshold Gain of 0.0003).....	62
Figure 4.7.	Experiment 4 Improved Results (Threshold Gain of 9e-5) .....	62
Figure 4.8.	Using the Moving Average Algorithm to Detect the Speech for the Recording Done in Experiment 3 .....	63
Figure 4.9.	Simulation Setup to Test the NN Based Adaptive Filter .....	64
Figure 4.10.	Comparison of MSE Results from the FC and NFC - IDNN Adaptive Filter with Different LR .....	65
Figure 4.11.	MSE Comparisons of Different Initial Weight Values (I-10, N-5, LR- 0.5) .....	66
Figure 4.12.	Experimental Setup Used to Approximate a Certain Linear Transfer Function .....	67
Figure 4.13.	Adaptive Filters - Weight Values after 4,000 Samples.....	69
Figure 4.14.	Adaptive Filters - Weight Values after 20,000 Samples.....	69
Figure 4.15.	Comparison of MSE for all the Adaptive Filters (NLMS, NN, APA, LMS and Volterra Filter) .....	70
Figure 4.16.	Experiment 2 Setup.....	71
Figure 4.17.	Experiment 2 Results with a Small Eigenvalue Spread of 1.22 .....	71
Figure 4.18.	MSE Comparisons of the Adaptive Filters - LMS, NLMS, NN and APA (Eigenvalue Spread of 1.22) .....	72
Figure 4.19.	Experiment 2 Results with a Large Eigenvalue Spread of 100 .....	73
Figure 4.20.	MSE Comparisons of the Adaptive Filters – LMS, NLMS, NN and APA (Eigenvalue Spread of 100) .....	73
Figure 4.21.	Experiment Setup.....	74
Figure 4.22.	Input and Output Signal from the White Noise Interference Experiment .....	77
Figure 4.23.	MSE (dB) Comparison of the Dual Microphone NLMS Based SGJBF and NN Based SGJBF Output Signals.....	77
Figure 4.24.	Comparison of SNR of the Input vs SNR of the Output Graph .....	78
Figure 4.25.	N Number of Microphone Neural Switched Griffiths-Jim Beamformer. .....	80
Figure 4.26.	Arrangement of the Three-Microphone Array.....	81

Figure 4.27. (a) Acquired Input Signal (b) Output Signal from the NLMS Based SGJBF, and (c) Output Signal from the Neural SGJBF .....	82
Figure 4.28. Comparison of the MSE for the Three-Channel NLMS and NN Based SGJBF's Output Signals .....	83
Figure 4.29. Shows the Arrangement of the Linear Four-Microphone Array .....	84
Figure 4.30. (a) Acquired Input Signal and VAD Output (b) Output Signal from the NLMS Based SGJBF (c) Output Signal from the NN Based SGJBF.....	85
Figure 4.31. Comparison of the MSE for the Four-Channel NLMS and NN Based SGJBF's Output Signals .....	86
Figure 5.1. Design Layout of the Proposed System .....	89
Figure 5.2. Hardware Setup of a Distant Talking Speech-Controlled VMS .....	91
Figure 5.3. Overview of the Software Implementation.....	92
Figure 5.4. Speech-Controlled Vibration Monitoring System .....	93
Figure 5.5. Example of a Grammar File used with the Speech Recognition System .....	93
Figure 5.6. Labview Implementation of the Data Acquisition of the Vibration Signal .....	94
Figure 5.7. <i>Start Acquire</i> Command .....	95
Figure 5.8. Labview Implementation of the Analyse Command .....	95
Figure 5.9. <i>Present Result</i> Command.....	95
Figure 5.10. Input and Output Signal from the Factory Noise Interference Experiment.....	97

## List of Tables

Table 2.1.	Comparison of Adaptive Algorithms.....	22
Table 4.1.	Variables Used for the Adaptive Filters.....	68
Table 4.2.	Variables Used for the Adaptive Filters.....	75
Table 4.3.	Comparison of NN and NLMS GJBF with Different Interferences ....	79
Table 4.4.	Variables Used for Three-Microphone Beamformer Experiment .....	81
Table 4.5.	SNR of the Input Signals and the Output Signal from the Three- Microphone NLMS and NN Based SGJBF .....	83
Table 4.6.	SNR of the Input Signals and the Output Signal from the Four- Microphone NLMS and NN Based SGJBF .....	86
Table 5.1.	SNR of the Input Signals and the Output Signal from the NLMS and NN Based SGJBF .....	97
Table 5.2.	Evaluation of the Speech-Controlled VMS with the Speech Enhancement System in Factory Noise Environment .....	99
Table 5.3.	Evaluation of the Speech-Controlled VMS with the Speech Enhancement System in White Noise Environment.....	99

## List of Abbreviations

ADALINE	Adaptive Linear Element
ANC	Adaptive Noise Canceller
ANN	Artificial Neural Network
APA	Affine Projection Algorithm
DLL	Direct Link Library
FB	Feed-Backward
FC	Fully Connected
FF	Feed-Forward
FIR	Finite Impulse Response
GCC	Generalised Cross Correlation
GJBF	Griffiths-Jim Beamformer
GSC	Generalised Sidelobe Canceller
IDNN	Input Delay Neural Network
LMS	Least Mean Squares
MLP	Multi-Layer Perceptron
MSC	Magnitude Squared Coherence
MSE	Mean Square Error
NFC	Non-Fully Connected
NLMS	Normalised Least Mean Squares
NN	Neural Network
RLS	Recursive Least Square
SDK	Software Development Kit
SGJBF	Switched Griffiths-Jim Beamformer
SNR	Signal-to-Noise Ratio
SRHR	Speech Recognition Hit Rate
TDNN	Time Delay Neural Network
VAD	Voice Activity Detector
VMS	Vibration Monitoring System

## List of Symbols

$d(n)$	Primary input sample value at time $n$
$e(n)$	Output error sample value at time $n$
$G_1, G_2$	FIR Transfer functions of the speech signal
$h_1, h_2$	Learning rates for the Neural Network
$H_1, H_2$	FIR Transfer functions of the noise signal
$k(n)$	Gain vector in the RLS algorithm
$n, k$	Discrete-time or number of iterations
$N$	Number of filter weights
$M$	Projection order for APA
$x(n)$	Reference input sample value at time $n$
$\vec{x}(n)$	Tap-input regression vector consisting of $x(n), x(n-1), \dots$ , as elements
$w(n)$	Weight value at time $n$
$\vec{w}(n)$	Weight vector consisting of $w(n), w(n-1), \dots$ , as elements
$y(n)$	Adaptive filter's output value at time $n$
$\mu$	Step-size parameter in LMS/NLMS algorithm
$\gamma$	Small positive constant in LMS algorithm
$\delta$	Regularisation parameter in APA
$\lambda$	Forgetting factor used in RLS algorithm