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

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

List of Tables

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
h1, h2	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),…., as elements
$w(n)$	Weight value at time *n*
$\vec{w}(n)$	Weight vector consisting of w(n), w(n-1),…., as elements
$y(n)$	Adaptive filter's output value at time *n*
μ	Step-size parameter in LMS/NLMS algorithm
γ	Small positive constant in LMS algorithm
δ	Regularisation parameter in APA
λ	Forgetting factor used in RLS algorithm