A Kepstrum Approach to Real-Time Speech Enhancement

Thesis

for

The Degree of Doctor of Philosophy

Information Engineering

Institute of Technology and Engineering

Massey University at Albany

Jinsoo Jeong

B.E., M.E., M.Sc.

2007

Abstract

This research is mainly concerned with a robust method for an improved performance of a real-time speech enhancement and noise cancellation in a real reverberant environment. Therefore, the thesis titled, **"A Kepstrum Approach to Real-Time Speech Enhancement"** presents an application technique of a kepstrum method to a speech enhancement method. The kepstrum approach is based on a fundamental theory of kepstrum analysis, which gives a mathematical construct to the application of a speech enhancement. Kepstrum analysis is applied to the system identification application of unknown acoustic transfer functions between two microphones. This kepstrum method provides a mathematical representation with FFT based processing and is independent of acoustic path model order. The front-end application of the kepstrum method to speech enhancement methods provides an improved performance in speech enhancement and noise cancellation with several favourable effects.

■

Table of Contents

List of Figures

■

List of Tables

■

List of Abbreviations and Acronyms

ANC Adaptive Noise Canceller

BSS Blind Source Separation

CRLB Cramér-Rao Lower Bound

CTRANC CrossTalk Resistant Adaptive Noise Canceller

DOA Direction of Arrival

DFT Discrete Fourier Transform

DS Delay and Sum

FFT Fast Fourier Transform

FIR Finite Impulse Response

GCC Generalized Cross Correlation

G-J Griffiths and Jim

GSC Generalized Sidelobe Canceller

GSD Generalized Sidelobe Decorrelator

HT Hannan Thomson

IDFT Inverse Discrete Fourier Transform

IFFT Inverse Fast Fourier Transform

IIR Infinite Impulse Response

KEPS Kolmogorov Equation Power Series

LabVIEW Laboratory Virtual Instrument Engineering Workbench

LCMV Linearly Constrained Minimum Variance

LMS Least Mean Square

MISO Multiple Input Single Output

ME	Maximum Entropy
ML	Maximum Likelihood
MMSBA	Multi-Microphone Sub-Band Adaptive
MMSE	Minimum Mean-Square Error
MSC(1)	Magnitude Squared Coherence
MSC(2)	Multiple Sidelobe Canceller
MUSIC	MUltiple SIgnal Classification
MVDR	Minimum Variance Distortionless Response
NLMS	Normalized Least Mean Squares
NRN	Normalized Residual Noise
PHAT	PHAse Transform
PSD	Power Spectral Density
RLS	Recursive Least Square
SAD	Symmetric Adaptive Decorrelation
SBAGJ	Sub-Band Adaptive Griffiths and Jim
SCOT	Smoothed COherence Transform
SD	Signal Distortion
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
TDOA	Time Difference Of Arrival
VAD	Voice Activity Detector
WOSA	Weighted Overlapped Segment Averaging

■

List of Symbols

μ	Step-size parameter for LMS
μ_n	A modified input dependent step size for NLMS
∇_n	Gradient vector at time n
$\hat{\nabla}_n$	Instantaneous estimate of the gradient vector at time n
$\mathbf{h}_n$	Tap weight vector at time n of LMS or NLMS
$\hat{\mathbf{h}}_n$	Instantaneous estimate of the tap weight vector at time n
$J(h)$	Mean square value of the estimation error
$E[\cdot]$	Expectation operator
$R_{xx}(k)$	Discrete autocorrelation function of the input signal x_n
$R_{xd}(k)$	Discrete cross-correlation function between x_n and the desired response d_n
$\Phi_{xx}(z)$	Z-transform auto power spectrum of the input signal x_n
$\Phi_{xd}(z)$	Z-transform cross power spectrum between the input signal x_n and a desired response d_n
$\mathbf{R}$	$E[X_n X_n^H]$, autocorrelation vector of tap input vector $\mathbf{x}_n$
$\mathbf{P}$	$E[X_n d_n^*]$, cross-correlation vector between the tap input vector $\mathbf{x}_n$ and the desired response d_n
$\mathbf{x}_n^T$	Transposition input vector $\mathbf{x}_n$ at time n
$\mathbf{x}_n^H$	Hermitian transposition input vector $\mathbf{x}_n$ at time n
$SNR_d(z)$	Signal-to-noise density ratio at the primary input

$SNR_x(z)$	Signal-to-noise density ratio at the reference input
$SNR_e(z)$	Signal-to-noise density ratio at the output
$H(z)$	Causal FIR filter, convergent within $\|z\|<1$
$H(z^{-1})$	Uncausal FIR filter, convergent in $\|z\|>1$
h_n	Impulse response of transfer function $H(z)$
$\mathbf{p}(\theta, w)$	Array response vector, steering vector or direction vector
$\delta(t)$	Dirac delta function
$\psi_g(f)$	General frequency weighting function
$\mathrm{R}_{d'x'}^{(g)}(\tau)$	Generalized cross correlation function between $d'(t)$ and $x'(t)$
$\hat{\gamma}_{dx}(f)$	Coherence estimate between $x_d(t)$ and $x_x(t)$
$\|\gamma_{dx}(f)\|^2$	Magnitude squared coherence function
$\lambda_{\max}$	The largest eigenvalue of the tap input auto correlation matrix **R**
$tr(\mathbf{R})$	The trace of the tap input auto correlation matrix **R**
$H_A(z)$	All-pass filter
$H_L(z)$	NLMS filter
$H_{\pm}(z)$	A double sided transfer function
$H_+(z), H_-(z)$	A positive sided and negative sided transfer functions
$H_M(z), H_N(z)$	Minimum phase and nonminimum phase transfer functions
$H^+(z), H^-(z)$	Spectral factors from the double sided z-transform, corresponding to a minimum phase part and a non-minimum phase part respectively
$K^+(z), K^-(z)$	Kepstrum minimum phase causal part with zeros inside the unit circle

of the z-plane and its 'mirror image' non minimum phase counterpart of $K^{+}(z)$

$z^{-n}H(z^{-1})$ n^{th} order reciprocal polynomial

γ Euler's constant (0.577215…)

β Forgetting factor

k_n Kepstrum coefficients

$E_y^{n_0}$, $E_m^{n_0}$ Output energy and the input energy to an all-pass filter, truncated at time n_0

$h_M(n), h_N(n)$ Minimum phase and non minimum phase impulse responses

$E_{hm}^{n_0}$, $E_{hn}^{n_0}$ Energies of $h_M(n)$ and $h_N(n)$

$H_M^1(z)H_N^2(z)$ The example of transfer function showing that the superscripts indicate the number of roots based on the lower subscripts, *M* and *N* corresponding to minimum phase and non minimum phase terms respectively

■

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, **Dr. Tom J. Moir** for his invaluable guidance in his position as a top class of world researcher in this field. From beginning to end, he has provided many opportunities for me to develop my research interests as well as his solid background in research expertise. Secondly, I would also like to give thanks to the department and institute for providing an excellent research and study environment, and financial support for the participation in international conferences. Finally, I have to express many thanks to my beloved wife, **Seungsook Jeong** for her endurance and sacrifice even in her severe illness, and also special thanks to my mother-in-law, **Jongsil Kim**, who has given us a consistent support with unlimited love in Korea.

■

Declaration

I declare that the thesis is based on my own research work under the supervision of Dr. T. J. Moir during the Ph.D. study in Information Engineering, Institute of Technology and Engineering, Massey University at Albany.

The research work has produced conference proceedings and presentations during the Ph.D. study. The contents of this thesis therefore contain theory, procedure, application and experimental outputs from the research papers published during the research period as listed below.

1. **J. Jeong** and T. J. Moir, "A real-time kepstrum approach to speech enhancement and noise cancellation" Accepted with a minor revision and submitted the final revision for a *special issue of Neurocomputing Journal in 2007* (will be published by Elsevier)

2. T. J. Moir and **J. Jeong**, "Identification of non-minimum phase transfer function components" *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology* (ISSPIT), pp 380-384, August 27-30, 2006, Vancouver, Canada

3. **J. Jeong** and T. J. Moir, "Two-microphone kepstrum approach to real-time speech enhancement methods" *Proceedings of the IEEE International Conference on Engineering of Intelligent Systems* (ICEIS), pp 392-397, April 22-23, 2006, Islamabad, Pakistan

4. **J. Jeong** and T. J. Moir, "Kepstrum approach to real-time speech enhancement methods using two microphones", *Proceedings of the International Conference on Sensing Technology* (ICST), pp 691-695, November 21-23, 2005, Palmerston North, New Zealand

■