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 $\mathbf{R}$
$tr(\mathbf{R})$	The trace of the tap input auto correlation matrix $\mathbf{R}$
$H_A(z)$	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

■

Chapter 1

Introduction

During the last several decades, a speech signal and array processing technology has been developed in various application areas. Therefore, improved performance has resulted from several modified or hybrid methods mostly based on conventional approaches, such as adaptive noise cancelling, spectral subtraction, beamforming technique and cepstrum method. However, a poor performance normally comes from signal distortions due to signal leakage into the reference input, uncorrelated noises between two microphones, and also reverberation. In addition, we have to consider the physical dimensions in the size of the microphone array and computational complexity of the software. Therefore, for a real-time application in real environments, three main limiting factors have been considered. Those are signal distortions, computational complexity in processing and a large dimension of microphone array size.

- **Research objective**

The objective of this thesis is to simplify the use of microphones and reduce the complexity of processing so as to provide a robust method with an improved performance for real-time processing in a real reverberant environment. In the thesis, the kepstrum approach is proposed, which uses the kepstrum method and its front-end application to a speech enhancement method with a modified application from theoretical assumptions of adaptive noise cancelling and beamforming. It provides an improved speech enhancement with efficient, robust and real time processing in a real environment as shown in Fig. 1-1.

- **Kepstrum approach**

The kepstrum approach uses a modified application from a conventional adaptive noise canceller (ANC) and the Griffiths and Jim (G-J) beamformer. The modified approach to an adaptive noise cancelling is to use a small separation (say 20cm) between two microphones and a voice activity detector (VAD) (to get access to the noise statistics) during the noise periods. It gives benefits of a reduced adaptive filter size and minimized reverberation. In addition, the modified application to G-J beamformer uses direct speech path in front of two microphones and employs sum and subtract functions (after phase alignment) for signal

blocking. This gives increased speech directivity in the primary input and a refined noise reference in a reference input (Fig. 1-2).

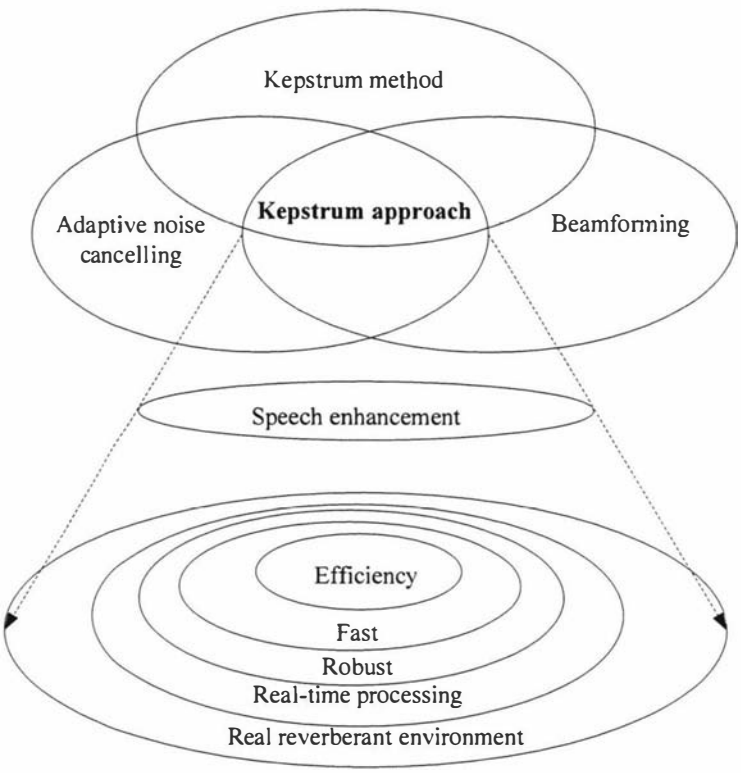


Fig.1-1 Research objective

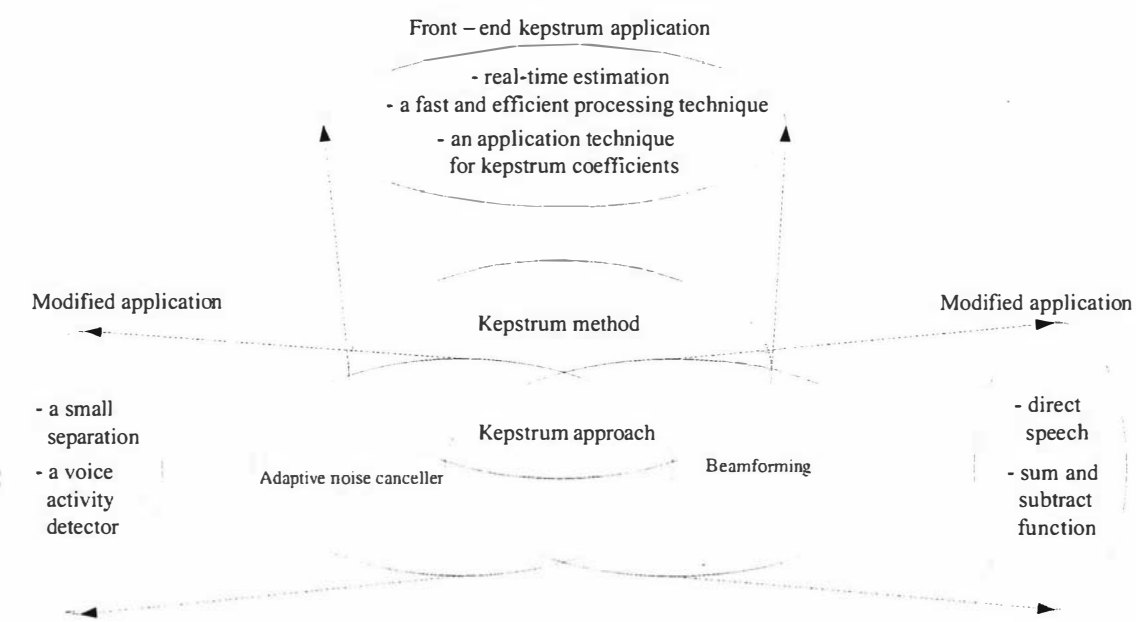


Fig. 1-2 Kepstrum approach:
Modified application from conventional main approaches and kepstrum method.

Based on these modifications, the front-end application of a FFT based kepstrum method provides a fast and efficient real-time estimation and processing technique for a system identification of unknown acoustic transfer functions between two microphones, and also an application technique for kepstrum coefficients by virtue of the VAD.

- **Contributions to knowledge**

In the thesis, the following original main contributions have been made:

1. The front-end kepstrum application has shown an improved performance for a speech enhancement with several favourable effects.

2. It has been implemented in real-time and its performance has been compared in a real reverberant environment.

3. Nonminimum phase terms can be directly identified from the cascaded RLS filter after front-end minimum phase kepstrum application for the noise-free case.

- **Performance with favourable effects**

The kepstrum approach has provided an improved performance with several favourable effects to the two-microphone approach, such as:

1. The system allows invertibility, which gives stable minimum phase transfer function on application of the inverse of a room impulse response.

2. A highly reduced adaptive filter size can be used because most of the minimum phase part is absorbed in the front-end kepstrum estimate.

3. A small number of kepstrum coefficients are required because the minimum phase property shows highly concentrated energy around time zero.

■

Chapter 2

Literature review

1. Introduction

A speech signal and array processing technology has been introduced and progressively developed with further modifications to achieve a goal of a speech enhancement and noise cancellation as a far-field solution in various environments over the last several decades.

For the near-field applications where the user can wear a headset, close-talking microphones can provide a noise cancelling solution using omnidirectional, unidirectional or bidirectional microphones. As new microphone technologies are also being developed, microphones do not only provide a near-field solution for a noise cancellation but also can add a better performance in far-field applications.

Therefore, applications may be considered as two main solutions (Fig. 2-1). One is a near-field solution, which uses noise cancelling microphones. The other is a far-field solution, which utilizes a speech and array signal processing technique. It is mostly based on main approaches, such as:

- Adaptive noise cancelling
- Spectral subtraction
- Beamforming
- Complex cepstrum or kepstrum

The chapter will cover a general review of microphone technology, main approaches, adaptive algorithms and spectral estimation algorithms.

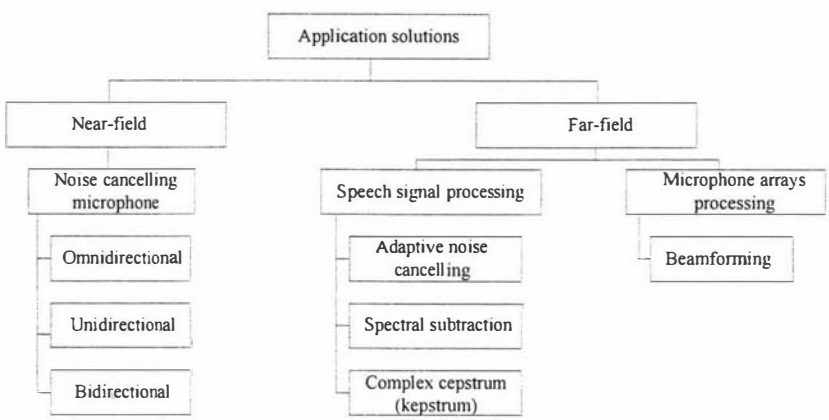


Fig. 2-1 Application solutions ■

2. Microphone technology

A microphone is a generic term that refers to any element that transforms an acoustic energy (sound) into an electrical energy (electricity (audio signal)). The word 'microphone' comes from Greek words, *micros* (small) and *phone* (sound) and the short form of the word is *mike* (1927) or *mic* (1961) (**Online etymology dictionary**).

The first microphone was invented by Emile Berliner but the first practical carbon microphone was commercialized by Thomas Edison in 1876. It has been developed as several different types but the most commonly used element is a thin and flexible diaphragm in it (Fig. 2-2). This thin piece of material vibrates when it is struck by sound waves. These vibrations are converted into electrical signals from the acoustic wave sound by various methods (**Media college**).

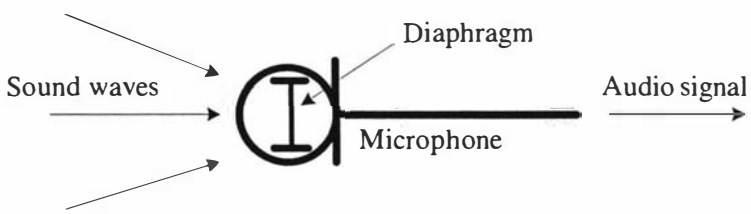


Fig. 2-2 Structure of microphone

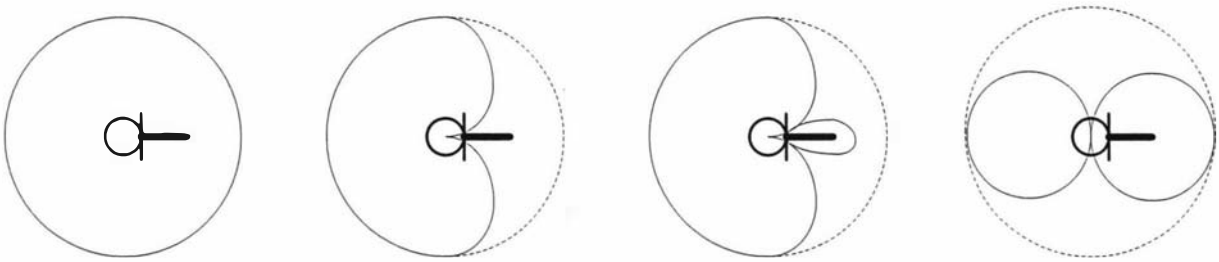
There are a number of different types of microphone. However, they can be broadly classified in two ways. One uses sensitivity polar patterns and the other uses physical materials.

2.1 Classification by sensitivity polar pattern

Microphone types are classified as non-directional (omni-directional) and directional (bi-directional and unidirectional) according to sensitivity (pickup) patterns as illustrated in Fig. 2-3 (**Sweetwater**).

- **Omni-directional**

Its diaphragm is designed to be sensitive to signals emanating from any direction (Fig. 2-3(A)). It can not discriminate between a direction of arrival (DOA) and a distant sound source. On a close-talking application, such as in a headset, it will give an output directly proportional to the acoustic signal it measures.



(A) (B) (C) (D)

Fig. 2-3 Sensitivity polar patterns: (A) omnidirectional (B) unidirectional (cardioid) (C) unidirectional (hypercardioid) (D) bidirectional (Figure-8)

- **Unidirectional**

A unidirectional microphone, which is similar to a pressure gradient microphone (bi-directional), incorporates an internal acoustic time delay which biases the directional response of the microphone towards a preferred direction. Their noise cancelling properties come from the fact that they are more sensitive to acoustic waves arriving from a forward direction and are less sensitive to waves arriving from behind the microphone. This microphone typically exhibits cardioid (Fig. 2-3(B)) or hypercardioid directivity (polar) patterns (Fig. 2-3(C)). It is popular in computer applications involving desktop microphones, similar to a close-talking microphone.

- **Bi-directional (Figure-8)**

This is a noise-cancelling microphone, also known as a pressure gradient microphone and uses properties of a gradient microphone to achieve a noise cancellation (Fig. 2-3(D)).
Sound pressure never arrives at the front and the back of the microphone at the same time. However, any noise which arrives from the side is cancelled.

This microphone is very sensitive to placement and cannot be used interchangeably with recognition systems.

■

2.2 Classification by physical material

A variety of physical materials can be used in building microphones. The two most commonly encountered microphones are made by dynamic (electromagnetic) and variable condenser (electrostatic) models (**Sound on Sound**; **University of Salford**).

- **Dynamic microphone**

One of the most commonly used microphones, a dynamic microphone, is much like a miniature loudspeaker. A flexibly-mounted diaphragm is coupled to a coil of fine wire. The coil is mounted in the air gap of a magnet such that it is free to move back and forth within the gap. When sound strikes the diaphragm, the diaphragm surface vibrates in response. The motion of the diaphragm couples directly to the coil, which moves back and forth in the field of the magnet. As the coil cuts through lines of magnetic force in the gap, a small electrical current is induced in the wire. The magnitude and direction of that current is directly related to the motion of the coil, and the current thus is an electrical representation of the incident sound wave.

This is highly dependable, rugged, and reliable and extremely common in stage use and is used extensively in outdoor applications.

- **Condenser microphone**

This is the most common type of microphone and has a capacitor consisting of a pair of metal plates separated by an insulating material called a dielectric. One of its plates is free to move in response to change in sound pressure. The sensitivity of the microphone is related to its polarizing voltage and the distance of separation between these plates. This microphone needs an external power supply. This power system is known as phantom power, sending a voltage of +48Vdc through pins two and three of the microphone cable to the microphone capsule. The return voltage comes back via the shield.

This is slightly less durable than a dynamic microphone. It is mostly used in a studio application. It is a bit cleaner and more predictable than a dynamic microphone, but generally more expensive.

- **Electret condenser microphone**

This is a variant of the condenser microphone, which does not need phantom power to charge the diaphragm, but does require a power supply for the in-microphone preamplifier. An electret is a dielectric material that has been permanently electrically charged or polarized. This is small, cheap, durable, and offers a good performance at high frequencies. Most modern telephone handsets use electrets.

- **Others**

There are several other types of microphones, but they are rarely used in sound reinforcement.

Ribbon microphones are very fragile and carbon microphones have poor sound quality.

■

3. Main approaches

This section will review main approaches, which cover adaptive noise cancelling, spectral subtraction, complex cepstrum and kepstrum, and beamforming techniques.

3.1 Adaptive noise cancelling

The principles and analysis of adaptive noise cancelling are reviewed from the reference of Widrow et al. (**Widrow et al.,** 1975) and Widrow and Stearns (**Widrow and Stearns,** 1985). This covers the concept of adaptive noise cancelling, Wiener solution to statistical noise cancelling problems and the effect of signal components in the reference input.

3.1.1 The concept of adaptive noise cancelling

An adaptive noise canceller (ANC) has been proposed by Widrow et al. (**Widrow et al.,** 1975). It uses two or more microphones based on the availability of reference channel(s). The method uses noise statistics in reference channel(s), which are characteristics of correlated samples or references of the contaminating noise.

An adaptive filter operates on the reference microphone output and produces an estimate of the noise. Its output is then subtracted from the primary microphone output (signal plus noise). The overall output of the canceller is used to control the adjustments applied to the tap weights in the adaptive filter. Using an adaptation algorithm, the ANC tends to minimize the mean square value of the overall output. It gives the output that is the best estimate of the desired signal in the minimum mean square error (MMSE) sense.

The ANC operates on the outputs of two microphones, a primary microphone that supplies a desired signal of interest buried in noise, $s + n_0$, and a reference microphone that supplies noise alone, n_1 as shown in Fig. 2-4.

The basic concept of adaptive noise cancelling is available under two main assumptions that

- The signal and noise at the output of the primary microphone are uncorrelated
- The noise at the output of the reference microphone is correlated with the noise component of the primary microphone output.

The reference microphone provides the reference input to the canceller. The noise n_1 is filtered to produce an output $n_0^{'}$ that is as close a replica as possible of n_0. This output is subtracted from the primary input $s+n_0$ to produce the system output $e=s+n_0-n_0^{'}$.

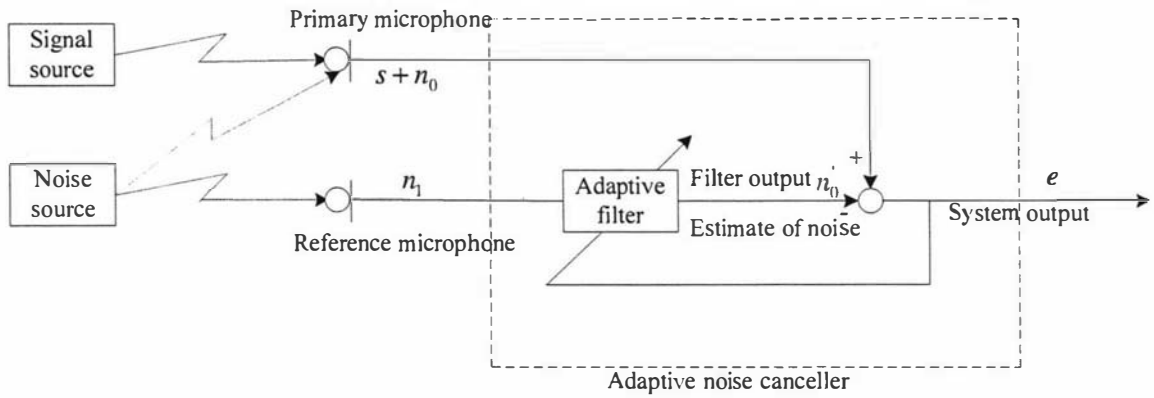


Fig. 2-4 The concept of ANC

Assuming that s, $n_0^{'}$, n_0 and n_1 are statistically stationary and have zero mean, and by squaring and expectation of $e=s+n_0-n_0^{'}$, the system output power $E[e^2]$ is written as (2-1).

$$E[e^2]=E[s^2]+E[(n_0-n_0^{'})^2]+2E[s(n_0-n_0^{'})]=E[s^2]+E[(n_0-n_0^{'})^2] \quad (2\text{-}1)$$

The signal power $E[s^2]$ will not be affected because the filter is adjusted to minimize $E[e^2]$. Accordingly, the minimum output power is written as (2-2).

$$\min\ E[e^2]=E[s^2]+\min E[(n_0-n_0^{'})^2] \quad (2\text{-}2)$$

When the filter is adjusted so that $E[e^2]$ is minimized, $E[(n_0-n_0^{'})^2]$ is therefore, also minimized. The filter output *y* is then a best least squares estimate of the primary noise n_0. Moreover, when $E[(n_0-n_0^{'})^2]$ is minimized, $E[(e-s)^2]$ is also minimized since $(e-s)=(n_0-n_0^{'})$.

Adjusting or adapting the filter to minimize the total output power thus provides the system output e to be a best least squares estimate of the signal s for the given structure and adjustability of the adaptive filter and for the given reference input.

The output will contain the signal s plus noise n. The output noise is given by $(n_0-n_0^{'})$. Since minimizing $E[e^2]$ minimizes $E[(n_0-n_0^{'})^2]$, minimizing the total output power

minimizes the output noise power. Since the signal in the output remains constant, minimizing the total output power maximizes the output SNR.

The smallest possible output power is $E[e^2]=E[s^2]$. When this is achievable, $E[(n_0-\hat{n_0})^2]=0$. Therefore, $\hat{n_0}=n_0$ and $e=s$. In this case, minimizing output power causes the output signal to be perfectly noise free.

3.1.2. Wiener solution to statistical noise cancelling problems

The objective of an optimal unconstrained Wiener solution to certain statistical noise cancelling problems is to increase performance in SNR using a noise cancelling technique. The analysis is based on the assumption of an infinitely long, two-sided noncausal tapped-delay line. However, finite length of filter and causalty are important in practical applications.

Fixed filters are not applicable in noise cancelling because the auto correlation and cross correlation functions of the primary and reference inputs are generally unknown and often variable with time. On the other hand, adaptive filters are required to 'learn' the statistics initially and to track them if they vary slowly.

For stationary stochastic inputs, the steady-state performance of adaptive filters shows close approximation to that of fixed Wiener filters. Wiener filter theory thus provides a convenient method of mathematically analyzing statistical noise cancelling problems.

1) The Wiener filter theory

For the least mean square filtering problem, Kolmogorov (**Kolmogorov,** 1941) studied the discrete time problem and originally solved it by using the Wold decomposition (**Wold,** 1938). On the other hand, Wiener (**Wiener,** 1949) studied the continuous time problem and derived the famous Wiener-Hopf integral equation, which requires knowledge of correlation functions of the signal processes of interest.

The equivalent of this integral equation in the discrete time case is known as the normal equation. The continuous time solution to the Wiener-Hopf integral equation and the discrete time solution to the normal equation are known collectively as Wiener filters.

Fig. 2-5 shows that for the classic single-input single-output (SISO) Wiener filter with input signal x_n, the output signal y_n, and desired response d_n, the input and output signals are assumed to be discrete in time, and the input signal and desired response are assumed to

be statistically stationary. The error signal is $e_n = d_n - y_n$. The filter is linear, discrete and designed to be optimum in the MMSE sense.

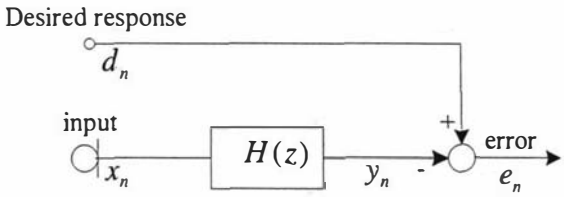


Fig. 2-5 Single-channel Wiener filter

It is composed of an infinitely long, two sided tapped delay line.

The discrete autocorrelation function of the input signal x_n, and the cross-correlation function between x_n and the desired response d_n are defined as (2-3) and (2-4) respectively.

$$R_{xx}(k) \cong E[x_n x_{n+k}] \tag{2-3}$$

$$R_{xd}(k) \cong E[x_n d_{n+k}] \tag{2-4}$$

The optimal impulse response $h(k)$ can then be obtained from the discrete Wiener-Hopf equation (2-5).

$$R_{xd}(k) = \sum_{l=-\infty}^{\infty} h(l) R_{xx}(k-l) = h(k) * R_{xx}(k) \tag{2-5}$$

This form of the Wiener solution is unconstrained in that the impulse response $h(k)$ may be causal or noncausal and of finite or infinite duration.

The transfer function of Wiener filter may now be derived as the ratio between the auto power spectral density (PSD) (2-6) of the input signals and the cross PSD (2-7) between the input signal and a desired response are the z-transforms of $R_{xx}(k)$ and $R_{xd}(k)$ respectively.

$$\Phi_{xx}(z) \cong \sum_{k=-\infty}^{\infty} R_{xx}(k) z^{-k} \tag{2-6}$$

$$\Phi_{xd}(z) \cong \sum_{k=-\infty}^{\infty} R_{xd}(k) z^{-k} \tag{2-7}$$

The transfer function of the Wiener filter is written as (2-8).

$$H(z) \cong \sum h(k) z^{-k} = \frac{\Phi_{xd}(z)}{\Phi_{xx}(z)} \tag{2-8}$$

2) The application of Wiener filter theory to adaptive noise cancelling.

An example of a single-channel ANC with correlated and uncorrelated noises in the primary input (2-9) and reference input (2-10) is considered as shown in Fig. 2-6.

$$d_n = s_n + (m_{1n} + n_n) \tag{2-9}$$

$$x_n = m_{2n} + (n_n * h_{2n}) \tag{2-10}$$

where h_{2n} is impulse response of channel whose transfer function is $H_2(z)$. The noises n_n and $n_n * h_{2n}$ have common origin, and they are correlated with each other, and are uncorrelated with s_n. The noises m_{1n} and m_{2n} are uncorrelated with each other, with s_n, and with n_n and $n_n * h_{2n}$.

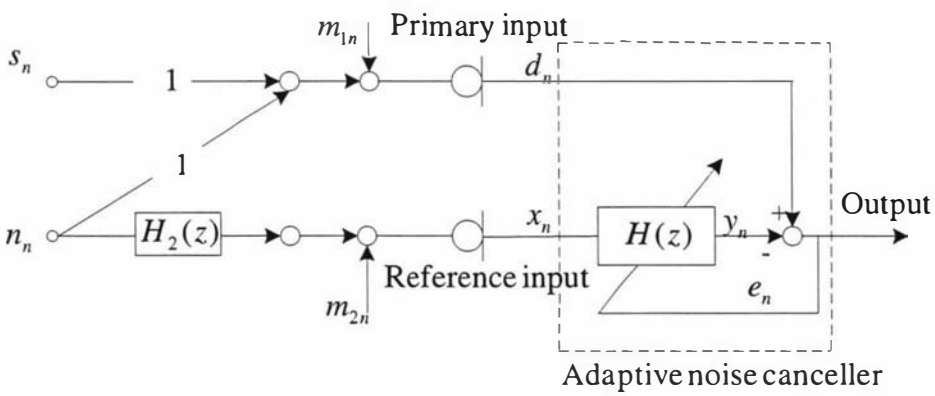


Fig. 2-6 Single channel ANC with correlated and uncorrelated noises in the primary and reference inputs (**Widrow et al.,** 1975)

If one assumes that the adaptive process has converged and the MMSE solution has been found, then the adaptive filter is equivalent to a Wiener filter.

$$\Phi_{xx}(z) = \Phi_{m_2 m_2}(z) + \Phi_{nn}(z)\left|H_2(z)\right|^2 \tag{2-11}$$

$$\Phi_{xd}(z) = \Phi_{nn}(z)H_2(z^{-1}) \tag{2-12}$$

Therefore, the Wiener transfer function is

$$H(z) = \frac{\Phi_{xd}(z)}{\Phi_{xx}(z)} = \frac{\Phi_{nn}(z)H_2(z^{-1})}{\Phi_{m_2 m_2}(z) + \Phi_{nn}(z)\left|H_2(z)\right|^2} \tag{2-13}$$

Note that this is the two-sided Wiener filter, i.e., it has a part of its impulse response which is uncausal though this does not affect the arguments which follow.

From the (2-13), $H(z)$ is independent of the primary signal spectrum $\Phi_{ss}(z)$ and of the primary uncorrelated noise spectrum $\Phi_{m_1 m_1}(z)$.

An interesting special case occurs when the additive noise m_{2n} in the reference input is zero. Then $\Phi_{m_2m_2}(z) = 0$. Therefore, the optimum transfer function is

$$H(z) = \frac{1}{H_2(z)} \tag{2-14}$$

provided of course that $H_2(z)$ is a minimum phase.

The adaptive filter, as in the balancing of a bridge, causes the noise n_n to be perfectly nulled at the noise canceller output. However, it shows that the primary uncorrelated noise m_{1n} remains uncancelled.

From the above analyses, it shows that the application violates main assumptions to be perfectly applied to a noise cancellation. Therefore, it cannot be cancelled if an uncorrelated noise in primary input exists. In addition, it does not guarantee a stable performance because of a nonminimum phase property.

3.1.3 Effect of signal components in the reference input

In addition to analyses of a noise cancellation, an effect of speech signal components in the reference input is now investigated when it is applied with a speech signal.

It is based on an application in a real environment, where in a certain circumstance the available reference input to an ANC may contain low-level unwanted signal components including the usual correlated and uncorrelated noise components, which will cause some cancellation of the primary input signal.

Three main analyses (an unconstrained Wiener filter theory as SNR, signal distortion, and noise spectrum at the canceller output) have been investigated (**Widrow et al.,** 1975).

Fig. 2-7 shows an ANC, where reference input contains signal components and where primary and reference inputs contain additive correlated noises. The z-transforms of auto and cross PSD are expressed as (2-15) and (2-16) respectively.

$$\Phi_{xx}(z) = \Phi_{ss}(z)\left|G_2(z)\right|^2 + \Phi_{nn}(z)\left|H_2(z)\right|^2 \tag{2-15}$$

$$\Phi_{xd}(z) = \Phi_{ss}(z)G_2(z^{-1}) + \Phi_{nn}(z)H_2(z^{-1}) \tag{2-16}$$

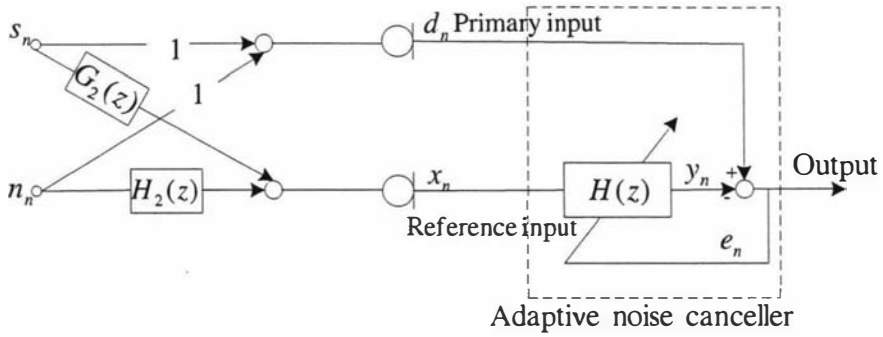


Fig. 2-7 ANC with signal components in the reference input (**Widrow et al.,** 1975)

When the adaptive process has converged, the unconstrained Wiener transfer function of the adaptive filter is

$$H(z)=\frac{\Phi_{ss}(z)G_2(z^{-1})+\Phi_{nn}(z)H_2(z^{-1})}{\Phi_{ss}(z)|G_2(z)|^2+\Phi_{nn}(z)|H_2(z)|^2} \tag{2-17}$$

1) Analysis on SNR at the noise canceller output, $SNR_e(z)$

The transfer function of the propagation path from the signal input to the noise canceller output is $1-G_2(z)H(z)$ and that of the path from the noise input to the canceller output is $1-H_2(z)H(z)$.

The spectrum of the signal component in the output is

$$\Phi_{ss_e}(z)=\Phi_{ss}(z)|1-G_2(z)H(z)|^2=\Phi_{ss}(z)\left|\frac{[H_2(z)-G_2(z)]\Phi_{nn}(z)H_2(z^{-1})}{\Phi_{ss}(z)|G_2(z)|^2+\Phi_{nn}(z)|H_2(z)|^2}\right|^2 \tag{2-18}$$

and that of the noise component is

$$\Phi_{nn_e}(z)=\Phi_{nn}(z)|1-H_2(z)H(z)|^2=\Phi_{nn}(z)\left|\frac{[G_2(z)-H_2(z)]\Phi_{ss}(z)G_2(z^{-1})}{\Phi_{ss}(z)|G_2(z)|^2+\Phi_{nn}(z)|H_2(z)|^2}\right|^2 \tag{2-19}$$

Therefore, the output SNR is

$$SNR_e(z)=\frac{\Phi_{ss}(z)}{\Phi_{nn}(z)}\left|\frac{\Phi_{nn}(z)H_2(z^{-1})}{\Phi_{ss}(z)G_2(z^{-1})}\right|^2=\frac{\Phi_{nn}(z)}{\Phi_{ss}(z)}\frac{|H_2(z)|^2}{|G_2(z)|^2} \tag{2-20}$$

The spectrums of the signal component in the reference input and that of noise components in the reference input are (2-21) and (2-22) respectively.

$$\Phi_{ss_x}(z)=\Phi_{ss}(z)|G_2(z)|^2 \tag{2-21}$$

$$\Phi_{nn_x}(z)=\Phi_{nn}(z)|H_2(z)|^2 \tag{2-22}$$

The SNR at the reference input is thus

$$SNR_x(z) = \frac{\Phi_{ss}(z)|G_2(z)|^2}{\Phi_{nn}(z)|H_2(z)|^2} \tag{2-23}$$

Therefore, the output SNR is

$$SNR_e(z) = \frac{1}{SNR_x(z)} \tag{2-24}$$

Assuming that the adaptive solution is to be unconstrained and the noise in the primary and reference inputs are mutually correlated, it shows that the SNR at the noise canceller output provides the reciprocal relations with one at the reference input at all frequencies.

2) Analysis on signal distortion (SD) at the noise canceller output

The transfer function of the propagation path through the filter is

$$-G_2(z)H(z) = -G_2(z)\frac{\Phi_{ss}(z)G_2(z^{-1}) + \Phi_{nn}(z)H_2(z^{-1})}{\Phi_{ss}(z)|G_2(z)|^2 + \Phi_{nn}(z)|H_2(z)|^2} \tag{2-25}$$

When $|G_2(z)|$ is small because of small amount of signal leakage, this function can be approximated as

$$-G_2(z)H(z) \cong -\frac{G_2(z)}{H_2(z)} \tag{2-26}$$

The spectrum of the signal component propagated to the noise canceller output through the adaptive filter is thus approximately

$$\Phi_{ss}(z)\left|\frac{G_2(z)}{H_2(z)}\right|^2 \tag{2-27}$$

The combining of this component with the signal component in the primary input involves complex addition and is the process that results in signal distortion. The worst case occurs when the two signal components are of opposite phase.

Let 'signal distortion' $SD(z)$ be defined as a dimensionless ratio of the spectrum of the output signal component propagated through the adaptive filter to the spectrum of the signal component at the primary input.

$$SD(z) \cong \frac{\Phi_{ss}(z)|G_2(z)H(z)|^2}{\Phi_{ss}(z)} = |G_2(z)H(z)|^2 \tag{2-28}$$

When $G_2(z)$ is small, (2-28) reduces to (2-29).

$$SD(z) \cong \left| \frac{G_2(z)}{H_2(z)} \right|^2 \tag{2-29}$$

The expression may be rewritten in a more useful form by combining the expressions for the SNR at the primary input,

$$SNR_d(z) \cong \frac{\Phi_{ss}(z)}{\Phi_{nn}(z)} \tag{2-30}$$

and the SNR at the reference input ,

$$SNR_x(z) = \frac{\Phi_{ss}(z)|G_2(z)|^2}{\Phi_{nn}(z)|H_2(z)|^2} \tag{2-31}$$

Therefore, (2-29) can be expressed as (2-32).

$$SD(z) \cong \frac{SNR_x(z)}{SNR_d(z)} \tag{2-32}$$

With an unconstrained adaptive solution and mutually correlated noises at the primary and reference input, (2-32) shows that the condition of a low signal distortion results from a high SNR at the primary input and a low SNR at the reference input.

3) Analysis on the spectrum of the output noise at the noise canceller output

The noise n_n propagates to the output with a transfer function.

$$1 - H_2(z)H(z) = 1 - H_2(z)\left[\frac{\Phi_{ss}(z)G_2(z^{-1}) + \Phi_{nn}(z)H_2(z^{-1})}{\Phi_{ss}(z)|G_2(z)|^2 + \Phi_{nn}(z)|H_2(z)|^2} \right] \tag{2-33}$$

$$= \frac{\Phi_{ss}(z)G_2(z^{-1})[G_2(z) - H_2(z)]}{\Phi_{ss}(z)|G_2(z)|^2 + \Phi_{nn}(z)|H_2(z)|^2}$$

When $|G_2(z)|$ is small,

$$1 - H_2(z)H(z) \cong \frac{-\Phi_{ss}(z)G_2(z^{-1})}{\Phi_{nn}(z)H_2(z^{-1})} \tag{2-34}$$

The output noise spectrum is

$$\Phi_{ON}(z) = \Phi_{nn}(z)|1 - H_2(z)H(z)|^2 \tag{2-35}$$

When $|G_2(z)|$ is small,

$$\Phi_{ON}(z) \cong \Phi_{nn}(z)\left|\frac{\Phi_{ss}(z)G_2(z^{-1})}{\Phi_{nn}(z)H_2(z^{-1})}\right|^2 \cong \Phi_{nn}(z)\left|SNR_x(z)\right|\left|SNR_d(z)\right| \quad (2\text{-}36)$$

From the above result, it can be understood that the first factor 1) implies that the output noise spectrum depends on the input noise spectrum and the second factor 2) implies that if SNR at the reference input is low, output noise will be low. That is, the smaller the signal component in the reference input, the more perfectly the noise will be cancelled. The third factor 3) implies that if the SNR in the primary input is low, the filter will be trained most effectively to cancel the noise rather than the signal and consequently output noise will be low.

■

3.2 Spectral subtraction

The spectral subtraction is a noise reduction method in a frequency domain. It is based on restoration of the magnitude spectrum or power spectrum of signal observed in an additive noise, by subtracting an estimate of the noise spectrum from the noisy signal spectrum (Fig. 2-8). It uses a simple underlying concept effectively by using one microphone. Therefore, it may be particularly of interest in the application areas of mobile phones or hearing aids where the applications are limited because of the number and position of microphones.

Assuming that the discrete noisy input signal x_n is composed of the clean speech signal s_n and the uncorrelated additive noise signal n_n, then it can be expressed in time and frequency domain forms as (2-37) and (2-38) respectively.

$$x_n = s_n + n_n \quad (2\text{-}37)$$

$$X_k = S_k + N_k \quad (2\text{-}38)$$

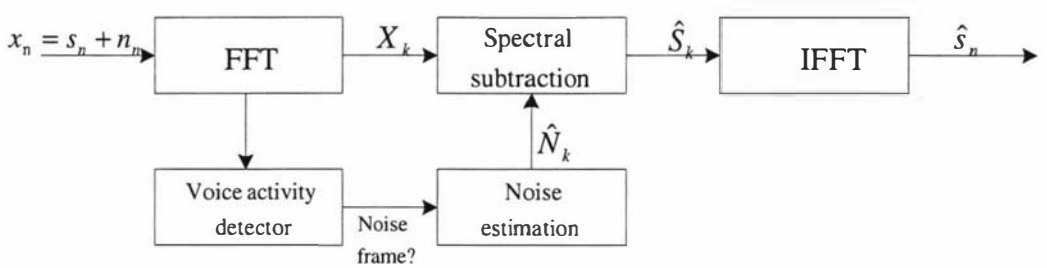


Fig. 2-8 block diagram of spectral subtraction algorithm

The equation describing the spectral subtraction may be expressed as

$$\left|\hat{S}_k\right|^b = \begin{cases} \left|X_k\right|^b - \alpha\left|\hat{N}_k\right|^b & \text{if } \left|X_k\right|^b - \alpha\left|\hat{N}_k\right|^b > \beta\left|X_k\right|^b \\ \beta\left|X_k\right|^b & \text{otherwise} \end{cases} \quad (2\text{-}39)$$

where $\left|\hat{S}_k\right|^b$ is an estimate of the original signal spectrum and $\left|\hat{N}_k\right|^b$ is an estimate of the time-averaged noise spectrum. The subtraction factor α controls the amount of noise subtracted from the noisy signal. The flooring factor β is a noise floor that is a positive constant. For a magnitude spectral subtraction (**Boll,** 1979), the exponent b is 1, and for a power spectral subtraction (**Lim,** 1979), the exponent b is 2.

From the above process, the performance could be achieved in a better SNR but found the subtraction process introduces an annoying artifact called musical noise, due to a residual noise in the enhanced speech. The other disadvantages are an application limitation to environmental conditions where: 1) noise signal is greater than speech signal, 2) nonstationarity and 3) speech-like noise environments.

■

3.3 Beamforming techniques

This section reviews beamforming techniques, which cover a basic beamforming operation and analysis of classified typical beamformers. It has been reviewed from the references of Van Veen and Buckley (**Van Veen and Buckley,** 1988), Haykin et al. (**Haykin et al.,** 1985) and Widrow and Stearn (**Widrow and Stearn,** 1985).

3.3.1 Beamforming operation and spatial filtering

The term beamforming, "forming beams" seems to indicate a radiation of energy of a signal, but it can be applicable to either a radiation or a reception of energy (**Van Veen and Buckley,**1988). It derives from the fact that early spatial filters has been designed to form pencil beams to receive a signal radiating from a specific location and attenuate signals from other locations.

The beamforming structure consists of an array of sensors and beamformer. The array of sensors collects spatial samples of propagating wave fields and a beamformer is a processor to provide a versatile form of spatial filtering, which separates signals that have overlapping frequency content but originate from different spatial locations. The objective of a

beamformer is to estimate the signal arriving from the 'look' direction (i.e., a desired speech direction) in the presence of noise and interfering signals arriving at a different angle.

The systems designed to receive spatially propagating signals often have problems with interfering signals. If the desired signal and interferers occupy the same temporal frequency band, then temporal filtering cannot be used to separate signal from interference. However, provided that the desired and interfering signals originate from different spatial locations, spatial separation can be used to separate a signal from interference using a spatial filter at the receiver.

Spatial filter is used to process data collected over a spatial aperture. A beamformer processes the spatial filtering with a discrete spatial sampling.

The output of a typical beamformer linearly combines the spatially sampled time series from each sensor to obtain a scalar output in a time domain. However, two principal advantages of spatial sampling with an array of sensors are spatial discrimination capability and versatility offered by discrete sampling.

1) Beamformers for narrowband and broadband signals

There are two typical beamformers. The first one is called narrowband beamformer (Fig. 2-9). It samples the propagating wave field in space for processing narrowband signals. The output y_k at time k is given by a linear combination of the data at n sensors at time k. The output in this case can be expressed as (2-40).

$$y_k = \sum_{i=1}^{n} x_{ki}^{H} h_i \tag{2-40}$$

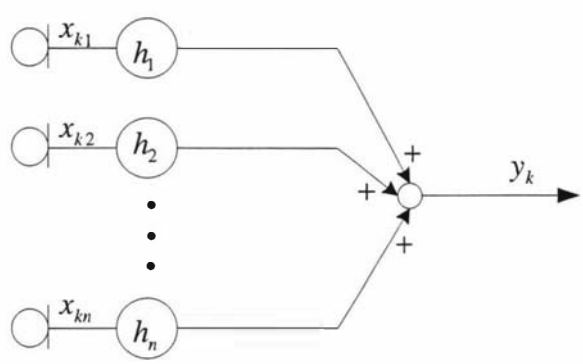


Fig. 2-9 Typical narrowband beamformer

where H represents Hermitian transposition.

Most propagated RF (radio frequency) waveforms are narrowband signals and have well defined nominal wavelength. Time delay can be compensated by a simple phase shift.

The second one is called broadband beamformer (Fig. 2-10). It samples the propagating wave field in both space and time and it is often used when signals of significant frequency extent (broadband) are of interest.

Audio (30-15KHz), acoustic vehicles (20-2KHz), seismic vehicles (5-500Hz) and vibrating machinery are broadband signals and have no characteristic wavelength. Time delay must be obtained by interpolation of the waveform.

The output in this case can be expressed as (2-41).

$$y_k = \sum_{i=1}^{n}\sum_{j=0}^{p-1} h_{i,j}^{H} x_{i(p-j)} \tag{2-41}$$

where $p-1$ is the number of delays in each of the n sensor channels. If the signal at each sensor is viewed as an input, then a beamformer represents a MISO (multiple-input single-output) system.

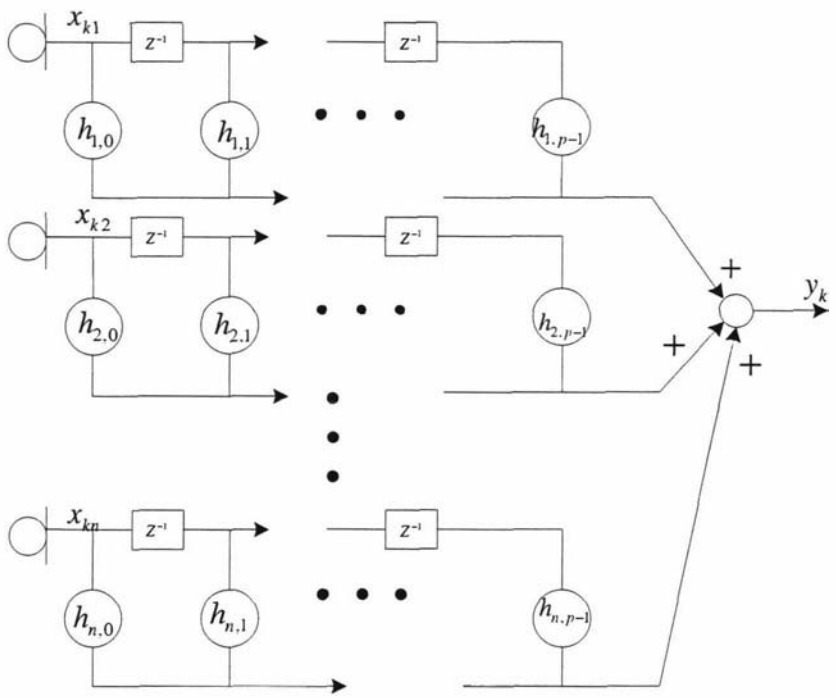


Fig. 2-10 Typical broadband beamformer

Assume that h and x_k are N dimensional, which implies that $N = pn$ is for broadband and $N = n$ is for narrowband.

2) FIR (finite impulse response) filter response and beamformer response

The frequency response of an FIR filter with tap weights $h_j, 0 \le j \le p-1$ and a tap delay of T seconds is given by

$$H(w) = \sum_{j=1}^{p} h_j^H e^{-jwT(j-1)} \tag{2-42}$$

alternatively, it can be expressed in a vector form.

$$H(w) = \mathbf{h}^H \mathbf{p}(w) \tag{2-43}$$

where $\mathbf{h}^H = [h_1^*, h_2^*, ...h_p^*]$ and $\mathbf{p}(w) = [1, e^{jwT}, e^{j2wT}, ...e^{j(p-1)wT}]$.

$H(w)$ represents the response of the filter to a complex sinusoid of frequency w and $\mathbf{p}(w)$ is a vector describing the phase of the complex sinusoid at each tap in the FIR filter relative to the tap associated with h_1.

Similarly, beamformer response is defined as the amplitude and phase presented to a complex plane wave as a function of location and frequency. Fig. 2-11 illustrates the manner in which an array of sensors samples a spatially propagating signal.

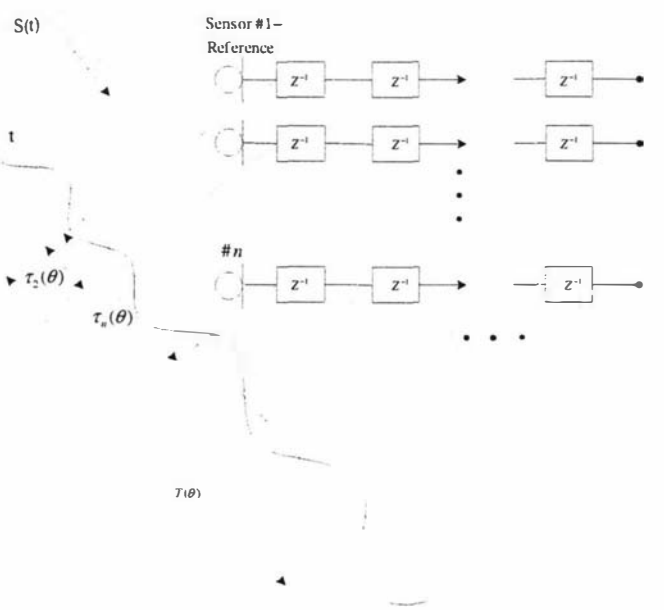


Fig. 2-11 Illustration of spatial and temporal sampling

Assume that the signal is a complex plane wave with a DOA θ and frequency w. For convenience let the phase be zero at the first sensor. This implies $x_{k1} = e^{jwk}$ and

$x_{ki} = e^{jw(k-\tau_i(\theta))}$, $2 \le i \le n$. $\tau_i(\theta)$ represents the time delay due to propagation from the first to the i-th sensor.

Therefore beamformer output y_k is

$$y_k = e^{jwk} \sum_{i=1}^{n} \sum_{j=0}^{p-1} h_{i,j}^{*} e^{-jw(\tau_i(\theta)+j)} = e^{jwk} H(\theta, w) \tag{2-44}$$

where $\tau_1(\theta)$ =0. $H(\theta, w)$ is the beamformer response and can be expressed in vector form as $H(\theta, w) = \mathbf{h}^H \mathbf{p}(\theta, w)$.

The elements of $\mathbf{p}(\theta, w)$ correspond to the complex exponentials $e^{-jw(\tau_i(\theta)+j)}$. In general, it can be expressed as $\mathbf{p}(\theta, w) = [1, e^{jw\tau_2(\theta)}, e^{jw\tau_3(\theta)}, ...e^{jw\tau_N(\theta)}]^H$ where the $\tau_i(\theta)$, $2 \le i \le N$, are the time delays due to propagation and any tap delays from the zero phase reference to the point at which the i-th weight is applied.

We refer to $\mathbf{p}(\theta, w)$ as the array response vector, also known as the steering vector or direction vector. Non-ideal sensor characteristics can be incorporated into $\mathbf{p}(\theta, w)$ by multiplying each phase shift by a function $a_i(\theta, w)$, which describes the associated sensor response as a function of frequency and direction.

The beampattern is defined as the magnitude squared of $H(\theta, w)$. Note that each weight in $\mathbf{h}$ affects both the temporal and spatial response of the beamformer. Historically, the use of FIR filters has been viewed as providing frequency dependent weights in each channel. However, as a MISO system, the spatial and temporal filtering that occurs is a result of mutual interaction between spatial and temporal sampling.

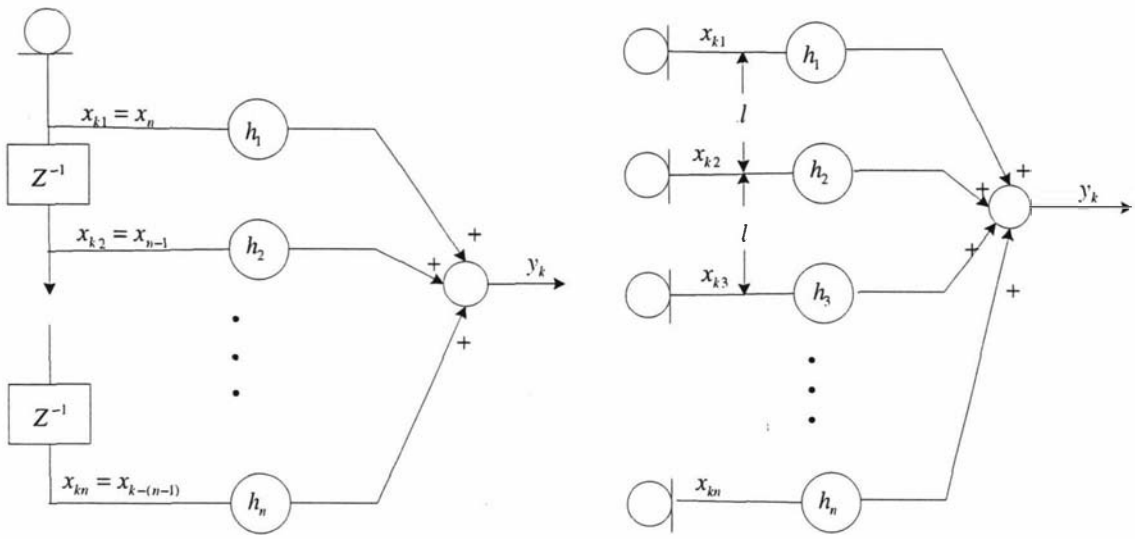
3) Correspondence between FIR filtering and beamforming

The correspondence between FIR filtering and beamforming is closest when the beamformer operates at a single temporal frequency w_0 and the array geometry is linear and equi-spaced as illustrated in Fig. 2-12.

Letting the sensor spacing be l, propagation velocity be c, and θ represent DOA relative to broadside (perpendicular to the array), we have $\tau_i(\theta) = (i-1)(l/c)\sin\theta$.

In this case we identify the relationship between temporal frequency w in $\mathbf{p}(w)$ (FIR filter) and direction θ in $\mathbf{p}(\theta, w_0)$ (beamformer) as $w = w_0(l/c)\sin\theta$. Thus, temporal frequency in a FIR filter corresponds to the sine of direction in a narrowband, linear equi-spaced

beamformer. Complete interchange of beamforming and FIR filtering methods is possible for this special case provided the mapping between frequency and direction is accounted for.



(a) (b)

Fig. 2-12 The analogy between an equi-spaced omni-directional narrowband line array and a single-channel FIR filter

3.3.2 Beamformer classification

Beamformers can be classified into data independent, statistically optimum or adaptive beamformer depending on how the weights are chosen (**Van Veen and Buckley,** 1988). The weights in a data independent beamformer do not depend on the array data or data statistics but are chosen to present a specified response for all signal/interference scenarios. The weights in a statistically optimum beamformer are chosen based on the statistics of the array data to 'optimize' the array response. In general, the statistically optimum beamformer places nulls in the directions of interfering sources to maximize the SNR at the beamformer output. The statistics of the array data are not usually known and may change over the time so adaptive algorithms are typically used to determine the weights. The adaptive algorithm is designed so the beamformer response converges to a statistically optimum solution.

Based on the method used to choose the weights, a delay and sum (DS) beamformer is classified as data independent, a minimum variance distortionless response (MVDR) beamformer is classified as statistically optimum, and a generalized sidelobe canceller (GSC) is classified as adaptive beamformer algorithm, where these will be reviewed respectively.

1) DS beamformer

DS beamformer is data independent and provides the simplest operation in beamforming. Delaying the microphone signals appropriately to bring all signals arriving from the angle θ in phase and summing the signals produce an increased performance in the SNR because in-phase components act constructively and the out-of-phase components act destructively.

The weights in a data independent beamformer are designed so the beamformer response approximates a desired response independent of the array data or data statistics. It could be considered that the design for approximating a desired response is the same as that for classical FIR filter design by exploiting the analogies between beamforming and FIR filtering in Fig. 2-13, which can be applied to the problem of separating a single complex frequency component from other frequency components using the n tap FIR filter.

If frequency w_0 is of interest, then the desired frequency response is unity at w_0 and zero elsewhere. A common solution to this problem is to choose $\mathbf{h}$ as the vector $\mathbf{p}(w_0)$. Since $\mathbf{h} = \mathbf{p}(w_0)$, each element of $\mathbf{h}$ has unit magnitude. Tapering or windowing the amplitudes of the elements of $\mathbf{h}$ permits trading of main lobe or beam width against sidelobe levels to form the response into a desired shape. Assuming that array response vector $\mathbf{p}(\theta_0, w_0)$ is from the signal arriving from a known location point θ_0 and narrowband (frequency w_0), the resulting array and beamformer is termed a phased array since the output of each sensor is phase shifted prior to summation. Amplitude tapering can be used to control the shape of the response, i.e., to form the beam.

If the array is narrowband and the sensors lie on a line, then methods evolving from continuous spatial aperture design can be employed. And if the array is planar and factorable, the line array techniques can be used to synthesize the overall response as the product of two linear array responses.

If the beamformer is broadband and employs FIR filters, the tapering can be applied independently to both the sensor outputs and FIR filters as shown in Fig. 2-13. The taper weights are chosen to shape the spatial response and the FIR filter coefficients to present a desired temporal response. This shows that a spatial and temporal response interacts so it can not be synthesized completely independently.

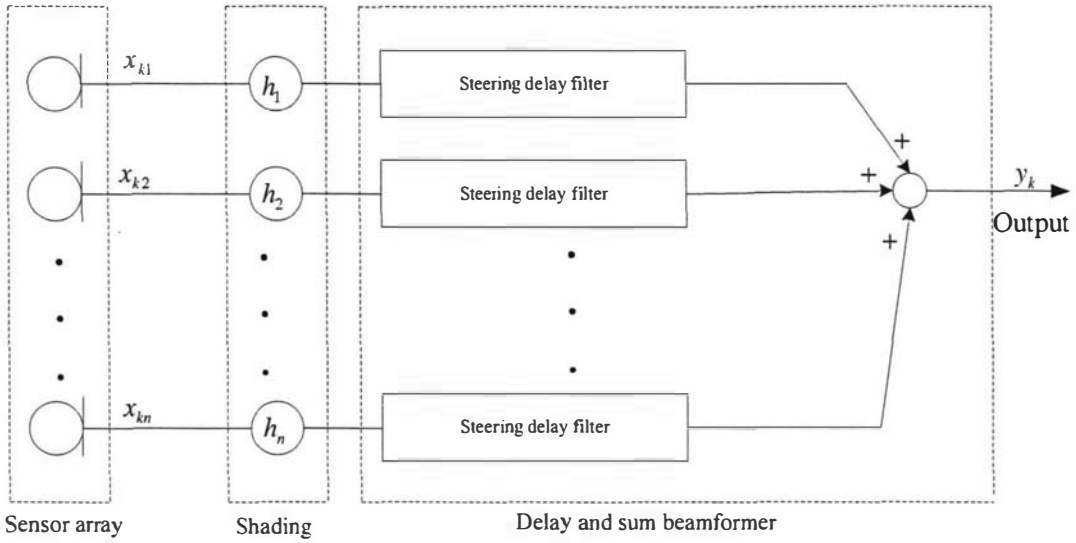


Fig. 2-13 A classical broadband beamformer for line arrays

DS beamformer is one example using the structure of Fig. 2-13, where the FIR filters approximate the propagation delays (linear phase over the frequency band of interest) and the taper weights are chosen to shape the main beam and sidelobe structure of the spatial response.

2) MVDR and LCMV beamformer

The poor performance of the DS beamformer is mainly due to the fact that its response along a direction of interest depends not only on the power of the incoming targeet signal coming from desired source but also depends on interfering noise signal coming from undesired sources. To overcome this limitation of the DS beamformer, the MVDR beamformer has been proposed by Capon (**Capon,** 1969). The weights at each element are chosen to minimize the variance (i.e., average power) of the output beamformer while the gain to the looking direction (i.e., direction of target source) is maintained at a specific level.

In 1972, a linearly constrained minimum variance (LCMV) beamformer has been alternatively proposed by Frost (**Frost,** 1972) to compensate for sensitivity of sensor gain and delay errors from MVDR beamformer.

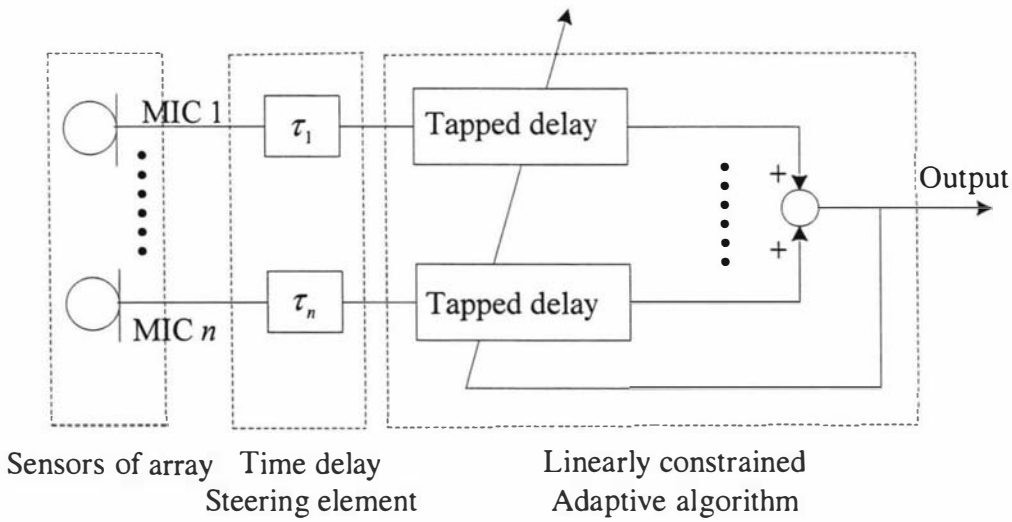


Fig. 2-14 Direct form implementation of linearly constrained adaptive array processing algorithm

The basic idea behind LCMV beamformer is to constrain the response of the beamformer so signals from the direction of interest are passed with specified gain and phase. The weights are chosen to minimize output variance or power subject to the response constraint. This has the effect of preserving the desired signal while minimizing undesirable contributions from interfering signals and noise arriving from directions other than the direction of interest. The analogous FIR filter has the weight chosen to minimize the filter output power subject to the constraint that the filter response to signals of frequency w_0 be unity.

As the beamformer response to a source at angle θ and temporal frequency w is given by $\mathbf{h}^H\mathbf{p}(\theta, w)$, thus, by linearly constraining the weights to satisfy $\mathbf{h}^H\mathbf{p}(\theta, w) = g$, where g is a complex constant, we ensure that any signal from angle θ and frequency w is passed to the output with response g.

Minimization of undesirable contributions to the output from interference (signal not arriving from θ with frequency w) is accomplished by choosing the weights to minimize the expected value of output power or variance $E[|y|^2] = \mathbf{h}^H\mathbf{R}_x\mathbf{h}$. The LCMV problem for choosing the weights is thus written as (2-45).

$$\min_{\mathbf{h}}[\mathbf{h}^H\mathbf{R}_x\mathbf{h}] \quad \text{subject to} \quad \mathbf{p}^H(\theta, w)\mathbf{h} = g^* \tag{2-45}$$

The equation (2-45) can be solved by using the method of Lagrange multipliers, which results in (2-46).

$$\mathbf{h} = g^* \frac{\mathbf{R}_X^{-1}\mathbf{p}(\theta, w)}{\mathbf{p}^H(\theta, w)\mathbf{R}_X^{-1}\mathbf{p}(\theta, w)} \tag{2-46}$$

The resulting weighting vector is often referred to as the maximum likelihood (ML) weight vector (**Hodgkiss,** 1979).

It shows that if $g = 1$, then above weight is often termed MVDR beamformer.

3) GSC

The GSC represents an alternative formulation of the LCMV problem, which essentially provides a mechanism for changing a constrained minimization problem into unconstrained form. The first application of this concept can be found in Hanson and Lawson (**Hanson and Lawson,** 1969), where a procedure for transforming constrained least squares problems to unconstrained least squares problems is given. Griffiths and Jim (**Griffiths and Jim,** 1982) has applied the same concept to LCMV beamforming and named the term GSC.

It consists of a fixed beamformer, $\mathbf{f}$, a signal blocking matrix, $\mathbf{B}_n$, and an unconstrained adaptive weight vector $\mathbf{h}_n$ as illustrated in Fig. 2-15.

Suppose we decompose the weight vector $\mathbf{h}$ into two orthogonal components $\mathbf{f}$ and $-\mathbf{y}$ that lie in the range and null space of $\mathbf{B}$, respectively. Since $\mathbf{B}^H\mathbf{y} = 0$, a fixed beamformer can be expressed as (2-47) if we assume that the constraints are as given in (2-45)

$$\mathbf{f} = g^* \frac{\mathbf{p}(\theta, w)}{\mathbf{p}^H(\theta, w)\mathbf{p}(\theta, w)} \tag{2-47}$$

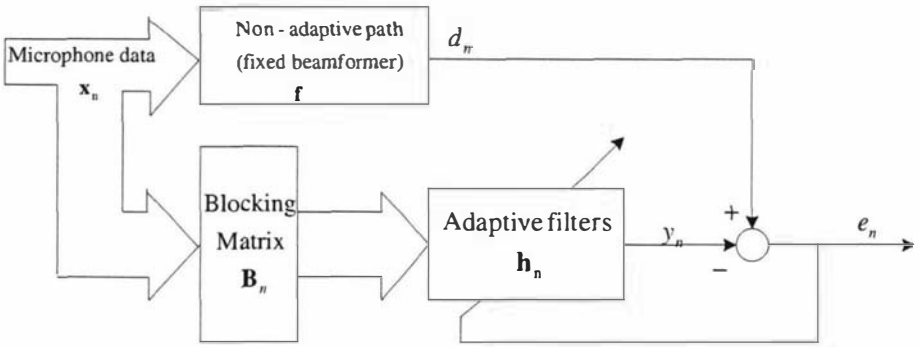


Fig. 2-15 Block diagram of the GSC

The vector $\mathbf{y}$ is a linear combination of the columns of an N by $N-L$ matrix $\mathbf{B}_n$ ($\mathbf{y} = \mathbf{B}_n\mathbf{h}_n$) provided the columns of $\mathbf{B}_n$ form a basis for the null space of $\mathbf{B}$. $\mathbf{B}_n$ can be obtained from $\mathbf{B}$ using any of several orthogonalization procedures such as Gram-Schmidt,

QR decomposition, or singular value decomposition. The weight vector $\mathbf{h} = \mathbf{f} - \mathbf{B}_n \mathbf{h}_n$ is depicted in block diagram form in Fig. 2-15. The choice for $\mathbf{f}$ and $\mathbf{B}_n$ implies that $\mathbf{h}$ satisfies the constraints independent of $\mathbf{B}_n$ and reduces the LCMV problem to the unconstrained problem.

$$\min_{\mathbf{h}_n} [\mathbf{f} - \mathbf{B}_n \mathbf{h}_n]^H \mathbf{R}_x [\mathbf{f} - \mathbf{B}_n \mathbf{h}_n]$$

The solution is

$$\mathbf{h}_n = \mathbf{f} \frac{\mathbf{B}_n^H \mathbf{R}_x}{\mathbf{B}_n^H \mathbf{R}_x \mathbf{B}_n} \tag{2-48}$$

where the weights $\mathbf{h}_n$ are unconstrained and a data independent beamformer. $\mathbf{f}$ is implemented as an integral part of the adaptive beamformer.

$\mathbf{B}_n$ satisfies $\mathbf{p}^H(\theta, w)\mathbf{B}_n = 0$ so each column $[\mathbf{B}_n]_i; 1 \le i \le N - L$, can be viewed as a data independent beamformer with a null in direction θ at frequency $w: \mathbf{p}^H(\theta, w)[\mathbf{B}_n]_i = 0$. Thus, a signal of frequency w and direction θ arriving at the array will be blocked or nulled by the matrix $\mathbf{B}_n$. In general, if the constraints are designed to present a specified response to signals from a set of directions and frequencies, then the columns of $\mathbf{B}_n$ will block those directions and frequencies. This characteristic has led to the term "blocking matrix" for $\mathbf{B}_n$. These signals are only processed by $\mathbf{f}$ and since $\mathbf{f}$ satisfies the constraints, they are presented with the desired response independent of $\mathbf{h}_n$. Signals from directions which are frequencies over which the response is not constrained, will pass through the upper branch with some response determined by $\mathbf{f}$. The lower branch chooses $\mathbf{h}_n$ to estimate the signals at the output of $\mathbf{f}$ as a linear combination of the data at the output of the blocking matrix.

■

3.4 Cepstrum and kepstrum

This section reviews a description of the fundamentals of cepstrum with a historical background and it is extended to complex cepstrum and kepstrum. It has been reviewed from the references of Oppenheim and Schafer (**Oppenheim and Schafer,** 1975), Silvia and Robinson (**Silvia and Robinson,** 1978) and Barrett and Moir (**Barrett and Moir,** 1986). Fig. 2-16 shows a historical background of cepstrum, complex cepstrum and kepstrum.

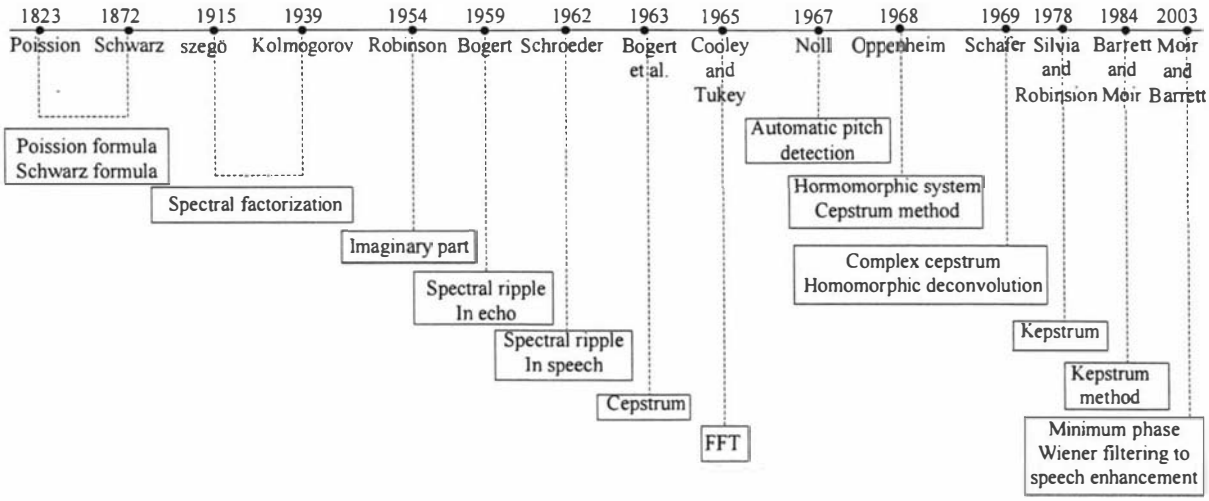


Fig. 2-16 A historical background of cepstrum, complex cepstrum and kepstrum

3.4.1 Cepstrum analysis

Temporal signals can be transformed into the frequency domain to be analyzed. But the combined signals with overlapping spectrum densities in the frequency domain may not be easily analyzed and characterized from the signal information (Fig. 2-17).

If the combined components can be separated from each other, they can be easily analyzed and so may be processed by a linear system. The cepstrum processing technique gives a solution to signals which have been convolved or multiplied in the time domain because operation by the nonlinear mapping can be processed by the generalized linear system.

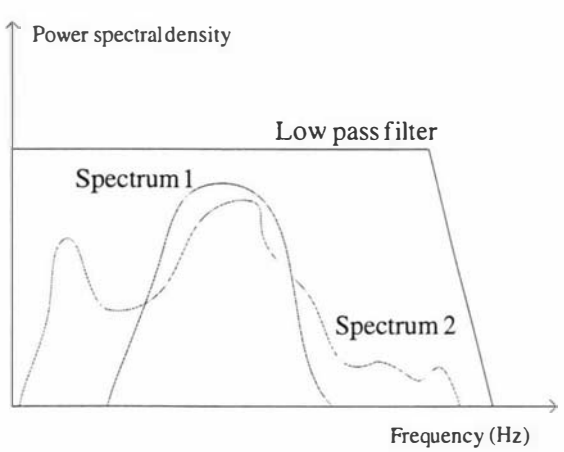


Fig. 2-17 Overlapping spectral densities

1) Cepstrum - historical background

Bogert found in 1959 that there was a spectral ripple in an echo waveform (**Noll,** 1967). Turkey realized that it came from the computation of logarithm of power spectrum and its power spectrum showed a strong peak in a time response as shown in Fig. 2-18 (**Noll,** 1967).

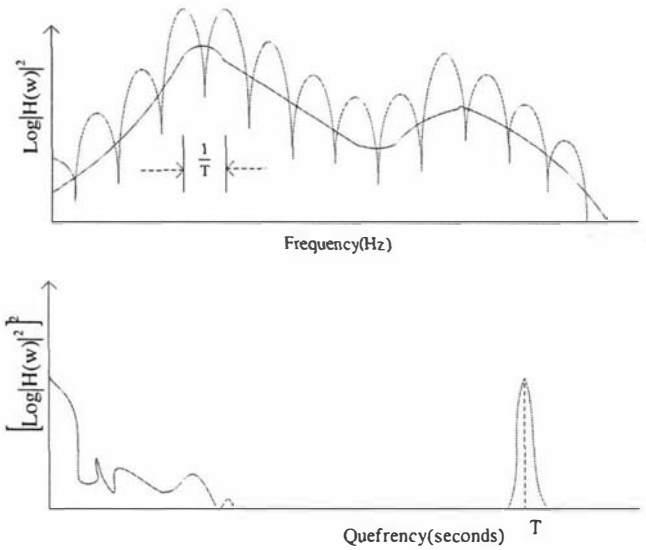


Fig. 2-18 Example of logarithmic power spectrum (top) and its power spectrum (bottom)

It provides a statistical variation about the frequency of spectral ripple. Schroeder in 1962 found that speech signals also have a spectral ripple (**Noll,** 1967). Bogert et al. (**Bogert et al.,** 1963) published a paper where they called this unusual information about frequency of spectral ripple, a cepstrum, and the frequency of the spectral ripples were referred to as quefrency, where two words, cepstrum and quefrency were prevalently being used until recent time. Cepstrum is defined as the power spectrum of the logarithm of the power spectrum so the output in its time response is called cepstrum in a 'quefrency' domain. These new words, 'cepstrum' and 'quefrency' come from the anagram of the words, 'spectrum' and 'frequency' respectively. Noll (**Noll,** 1967) introduced an automatic pitch determination method for speech analysis on a short-time processing with the use of a fast computation algorithm, a fast Fourier transform (FFT) for the discrete Fourier transform (DFT) introduced by Cooley and Tukey in 1965 (**Cooley and Tukey,** 1965).

Fundamentally, cepstrum techniques are suited to the analysis of data that contain echoes (wavelets) or reverberations of a fundamental wavelet (sometimes called signature) whose shapes need not be known a priori. Therefore, it can be effective in a pitch detection, but

ineffective by an autocorrelation function because of problems with multiple peaks or broad peaks. The result comes from the operational difference between a convolution of autocorrelation functions and an addition of cepstrum functions. It shows that the operation by a cepstrum gives a separation of signal components. But it has been realized that it is effective only in a single echo and was found to have limited application with multiple echoes.

Oppenheim and Schafer (**Oppenheim and Schafer,** 1968) have introduced the generalized linear system (homomorphic system). By the logarithmic (homomorphic) transformation of the z-transform of the observed process, convolution of two signal components is mapped into the addition of their cepstra.

Schafer (**Schafer,** 1969) has been concerned with the recovery of signals generated by a convolutional process and called his method 'homomorphic deconvolution' or 'homomorphic filtering'. This method is opposite to the cepstrum method using the determination of echo arrival times. Schafer has used the term 'complex cepstrum' to account for the fact that both the magnitude and phase spectra of the observed signal are used. The application can be found from seismic data (**Ulrych,** 1971; **Stoffa et al.,** 1974), speech (**Schafer,** 1968; **Oppenheim,** 1969; **Flanagan,** 1972) and image processing (**Oppenheim et al.,** 1968; **Gold and Rader,** 1969). Lim (**Lim,** 1979) has developed a new homomorphic deconvolution system, essentially the same as the logarithmic homomorphic deconvolution system except that the logarithmic and exponential operations are replaced with $(\cdot)^{r}$ and $(\cdot)^{\frac{1}{r}}$ operations. However, limitations have been found in a practical application due to the effect of segmentation errors on the evaluation of complex cepstrum (**Bees et al.,** 1991) and certain numerical errors associated with the use of exponential weighting (**Oppenheim and Schafer,** 1975).

For signal processing with analysis and synthesis, the phase information is needed. However, a cepstrum uses magnitude information only so it can not be inverted. On the other hand, complex cepstrum uses both magnitude and phase information so it can be inverted. Therefore, the minimum phase recovery technique from cepstrum information has been introduced and used for a homomorphic deconvolution.

The application areas are varied as follows.

- Radar and sonar (**Gold and Rader,** 1969; **Kemerait and Childers,** 1972)

- Speech processing (**Noll,** 1967; **Oppenheim and Schafer,** 1968; **Oppenheim et al.,** 1968; **Schafer,** 1968; **Gold and Rader,** 1969; **Oppenheim,** 1969; **Schafer and Rabiner,** 1970; **Rabiner et al.,** 1975; **Tribolet et al.,** 1977),

- Seismic data processing(**Bogert et al.,** 1963; **Cohen,** 1970; **Ulrych,** 1971; **Wood and Treitel,** 1975; **Oppenheim and Schafer,** 1976)

- Image processing (**Oppenheim et al.,** 1968; **Gold and Rader,** 1969) .

2) The generalized linear system (homomorphic system)

In a speech signal processing, a short segmented speech signal (Fig. 2-19) can be characterized by the output of a linear time invariant system modelled by an acoustic transfer function (**Rabiner and Schafer,** 1983). Therefore, a direct approach for signal separation can be considered by a deconvolution of the output signal.

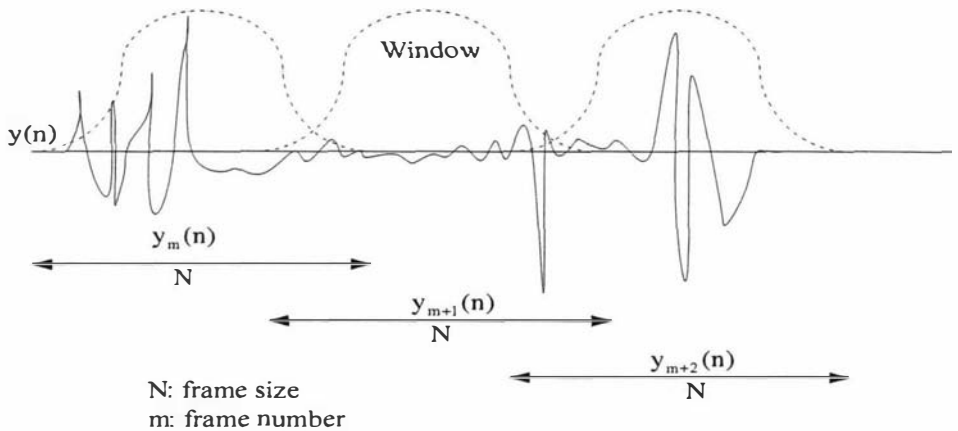


Fig. 2-19 Example of windowing

One of typical deconvolution methods in a speech signal processing is a homormorphic deconvolution and its system uses the property in a homomorphic signal processing where the desired speech component passes through the system essentially unaltered, while the undesired noise component is removed (**Oppenheim,** 1969; **Schafer and Rabiner,** 1973).

Oppenheim (**Oppenheim and Schafer,** 1968) has introduced a generalized linear system which gives an approximation to a linear system (Fig. 2-20). It shows that a linear system may be regarded as a special case of a generalized linear system for the processing of

additive signals. It shows that the multiplicative signal and the convolved signal can be processed in a logarithmic time and logarithmic frequency domain respectively. It is composed of three systems, which are called a characteristic system, a linear system and an inverse characteristic system. In a signal processing, the first part is used for a signal analysis, the second part for linear filtering and the third part for a signal synthesis.

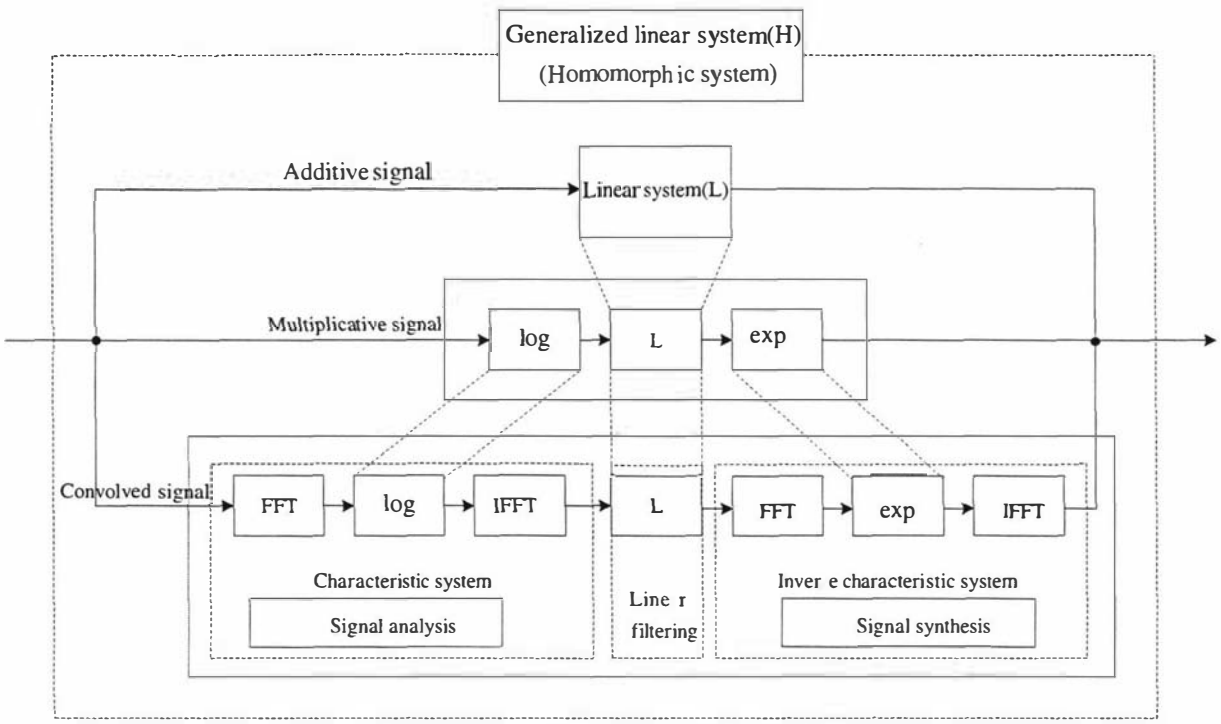


Fig. 2-20 The generalized linear system

The first part, a characteristic system is named because it can be characterized in logarithmic components in the frequency domain. Therefore, this part is used for a signal analysis and called the characteristic system for a homomorphic deconvolution. The convolved components in the time domain are transformed into additive log components in the frequency domain so resultant separated components are processed by a linear filtering.

The third part, an inverse characteristic system is named because it can be synthesized by exponential components in the frequency domain. Therefore, this part is used for a signal synthesis and called an inverse characteristic system for a homomorphic deconvolution. The processed components can be synthesized to an original desired signal by a linear filtering.

By analogy with the principle of superposition for conventional linear systems, the non linear system can be processed by a large class of generalized systems, which obey a

generalized principle of superposition where an addition is replaced by a convolution as shown in Fig. 2-21.

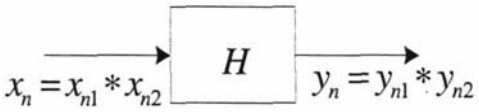


Fig. 2-21 Representation of homormophic system for convolution

That is,

$$H[x_n] = H[x_{n1} * x_{n2}] = H[x_{n1}] * H[x_{n2}] = y_{n1} * y_{n2} = y_n \quad (2\text{-}49)$$

where H represents the operator of a homomorphic system.

Systems having the property expressed by this equation are termed 'homomorphic systems for convolution'. This terminology stems from the fact that such transformations can be shown to be a homomorphic transformations in the sense of linear vector spaces.

A homomorphic filter is simply a homomorphic system having the property that one component (the desired component) passes through the system essentially unaltered, while the undesired component is removed. For example, if x_{n1} is the undesirable component, we would require that the output corresponding to x_{n1} be a unit sample, while the output corresponding to x_{n2} would closely approximate x_{n2}. This is entirely analogous to the situation with conventional linear systems where we are faced with the problem of separating a desired signal from an additive combination of signal and noise.

An important aspect of the theory of homomorphic systems is that any homomorphic system can be represented as a cascade of three homomorphic systems as depicted in Fig. 2-22.

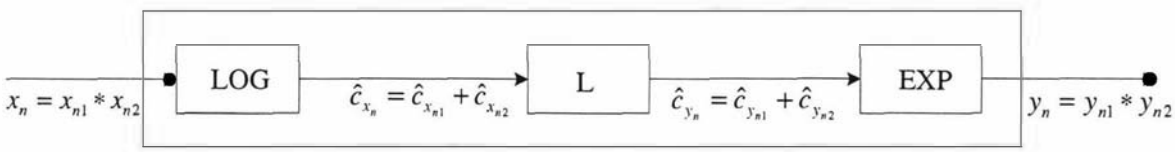


Fig. 2-22 Canonic form of system for homomophic deconvolution

The first system is called the characteristic system for homomorphic deconvolution, which obeys a generalized principle of superposition where the input operation is convolution and the output operation is ordinary addition as shown in Fig. 2-22. The internal operation of z-transform by FFT, logarithmic function and inverse z- transform by inverse fast Fourier transform (IFFT) is shown in Fig. 2-23.

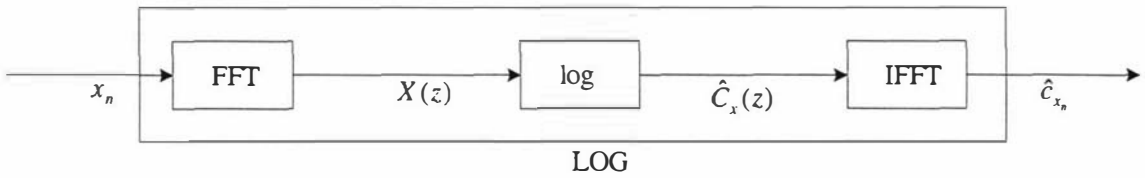


Fig. 2-23 Representation of the characteristic system for homomorphic deconvolution

This approach comes from the property that the logarithm of a product is equal to the sum of the logarithms, which is trivially true for real positive quantities. That is,

$$\hat{C}_x(z) = \log[X(z)] = \log[X_1(z)X_2(z)] = \log[X_1(z)] + \log[X_2(z)] \tag{2-50}$$

The second system is a conventional linear system obeying the principle of superposition and the third system is the inverse of the first system, i.e., it transforms signals combined by addition back into signals combined by convolution, which is shown in Fig. 2-24 by using z- transform by FFT and exponential function and then inverse z- transform by IFFT.

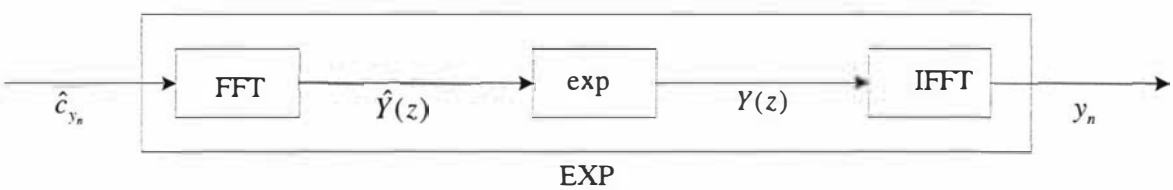


Fig. 2-24 Representation of the inverse of the characteristic system for homomorphic deconvolution

However, the z-transform is in general a complex quantity and there are important considerations of uniqueness when dealing with the logarithm of a complex number. For computational purposes we shall be primarily concerned with ensuring that $\hat{C}_x(z)$ is valid when evaluated upon the unit circle, $z = e^{jw}$. However, an appropriate definition of the complex logarithm is

$$\hat{C}_x(e^{jw}) = \log\left|X(e^{jw})\right| + j\ \arg[X(e^{jw})] \tag{2-51}$$

In this equation, the real part causes no particular difficulty. However, problems of uniqueness arise in defining the imaginary part, which is simply the phase angle of the z-transform evaluated on the unit circle. One approach to dealing with the problems of uniqueness of the phase angle is to require that the phase angle be a continuous odd function of w.

Fig. 2-25 shows a block diagram which can be used to process data in a real-time (**Childers et al.,** 1977).

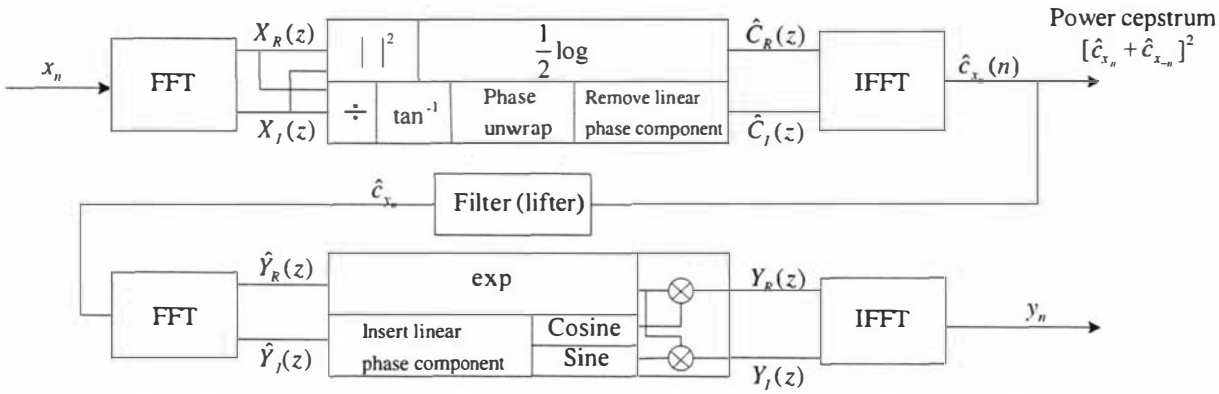


Fig. 2-25 Block diagram which can be used to process data in a real-time (**Childers et al.,** 1977)

Complex cepstrum, $\hat{c}_{x_n}$ is defined as the inverse transform of the complex logarithm of the Fourier transform of the input and it is the output of the characteristic system for a convolution, which can be written as

$$\hat{c}_{x_n} = \frac{1}{2\pi}\int_0^{2\pi}\hat{C}_x(e^{jw})e^{jwn}dw = \frac{1}{2\pi}\int_0^{2\pi}[\log|X(e^{jw})| + j\arg[X(e^{jw})]]\ e^{jwn}dw \qquad (2\text{-}52)$$

The output of the characteristic system, $\hat{c}_{x_n}$ is called the 'complex cepstrum' (the term 'complex cepstrum' implies that the complex logarithm is involved).

The term 'cepstrum' can be written as

$$c_{x_n} = \frac{1}{2\pi}\int_0^{2\pi}\log|X(e^{jw})|e^{jwn}dw \qquad (2\text{-}53)$$

where the sequence c_{x_n} can be shown to be equal to the even part of the complex cepstrum $\hat{c}_{x_n}$.

3.4.2 Kepstrum analysis

The idea of the kepstrum comes from the classical works on a potential theory by Poisson (**Poisson,** 1823) and Schwarz (**Schwarz,** 1872) for solving the problem of determining a potential function whose a real part is an assigned known value on the unit circle. It has been applied to echo detection, geophysics, speech processing and vibration signal analysis etc. by Robinson (**Robinson,** 1954), Bogert et al.(**Bogert et al.,** 1963), Schafer (**Schafer,** 1969), Oppenheim and Schafer (**Oppenheim and Schafer,** 1975), Tribolet (**Tribolet,** 1977), Silvia and Robinson (**Silvia and Robinson,** 1978) and others.

1) Word definition: cepstrum, complex cepstrum and kepstrum

The word 'cepstrum', which is an anagram of the word 'spectrum', has been first introduced and defined as power spectrum of logarithm of power spectrum by Bogert et al. (**Bogert et al.,** 1963). The word 'complex cepstrum' has been used as an extended word from cepstrum by Schafer (**Schafer,** 1969) in 1969. However, nowadays many texts state that the cepstrum is defined as the inverse Fourier transform of the logarithm of the spectrum. On the other hand, the word 'kepstrum', quite similar to the word 'cepstrum', has been coined by Silvia and Robinson (**Silvia and Robinson,** 1978) in 1978. It comes from the first letters of the *Kolmogorov equation power series time response*, and then by adding the Latin singular ending 'um' to denote one kepstrum. Its plural word 'kepstra' is also used to denote more than one kepstrum by adding the Latin plural ending 'a'. The complex cepstrum of a real sequence is also a real sequence because of its symmetry property. Therefore, the meaning of the word 'complex cepstrum' may be confused with the real-valued 'complex cepstrum', because 'complex' from complex cepstrum actually comes from the use of both magnitude and phase spectrum information in logarithm frequency domain. Therefore, the word 'kepstrum' has been used in the thesis throughout because it is the operation of a real sequence because of symmetry property and it comes from the Kolmogorov's fundamental work, which gives a mathematical construct.

2) Kepstrum analysis of a minimum phase transfer function

From the Schwarz's classical expression (**Schwarz,** 1872; **Silvia and Robinson,** 1978),

$$H_+(z)=\frac{1}{2\pi}\int_0^{2\pi}H_R(\lambda)\left(\frac{1+z^{-1}e^{j\lambda}}{1-z^{-1}e^{j\lambda}}\right)d\lambda, \qquad |z|<1 \tag{2-54}$$

where λ as the integration variable and $z = re^{jw}$. This equation gives the a potential function $H_+(z)$ whose a real part on the unit circle is $H_R(w)$. The Schwarz expression follows from the famous result of Poisson on potential theory. An interesting result arises from the logarithm of a potential function because the application of logarithm to a potential function gives changes to the minimum phase property. Therefore, it may be expressed as a logarithm of a minimum phase transfer function, $\log H_M(z)$.

$$\log H_M(z) = \log|H_M(z)| + j \arg H_M(z) \tag{2-55}$$

The real part of the logarithmic potential evaluated on the unit circle $z = e^{jw}$ is then $\log|H_M(z = e^{jw})| = \log|H_M(w)|$. In terms of the logarithmic potential, Schwarz's expression becomes

$$\log H_M(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|H_M(\lambda)| \left(\frac{1+z^{-1}e^{j\lambda}}{1-z^{-1}e^{j\lambda}}\right) d\lambda, \quad |z| < 1 \tag{2-56}$$

where $\log|H_M(w)|$ is the assigned or known value on the unit circle and the function $H_M(z)$ has zeroes or poles inside of the unit circle, i.e., $H_M(z)$ is a minimum phase and the z-transform of a sequence h_n, is defined by $H_{\pm}(z) = \sum_{n=-\infty}^{\infty} h_n z^{-n}$. As a result, $\log H_M(z)$ has a Taylor-Maclaurin series expansion in the unit circle.

$$\log H_M(z) = \sum_{n=0}^{\infty} k_n z^{-n}, \quad |z| < 1 \tag{2-57}$$

The coefficients k_n of this Taylor-Maclaurin series are the coefficients of a stable causal system. These coefficients are called the kepstrum of the minimum phase system $H_M(z)$ (**Silvia and Robinson,** 1978).

The complex variable z represents a point within the unit circle (i.e., $|z| < 1$). However, we can consider the limit of $\log H_M(z)$ as the interior point, z approaches a point on the circumference of the unit circle. That is, we consider the limit of $\log H_M(z)$ as z approaches to e^{jw}.

This limit is

$$\log H_M(w) = \sum_{n=0}^{\infty} k_n e^{-jnw} \tag{2-58}$$

Separating the above into real and imaginary part, we obtain

$$\log H_M(w) = \log|H_M(w)| + j\theta_M(w) = \left(k_0 + \mathrm{Re}\left[\sum_{n=1}^{\infty} k_n e^{-jnw}\right]\right) + j\left(\mathrm{Im}\left[\sum_{n=1}^{\infty} k_n e^{-jnw}\right]\right) \quad (2\text{-}59)$$

By equating real part, it can be shown that

$$k_n = \begin{cases} k_n = 2c_n & for\ n \geq 1 \\ k_0 = c_n & for\ n = 0 \\ k_{-n}^{*} = 2c_n & for\ n \leq -1 \end{cases} \quad (2\text{-}60)$$

where $c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|H_M(w)| e^{jnw} dw$ for $n = 0, \pm 1, \pm 2, ...$ and '*' denotes the complex conjugate.

By equating the imaginary parts, it can be shown that (**Silvia and Robinson,** 1978)

$$\theta_M(w) = \frac{-1}{2\pi} P \int_{-\pi}^{\pi} \log|H_M(\lambda)| \cot\left(\frac{w-\lambda}{2}\right) d\lambda \quad (2\text{-}61)$$

where the symbol P denotes that the integral has its Cauchy principal value. The kepstrum can be used as an intermediate step in deriving the above equation, which shows that for a minimum phase system $H_M(z)$, knowledge of only the log magnitude spectrum $\log|H_M(w)|$ is sufficient to determine the phase spectrum $\theta_M(w)$.

It can also be analyzed by a mathematical representation from a minimum phase causal kepstrum as follows.

Defining $z = e^{jw}$,

$$K_M(e^{jw}) = \log H_M(e^{jw}) = \log|H_M(e^{jw})| + j\arg[\log H_M(e^{jw})] \quad (2\text{-}62)$$

$$= k_0 + \mathrm{Re}\left(\sum_{n=1}^{\infty} k_n e^{-jnw}\right) + j\mathrm{Im}\left(\sum_{n=1}^{\infty} k_n e^{-jnw}\right)$$

where $\log|H_M(e^{jw})| = \left[k_0 + \mathrm{Re}\left(\sum_{n=1}^{\infty} k_n e^{-jnw}\right)\right]$ and $\arg[H_M(e^{jw})] = \left[\mathrm{Im}\left(\sum_{n=1}^{\infty} k_n e^{-jnw}\right)\right]$

In an exponential form,

$$K_M(e^{jw}) = k_0 + k_1 e^{-jw} + k_2 e^{-j2w} + ... \quad (2\text{-}63)$$

$$= k_0 + k_1(\cos w - j\sin w) + k_2(\cos 2w - j\sin 2w) + ...$$

$$= k_0 + k_1 \cos w + k_2 \cos 2w + ... + j[-(k_1 \sin w + k_2 \sin 2w + ...)]$$

$$= k_0 + \sum_{n=1}^{\infty} k_n \cos nw + j\left(- \sum_{n=1}^{\infty} k_n \sin nw \right)$$

where $\log\left|H_M(e^{jw})\right| = k_0 + k_1 \cos w + k_2 \cos 2w + ... = k_0 + \sum_{n=1}^{\infty} k_n \cos nw$

$$\arg\left[H_M(e^{jw})\right] = -(k_1 \sin w + k_2 \sin 2w + ...) = j\left(- \sum_{n=1}^{\infty} k_n \sin nw \right)$$

For the N-point discrete form,

The minimum phase kepstrum $K_M\left(\frac{2\pi}{N}k\right)$ is

$$K_M\left(\frac{2\pi}{N}k\right) = \log H_M\left(\frac{2\pi}{N}k\right) = \log\left|H_M(e^{j\frac{2\pi}{N}k})\right| + \mathrm{j}\arg H_M(e^{j\frac{2\pi}{N}k}) \tag{2-64}$$

From (2-62),

$$K_M\left(\frac{2\pi}{N}k\right) = \sum_{n=0}^{N-1}\left(k_n \cos\frac{2\pi}{N}kn\right) + j\left(-\sum_{n=0}^{N-1} k_n \sin\frac{2\pi}{N}kn\right) \tag{2-65}$$

$$= k_0 + \mathrm{Re}\left[\log H_M\left(\frac{2\pi}{N}k\right)\right] + j\mathrm{Im}\left[\log H_M\left(\frac{2\pi}{N}k\right)\right]$$

$$= k_0 + \sum_{n=1}^{N-1}\left(\frac{1}{2}k_{-n}e^{j\frac{2\pi}{N}kn} + \frac{1}{2}k_n e^{-j\frac{2\pi}{N}kn}\right) + j\left[\sum_{n=1}^{N-1} -\left(\frac{1}{2}k_{-n}e^{j\frac{2\pi}{N}kn} - \frac{1}{2}k_n e^{-j\frac{2\pi}{N}kn}\right)\right]$$

where the magnitude of the minimum phase kepstrum is

$$\log\left|H_M(e^{j\frac{2\pi}{N}k})\right| = \sum_{n=0}^{N-1} k_n \cos\frac{2\pi}{N}kn = k_0 + \frac{1}{2}\left[\sum_{n=1}^{N-1}\left(k_{-n}e^{j\frac{2\pi}{N}kn} + k_n e^{-j\frac{2\pi}{N}kn}\right)\right]$$

$$= k_0 + \sum_{n=1}^{N-1}\left(\frac{1}{2}k_{-n}e^{j\frac{2\pi}{N}kn} + \frac{1}{2}k_n e^{-j\frac{2\pi}{N}kn}\right)$$

and the phase of the minimum phase kepstrum is

$$\arg\left[H_M(e^{j\frac{2\pi}{N}k})\right] = -j\sum_{n=0}^{N-1}\left(k_n \sin\frac{2\pi}{N}k_n\right) = j\left[\sum_{n=1}^{N-1} -\left(\frac{1}{2}k_{-n}e^{j\frac{2\pi}{N}kn} - \frac{1}{2}k_n e^{-j\frac{2\pi}{N}kn}\right)\right]$$

For the N-point discrete form,

$$k_0 = \frac{1}{N}\sum_{k=0}^{N-1} \log\left|H_M(e^{j\frac{2\pi}{N}k})\right| \tag{2-66}$$

$$k_n = \frac{2}{N}\sum_{k=0}^{N-1} \log\left|H_M(e^{j\frac{2\pi}{N}k})\right| \cos\frac{2\pi}{N}kn \tag{2-67}$$

$$= \frac{1}{N}\sum_{k=0}^{N-1}\left(\log\left|H_M(e^{j\frac{2\pi}{N}k})\right| e^{j\frac{2\pi}{N}kn} + \log\left|H_M(e^{j\frac{2\pi}{N}k})\right| e^{-j\frac{2\pi}{N}kn} \right)$$

3.4.3. Analysis of causal kepstrum

Any causal signals (h_{+n}) with a minimum phase or non minimum phase can be expressed as a sum of even function (h_e) and odd function (h_o).

$$h_{+n} = h_e + h_o \tag{2-68}$$

So if we know the even function of the signal, it can be recovered by multiplying it two times in a positive time series except a zeroth coefficient (Fig. 2-26 (A)).

$$h_n = \begin{cases} 2h_e & for\ n \geq 1 \\ h_e & for\ n = 0 \\ 0 & for\ n \leq -1 \end{cases} \tag{2-69}$$

On the other hand, only a minimum phase signal (h_M) can be expressed as the sum of even and odd functions in a kepstrum time series (Fig. 2-26 (B)).

$$k_n = \begin{cases} 2k_e & for\ n \geq 1 \\ k_e & for\ n = 0 \\ 0 & for\ n \leq -1 \end{cases} \tag{2-70}$$

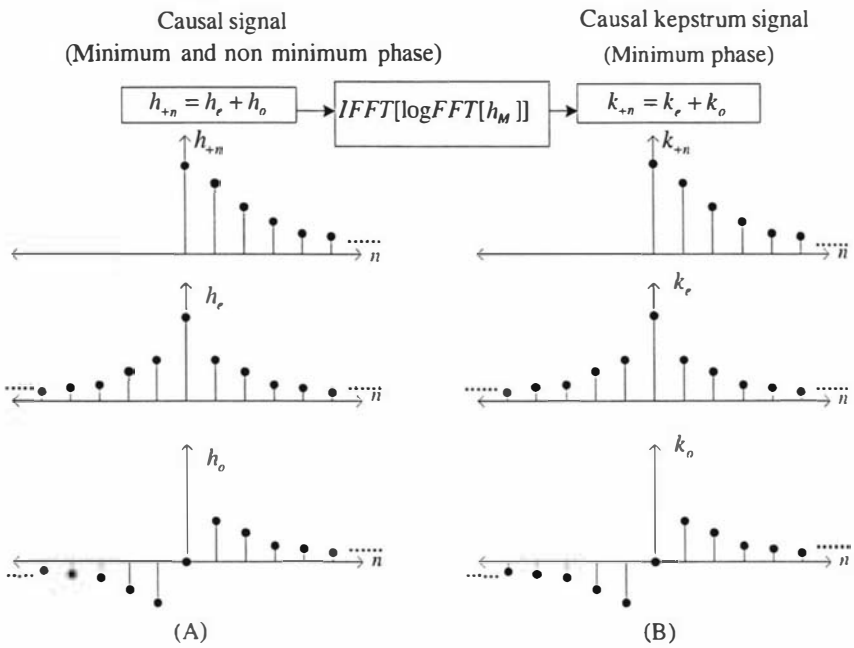


Fig. 2-26 Causal signal (A) and minimum phase kepstrum signal (B)

1) Minimum phase computation

Minimum phase term can be calculated from a Hilbert transform directly or kepstrum coefficients indirectly.

i) Direct computation by Hilbert transform relation

It can be calculated directly from Hilbert transform relation, which has been developed from Schwarz formula in 1872 (**Schwarz,** 1872).

From the description from even function only,

$$H_+(z) = \sum_{n=0}^{\infty} h_n z^{-n} = h_0^e + 2\sum_{n=1}^{\infty} h_n^e z^{-n} \tag{2-71}$$

where h_0^e and h_n^e are zeroth and n-th coefficients of even function respectively.

By using that $h_n^e = \frac{1}{2\pi}\int_0^{2\pi} H_R(w) e^{jnw} dw$ (2-72)

$$H_+(z) = \frac{1}{2\pi}\int_0^{2\pi} H_R(\lambda) d\lambda + 2\sum_{n=1}^{\infty}\left(\frac{1}{2\pi}\int_0^{2\pi} H_R(\lambda) e^{jn\lambda} d\lambda\right) z^{-n} \tag{2-73}$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\left(1 + 2\sum_{n=1}^{\infty} e^{jn\lambda} z^{-n}\right) H_R(\lambda) d\lambda$$

$$\text{where } 1+2\sum_{n=1}^{\infty}(e^{j\lambda}z^{-1})^n=\frac{1+e^{j\lambda}z^{-1}}{1-e^{j\lambda}z^{-1}} \quad for \ |z|<1 \tag{2-74}$$

Thus we have Schwarz expression

$$H_+(z)=\frac{1}{2\pi}\int_0^{2\pi}(\frac{1+e^{j\lambda}z^{-1}}{1-e^{j\lambda}z^{-1}})H_R(\lambda)d\lambda \tag{2-75}$$

In order to obtain a direct relationship between the real part and imaginary part on the unit circle,

$$h_{+n}=2h_n^e u_n-h_0^e \tag{2-76}$$

$$H_M(w)=H_R(w)+jH_I(w)=\frac{1}{\pi}\int_0^{2\pi}H_R(\lambda)U(w-\lambda)d\lambda-h_0^e \tag{2-77}$$

$$\text{where } U(w)=\pi\delta(w)+\frac{1}{1-e^{-jw}} \tag{2-78}$$

$$=\pi\delta(w)+\frac{1}{2}-j\frac{1}{2}\cot\frac{w}{2} \quad -\pi\le w\le\pi$$

Now we obtain the relation between $H_R(w)$ and $H_I(w)$,

$$H_I(w)=-\frac{1}{2\pi}\int_0^{2\pi}H_R(\lambda)\cot\left(\frac{w-\lambda}{2}\right)d\lambda \tag{2-79}$$

This integral is called Hilbert transform.

Now, for the application of kepstrum,

$$K_M(w)=\log H_M(w)=K_R(w)+jK_I(w) \tag{2-80}$$

$$=\log|H_M(w)|+j\theta_M(w)=\left(k_0+\mathrm{Re}\left[\sum_{n=1}^{\infty}k_n e^{-jnw}\right]\right)+j\left(\mathrm{Im}\left[\sum_{n=1}^{\infty}k_n e^{-jnw}\right]\right)$$

$$\theta_M(w)=-\frac{1}{2\pi}\int_0^{2\pi}\log|H_M(w)|\cot\left(\frac{w-\lambda}{2}\right)d\lambda \tag{2-81}$$

ii) Indirect computation by kepstrum coefficient

ii-1) By transfer function

For the minimum phase computation, the kepstrum coefficient can be used as an intermediate step to calculate the phase information. The procedure is shown as follows.

$$K_M(w) = \sum_{n=0}^{\infty} k_n e^{-jnw} = \log H_M(e^{jw}) = \log\left|H_M(e^{jw})\right| + j\arg H_M(e^{jw}) \quad (2\text{-}82)$$

where

$$\log\left|H_M(e^{jw})\right| = (k_0 + k_1\cos w + k_2\cos 2w + ...) \quad (2\text{-}83)$$

$$\arg H_M(e^{jw}) = -(k_1\sin w + k_2\sin 2w + ...) \quad (2\text{-}84)$$

From the cosine series for log magnitude we have

$$k_0 = \frac{1}{2\pi}\int_0^{2\pi} \log\left|H_M(e^{jw})\right| dw \quad (2\text{-}85)$$

$$k_n = \frac{1}{\pi}\int_0^{2\pi} \log\left|H_M(e^{jw})\right| \cos nw dw \quad (2\text{-}86)$$

Using these coefficients the phase can be computed from the sine series.

ii-2) By spectral factor

By spectral factor from the knowledge of power spectrum,

$$\log\left|H^+(e^{jw})\right| = \frac{1}{2}\log\Phi(w) \quad (2\text{-}87)$$

From this, the corresponding missing phase information $\arg H^+(e^{jw})$ may be reconstructed by means of the Fourier series method. Using equation (2-87),

$$k_0 = \frac{1}{4\pi}\int_0^{2\pi} \log\Phi(w) dw = \frac{1}{2\pi}\int_0^{2\pi} \log\left|H^+(e^{jw})\right| dw \quad (2\text{-}88)$$

$$k_n = \frac{1}{2\pi}\int_0^{2\pi} \log\Phi(w)\cos nw dw = \frac{1}{\pi}\int_0^{2\pi} \log\left|H^+(e^{jw})\right| \cos nw dw \quad (2\text{-}89)$$

It gives the same result with equations (2-85) and (2-86) by transfer function.
Therefore, using these coefficients the phase can be computed from the sine series.

2) Pole-zero analysis for causal kepstrum signal

Any transfer functions can be specified by numerator polynomial of zeros and denominator polynomial of poles with a gain factor and may be resolved as its polynomial factorizations.

$$H(z) = \frac{A_0 + A_1 z^{-1} + A_2 z^{-2} + ...A_m z^{-m}}{B_0 + B_1 z^{-1} + B_2 z^{-2} + ...B_n z^{-n}} \quad (2\text{-}90)$$

$$= \frac{A_0(1+\frac{A_1}{A_0}z^{-1}+\frac{A_2}{A_0}z^{-2}+....+\frac{A_m}{A_0}z^{-m})}{B_0(1+\frac{B_1}{B_0}z^{-1}+\frac{B_2}{B_0}z^{-2}+....+\frac{B_n}{B_0}z^{-n})} = \frac{A_0}{B_0}(\frac{1+a_1z^{-1}+a_2z^{-2}+....+a_mz^{-m}}{1+b_1z^{-1}+b_2z^{-2}+....+b_nz^{-n}})$$

$$= C_0(\frac{1+a_1z^{-1}+a_2z^{-2}+....+a_mz^{-m}}{1+b_1z^{-1}+b_2z^{-2}+....+b_nz^{-n}}) = C_0\frac{(1-z_1z^{-1})(1-z_2z^{-1}).....(1-z_mz^{-1})}{(1-p_1z^{-1})(1-p_2z^{-1}).....(1-p_nz^{-1})}$$

From the polynomial factorization (2-90), its logarithm of transfer function $\log H(z)$ is shown as a separate terms, i.e., constant term, positive and negative exponential terms (2-91).

$$\log H(z) = \log(C_0) + \log\left\{\frac{(1-z_1z^{-1})(1-z_2z^{-1}).....(1-z_mz^{-1})}{(1-p_1z^{-1})(1-p_2z^{-1}).....(1-p_nz^{-1})}\right\} \tag{2-91}$$

$$= \log(C_0) + \log(1-z_1z^{-1}) + \log(1-z_2z^{-1}) + ... + \log(1-z_mz^{-1})$$

$$-\log(1-p_1z^{-1}) - \log(1-p_2z^{-1}) - ..... - \log(1-p_nz^{-1})$$

$$= \log(C_0) + \sum_{k=1}^{m}\log(1-z_kz^{-1}) - \sum_{k=1}^{n}\log(1-p_kz^{-1})$$

$$= \log(C_0) - \sum_{k=1}^{m}\log\left(\frac{1}{1-z_kz^{-1}}\right) + \sum_{k=1}^{n}\log\left(\frac{1}{1-p_kz^{-1}}\right)$$

where $\log\left(\frac{1}{1-h}\right) = h + \frac{h^2}{2} + \frac{h^3}{3} + ... + \frac{h^k}{k} + ..., = \sum_{n=1}^{\infty}\frac{1}{n}h^n \qquad |h|<1$

so $\log\left(\frac{1}{1-z_kz^{-1}}\right) = \sum_{n=1}^{\infty}\frac{1}{n}(z_kz^{-1})^n, \quad |z|>|z_k|$ (2-92)

$$\log\left(\frac{1}{1-p_kz^{-1}}\right) = \sum_{n=1}^{\infty}\frac{1}{n}(p_kz^{-1})^n, \quad |z|>|p_k| \tag{2-93}$$

When all poles and zeros are inside the unit circle, the complex cepstrum is causal and can be expressed simply in terms of the filter poles and zeros as (2-94).

$$h_n = \begin{cases} \log(C_0), & n=0 \\ \sum_{k=1}^{N}\frac{1}{n}p_k^n - \sum_{k=1}^{M}\frac{1}{n}z_k^n & n=1,2,3,..... \end{cases} \tag{2-94}$$

where N is the number of poles and M is the number of zeros. Note that when $N > M$, there are really $N-M$ additional zeros at $z=0$, but these contribute zero to the complex cepstrum. When $M > N$, there are $M-N$ additional poles at $z=0$ which also contribute zero.

In summary, each stable pole contributes a positive decaying exponential (weighted by $1/n$) to the complex cepstrum, while each zero inside the unit circle contributes a negative weighted exponential of the same type. On the other hand, poles and zeros outside the unit circle contribute anticausal exponentials to the complex cepstrum, negative for the poles and positive for the zeros.

However, any spectrum can be converted to minimum phase form without affecting the spectral magnitude by computing its cepstrum and replacing any anticausal components with corresponding causal components. It can be done by reversing the anticausal part cepstrum to the positive part of time zero so that it adds to the causal part. This corresponds to reflecting non minimum phase zeros inside the unit circle that preserves spectral magnitude. The original spectral phase is then replaced by the unique minimum phase corresponding to the given spectral magnitude.

3) Reflection of phase zeros by reciprocal polynomial

Any non minimum phase signal can be transformed into minimum phase signal by using reciprocal polynomial in the z-transform domain. The example shows that one non minimum phase zero outside of unit circle is reflected into the unit circle with same spectral magnitude (Fig. 2-27).

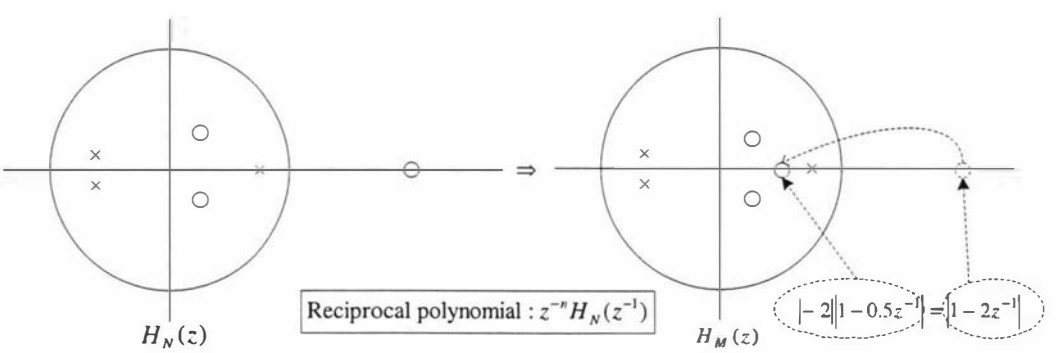


Fig. 2-27 Phase zeros reflection into minimum phase

On the other hand, it can also be transformed into non minimum phase signal by using reciprocal polynomial in z-transform domain. The example shows that one minimum phase zero inside of unit circle is reflected to the outside of unit circle with same spectral magnitude (Fig. 2-28).

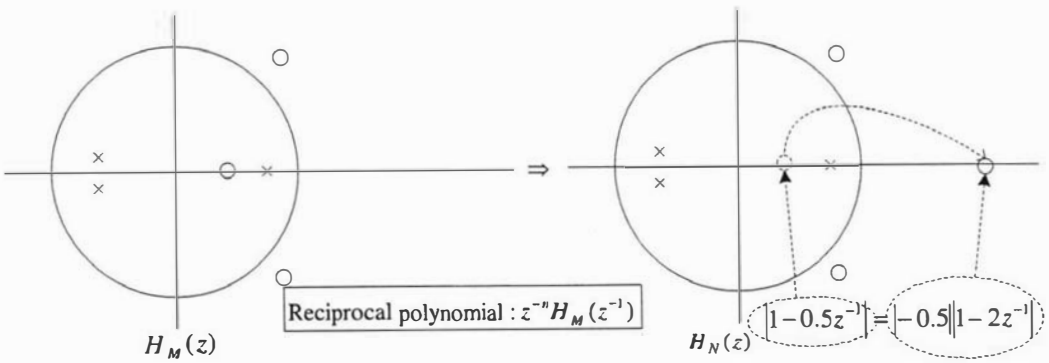


Fig. 2-28 Phase zeros reflection into non minimum phase

■

4. Adaptive algorithms

The adaptive algorithms, LMS, NLMS and RLS are reviewed. It is well known from the references of Haykin (**Haykin,** 1996) and Widrow and Stearns (**Widrow and Stearns,** 1985).

4.1 LMS algorithm

The operation of an adaptive filter is mainly composed of two processes, the adaptive and filtering process. The former involves the automatic adjustment of the tap weights of the filter in accordance with an adaptive algorithm. The latter involves 1) a multiplication of the tap inputs with the corresponding set of tap weights resulting from the adaptive process to produce an estimate of the desired response, and 2) a generation of an estimation error by comparing this estimate with the actual value of the desired responses. The estimation error is then used to actuate the adaptive process, thereby closing the feedback loop.

Fig. 2-29 shows the linear transversal filter in a block diagram form. The filter output $\hat{y}_n$ is expressed in the convolution form as (2-95)

$$\hat{y}_n = \sum_{k=1}^{P} h_k^* x_{n-k+1} \tag{2-95}$$

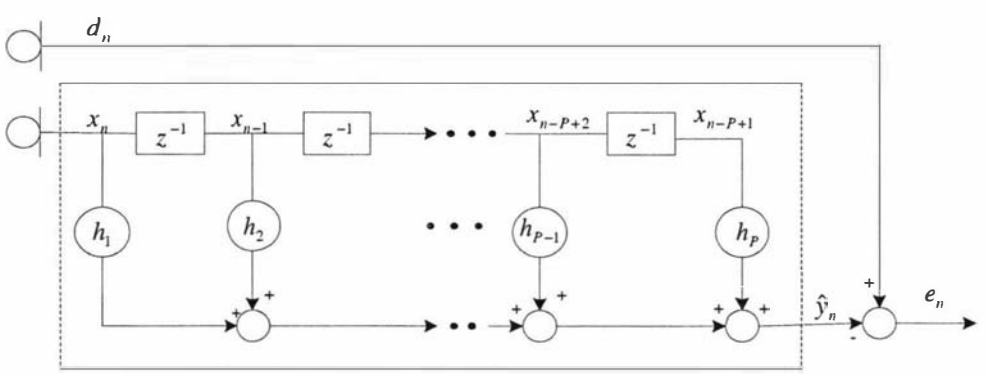


Fig. 2-29 Transversal filter

The estimation error e_n is

$$e_n = d_n - \hat{y}_n \tag{2-96}$$

The mean square value of the estimation error (the mean square error), $J(h)$ is written as (2-97).

$$J(h) = E[e_n\, e_n^*] \tag{2-97}$$

The mean square error is a real and positive scalar. By minimizing $J(h)$, we obtain the optimum linear filter in the minimum mean square sense.

The composition of the $P-\text{by}-1$ tap weight vector $\mathbf{h}_n$ of the filter at time n and that of the $P-\text{by}-1$ input vector $\mathbf{x}_n$ at time n, respectively are shown by

$$\mathrm{h}_n^T = [h_1, h_2, \ldots\ldots h_P] \tag{2-98}$$

$$\mathrm{x}_n^T = [x_n, x_{n-1}, \ldots\ldots, x_{n-P+1}] \tag{2-99}$$

where the superscript T signifies transposition. We may rewrite in the form of an inner product of vectors as

$$\hat{y}_n = \mathbf{h}_n^H \mathbf{x}_n \text{ or equivalently, } \hat{y}_n^* = \mathbf{x}_n^H \mathbf{h}_n \tag{2-100}$$

where the superscript H signifies Hermitian transposition.

Accordingly, we may express the estimation error as $e_n = d_n - \mathbf{h}_n^H \mathbf{x}_n$ or, in complex conjugated form, as $e_n^* = d_n^* - \mathbf{x}_n^H \mathbf{h}_n$.

Assuming that the tap input vector $\mathbf{x}_n$ and the desired response d_n are jointly stationary, mean square error $J(w)$ is precisely a second order function of the tap weight vector $\mathbf{h}$.

$$\begin{aligned} J(h) &= E[e_n e_n^*] = E[(d_n - \mathbf{h}_n^H \mathbf{x}_n)(d_n^* - \mathbf{x}_n^H \mathbf{h}_n)] \\ &= \sigma_d^2 - \mathbf{P}^H \mathbf{h}_n - \mathbf{h}_n^H \mathbf{P} + \mathbf{h}_n^H \mathbf{R} \mathbf{h}_n \end{aligned} \tag{2-101}$$

where σ_d^2 is $E[d_n d_n^*]$, the va riance of the desired response d_n assuming zero mean. $\mathbf{R}$ is $E[\mathbf{x}_n \mathbf{x}_n^H]$, autocorrelation vector of tap input vector $\mathbf{x}_n$. $\mathbf{P}$ is $E[\mathbf{x}_n d_n^*]$, cross-correlation vector between the tap input vector $\mathbf{x}_n$ and the desired response d_n.

The requirement is to design the filter so that it operates at the minimum point of the error performance surface (denoted by $J_{\min}$) making the transversal filter optimum in the minimum mean square sense

To determine the optimum tap weight vector $\mathbf{h}$, we first differentiate the mean square error $J(h)$ with respect to the tap weight vector $\mathbf{h}$, called gradient vector and denoted by ∇, and then set the result equal to zero as follows.

$$\nabla = \frac{dJ(h)}{dh} = -2\mathbf{P} + 2\mathbf{R}\mathbf{h} \tag{2-102}$$

Now, we get the discrete form of the Wiener-Hopf equation, which is also called the normal equation as (2-103).

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{P} \tag{2-103}$$

From the above equation, the computation of the optimum tap weight vector $\mathbf{h}$ requires knowledge of two quantities, the correlation matrix $\mathbf{R}$ of the tap input vector $\mathbf{x}_n$ and the cross correlation vector $\mathbf{P}$ between the tap input $\mathbf{x}_n$ and the desired response d_n.

Let ∇_n denote the value of the gradient vector at time n and $\mathbf{h}_n$ denotes the value of the tap weight vector at time n.

According to the method of steepest descent, the updated value of the tap weight vector at time $n+1$ is computed by using the simple recursive relation

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu[-\nabla_n] = \mathbf{h}_n + 2\mu[\mathbf{P} - \mathbf{R}\mathbf{h}_n], \qquad n = 0, 1, 2, \ldots\ldots, \tag{2-104}$$

where μ is a positive real valued constant.

We call (2-104) the steepest descent algorithm and μ is the step-size parameter or weighting constant.

To develop an estimate of the gradient vector $\nabla_n = -2\mathbf{P} + 2\mathbf{R}\mathbf{h}_n$, the instantaneous estimates of $\mathbf{R}$ and $\mathbf{P}$, based on sample values of the tap input vector and desired response are used.

$$\hat{\mathbf{R}}_n = \mathbf{x}_n\mathbf{x}_n^H \tag{2-105}$$

$$\hat{\mathbf{P}}_n = \mathbf{x}_n d_n^* \tag{2-106}$$

Correspondingly, the instantaneous estimate of the gradient vector is

$$\hat{\nabla}_n = -2\mathbf{x}_n d_n^* + 2\mathbf{x}_n\mathbf{x}_n^H\hat{\mathbf{h}}_n \tag{2-107}$$

Note also that the estimate $\hat{\nabla}_n$ equals the gradient of the instantaneous square error $|e_n|^2$. Substituting the estimate of above equation (2-107) in the steepest descent algorithm (2-104), we get a new recursive relation for updating the tap weight vector.

$$\hat{\mathbf{h}}_{n+1} = \hat{\mathbf{h}}_n + 2\mu\mathbf{x}_n[d_n^* - \mathbf{x}_n^H\hat{\mathbf{h}}_n] = \hat{\mathbf{h}}_n + 2\mu\mathbf{x}_n e_n^* \tag{2-108}$$

where $e_n = d_n - \hat{\mathbf{h}}_n^H\mathbf{x}_n$ and e_n defines the estimation error and note that the second term $2\mu\mathbf{x}_n e_n^*$ represents a correction that is applied to the current estimate of the tap weight vector, $\hat{\mathbf{h}}_n$ and the iterative procedure is started with the initial guess , $\hat{\mathbf{h}}_0$.

The above equation (2-108) is called the adaptive LMS algorithm, which was proposed by Widrow and Hoff in 1960 (**Widrow and Hoff,** 1960) as one of the stochastic gradient algorithms.

It is widely used as the simplest algorithm through various applications because of the non-requirement of correlation function and matrix inversion.

The major limitations are a relatively slow rate of convergence and sensitivity to variations in the eigenvalue spread, defined as the ratio of the maximum to minimum eigenvalues of the correlation matrix of the tap inputs.

For the computational complexity, LMS algorithm with N real weights requires N multiplications to compute its output and the other N multiplications to update the weight vector. Therefore, a total of $2N$ real multiplications are needed to produce each output sample. Thus $2N^2$ real multiplications are required for every N output samples (**Shynk,** 1992). In addition, LMS algorithm requires approximately $20N$ iterations to converge in mean square, where N is the number of tap coefficients contained in the tapped-delay-line filter (**Haykin,** 1996). The LMS algorithm always exhibits a nonzero misadjustment. However, this misadjustment may be made arbitrarily small by using a sufficiently small step size parameter μ .

■

4.2 NLMS (normalized least mean square) algorithm

It is well known from Haykin (**Haykin,** 1996) that the performance of LMS may be evaluated in terms of convergence rate and stability and it is directly related to the proper selection of step size (μ). It is well known that a compromising effect takes place when μ is too large resulting in faster convergence but less stability. On the other hand, if μ is smaller, then we get slower convergence but greater stability. Therefore conventional LMS has difficulties when applied to applications in a real-time processing in a real environment since a speech by its very nature has a wide dynamic range and may give rise to stability in quiet utterances and conversely instability when the utterances are louder or when non stationary noise is added. Increasing the step size only serves to make the risk of instability even greater whereas a conservative estimate can have slow convergence.

It has been shown from the performance analysis of the LMS algorithm (**Haykin,** 1996) that it can be convergent or stable in the mean, if and only if $0<\mu<\frac{2}{\lambda_{max}}$. Using the analysis that the maximum value of μ depends on the largest eigenvalue λ_{max} of the input autocorrelation R, which can be approximated to $tr(\mathbf{R})$ and then in the same way to $\|\mathbf{x}_n\|^2$

(i.e., $\lambda_{max} \approx tr(\mathbf{R}) \approx \|\mathbf{x}_n\|^2$), it can be induced that the maximum value of μ depends on the input power signal. Accordingly, the step size for the stable adaptation has to be constrained according to

$$0 < \mu < \frac{2}{\lambda_{max}} \approx \frac{2}{tr(\mathbf{R})} \approx \frac{2}{\|\mathbf{x}_n\|^2} \tag{2-109}$$

where, λ_{max} is the largest eigenvalue of the tap input auto correlation matrix $\mathbf{R}$, $tr(\mathbf{R})$ is trace of $\mathbf{R}$, which is the sum of the elements on its diagonal ($\sum_{i=1}^{P} \lambda_i$), and $\|\mathbf{x}_n\|^2$ is the input power. (2-109) is often referred to as the condition for convergence in the mean-square.

Based on the above, the NLMS (2-110 and 2-111), which is called the variant algorithm of LMS uses the input-dependent adaptation step size so it provides a faster convergence and greater stability than ordinary LMS.

$$h_{n+1} = h_n + \mu_n \mathbf{x}_n e_n \tag{2-110}$$

$$e_n = d_n - y_n = d_n - \mathbf{h}_n \mathbf{x}_n \tag{2-111}$$

$$\mu_n = \frac{\tilde{\mu}}{\alpha + \|\mathbf{x}_n\|^2}, 0 < \tilde{\mu} < 2 \tag{2-112}$$

where μ_n is a modified input dependent step size and α is an infinitesimal positive value added to prevent the possibility of zero division in the event of a very small input value.

For the computational complexity, NLMS requires $3N + 2$ multiplications per sample interval (MPSI) (**Homer,** 2000). Thus $3N^2 + 2N$ real multiplications are required for every N output samples.

■

4.3 RLS (recursive least square) algorithm

The RLS algorithm uses a different adaptive method to determine the coefficients of an adaptive filter. The method utilizes information from all the previous input data to estimate the inverse of the autocorrelation matrix of the input vector (**Haykin,** 1996). It has been derived independently by several investigators, but the original reference on the RLS algorithm is appeared to be Plackett in 1950 (**Plackett,** 1950).

The recursive method of least squares is to minimize the residual sum of squares of the error signal (e_n). To adjust of influence of input samples from the far past, the weighting factor is used in the cost function (J_n).

$$J_n = \sum_{k=1}^{n} \beta^{n-k} e_k^2 \tag{2-113}$$

where β is called the exponentially weighted forgetting factor to be selected between $0 < \beta < 1$.

In a analogous way with LMS method, we now find the gradient of the cost function with respect to the current weights h_n

$$\nabla_h(J_n) = \sum_{k=1}^{n} \beta^{n-k}(-2\mathbf{P} + 2\mathbf{R}\mathbf{h}_n) \tag{2-114}$$

where $\mathbf{R}$ is $E[\mathbf{x}_n \mathbf{x}_n^H]$, autocorrelation vector of tap input vector $\mathbf{x}_n$. $\mathbf{P}$ is $E[\mathbf{x}_n d_n^*]$, cross-correlation vector between the tap input vector $\mathbf{x}_n$ and the desired response d_n.
However, this method does not use a gradient descent method instead it uses an immediate search for the minimum of the cost function by setting its gradient to zero, $\nabla_h(J_n) = 0$.

The resulting equation for the optimum filter coefficients at time n is,

$$\mathbf{h}_n = \mathbf{R}_n^{-1}\mathbf{P}_n \tag{2-115}$$

where $\mathbf{R}_n = \sum_{k=1}^{n} \beta^{n-k}\mathbf{x}_k\mathbf{x}_k^H$ and $\mathbf{P}_n = \sum_{k=1}^{n} \beta^{n-k} d_k \mathbf{x}_k^H$.

Both $\mathbf{R}_n$ and $\mathbf{P}_n$ can be computed recursively:

$$\mathbf{R}_n = \beta\,\mathbf{R}_{n-1} + \mathbf{x}_n\mathbf{x}_n^H \tag{2-116}$$

$$\mathbf{P}_n = \beta\,\mathbf{P}_{n-1} + d_n\mathbf{x}_n \tag{2-117}$$

To find the coefficient vector $\mathbf{h}_n$, we need the inverse matrix $\mathbf{R}_n^{-1}$. Using a matrix inversion lemma (**Haykin,** 1996), a recursive update equation for $\mathbf{R}_n^{-1}$ is found as

$$\mathbf{R}_n^{-1} = \beta^{-1}\mathbf{R}_{n-1}^{-1} + \beta^{-1}\mu_n'\mathbf{x}_n \tag{2-118}$$

where $\mu_n' = \dfrac{\beta^{-1}\mathbf{R}_{n-1}^{-1}\mathbf{x}_n}{1 + \beta^{-1}\mathbf{x}_n^H\mathbf{R}_{n-1}^{-1}\mathbf{x}_n}$

Therefore, we find the weights update equation as

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mu_n'(d_n - \mathbf{x}_n\mathbf{h}_{n-1}) \tag{2-119}$$

The RLS algorithm is found more computationally complex but it has faster convergence than the LMS or its variant NLMS. For the computational complexity, RLS requires $5N^2 + 2N + 2$ multiplications (**Lim and Macleod,** 1994). Table 2-I shows summary of LMS, NLMS and RLS adaptive algorithm.

Table 2-I Summary of LMS, NLMS and RLS adaptive algorithm

LMS	NLMS	RLS
: Steepest decent algorithm for the minimum of cost function $\mathbf{h}_{n+1} = \mathbf{h}_n + \mu \mathbf{x}_n e_n \quad 0 < \mu < 1$ - convergence rate and stability depends on step size μ is large: faster convergence but unstable μ is small: stable but slower convergence	: Variant of LMS $\mathbf{h}_{n+1} = \mathbf{h}_n + \mu_n \mathbf{x}_n e_n$ $\mu_n = \frac{\tilde{\mu}}{\alpha + \lVert \mathbf{x}_n \rVert^2} \quad 0 < \tilde{\mu} < 2$ - Stable and faster convergence than LMS - Input dependent step size - More complexity than LMS	: Recursive algorithm: immediate search for the minimum of the cost function by setting its gradient to zero $\mathbf{h}_n = \mathbf{R}_n^{-1}\mathbf{P}_n$ $\mathbf{R}_n^{-1} = \beta^{-1}\mathbf{R}_{n-1}^{-1} + \beta^{-1}\mu_n' \mathbf{x}_n$ $\mathbf{h}_n = \mathbf{h}_{n-1} + \mu_n' \mathbf{x}_n \eta_n$ $\mu_n' = \frac{\beta^{-1}\mathbf{R}_{n-1}^{-1}}{1 + \beta^{-1}\mathbf{x}_n^H \mathbf{R}_{n-1}^{-1}\mathbf{x}_n}$ $e_n = d_n - \mathbf{x}_n^H \mathbf{h}_n$: apriori error $\eta_n = d_n - \mathbf{x}_n^H \mathbf{h}_{n-1}$: aposteriori error - Faster convergence than NLMS - Computational burden

Both algorithms have compromised effects among simplicity, performance or computational complexity. For a comparison test between RLS and LMS, Harrison et al. (**Harrison et al,** 1986) show that an exact least square method, RLS gives, on average, two-tenths decibel of a SNR improvement over the LMS algorithm. Goubran and Hafez (**Goubran and Hafez,** 1986) describe that the advantage of using a pole-zero model is a significant reduction in the number of taps required, this number is nearly 1/5 the number of taps in a conventional transversal filter. Pole-zero adaptive filters are, however, less stable and more susceptible to truncation errors.

■

5. WOSA, MSC and TDOA estimation algorithms

This section reviews WOSA (weighted overlapped segment averaging), MSC (magnitude squared coherence) and TDOA (time delay of arrival) estimation algorithms. It can be used for power spectrum estimation and its analysis, and an application to a noise reduction scheme using a time delay, bearing and coherence estimate. It has been reviewed from the references of Agaiby (**Agaiby,** 1999), Knapp and Carter (**Knapp and Carter,** 1976), Carter (**Carter,** 1973) and Nuttall and Carter (**Nuttall and Carter,** 1982).

5.1 WOSA estimation algorithm

A nonparametic method, which is based on a time averaging over short, modified periodogram (**Welch,** 1967) has a benefit of using an FFT for a power spectrum estimation. The periodogram was originally introduced by Arthur Schuster in 1898 to study periodicity of sunspots and developed by Bartlett (**Bartlett,** 1948), Blackman and Tukey (**Blackman and Tukey,** 1958), Welch (**Welch,** 1967) and Nuttall (**Nuttall and Carter,** 1982). These spectral based methods make no assumption about how the data were generated and hence are called a nonparametric.

1) A generalized framework for spectral based estimation

A generalized framework for spectral based estimation has been described by Nuttall and Carter (**Nuttall and Carter,** 1982). It provides the following combined five steps (two-stage) time and lag weighting method for a nonparametric spectral estimation.

i) The signal record is partitioned into L frames. Each frame is multiplied by a time weighting function $w_1(t)$ such that the k-th frame becomes

$$y_k(t) = x(t)w_1(t - \frac{N}{2} - k\tau) \tag{2-120}$$

where N is the frame length and τ is the time shift. By a proper selection of τ, the segments may be disjointed or overlapped.

ii) The FFT is calculated for each segment, and multiplied by the complex conjugate of the corresponding FFT of the other signal according to the type of function required whether auto or cross spectrum. The computed values are then averaged over the available segments. In particular, the first-stage spectral estimate is

$$\hat{\Phi}_1(f) = \frac{1}{L}\sum_{p=0}^{L-1}\left|\mathrm{FFT}[y_p(t)]\right|^2 \tag{2-121}$$

iii) The resultant spectral estimates are inverse Fourier transformed into the correlation or lag domain. We have the third step (2-122).

$$\hat{R}_1(\tau) = \mathrm{IFFT}[\hat{\Phi}_1(f)] \tag{2-122}$$

iv) The correlation estimate is multiplied by a lag weighting function $w_2(\tau)$.

$$\hat{R}_2(\tau) = \hat{R}_1(\tau) w_2(\tau) \tag{2-123}$$

v) The result of step (iv) is transformed back again to the frequency domain to yield the second-stage spectral estimate.

$$\hat{\Phi}_2(f) = \mathrm{FFT}[\hat{R}_2(\tau)] \tag{2-124}$$

2) Analysis of nonparametric estimation technique

According to the generalized framework, three main techniques, Blackman and Tukey method (**Blackman and Tukey,** 1958), WOSA technique (**Welch,** 1967) and the lag-reshaping method (**Nuttall and Carter,** 1982) have been analyzed.

Blackman and Tukey method shows that it allows for only one segment with time weighting, and a smooth lag-weighting function in the correlation domain. This method provides a good spectral resolution but requires a large number of data points to achieve a low estimate variance or stability.

In the WOSA method, a large number of overlapped segments are multiplied by a smooth time weighting function, and the FFT products of these segments are averaged to obtain the final spectral estimate. No lag weighting is used in the WOSA method. This method provides a reduced estimate variance but at the expense of a lost spectral resolution.

In the lag-reshaping method, the data is segmented with a unit gain rectangular time weighting and no overlapping. After averaging, the resultant power spectrum is transformed into the correlation domain, where a smooth multiplicative lag-weighting function is applied before transforming back into the frequency domain. It provides virtually the same stability as the WOSA method while requiring fewer computations.

However, It has been found that a fast response time is important especially in a real time processing but a high frequency resolution is not particularly necessary in the wide speech spectrum (**Agaiby,** 1999). For the computational complexity, a total computation of 50% overlapping windowed WOSA algorithm requires $N\log_2(5.12/\Delta f)$, where N is data length and Δf is frequency resolution (**Proakis and Manolakis,** 1992).

Therefore, the WOSA method can be modified to give a much faster response time by using a moving average similar to that used by Allen et al. (**Allen et al.,** 1977) instead of a simple average on the spectral estimates although the lag-reshaping method provides fewer overall computations.

Using a moving average, it provides an almost instant estimate of the coherence function as an estimate, which can be made at the end of each segment instead of at the end of the whole record. Using such a moving average, the auto (2-125 and 2-126) or cross-spectral densities (2-127) are estimated using the following recursive formula (**Allen et al.,** 1977).

$$\Phi_{dd}(i) = \beta\Phi_{dd}(i-1) + (1-\beta)X_d(i)X_d^*(i) \tag{2-125}$$

$$\Phi_{xx}(i) = \beta\Phi_{xx}(i-1) + (1-\beta)X_x(i)X_x^*(i) \tag{2-126}$$

$$\Phi_{dx}(i) = \beta\Phi_{dx}(i-1) + (1-\beta)X_d(i)X_x^*(i) \tag{2-127}$$

where i is the frame number, β is a forgetting factor ($0<\beta<1$) and X_d and X_x are the DFTs of the signals $x_d(t)$ and $x_x(t)$ respectively. The symbol β is a forgetting factor $0<\beta<1$, β=0.8 will be used in the experiments, which was found to be adequate for 50% overlapping processing (**Le Bouquin Jeannes and Faucon,** 1994).

Based on above, the modified WOSA method can be applicable to real-time power spectrum estimation. It may also be extended to MSC and TDOA estimation accordingly.

3) Parameters selection for WOSA, MSC and TDOA algorithm

A proper selection of the parameter values is important because it can significantly increase the performance of the WOSA method as a spectral estimator. It has been shown that a significant improvement in the estimate variance can be achieved by a proper parameters selection (**Carter et al.,** 1973). Other measures such as a computation cost, response time, and frequency resolution can also be considered for a parameters selection.

i) Weighting function, WOSA

For any spectral analysis with finite duration records, the sampled signals are multiplied by a window in the time domain. To reduce the sidelobes, a smooth window can be used. The smoother the weighting function, the more rapidly the side lobes of its Fourier transform is decayed. Hence, smooth weighting functions give rise to a better power spectrum estimate, and consequently, better coherence and TDOA estimates.

However, smooth weighing function has a wider mainlobe, which results in poorer frequency resolution for the same segment length. For better resolution, more data points per segment are required (**Proakis and Manolakis,** 1992).

ii) Overlapping, WOSA

For the estimate of MSC, it has been shown that the estimate variance and bias vary inversely with the number of segments (**Carter et al.,** 1973). Therefore for a better estimate of the coherence, we need to increase the number of segments.

By increasing the number of segments while maintaining the same segment length, segments are overlapped to maintain the same frequency resolution. The amount of overlapping is a function of the used weighting function (**Knapp and Carter,** 1976). Most of the reduction in the bias and variance of the estimate may be achieved through 50 percent overlap and it shows good compromise between accuracy and computational complexity (**Carter et al.,** 1973).

iii) Segment length, WOSA

A selection of the frame length is one of consideration for various factors. The longer the frame length, the longer the response time because no processing can be completed until the full frame is acquired and the processing time itself is longer for longer frame length. On the other hand, the longer frame length gets a better frequency resolution.

With a frame length of N and sampling frequency f_s, the highest achievable frequency resolution is Δf, which is from the equation $\Delta f = c\frac{f_s}{N}\text{Hz}$, where c is a constant that depends on the window used and is always greater that one (**Orfanidis,** 1996).

Assuming that the processed signals are stationary random processes, segment length could be considered only for short time intervals, typically $20-40\,ms$ for a speech.

However, the objective for a suitable selection of segment length is to find the minimum length that provides the required frequency resolution.

iv) Averaging, WOSA

It has been described how the bias and variance of the coherence function vary inversely with the number of segments. For that reason, a large number of segments are needed to generate a good estimate. However, a large number of segments requires long data records and it indicates that this makes the response time of the estimator longer.

A segment overlapping can be used to reduce the required data record length but to a very limited extent. For example, the 50% overlap reduces the number of segments to a half. Minimizing the estimator response time is an important factor for a real time application. For the minimization of the response time while achieving low estimator variance and bias, a moving average may be used instead of a simple average.

■

5.2 The MSC estimation

While the spectrum of the noise field is very important, it does not give sufficient information by itself for predicting the expected performance of speech enhancement algorithms. The coherence function is a measure that does enable such predictions to be made. The coherence between two signals is defined by

$$\gamma_{dx}(f) = \frac{\Phi_{dx}(f)}{[\Phi_{dd}(f)\Phi_{xx}(f)]^{1/2}} \tag{2-128}$$

where $\Phi_{dx}(f)$ is the cross PSD between the signals $x_d(t)$ and $x_x(t)$, $\Phi_{dd}(f)$ and $\Phi_{xx}(f)$ are the auto PSD for $x_d(t)$ and $x_x(t)$ respectively.

The MSC gives a measure of the correlation between the two signals, ranging from a value of zero (0) when the signals are mutually uncorrelated to one (1) when they are perfectly correlated. The MSC function is defined as (2-129).

$$|\gamma_{dx}(f)|^2 = \frac{|\Phi_{dx}(f)|^2}{\Phi_{dd}(f)\Phi_{xx}(f)} \tag{2-129}$$

It shows that an implementation of the MSC function requires both auto and cross spectral estimation for the signals $x_d(t)$ and $x_x(t)$.

The MSC function is widely used in signal detection (**Carter,** 1977 a), time delay estimation (**Carter,** 1987), SNR estimation (**Fay,** 1980), a noise reduction scheme (**Cron**

and Sherman, 1962; **Rodriquez,** 1987; **Zelinski,** 1988; **Goulding and Bird,** 1990; **Le Bouquin Jeannes and Faucon,** 1994) and VAD application (**Zelinski,** 1988; **Le Bouquin Jeannes and Faucon,** 1994; **Agaiby and Moir,** 1997, b).

■

5.3 The TDOA estimation

A ML estimator has been developed for determining a time delay between signals received from two spatially separated sensors on the presence of uncorrelated noise. This ML estimator can be realized as a pair of receiver prefilters followed by a cross correlator. It is defined that the time argument at which the correlator achieves a maximum is the delay estimate.

The role of the prefilters is to accentuate the signal passed to the correlator at frequencies for which the SNR is highest and, simultaneously, to suppress the noise power.

Various prefilter weighting functions have been proposed, including the Roth filter, the smoothed coherence transform (SCOT), the phase transform (PHAT), the Eckart filter and the Hannan and Thomson (HT) filter and the most suitable one may be applied according to a specific application.

However, among various prefilter weighting functions for the GCC method, HT weighting function has been found to be the most advantageous in terms of reducing the spreading effect resulting from reverberation, and achieving good estimation accuracy.

5.3.1 The concept of the cross correlator

Parameter estimation techniques are based on modelling the time delay as an FIR filter. With this formulation, a time delay estimation becomes a parameter estimation problem, which is estimating the coefficients of the FIR filter (**Chan et al.,** 1979). This gives the parameter estimation techniques which are applicable to the time delay estimation problem.

The most common method for estimating a time delay is the GCC method, proposed by Knapp and Carter in 1976 (**Knapp and Carter,** 1976). Fig. 2-30 shows the basic concept of a cross correlator. A hypothesized delay $(\hat{D})$ is adjusted for one of the signals to align it with the other received signal. After that, the two signals are multiplied and averaged, then the hypothesized delay $(\hat{D})$ is adjusted to maximize the output $J(\hat{D})$.

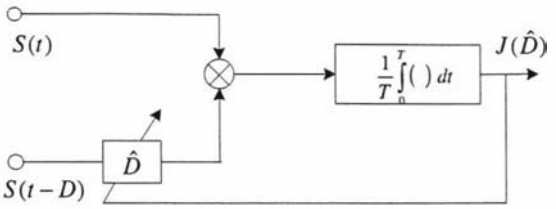


Fig. 2-30 Concept of cross correlator

Another similar type of method can be found from the beamformer time delay estimator (**Carter,** 1977, b; **Carter and Robinson,** 1993). It has shown that the application of beamformer approach is better than the GCC approach in the presence of noise only for detection purposes and not for estimation (**Carter,** 1981).

5.3.2 The concept of the GCC Method

Considering the general case where $s(t)$ is the signal emanating from the desired speech source which passes through two acoustic transfer functions and noises are added to the signals before reaching the microphones, and then assuming that $x_d(t)$ and $x_x(t)$ are the signals received at the two microphones then by reducing the effect of the acoustic transfer function to a delay D and an attenuation factor α, the signals can be expressed as (2-130) and (2-131) respectively.

$$x_d(t) = s(t) + n_1(t) \tag{2-130}$$

$$x_x(t) = \alpha\, s(t-D) + n_2(t) \tag{2-131}$$

and the cross correlation between the two signals is defined as (2-132).

$$R_{dx}(\tau) = E[x_d(t)x_x(t-\tau)] \tag{2-132}$$

where $E[\cdot]$ denotes statistical expectation and the argument τ that maximise the above provides an estimate of delay.

Assuming that $s(t), n_1(t)$ and $n_2(t)$ are real, jointly stationary random processes and that $s(t)$ is uncorrelated with $n_1(t)$ and $n_2(t)$, then a cross correlation function and a cross power spectrum are written as (2-133) and (2-134) respectively.

$$R_{dx}(\tau) = \alpha R_{ss}(\tau - D) + R_{n_1n_2}(\tau) \tag{2-133}$$

$$\Phi_{dx}(f) = \alpha\Phi_{ss}(f)\mathrm{e}^{-\mathrm{j}2\pi fD} + \Phi_{n_1n_2}(f) \tag{2-134}$$

If $n_1(t)$ and $n_2(t)$ are uncorrelated then $\Phi_{n_1n_2}(f)$=0 and it reduces to (2-135).

$$\Phi_{dx}(f) = \alpha\Phi_{ss}(f)\mathrm{e}^{-\mathrm{j}2\pi fD} \tag{2-135}$$

Then, inverse Fourier transform gives cross correlation between $d(t)$ and $x(t)$ as (2-136).

$$R_{dx}(\tau) = \alpha R_{ss}(\tau) * \delta(t - D) \tag{2-136}$$

where $*$ denotes convolution and $\delta(t - D)$ is the delta function, which shows that it has been spread or smeared by the Fourier transform of the signal spectrum.

If the spectral characteristic of $s(t)$ is white, then its Fourier transform will be a delta function. However, a property of autocorrelation functions is that $R_{ss}(\tau) \leq R_{ss}(0)$. It has been shown that equality holds only for a certain τ for periodic functions, but for most practical applications, equality does not hold for $\tau \neq 0$, and the cross correlation will peak at D regardless of any spreading.

The above is considered for single delay. However, we can expect that the situation changes when multiple delays are concerned. The cross correlation in the multiple delay case is given by (2-137).

$$R_{dx}(\tau) = R_{ss}(\tau) * \sum_{i} \alpha_i \delta(t - D_i) \tag{2-137}$$

In this case, the convolution with $R_{ss}(\tau)$ may spread one delta function into another so it becomes impossible to distinguish peaks. To reduce this spreading effect, the cross power spectrum is prefiltered using a filter $\psi_g(f)$ called the general frequency weighting. The time delay is then estimated using the GCC function (2-138).

$$R_{d'x'}^{(g)}(\tau) = \int_{-\infty}^{\infty} \psi_g(f) \Phi_{dx}(f) e^{j2\pi f\tau} df \tag{2-138}$$

Fig. 2-31 shows a block diagram of time delay estimator showing the general frequency weighting function as prefilters.

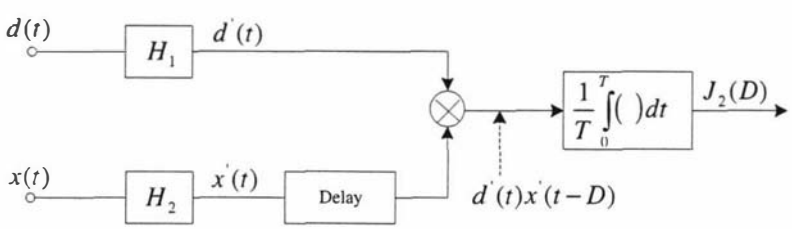


Fig. 2-31 A block diagram of the generalized cross correlation as time delay estimator

The selection of $\psi_g(f)$ should ensure a sharp peak in $R_{d'x'}^{(g)}(\tau)$ rather than a broad one to ensure a good time delay resolution. However, sharp peaks are more sensitive to errors

introduced by a finite observation time, particularly in the case of a low SNR, and therefore the choice of $\psi_g(f)$ is a compromise between a good resolution and stability.

In addition to reduce the spreading effect, $\psi_g(f)$ can be selected to optimize a certain performance criteria of the time delay estimator. An important performance criterion is to accentuate the signal at those frequencies at which the SNR is highest.

Various weighting functions have been proposed and therefore these functions are reviewed.

5.3.3 The classification of prefilter weighting function

1) The Roth Filter

The weighting function as proposed by Roth in 1971 (**Roth,** 1971) is

$$\psi_g(f) = \frac{1}{\Phi_{dd}(f)} \tag{2-139}$$

where $\Phi_{dd}(f) = \Phi_{ss}(f) + \Phi_{n_1 n_1}(f)$, and assuming $n_1(t)$ and $n_2(t)$ are uncorrelated,

The resulting GCC function is

$$R_{dx}^{(g)}(\tau) = \delta(\tau - D) * \int_{-\infty}^{\infty} \frac{\alpha \Phi_{ss}(f)}{\Phi_{ss}(f) + \Phi_{n_1 n_1}(f)} e^{j2\pi ft} df \tag{2-140}$$

It can be seen that the Roth filter is simple to calculate and has the favourable effect of suppressing those frequency where $\Phi_{n_1 n_1}(f)$ is large. However, it shows that the delta function will again be spread out unless $\Phi_{n_1 n_1}(f)$ equals any constant times $\Phi_{ss}(f)$ or that the noise term $n_1(t)$ does not exist.

2) The Smoothed Coherence Transform (SCOT)

Like a Roth filter, the SCOT (**Carter et al.,** 1972; **Carter et al.,** 1973) tries to suppress a part of the cross spectrum $\Phi_{dx}(f)$ that are more likely to be in error.

However, the SCOT filter tries to give equal weight to the effect of both noise signals $n_1(t)$ and $n_2(t)$ in order to reduce errors by either noise signals. To solve it, the SCOT uses the weighting function as (2-141).

$$\psi_g(f) = \frac{1}{\sqrt{\Phi_{dd}(f)\Phi_{xx}(f)}} \tag{2-141}$$

This weighting function gives the SCOT

$$\hat{R}_{d'x'}^{(g)}(\tau) = \int_{-\infty}^{\infty} \hat{\gamma}_{dx}(f) e^{2\pi f\tau} df \tag{2-142}$$

where the coherence estimate is

$$\hat{\gamma}_{dx}(f) \cong \frac{\hat{\Phi}_{dx}(f)}{\sqrt{\Phi_{dd}(f)\Phi_{xx}(f)}} \tag{2-143}$$

This gives the following cross correlation function

$$R_{d'x'}^{(g)}(\tau) = \delta(\tau - D) * \int_{-\infty}^{\infty} \frac{\alpha\Phi_{ss}(f)}{\sqrt{\{\Phi_{ss}(f)+\Phi_{n_1n_1}(f)\}\{\alpha^2\Phi_{ss}(f)+\Phi_{n_2n_2}(f)\}}} e^{j2\pi f\tau} df \tag{2-144}$$

Like the Roth filter, the SCOT tries to suppresses frequencies where the noise level is high, although it does that for $n_2(t)$ as well as $n_1(t)$. The SCOT also tries to reduce the spreading effect as well by acting as pre-whitening filter for $d(t)$ and $x(t)$.

Considering the case where the received signals $d(t)$ and $x(t)$ are filtered through H_1 and H_2 respectively before the cross correlation is calculated as shown in Fig. 2-31, the cross spectrum between $d'(t)$ and $x'(t)$ can be written as (2-145).

$$\Phi_{d'x'}(f) = H_1(f)H_2^*(f)\Phi_{dx}(f) \tag{2-145}$$

where $H^*(f) = H(-f)$ and the GCC is then

$$R_{d'x'}^{(g)}(\tau) = \int_{-\infty}^{\infty} H_1(f)H_2^*(f)\Phi_{dx}(f)e^{j2\pi f\tau} df \tag{2-146}$$

where $H_1(f)H_2^*(f) = \frac{1}{\sqrt{\Phi_{dd}(f)\Phi_{xx}(f)}}$

A possible realization of the $\psi_g(f)$ is to take each value of $H_1(f)$ and $H_2(f)$ as (2-147).

$$H_1(f) = \frac{1}{\sqrt{\Phi_{dd}(f)}} \text{ and } H_2(f) = \frac{1}{\sqrt{\Phi_{xx}(f)}} \tag{2-147}$$

Using the transfer functions $H_1(f)$ and $H_2(f)$, the SCOT can be interpreted as pre-whitening filters followed by a cross correlation. However, even if ideal pre-whitening filters could be achieved, these filters act on $d(t)$ and $x(t)$ and do not pre-whiten $s(t)$. Only if $s(t)$ has a white spectrum, the spreading can be eliminated and therefore pre-whitening $d(t)$ and $x(t)$ does not guarantee the removal of spreading unless $n_1(t) = n_2(t) = 0$.

3) The Phase Transform (PHAT)

To solve the problem of spreading, the PHAT (**Jenkins** 1968; **Knapp** 1976) uses the weighting function as (2-148).

$$\psi_g(f) = \frac{1}{\left|\Phi_{dx}(f)\right|} \tag{2-148}$$

which gives the following GCC function for the outputs $d'(t)$ and $x'(t)$.

$$R_{d'x'}^{(g)}(\tau) = \int_{-\infty}^{\infty} \frac{\hat{\Phi}_{dx}(f)}{\left|\Phi_{dx}(f)\right|} e^{j2\pi ft} df \tag{2-149}$$

When $\hat{\Phi}_{dx}(f) = \Phi_{dx}(f)$ (i.e., exact estimate), the above GCC function becomes (2-150).

$$R_{d'x'}^{(g)}(\tau) = \delta(t - D) \tag{2-150}$$

This indicates that the PHAT ideally does not suffer the spreading effect, however practically the cross spectrum estimate cannot be exact, and $R_{d'x'}^{(g)}(\tau)$ will not be a delta function. On the other hand, as the cross spectrum estimate $\hat{\Phi}_{dx}(f)$ is weighted by the inverse of $\Phi_{dx}(f)$, any errors in the estimate are accentuated where a signal power is low.

4) The Eckart Filter

The Eckart filter (**Eckart,** 1952) uses the weighting function as (2-151).

$$\psi_g(f) = \frac{\alpha \Phi_{ss}(f)}{\Phi_{n_1 n_1}(f)\Phi_{n_2 n_2}(f)} \tag{2-151}$$

The Eckart filter has been developed to maximize the deflection criteria, which is defined as the ratio of the change in mean correlator output due to signal present to the standard deviation of correlator due to noise alone. Maximizing the deflection criteria will make it easier to detect the peak in the correlator output and thus estimate the time delay.

Like the SCOT, the Eckart filter suppresses frequency bands with high level of noise. On the other hand, unlike the PHAT, it attaches zero weight to bands where $\Phi_{ss}(f) = 0$. The Eckart filter however requires knowledge of the signal and noise spectra. Only when $\alpha = 1$ in equation (2-131), the filter weighting function becomes as (2-152).

$$\psi_g(f) = \left|\hat{\Phi}_{dx}(f)\right| [\{\hat{\Phi}_{dd}(f) - \left|\hat{\Phi}_{dx}(f)\right|\}\{\hat{\Phi}_{xx}(f) - \left|\hat{\Phi}_{dx}(f)\right|\}] \tag{2-152}$$

which can be estimated using the available signals $d(t)$ and $x(t)$.

5) The Hannan and Thomson (HT) Filter

Hannan and Thomson (**Hannan and Thomson,** 1971) and Hahn and Tretter (**Hahn and Tretter,** 1973) have derived a ML estimator criterion of the time delay under general conditions.

Using the ML, Knapp and Carter (**Knapp and Carter,** 1976) and Hahn and Tretter (**Hahn and Tretter,** 1973) have shown that this criterion leads to a generalized correlator under the conditions that $s(t)$, $n_1(t)$ and $n_2(t)$ are Gaussian and uncorrelated.

The generalized correlator has the general frequency weighting function as (2-153).

$$\psi_g(f) = \frac{|\gamma_{dx}(f)|^2}{|\Phi_{dx}(f)|\,[1-|\gamma_{dx}(f)|^2]} \tag{2-153}$$

The filter is also shown to be optimum in the sense of maximizing the expected signal peak at the time delay D relative to the background noise, which facilitates the detection of the delay. This shows that the weighting function of the HT filter is that of the PHAT filter (2-148) multiplied by the function (2-154).

$$\frac{|\gamma_{dx}(f)|^2}{[1-|\gamma_{dx}(f)|^2]} \tag{2-154}$$

This makes the HT filter, while trying to remove the spreading effect as the PHAT, does not suffer from the PHAT problem of accentuating the errors where signal power is smallest. The reason behind this is that the HT filter weights the phase according to the strength of the coherence.

Following the initial assumption that the noises are uncorrelated, this will ensure that the cross spectrum is strengthen when the desired signal exists. Furthermore, with a multiplication factor of unity ($\alpha=1$) and for a low SNR, it can be shown that the HT filter reduces to the Eckart filter, which is an optimum signal detection at a low SNR. Also for a low SNR and for a sufficiently large observation time, HT asymptotically achieves the Cramér-Rao lower bound (CRLB) (**Van Trees,** 1968; **Kay,** 1993), which is the minimum variance of any time delay estimator so it commonly used as the performance standard for the time delay estimation.

With the HT prefiltering weighting function (2-153), TDOA d samples can be estimated by

$$d = \max \text{IFFT}[\psi(i)\Phi_{dx}(i)] \tag{2-155}$$

where the term in brackets is the GCC (**Knapp and Carter,** 1976).

The GCC gives a better estimate of time-delay than ordinary cross-correlation, particularly in reverberant environments. In fact, ordinary cross-correlation (when $\psi(i)=1$) only works well (that is, gives a well defined peak at the time-delay) with white-noise signals.

■

6. Summary

The main approaches (ANC, spectral subtraction, beamforming and complex cepstrum) have been reviewed with adaptive algorithms (LMS, NLMS and RLS) and spectral estimation algorithms (WOSA, MSC and TDOA). It has shown that each approach provides advantages under its own theoretical fundamentals but has limitations for the applications in a certain environment.

The typical one-microphone approach, a spectral subtraction method is based on a simple and underlying concept, which removes an estimated noise spectrum from a noisy spectrum in a frequency domain with a VAD application but it has limitations according to an environmental condition (see p. 18-19).

Another one-microphone approach, the cepstrum method is based on a pitch detection and spectral envelope estimate for a speech analysis. On the other hand, the complex cepstrum method is based on a homomorphic deconvolution (homomorphic filter) and it is available for signal recovery from the signal analysis. This method gives advantages of a signal separation and a dereverberation effect but it may give an unstable performance because of a segmentation error (see p. 29-48).

The typical two-microphone approach, an ANC method uses a reference microphone for a speech free noise signal, where it is applied as a reference input to an adaptive filter. The theoretical assumptions should be met for the application in environmental conditions (see p. 9-18).

Beamforming method uses a multiple microphones array to maximize the speech directivity. This method also has a problem with a speech distortion because of a mismatch of microphones and environmental conditions, such as a reverberation (see p. 19-29).

For the spectral estimation algorithms, it has been shown that the GCC method using HT weighting function is appropriate for the application in a real reverberant environment. In addition, MSC estimate based on a power spectrum estimation method, WOSA can be used as a noise reduction scheme by measuring and hence predicting a coherence level between two microphones. Thus, TDOA and MSC functions may contribute to provide better performance to a speech enhancement and noise cancellation system (see p. 60-68).

■

Chapter 3

Analysis of approaches and practical applications

1. Introduction

Based on a general review, an analysis of exemplary approaches and several factors for practical applications are described in detail. This chapter will show that several modified approaches have been developed by remedying shortcomings from the main approaches with the application of adaptive algorithms, TDOA and MSC algorithms in various application areas. The applications encompass various areas such as:

- Teleconferencing (**Kellermann,** 1991; **Lleida et al.,** 1998; **Per,** 2001)

- Hearing aids for the impaired (**Weiss,** 1987; **Greenberg and Zurek,** 1992; **Kompis and Dillier,** 1994; **Kates and Weiss,** 1996; **Berghe and Wouters,** 1998; **Widrow and Luo,** 2003), Hearing aids with a binaural output (**Soede et al.,** 1993; **Welker et al.,** 1997; **Zurek and Greenberg,** 2000)

- Hands-free telephony in automobiles (**Grenier,** 1992; **Elko,** 1996; **Bendjima et al.,** 1999)

- Automatic verification on an ATM (**McCowan et al.,** 2001)

- Robotics (**Stergiopoulos and Dhanantwari,** 1998)

- Communications (**Griffiths and Jim,** 1982)

With the wide range of far-field applications, speech and array processing technology can be classified into such research areas as:

- speech enhancement (**Widrow and Luo,** 2003)
- speaker/speech recognition (**Lin et al.,** 1994; **Lleida et al.,** 1998)
- speech acquisition (**Grenier,** 1992).

The first conventional approach to get rid of problems caused by noise was simple amplification of the signal. This amplified the noise, too. This implies that all sounds were amplified without discriminating between a desired speech and a background noise. Therefore, to overcome this problem and achieve a better performance, several approaches have been considered in the way of not only amplifying a desired signal but also improving the SNR, by even reducing the effect of reverberation.

Some exemplary approaches and application algorithms are shown in Fig. 3-1.

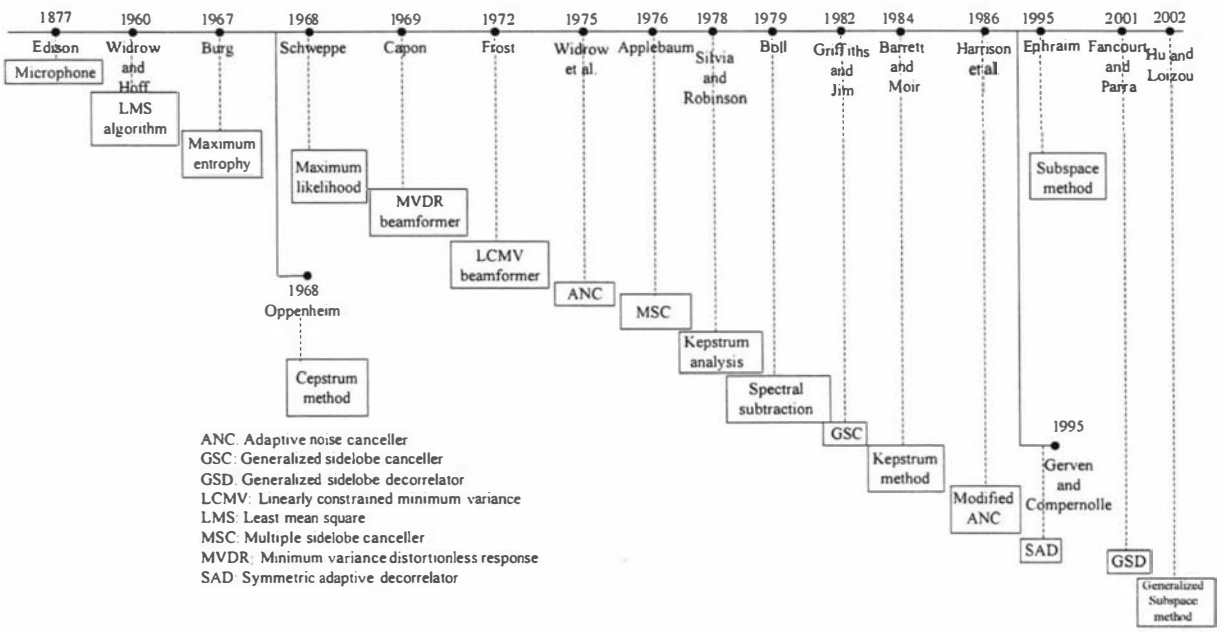


Fig. 3-1 Some exemplary approaches and application algorithms of a speech signal and array processing technology

■

2. Single microphone approach

As a single microphone approach, spectral subtraction method (**Boll,** 1979) offers the attraction of simplicity of processing and compactness in size but has shown limitations because of speech distortion by a certain level of musical and residual noise. For the reduction of this residual noise, the over-subtraction method (**Berouti et al.,** 1979) used over-subtraction factor and the non-linear spectral subtraction (NSS) method (**Lockwood and Boudy,** 1992) used frequency dependent over-subtraction and over-estimation factors.

The other application limitation is because of an ineffectiveness in processing especially when the noise power is equal to or greater than the signal power, when the noise spectral characteristics change rapidly in time or when the noise source is a speech-like noise, e.g., cocktail party. Several variant and modified methods (**Sovka et al.,** 1996; **Bendjima et al.,** 1999; **Virag,** 1999; **Kamath and Loizou,** 2002) have been proposed accordingly.

Following is an example of single microphone approaches which has been developed as an independently modified method or hybrid of other methods.

- Single microphone approach
  1) Multi-band spectral subtraction (**Kamath and Loizou,** 2002)
  2) Kalman filtering (**Gannot et al.,** 1998)
  3) Signal subspace based technique (**Ephraim and Van Trees,** 1995; **Jensen et al.,** 1995)
  4) Adaptive filtering approach for removing noise from speech (**Sambur,** 1978)
  5) Cepstrum method (**Oppenheim,** 1969)

■

3. ANC based approach

For the two-microphone approach, the initial work on adaptive echo cancellers was originally been started around 1965 at the Bell telephone laboratory and proposed the use of an adaptive filter for echo cancellation. It was useful in applications for the environment where a reference signal which contains only a correlated version of the noise in the primary input is easily obtained.

Examples of such situations are echo cancellation for long distance telephone calls (**Sondhi,** 1967), adaptive line enhancement (**Treicher,** 1977), and adaptive antenna array processing (**Griffiths,** 1976).

The typical two-microphone approach, a LMS based ANC (**Widrow et al.,** 1975) has been proposed. According to the analyses, this method works well under the satisfaction of the two main assumptions (see p. 9).

However, the method easily violates above assumptions and it was found that this typical approach does not work properly because of:

- Speech leakage in the reference input
- Uncorrelated noise between two microphones and even reverberation.

Considering the signal leakage to noise reference input (Fig. 3-2), Widrow et al. (**Widrow et al.,** 1975) show an important reciprocal relationship in SNR between reference input and output (2-24). For a low signal distortion, a high SNR at the primary input is needed and a low SNR at the reference input (2-32).

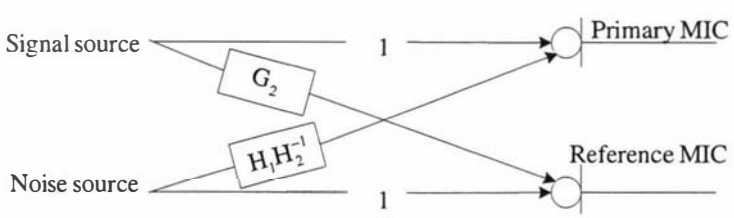


Fig. 3-2 Data generation model with signal leakage

It implies that 1) the more noise existing in a reference input, the more the noise in the output is reduced reciprocally and 2) the less speech signal existing in a reference input, the less signal distortion is achieved in the output. This may be realistically applied in a

simulation set-up but it can result in another limitation for the application in certain real environments. The relationship between a coherence level of noise and the amount of signal leakage shows an opposite effect in a distance relation between two microphones.

Following are applications which should be considered for a better performance in ANC.

3.1 Application of a longer adaptive filter in a low SNR

A longer filter may be applied to improve the performance of the system with the use of a delay in the primary path, especially in a low SNR environment. It will cancel noise reflections more effectively, but at a high SNR, it will be more prone to speech cancellation from speech reflection. The delay filter allows both past and future reference samples to contribute to the current output. However, following are problems and possible solutions to be considered when a longer adaptive filter is used.

- Boll and Pulsipher (**Boll and Pulsipher,** 1980) needed a tapped delay line of 1500 taps to cancel the noise in the primary microphone adequately. Longer filter lengths require more computation in both estimating the filter coefficients and filtering the signals. It was also found that long filter lengths tend to introduce reverberation in the processed speech due to the feedback of the speech through the adaptive filter.

- Harrison et al. (**Harrison et al.,** 1986) showed that updating the adaptive filter continuously can produce a significant speech distortion because an adaptive filter will attempt to reproduce a signal containing speech as well as noise. The use of VAD could resolve the problem by updating the adaptive filter coefficients only during intervals when there is no speech present. In this case, the correct operation of VAD is needed otherwise this approach will treat any speech as noise, and cancel that too.

- A simulation test in a laboratory environment (**Boll and Pulsipher,** 1980) shows the two important design factors of misadjustment and noncausality, where approximating inverse transfer function adequately requires a large amount (1500 taps) of an all-zero FIR filter size. Such a large size of tap weights result in misadjustment (i.e., the ratio of excess to MMSE) (**Widrow et al.,** 1976). A noncausal filter is also required in a situation where the noise reaches the primary

microphone before reaching the reference microphone. Large misadjustment manifests itself as pronounced echo in the speech. This echo is removed by reducing the adaptation step size. Step sizes were used corresponding to misadjustments of 1, 5, and 10 percent. The results showed that the algorithm converges to a steady-state noise power reduction of -20dB in approximately 15 seconds for 10 percent misadjustment and 21seconds for 5 percent misadjustment.

- A longer filter system needs a large settling time because of the large number of taps and may need a multiprocessor system with extra hardware, which increases the complexity and the cost, and decreases the reliability.

■

3.2 Application by physical environmental set-up

- Sound absorbing materials: By using this around the microphone position, echo could considerably be reduced. It implies that the required number of taps for the adaptive filter, which affect its settling time, could be also reduced. Goubran and Hafez (**Goubran and Hafez,** 1986) found that the number of taps can be reduced when the reference microphone is placed inside the cabin in automobile application. This gives the effect of reducing sound reflection. However, if some speech is present at the reference microphone, the filter may tend to cancel the speech instead of the noise. It can be solved if VAD is activated during the noise period.

- Location of the reference microphone: The performance of the adaptive noise reduction system may deteriorate if the noise reaches the primary microphone before reaching the reference microphone. Therefore this requires the use of a noncausal filter. It is easily generated by delaying the primary input. But, long delays may be annoying to the users and may require some extra hardware.

- Swapping the microphones or placing a reference microphone next to the noise source gives an alternative solution (**Campbell and Shields,** 2003) but this is not a practical application in a real environment.

■

3.3 Application of a directional microphone

It is evident that a proper selection of microphone contributes to a better performance in noise reduction, so it can be applied according to the application, the acoustic conditions, the working distance required and the kind of sound required.

Generally, directional microphones can suppress unwanted noise, reduce the effects of reverberation and increase gain-before-feedback. The more the microphone is directional, the less it is going to pick up background noise. It could improve performance but does not give a satisfactory solution in most noisy environments (**Goubran and Hafez,** 1986).

- Mikhael and Hill (**Mikhael and Hill,** 1988) have experimented using a different microphone types (directional and omnidirectional) as reference inputs and showed that the ANC using directional microphones consistently performed better than the ANC using omnidirectional microphones.

- Wenger (**Wenger,** 2003) describes that unidirectional microphones can also be utilized as far-field microphones and show approximately 4.8dB improvement in SNR compared to omni-directional ones for a single speaker in a reverberant environment.

- Soede et al. (**Soede et al.,** 1993) provide test results that directional microphones have shown improved attenuation of diffuse background noise by about 2.5dB compared to omnidirectional microphones.

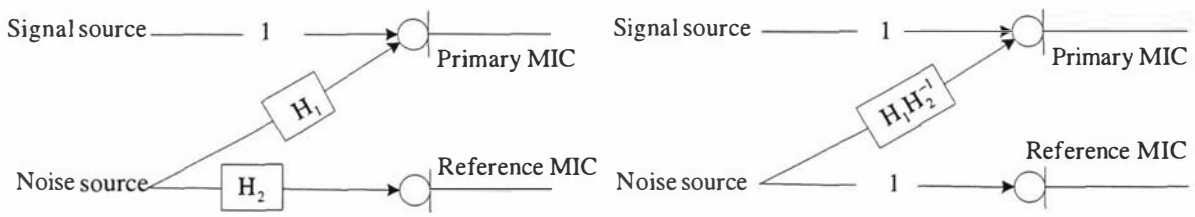
- In an application for the fighter cockpit environment, Harrison et al. (**Harrison et al.,** 1986) use two different types of microphone, one bi-directional microphone inside and one omnidirectional microphone on the outside of the mask.

However, in good acoustic surroundings, omnidirectional microphones can preserve the sound of the recording location, and are often preferred for their flatness of response and freedom from proximity effects. Omnidirectional microphones are normally better at resisting wind noise and mechanical or handling noise than directional microphones and also less susceptible to 'popping' caused by certain explosive consonants in speech, such as 'p', 'b' and 't' (**Audio-technica**). ■

3.4 Application on estimated acoustic transfer function

Pulsipher et al. (**Pulsipher et al.,** 1979) showed that using LMS ANC, a nonstationary acoustic noise in speech can be cancelled effectively by estimated acoustic transfer functions when noise energy is possibly equal to or greater than the speech. The effectiveness of noise cancellation depends directly on the ability of the filter to estimate the acoustic transfer function relating the primary and reference noise channels.

A data generation model shows that the estimated adaptive filter in the absence of uncorrelated noise represents an acoustic transfer function ($H_1 H_2^{-1}$) equal to the product of the transfer function (H_1) from the noise source to the primary microphone multiplied by the inverse of the transfer function (H_2) from the noise source to the reference microphone (Fig. 3-3) (**Pulsipher et al.,** 1979).



(a) A signal generation model (b) its equivalence

Fig. 3-3 A data generation model without signal leakage and its equivalence

The difficulty in estimating the optimal filter arises from a non-minimum phase problem because a data generation model is needed to effectively invert transfer function (H_2) if both of the transfer functions (H_1) and (H_2) are modelled as all-zero filters. It is expected to use a large size of filter length when a classic ANC is used, and pure delay often needs to be introduced in the primary path for the causality. Many pole-zero models have been proposed for modelling acoustic signal paths, and have been used in applications such as acoustic echo cancellers (**Long,** 1986). The advantage of using a pole-zero model is a significant reduction in the number of taps required. This number is nearly 1/5 the number of taps in a conventional transversal filter (**Hosoda et al.,** 1985). Pole-zero adaptive filters are, however, less stable and more susceptible to truncation errors.

In addition to the estimation of acoustic transfer functions for a noise cancellation, following is theoretical analysis for a noise cancellation and a non-speech distortion in ANC method.

- **Analysis of acoustic transfer functions to a two-microphone ANC**

In the periods of no-speech of Fig. 3-4,

$$d_n = H_1(z)n_n, \quad x_n = H_2(z)n_n \tag{3-1}$$

$$e_n = d_n - y_n = d_n - H(z)x_n = (H_1(z) - H(z)H_2(z))n_n \tag{3-2}$$

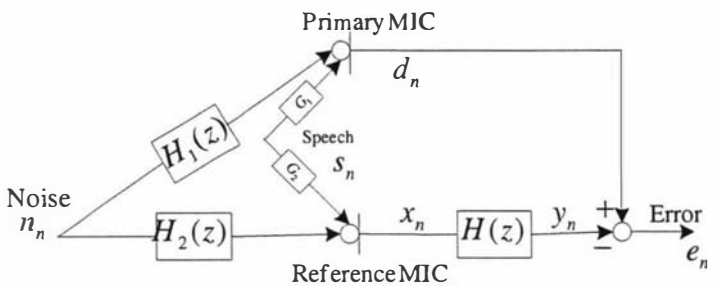


Fig. 3-4 Block diagram of basic noise cancellation method

This shows that noise is cancelled if $H_1(z) - H(z)H_2(z)$ becomes zero, so the estimated acoustic transfer function is $H(z) = H_1(z)H_2^{-1}(z)$ (provided $H_2(z)$ is minimum phase).

In periods of speech with noise and with $H(z) = H_1(z)H_2^{-1}(z)$,

$$d_n = H_1(z)n_n + G_1(z)s_n, \quad x_n = H_2(z)n_n + G_2(z)s_n \tag{3-3}$$

$$e_n = d_n - y_n = d_n - H(z)x_n = \{G_1(z) - H_1(z)H_2^{-1}(z)G_2(z)\}s_n \tag{3-4}$$

This indicates that to increase SNR in speech periods by reducing noise, we could estimate the ratio of unknown acoustic transfer functions, $H(z) = H_1(z)H_2^{-1}(z)$, it can effectively cancel noise and if the speech can be delivered in an equal distance to both of two microphones with a minimal attenuation, $G_1(z) = G_2(z) \cong 1$, the resulting speech distortion will be negligible. The latter condition can be taken to mean that the speech is both close and directly in front of the two microphones. We must also have $H_1(z) \neq H_2(z)$ so that the noise can never be directly in front of or behind the two microphones. However, the condition of $H_1(z) = H_2(z)$ is very unlikely to occur in a real reverberant environment (**Campbell and Shields,** 2003).

For a stable performance, $H_2(z)$ should not be nonminimum phase. However, it is found that we can easily have nonminimum phase in a room reverberant environment. A direct speech application in ANC can be found in Hussain et al. (**Hussain et al.,** 1997)

From the above analysis, it shows that the condition for a noise cancellation and non-speech distortion in ANC method is an estimation of the ratio of unknown acoustic transfer functions during the noise period, and direct speech in front of two microphones.

■

3.5 Small separation between two microphones using a VAD

Harrison, Lim and Singer (**Harrison et al.,** 1986) have introduced a new method in a modified ANC. This uses a small separation between the two microphones with the use of VAD during the silence periods of speech simply by the use of delay in primary input as a noncausal filter. Application of a small separation between microphones, a short distance from the speech source and use of VAD give favourable effects that significantly reduce the filter length required for noise cancellation and minimize the presence of reverberation.
Following is a comparison between distance of two microphones and taps used.

- 367cm inter-distance in a reverberant room environment, 1500 taps used without VAD (**Boll and Pulshpher,** 1980)

- 3cm inter-distance in a fighter cockpit environment, 100 taps are used with VAD using two different microphone, one is directional and the other one is omnidirectioal microphone (**Harrison et al.,** 1986)

- 6cm inter-distance in an application of mobile telephony, with the car stopped and the engine on, 128 taps used with sound absorbing materials inside the cabin using two directional microphones (**Goubran and Hafez,** 1986)

From the above analysis, it shows that a small separation with VAD application in a sound absorbing environment can give an improved performance using a highly reduced filter size.

Small separation also gives a possible improved performance in a multiple noise source environment. The theoretical basis is that if the microphones are close together by zero length, they would both record the same signal and would be perfectly correlated with each other.

While several applications are considered and analyzed, there is a certain limitation for the maximum cancellation in ANC according to the coherence level of noise between two microphones and also analysis of SNR between the output and the reference input.

1) Analysis of a theoretical maximum cancellation using a coherence function

Following is a theoretical maximum cancellation calculated using coherence function (**Goulding and Bird,** 1990) and analysis from output SNR (**Widrow et al.,** 1975) in ANC.

- Limitation in maximum cancellation from coherence function:

The theoretical maximum cancellation for ANC as function of MSC comes from the well-known equation (**Rodriquez,** 1987). It shows the maximum cancellation possible with a linear filter as a function of the MSC between the noise in the primary and reference inputs.

$$\text{Cancellation } (e^{jw}) \text{ in dB} = 10\log\left(\frac{1}{1-\left|\gamma_{xd}(e^{jw})\right|^2}\right) \tag{3-5}$$

According to (3-5), it is found that a significant coherence is required for even modest noise cancelling performance. It shows that values of MSC near 0.7 are required for even 5dB of attenuation (Fig. 3-5).

It also shows that the noise cancellation performance of ANC is highest, when the microphones point in the same direction (**Goulding and Bird,** 1990).

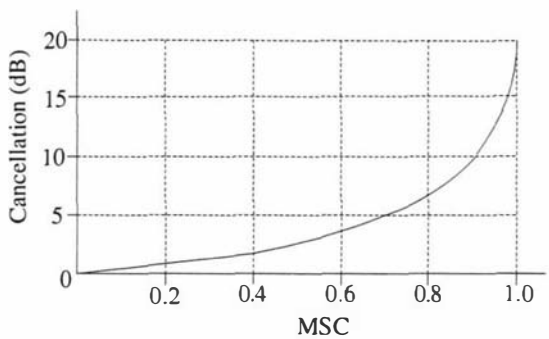


Fig. 3-5 Theoretical maximum cancellation for ANC as function of MSC

- Limitation in maximum cancellation from output SNR

To increase the coherence, especially at higher frequencies, we may decrease the length between two microphones. However, this tends to increase the speech present in the reference microphone. Widrow et al. has analyzed the problem of speech in

the reference microphone and has shown that the resulting output SNR is given by (2-24). This is called 'power inversion', which results in the cancellation of speech at the output (**Widrow et al.,** 1975).

It also can be analyzed by using a coherence function according to microphone types.

2) Coherence analysis to microphone types

It has been shown (**Cron and Sherman,** 1962) that the coherence between the signals received by two omnidirectional microphones in a diffuse noise field, separated by a distance l and at frequency w is given by

$$\gamma_{xy}(w) = \frac{\sin(wl/c)}{wl/c} \tag{3-6}$$

Thus, the coherence between two signals received by omnidirectional microphones separated by l in a diffuse noise field is given by the familiar sinc function.

It is also shown that for two cardioid unidirectional microphones with axes along unit vectors (x_1, y_1, z_1) and (x_2, y_2, z_2), and separated by a distance l, the diffuse field coherence is given by

$$\gamma_{xy}(w) = \frac{3}{4}\left[\frac{\sin(wl/c)}{wl/c} + (x_1x_2 + y_1y_2)\left(\frac{\sin(wl/c)}{(wl/c)^3} - \frac{\cos(wl/c)}{(wl/c)^2}\right)\right. \tag{3-7}$$

$$+ z_1z_2\left(\frac{\sin(wl/c)}{wl/c} + \frac{2\cos(wl/c)}{(wl/c)^2} - \frac{2\sin(wl/c)}{(wl/c)^3}\right)$$

$$\left. + j(z_1 + z_2)\left(\frac{\cos(wl/c)}{wl/c} - \frac{\sin(wl/c)}{wl/c^2}\right)\right]$$

■

3.6 Application of signal separation algorithm

A close placement of the two microphones would reduce the order of the filter but correlated speech or crosstalk is induced in the reference signal. The effect of crosstalk in ANC has been determined for the unconstrained Wiener solution (**Widrow et al.,** 1975)

A CTRANC (crosstalk resistant adaptive noise canceller) have been introduced because of the problem of a small separation between two microphones (**Zinser et al.,** 1985; **Mirchandani et al.,** 1986) (Fig. 3-6).

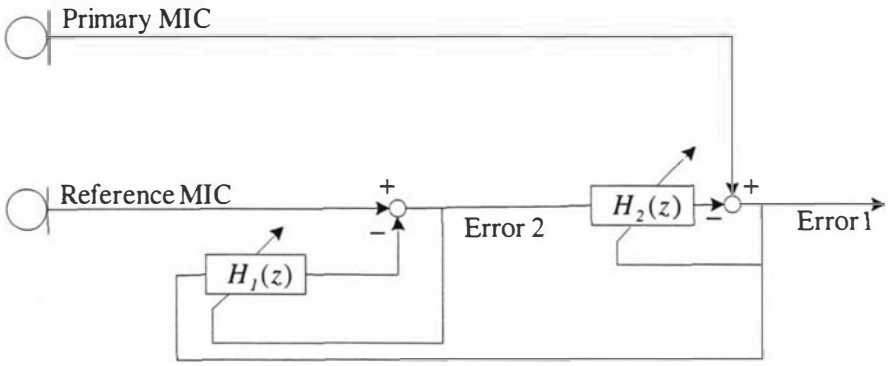


Fig. 3-6 CTRANC method

The structure of CTRANC has been used in a SAD (symmetric adaptive decorrelation) algorithm (Fig. 3-7). The SAD (**Compernolle and Gerven,** 1992; **Gerven and Compernolle** 1995) is an ANC based signal separation method and deals with the problem of speech signal leakage. It also uses a decorrelation estimate based on an intermittent adaptation algorithm, where a decorrelation may be done between a noise free signal estimate and a signal free noise estimate. The least squares criterion is replaced by the decorrelation criterion.

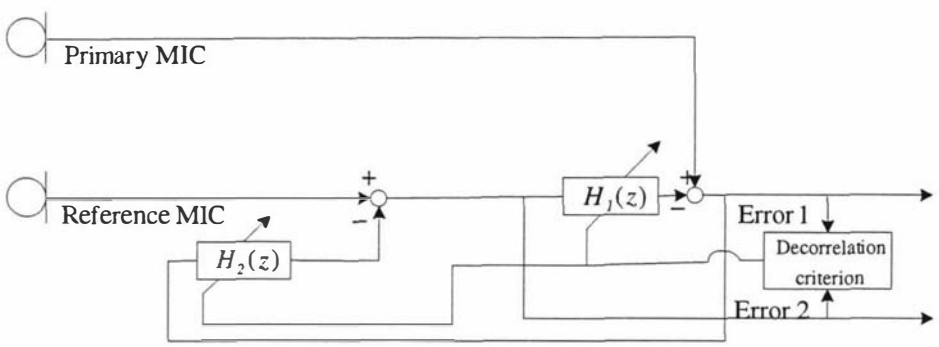


Fig. 3-7 SAD algorithm

3.7 Application using sub-band method to multiple-microphone ANC

The method using multiple adaptive filters has been introduced. Wallace and Goubran (**Wallace and Goubran,** 1992) provide methods using ANC in parallel in multiple reference inputs (Fig. 3-8) and adding sub-band processing to its algorithm (Fig. 3-9).

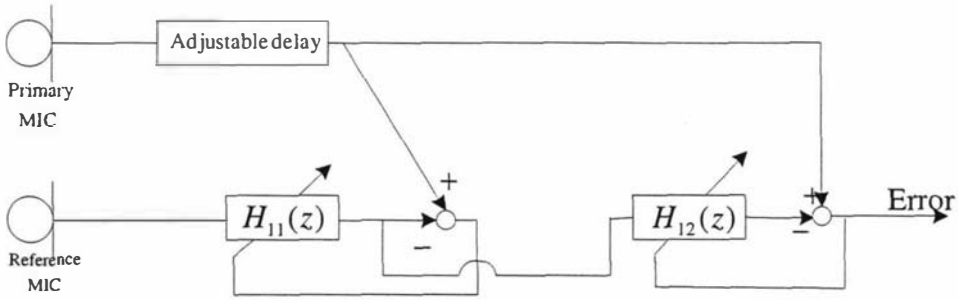


Fig. 3-8 Two-microphone representation from two stage beamforming multi-reference ANC

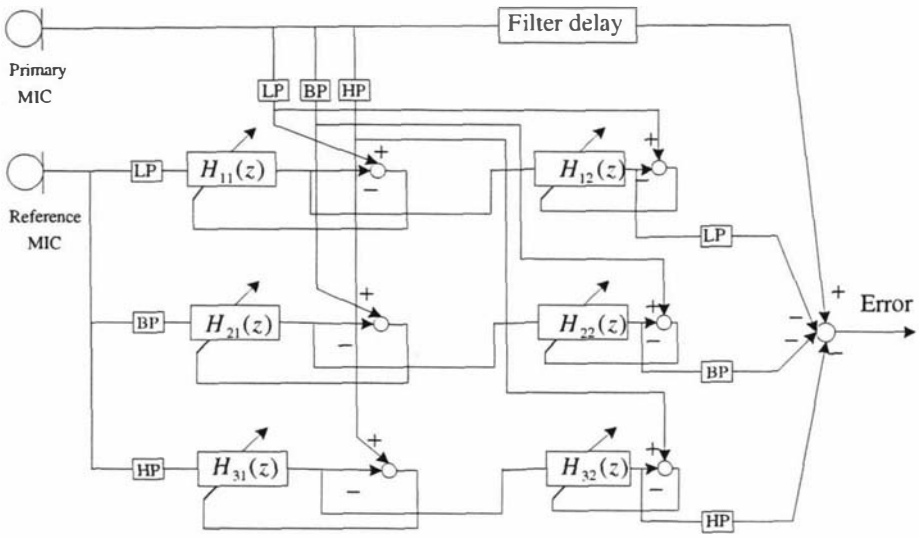


Fig. 3-9 Two-microphone representation from sub-banded two stage beamforming multi-reference ANC with sub-banded second stage

Fig. 3-9 shows that for the sub-band method, a two-microphone ANC approach needs the use of six adaptive filters and a sub-band processing so highly complexity in processing can be expected when it is extended to a multiple-microphone based ANC approach to increase the performance in SNR.

■

4. Approach using multiple microphones array

The main drawback of spectral subtraction and limiting factors of adaptive noise cancelling may be overcome by using microphone arrays based technology (beamforming), which mainly employs the difference in spatial domain (in location and direction) between the desired speech signal and the noise. Fig. 3-10 shows historical background for the beamforming technique.

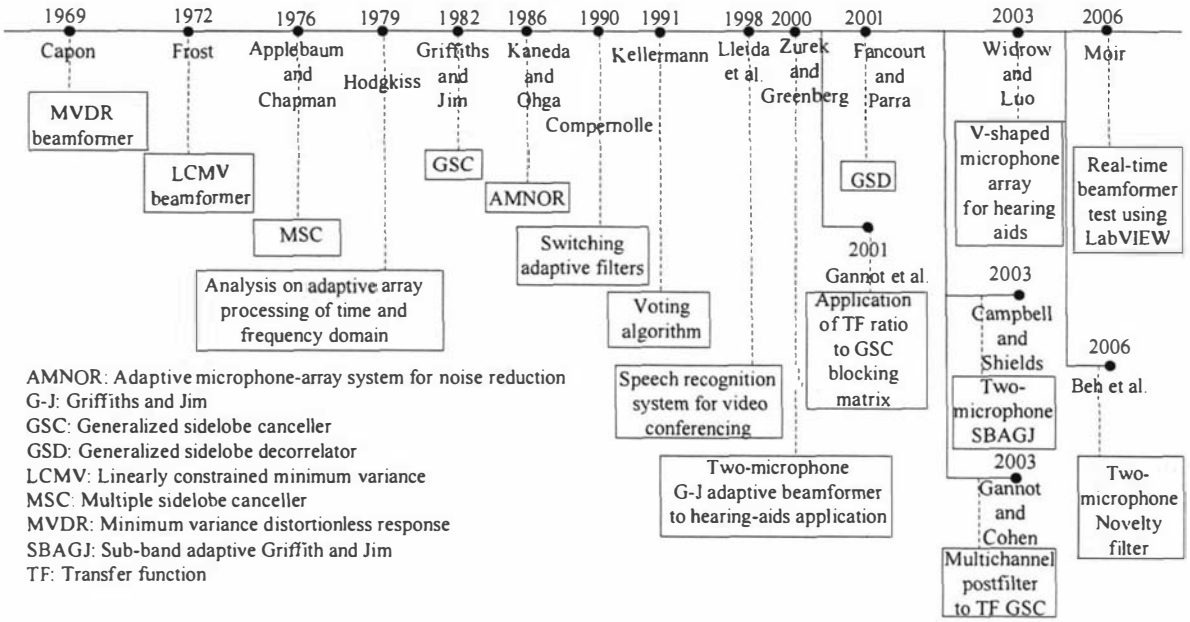


Fig. 3-10 Historical background for beamforming technique

4.1 Application using speech directivity function

With the use of multiple microphones, adaptive beamforming technologies using a spatial characteristics (i.e., TDOA) as well as spectral information have been introduced.

As one of the simple methods, DS beamformer (see p. 25-26) has been introduced and later, it becomes an important component with an adaptive algorithm in modified adaptive beamforming techniques. For the application in an ideal anechoic environment, SNR can be theoretically increased by increasing the number of microphones to a certain level of SNR improvement.

$$\text{SNR with one microphone is } \mathrm{SNR}_{\text{one}} = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \frac{E[\mathrm{s}^2]}{\sigma_n^2}, \tag{3-8}$$

so SNR with N numbered microphone arrays is

$$\mathrm{SNR}_{\text{array}} = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \frac{E[\mathrm{N}^2\mathrm{s}^2]}{\mathrm{N}\sigma_n^2} = \mathrm{N}\frac{E[\mathrm{s}^2]}{\sigma_n^2} = \mathrm{N}(\mathrm{SNR}_{\text{one}}), \tag{3-9}$$

where noise variance of each microphone is σ_n^2.

It indicates that the more microphones with a fixed equidistance use, the greater the improvement in SNR proportionally. Note that this theoretical assumption is based on non-reverberant anechoic environment. In a certain reverberant environment, it shows that a typical SNR improvement for a 4-microphone DS beamformer achieves approximately 4 dB improvement while a G-J beamformer can expect to give up to 8 dB improvement (**Wenger,** 2003).

Practically, the method of maximizing speech directivity by delaying in TDOA and summing its outputs gives a poor performance due to the spatial misalignment of microphone signals to the direction of the speech source and reverberation in real environments.

As a derived version of DS beamformer, the adaptive Frost beamformer with a constrained adaptive algorithm has been proposed (**Frost,** 1972) (see p. 26-28). The purpose is to preserve desired signals from straight ahead adaptively and also to minimize noise signals from other directions. So its algorithm is constrained to a chosen frequency response in the look direction, i.e., the direction of the desired speech source, and then it iteratively adapts the weights to minimize the noise power at the output. Hodgkiss (**Hodgkiss,** 1979) has analyzed this in both the time domain and frequency domain for the adaptive signal and array processing. It shows a typical time domain (Fig. 3-11) and frequency domain adaptive array processors (Fig. 3-12).

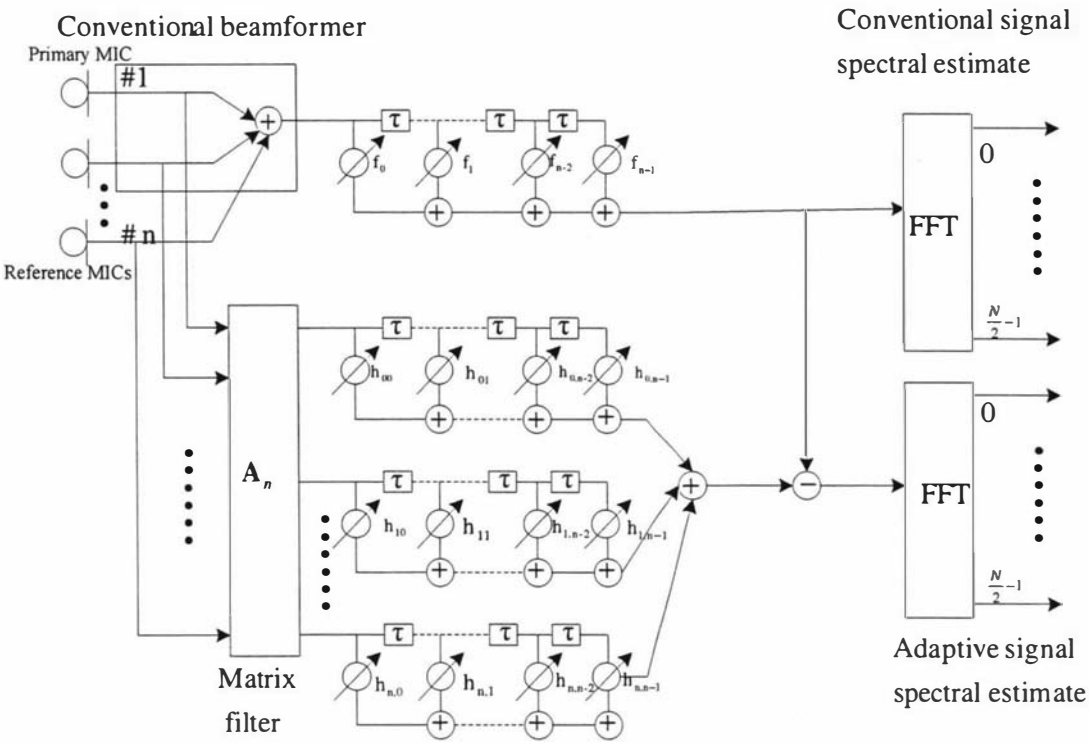


Fig. 3-11 Time domain adaptive array processor (**Hodgkiss,** 1979)

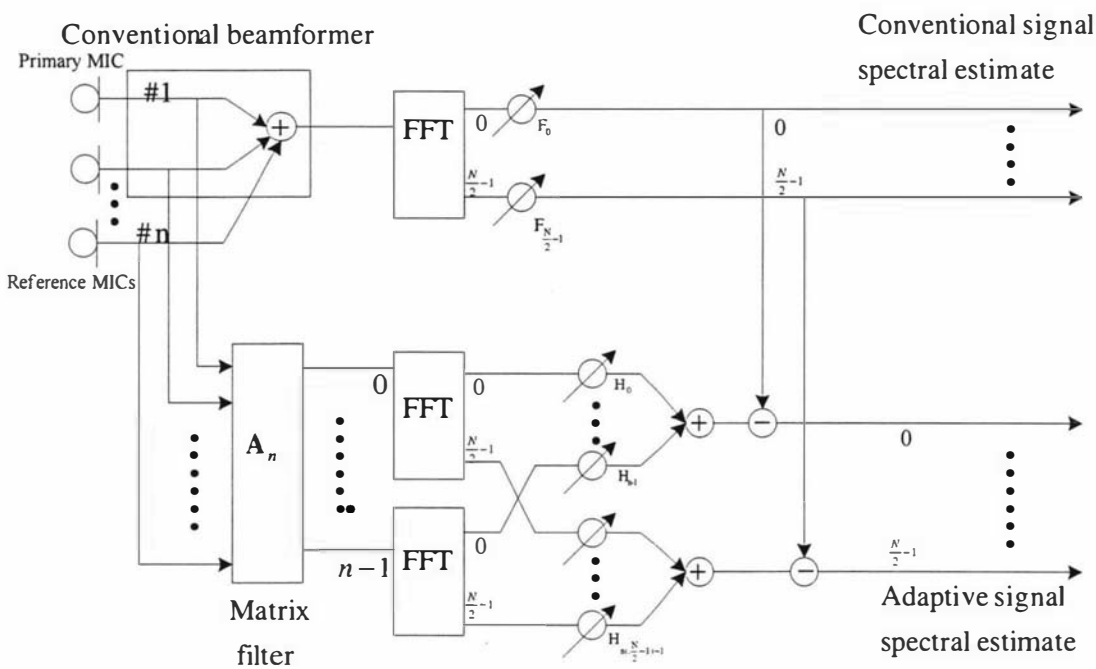


Fig. 3-12 Frequency domain adaptive array processor (**Hodgkiss,** 1979)

■

4.2 Application using a signal blocking function

As an alternative implementation of the adaptive Frost beamformer, the G-J beamformer has been proposed (**Griffiths and Jim,** 1982) (Fig. 3-13). The basic structure consists of three building blocks, 1) constrained, fixed beamformer in a primary input 2) blocking matrix to provide a noise reference input and 3) unconstrained ANC.

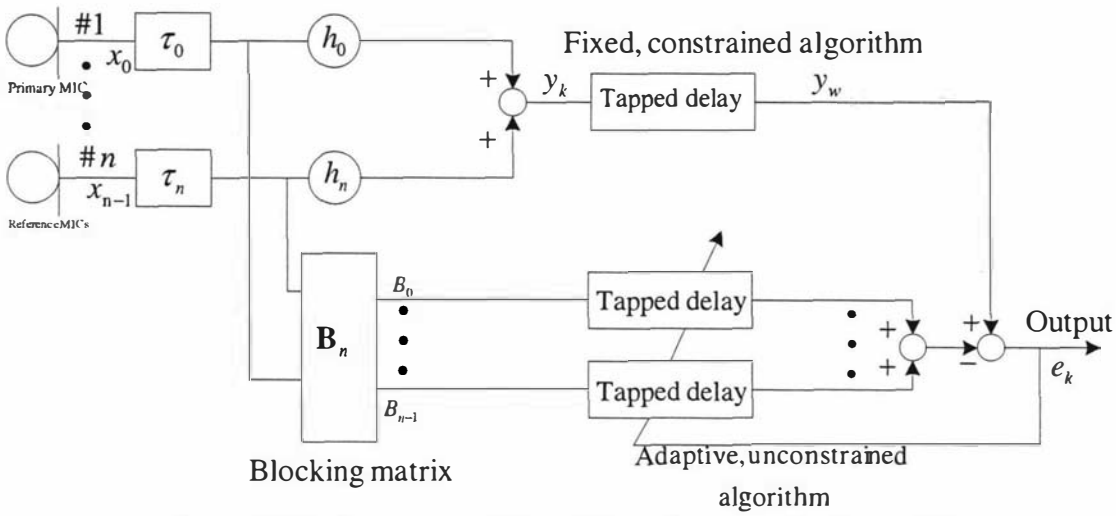


Fig. 3-13 GSC representation of linearly constrained adaptive array processing algorithm

Fixed beamformer is typically designed to maximize directionality and enhance the signal via the spatial filter. Blocking matrix tries to diminish the noisy acoustic power in a specific direction. An ANC cancels the undesired noise signal band with the predefined FIR filter coefficients derived from an adaptive algorithm, therefore provides additional benefits in time varying acoustic conditions. It alleviates problems of signal cancellation and misadjustment that arise in the presence of strong desired signals.

i) Fixed beamformer

Spatially aligns the microphone signals to the direction of the speech source by delaying and summing the microphone signals, i.e.,

$$y_k = \frac{1}{n}\sum_{i=0}^{n-1} h_i^H x_{i(k-\tau_i)} \tag{3-10}$$

where τ_i is delays.

The Wiener postfiltering (**Fischer and Simmer,** 1996), realizing an estimation of a desired signal in a MMSE sense, sets its weights according to a well-known expression

$$W(e^{jw}) = \frac{\Phi_{ss}(e^{jw})}{\Phi_{ss}(e^{jw}) + \Phi_{nn}(e^{jw})} \tag{3-11}$$

where $W(e^{jw})$ is the transfer function of the Wiener filter, $\Phi_{ss}(e^{jw})$ is the PSD of the desired signal and $\Phi_{nn}(e^{jw})$ is the PSD of the noise at the output of the beamformer.

ii) Blocking matrix

The standard GSC uses the output signal of a DS beamformer as speech reference signal, and creates a noise reference by combining the delayed microphone signals using a blocking matrix, blocking out signals arriving from the direction of the speech source.

$$\mathrm{B_n} = \begin{bmatrix} x_{0(k-\tau_0)} - x_{1(k-\tau_1)} \\ x_{1(k-\tau_1)} - x_{2(k-\tau_2)} \\ \bullet \\ \bullet \\ x_{n-2(k-\tau_{n-2})} - x_{n-1(k-\tau_{n-1})} \end{bmatrix} \tag{3-12}$$

When using one reference signal for the two-microphone approach, only the first element of B_n is considered.

iii) Unconstrained adaptive filters

A multichannel adaptive filter then removes the correlation between the residual noise component in the speech reference signal and the noise reference signals.

The weights of the $H_i(e^{jw})$ filters are stated by formula (3-13).

$$H_i(e^{jw}) = \frac{\Phi_{B_i y_k}(e^{jw})}{\Phi_{B_i B_i}(e^{jw})} \tag{3-13}$$

where $H_i(e^{jw})$ is the transfer function of i-th adaptive filter of ANC, $\Phi_{B_i y_k}(e^{jw})$ is the cross PSD between the output of the Wiener filter and i-th output of the block matrix and $\Phi_{B_i B_i}(e^{jw})$ is the PSD of the relevant blocking matrix output.

The structure of a G-J beamformer provides a concrete basis for the versatile and refined methods of the modified two-microphone adaptive beamforming technology. However, a problem arises from the unknown transfer function between channels. Due to this function, the signal blocking part does not operate properly. Thus, a weak speech signal leaks into the reference channel. This violates the basic assumption of the ANC that the speech signal is

required to be uncorrelated with reference noise. In this case, speech distortion occurs. However, the GSC algorithm has been referred to as one of the more effective methods in solving this leaking problem in the two microphone approach (**Compernolle et al.,** 1990; **Campbell and Shields,** 2003).

Several modified methods based on adaptive noise cancelling and beamforming techniques as well as combined hybrid methods using both benefits in adaptive algorithm and beamformer have been proposed and show a significant improvement in maximizing SNR and minimizing mean square error.

■

4.3 Application using speech beamforming or TDOA function

Compernolle (**Compernolle,** 1990, a) has introduced a switching two stage adaptive filter using a multiple microphones. This method is based on DS beamformer and an ANC. The former cues in on the direct path only and neglects all multipath contributions, and the latter is used for the purpose of both beamsteering and a noise cancelling function by a switching VAD properly (Fig. 3-14). Delay and subtract operation can be used to eliminate the speech from the noise reference. In a real environment, there are reverberations and the unknown acoustic transfer function will be more complex, which is a pure delay and time delay plus a nonminimum phase transfer function. Delay 1 and 2 are to provide a physical realisability when there is the possibility of an uncausal solution if the microphones are in the wrong position with respect to each other or with nonminimum phase acoustic transfer functions.

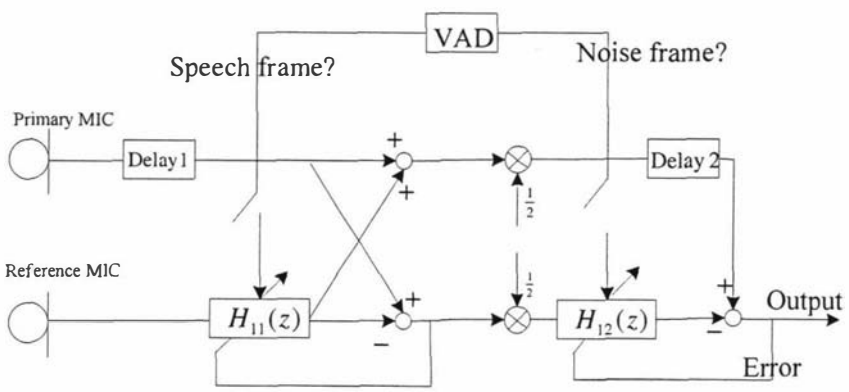


Fig. 3-14 Two-microphone representation from the modified G-J beamformer using switching multiple adaptive filters

As a similar method, Berghe and Wouters (**Berghe and Wouters,** 1998) have proposed a two-microphone G-J ANC approach in an endfire configuration for the application of

hearing aids. The first section of the adaptive filter serves at improving the noise reference by eliminating speech, and may therefore only adapt when speech peak energy is dominant. The second section consists of an unconstrained ANC, only allowed to adapt during absence of speech (Fig. 3-15). A three-microphone approach based on G-J ANC also has been introduced (**Cho and Ko,** 2004).

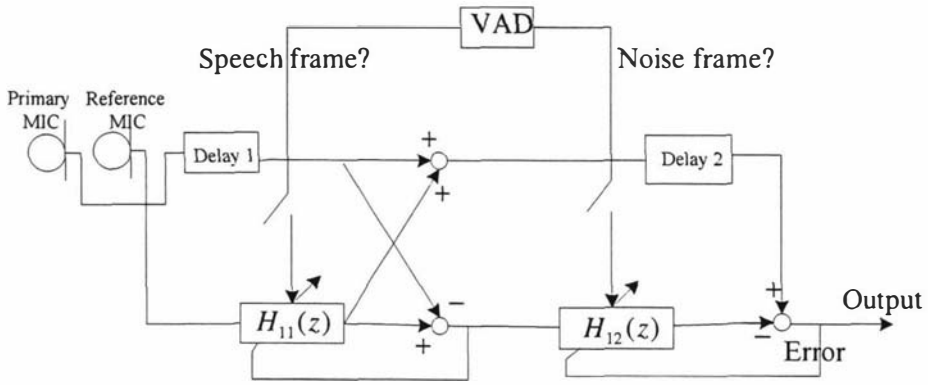


Fig. 3-15 Two-microphone G-J ANC based on switching adaptive filters for the hearing aids application

These two methods (Fig. 3-14 and Fig. 3-15) show a speech adaptive beam steering algorithm based on a multiple-microphone G-J beamforming (**Compernolle,** 1990, a) and a two-microphone G-J ANC (**Berghe and Wouters,** 1998). The two LMS works for signal free noise reference but needs a robust VAD. Alternatively a TDOA function may be used. However, this scheme can suffer from a signal leakage into the reference channel even when adaptation is inhibited during speech due to reverberation and target misalignment which could lead to uncontrollable speech distortion (**Compernolle,** 1990, b).

- Analysis of TDOA estimation to far field and near field sources

When an acoustic or seismic source is located close to the sensor, the waveform of the received signal is curved sound or a vibrating wave. So the curvature depends on the distance and the collection of all relative time delays of the source being used to determine the source location (Fig. 3-16).

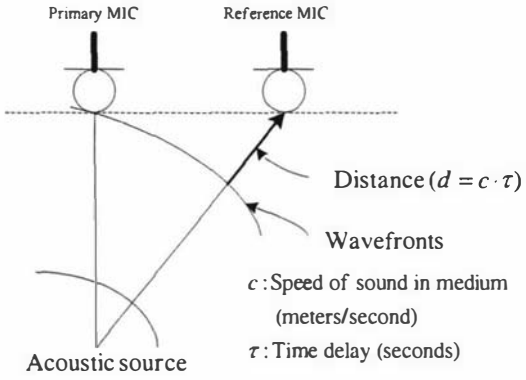


Fig. 3-16 Time delay associated with wavefronts emitted by an acoustic source (**Carter,** 1987; **Carter and Robinson,** 1993)

As the distance is longer, the waveform becomes planar and parallel. Only DOA in the coordinate system of the source is it observable (Fig. 3-17).

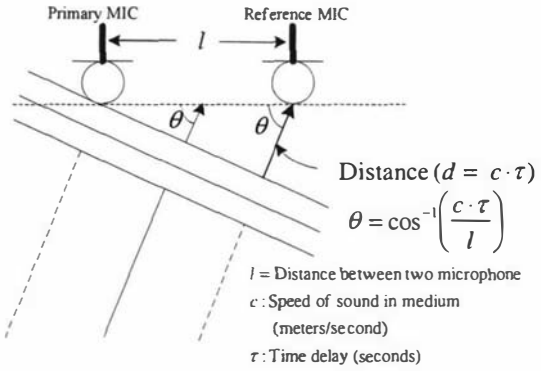


Fig. 3-17 Geometry used to estimate the range and bearing of an acoustic source (**Carter**, 1987; **Carter and Robinson,** 1993)

■

4.4 Application using a direct speech in front of microphones

Following is theoretical analysis for a speech enhancement and a non-speech distortion in G-J beamforming method.

- Analysis of acoustic transfer functions to a two-microphone G-J beamforming

In the periods of no speech of Fig. 3-18,

$$d_n = H_1(z)n_n, \quad x_n = H_2(z)n_n \tag{3-14}$$

$$e_n = \frac{1}{2}(d_n + x_n) - \frac{1}{2}H(z)(d_n - x_n) = \frac{1}{2}[\{H_1(z) + H_2(z)\} - H(z)\{H_1(z) - H_2(z)\}]s_n \tag{3-15}$$

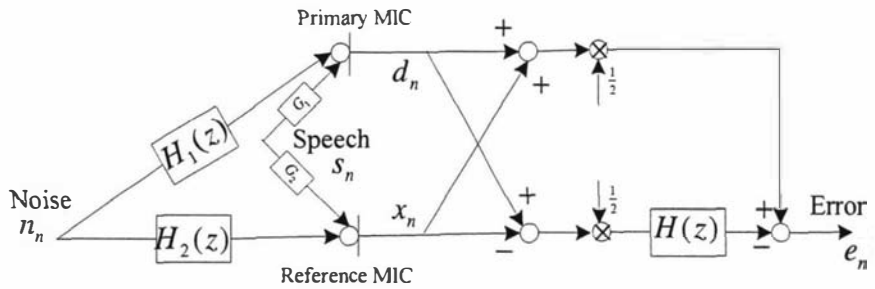


Fig. 3-18 Block diagram of speech enhancement method: Modified G-J beamformer

This indicates that the error is zero when acoustic transfer function is $H(z)=\dfrac{H_1(z)+H_2(z)}{H_1(z)-H_2(z)}$ (assuming that $H_1(z)\neq H_2(z)$ and that $H_1(z)-H_2(z)$ is minimum phase).

In the periods of speech with noise using the above expression for $H(z)$,

$$d_n = H_1(z)n_n + G_1(z)s_n, \quad x_n = H_2(z)n_n + G_2(z)s_n \tag{3-16}$$

$$e_n = \frac{1}{2}(d_n + x_n) - \frac{1}{2}H(z)(d_n - x_n) = \frac{1}{2}[\{G_1(z)+G_2(z)\} - H(z)\{G_1(z)-G_2(z)\}]s_n \tag{3-17}$$

This shows that error is speech signal alone when

$$H(z)=\frac{H_1(z)+H_2(z)}{H_1(z)-H_2(z)} \text{ and} \tag{3-18}$$

$$G_1(z)=G_2(z)\cong 1 \tag{3-19}$$

It shows that if we could estimate the ratio of unknown acoustic transfer functions (3-18), it can effectively enhance a speech and if the speech can be delivered at an equal distance to both microphones with a minimal attenuation, $G_1(z)=G_2(z)\cong 1$, the resulting speech distortion will be negligible. For an estimation of unknown acoustic transfer functions, the denominator part of a transfer function in (3-18) should not be a nonminimum phase for a stable performance.

However, it indicates that for the non-speech distortion, we only require (3-19). For an application in a real environment, it shows that application of both direct speech in front of the two microphones and a directivity function of sum and subtract function can contribute to an increased SNR.

The direct speech application in G-J beamforming can be found from (**Greenberg and Zurek,** 1992; **Campbell and Shields,** 2003). In an application of a direct speech in a non-reverberant environment, a delay and subtract part (blocking signal part) cancels the speech signal and produces noise signal only and it is input to an adaptive FIR filter whose weights are adjusted to minimize the power in the output signal. This minimization is achieved by filtering the reference input to approximate the correlated signal in the primary path, and subtracting. Ideally, a direct speech application gives a solution but reverberation creates a speech leakage into the reference input and subsequently speech cancellation.

As described, the following needs to be considered for the application.

- SPL analysis to a direct speech application

A typical speaker speaks with the amplitude of approximately 96dB SPL (sound pressure level) at the speaker's lip. The amplitude of voice signal decreases significantly with distance according to an inverse square law. Attenuation ratio can be described in dB as (3-20) (**Speaks,** 1996; **Wenger,** 2003).

$$dB = -10\log_{10}\left(\frac{d_s}{d_n}\right)^2 \tag{3-20}$$

where d_s is the distance of interest, d_n is some reference distance and dB is the difference in acoustic intensity between d_s and d_n.

According to (3-20), it shows that as the distance from the speaker increases, the amplitude of the speech signal from the speaker decreases significantly from the speaker as shown in Fig. 3-19.

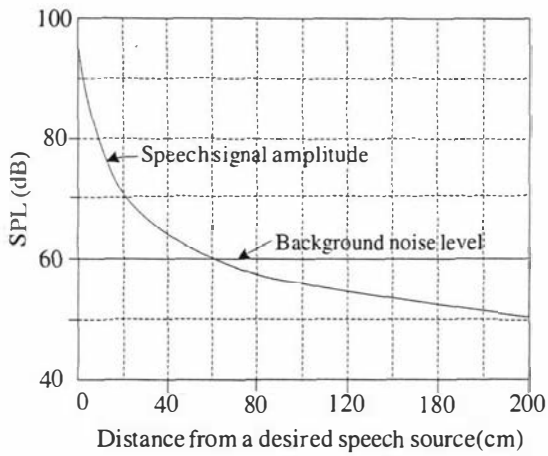


Fig. 3-19 Acoustic SPL of speech as a function of distance from the mouth of speaker and background noise level (**Wenger,** 2003)

This indicates that the further a speaker is from a microphone, the more information content of speech is lost by a background noise (indicates that as the SPL drop off, so does the SNR). Humans require a SNR of greater than 10dB for quality understanding. However, for computer based speech recognition, a SNR of greater than 20dB is typically required. For example, SPL in typical office environment is 60~70dB and a personal office is 45~50dB (**Juang,** 1991).

■

4.5 Application using an intermittent adaptation to beamforming

VAD application can be used for a filter adaptation during the noise period only. A good detection performance of a robust VAD may be evaluated from its capability from several environmental factors such as a noise type and a level with even low SNRs, a reverberation, and positions of the desired signal and noise sources. However, though it works well in a certain environment in accordance with above combined factors, it does not properly work in different environment conditions because an effect of each individual factor is independently changed. Therefore, the proper selection of a robust VAD is particularly essential in a real and adverse environment. Following are examples of estimation algorithms for an intermittent adaptation.

- Log energy (**Compernolle,** 1990, a; **Gerven and Xie,** 1997)
- MSC and TDOA (**Agaiby and Moir,** 1997, a; **Campbell and Shields,** 2003)
- Correlation estimator (**Greenberg and Zurek,** 1992)
- Coherence estimate (**Zelinski,** 1988; **Le Bouquin Jeannes and Faucon,** 1994)
- Decorrelation estimator (**Compernolle and Gerven,** 1992; **Gerven and Compernolle,** 1995)
- Cepstrum estimator (**Haigh and Mason,** 1993; **Pollák and Sovka,** 1995)
- Variance estimator (**Moir,** 2001)

■

4.6 Application using dereverberation techniques

Dereverberation techniques are required for enhancing the intelligibility of speech degraded through the addition of multiple echoes. Since the impulse responses of typical rooms are nonminimum phase and have therefore unstable inverses (**Neely and Allen,** 1979), inverse filtering based deconvolution methods have a limited scope in practice. The situation is further complicated by the difficulty of measuring and tracking the room impulse response in real time applications.

An alternative approach for the enhancement of reverberant speech is provided by cepstrum filtering techniques (**Oppenheim and Schafer,** 1975), where a low time filtering or peak picking in the quefrency domain is used to remove the echo's cepstrum. While cepstrum filtering has been applied successfully to the enhancement of speech degraded by simple echoes, its use for the enhancement of microphone speech affected by room reverberation poses several practical problems. These are due mainly to the effect of segmentation errors on the evaluation of complex cepstrum (**Bees et al.,** 1991) and to certain numerical errors associated with the use of exponential weighting.

Microphone array techniques have long been proposed for the removal of room reverberation. Allen et al. (**Allen et al,** 1977) have introduced a two-microphone technique to remove room reverberation from speech signals. This approach, which is a form of delay and sum beamforming, takes advantages of the uncorrelated nature of reverberant speech tails at different locations. Other multi-microphone techniques for room dereverberation can be found from the references (**Pirz,** 1979; **Flanagan et al.,** 1985).

Following are typical methods for dereverberation techniques.

- Sub-band method (**Goulding and Bird,** 1990; **Toner and Campbell,** 1993; **Darlington and Campbell,** 1996; **Hussain et al.,** 1997; **Campbell and Shields,** 2003; **Shields and Campbell,** 2003; **Zheng et al.,** 2004)

- Neural network method (**Lin et al.,** 1990; **Yin et al.,** 1993; **Knecht et al.,** 1995; **Beh et al.,** 2006)

- Inverse filtering (**Miyoshi and Kaneda,** 1988)

- Inverse filtering based deconvolution method (**Neely and Allen,** 1979)

- Cepstrum filtering (**Oppenheim and Schafer,** 1975; **Bees et al.,** 1991)

- Optimum Wiener filtering based minimum phase kepstrum method (**Barrett and Moir,** 1984; **Barrett and Moir,** 1986)

In addition, direction finding methods have been introduced.

- Subspace based method (**Schmidt,** 1981; **Schmidt,** 1986; **Viberg et al.,** 1991; **Ephraim and Van Trees,** 1995; **Hu and Loizou,** 2002)

- Maximum Entropy (ME) spectral estimation method (**Burg,** 1967)

- ML method (**Schweppe,** 1968; **Ziskind and Wax,** 1988; **Stoica and Sharman,** 1990)

■

4.7 Application using sub-band method to beamforming

For the problem of multiple noise sources, the prototypical cocktail party is probably the most common approximation of the worst case. It may be reduced by adding additional microphones so that the number of adaptive filters equals or exceeds the number of noises with increasing adaptive filter sizes. To reduce the number of microphones, a sub-band method could be alternatively applied to a two-microphone approach and it can add improvement in SNR. However, both approaches increase the computational complexity. For a reverberant environment, it shows that a sub-band method gives a better performance than the full-band method (**Neo and Farhang-Boroujeny,** 2002; **Campbell and Shields,** 2003).

■

4.8 Application using minimum phase and a cascaded adaptive filter

A minimum phase and all-pass filter application (**Liu et al.,** 1995) is based on the joint use of microphone array and cepstrum processing. In this technique, the microphone signals are first delay-steered and then decomposed into minimum phase and all-pass components. The former are processed in the cepstrum domain, where spatial averaging followed by a low time filtering is applied. The latter are processed in the frequency domain by performing spatial averaging and by retaining only the all pass component of the resulting output. The low quefrency filtering then removes the remaining echoes in the high quefrency region.

The recursive identification of nonminimum phase systems using an adaptive all-pass filter together with an adaptive transversal filter is introduced (**Lim and Macleod,** 1994). The algorithms basically identify the minimum phase and all-pass components of a nonminimum phase system separately, and should be less computationally expensive to implement than conventional unconstrained algorithm because of the mirror image property between the numerator and denominator polynomials of an IIR all-pass transfer function. The mirror image property of the numerator and denominator polynomials in an all-pass transfer function reduces the number of adaptive parameters needed compared to an unconstrained recursive identification scheme. (**Lim and Macleod,** 1994) The computational complexity of the all-pass algorithms and the conventional recursive least squares equation error technique is compared and it is shown that the all-pass algorithm possesses significant advantages in this area over the RLS algorithm (**Ljung and Soderstrom,** 1983). The applications can be found in the area of acoustic echo cancellation and nonminimum phase channel equation. It has shown that the digital all-pass filter is a computationally efficient signal processing block which is quite useful in many signal processing applications (**Regalia et al.,** 1988).

The similar method but a different multi-microphone approach shows that a recursive generalized singular value decomposition (GSVD) – based optimal filtering in a DS beamformer part of GSC structure can give the benefit of using short filter length in the post-processed ANC (**Doclo and Moonen,** 2005).

■

5. Analysis on sound sources, room reverberation and microphones array

There are three important factors to be considered in a real environment, which are dependent on sound sources, room reverberation and microphones array. Signal distortions normally come from 1) reverberation from walls and ceilings 2) the type of microphones 3) its position and orientation (speech leakage and microphone misalignment) 4) transmission media, such as telephone channel.

5.1 Analysis on sound sources

The movement of people, motor vehicles or vibrating machinery can generate either an acoustic or a seismic waveform. Both acoustic and seismic signal are common in a waveform. However, both are different at propagation speed.

There are commonly encountered noise source types in real environments. Ambient noise is a feature of any environmental condition and can be varied according to the location such as office, vehicle, factory or elsewhere. However, the movement of people, motor vehicles or vibrating machinery can generate either acoustic or seismic waveforms. Therefore, an application should be differently considered because of a difference at propagation speed between acoustic and seismic (i.e., vibrant) sources. Acoustic propagation speed in air is known to be $345 m/s$. On the other hand, the seismic speed is unknown and strongly depends on the propagation medium.

■

5.2 Analysis on a room reverberation

In an ideal empty room or outdoor field, there is no sound reverberation. Then the generated sound energy decreases ideally as the inverse of the distance squared. In a normal indoor room, there is fair amount of reverberation by reflecting sound from nearby walls or large objects (Fig. 3-20).

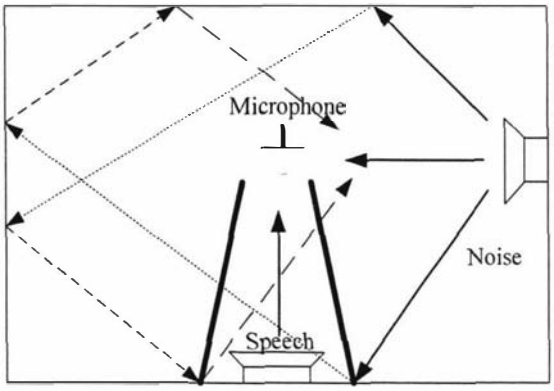


Fig. 3-20 Example of reverberation: reflected sound reaching a directional microphone (multiple reflections) (**Levitt,** 2001)

The heavy lines show the range of directions within which sound is picked up by the microphone without attenuation when a directional microphone is used.

A real acoustic room environment has a mixture of direct sound with the diffuse reflected sound (**Lehnert and Blauert,** 1992). These are dependent on 1) the location of the listener in the acoustic space, 2) the presence of objects and 3) the reverberant characteristics of the acoustic space that vary with room dimension and with absorptive wall materials.

A reverberation time is the time required for a steady-state sound to reach one millionth or -60dB of its original intensity. It can be calculated from Eyring's formula (**Ballow,** 1991)

$$T_{60} = -\frac{0.161V}{A\ln(1-\tilde{\Gamma})} \text{ with } \tilde{\Gamma} = \frac{1}{A}\sum_i A_i\Gamma_i \tag{3-21}$$

where T_{60} is the time required for SPL in the room to decay 60 dB, $\tilde{\Gamma}$ is the average absorption coefficient, V is the volume of the room and A is the total surface area of all the six walls from each wall numbered i, where i =1 to 6.

Some typical reverberation times have been measured by Ford (**Ford,** 1970) and they range from 0.5 seconds for a living room to a maximum of 2.1 seconds for a concert hall.

■

5.3 Analysis on a microphones array

5.3.1 Inter-distance between microphones

Distance between microphones is one of important factors in application. The general rule of inter-distance between microphones is shown as follows:

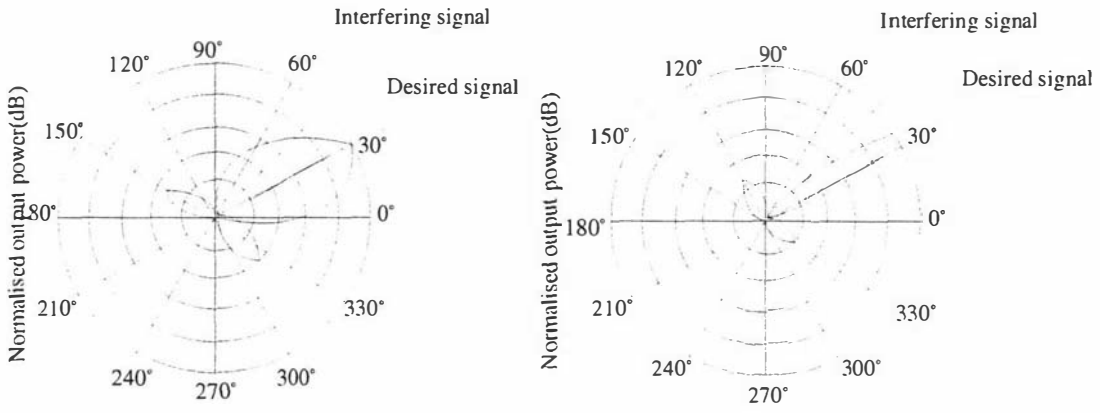
$$d \le \frac{\lambda_{min}}{2} \tag{3-22}$$

$$\lambda_{min} = \frac{c}{f_n} \tag{3-23}$$

where λ_{min} is minimum wavelength, c is the speech of sound and f_n is Nyquist frequency.

As a distance is shorter than $d = \frac{\lambda_{min}}{2}$, the application suffers from spatial discrimination problems because directional noise sources at small angles cannot be suppressed, compared to the desired signal source. If $d << \frac{\lambda_{min}}{2}$, it suffers from poor spatial resolution because of wide beamwidth (Fig. 3-21 (A)).

If a separation is greater than $d = \frac{\lambda_{min}}{2}$, it suffers from a spatial aliasing problem because of undersampling at the high frequency components of the signal but has narrower beamwidth at the low frequencies so there is better spatial resolution. If $d >> \frac{\lambda_{min}}{2}$, the main beamwidth is too narrow to detect the desired signal, therefore, the output of the beamformer could miss the desired signal with a non-optimal DOA estimation (Fig. 3-21 (B)).



(A) $d << \frac{\lambda_{min}}{2}$ (B) $d >> \frac{\lambda_{min}}{2}$

Fig. 3-21 Example of side effects resulting from too long and too short distances between microphones

A compromise must be made between a large spatial aperture, which provide good spatial resolution, and a small spatial aperture, which better conforms to the far-field assumption on which beamforming is based.

In an application using a modified ANC, shorter inter-distance is preferred because of the possibility of a short length of adaptive filter and improved performance in a multiple noise source environment (**Harrison et al.,** 1986).

On the other hand, a longer distance may be preferred in a beamforming technology from the fact that the nature of speech signals contains high energy at low frequencies and not much energy at high frequencies. Performance at high frequencies can be sacrificed in exchange for better rejection capability at low frequencies which results in better overall gain.

5.3.2 Analysis on array geometry

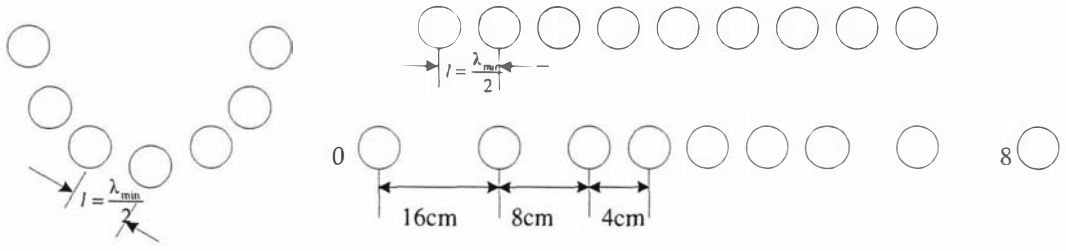
Spatial discrimination capability depends on the size of the spatial aperture. As the aperture increases, discrimination improves.

A single physical antenna (continuous spatial aperture) capable of providing the requisite discrimination is often practical for high frequency signals since the wavelength is short. However, when low frequency signals are of interest, an array of sensors can often synthesize a much larger spatial aperture than that practical with a single physical antenna.

There are several ways to present sensors of array according to their application and environment condition.

1) Planar V shape / linear-nonlinear broadside displacement

Widrow (**Widrow,** 2001; **Widrow and Luo,** 2003) uses a planar V shaped necklace type of array in application of aids for the hearing impaired which shows a dramatic improvement in speech perception over existing hearing aid designs, particularly in the presence of background noise, reverberation, and feedback (Fig. 3-22 (A)).



(A) Planar V shape displacement (B) linear/nonlinear broadside displacement

Fig. 3-22 Example of planar V shape / linear-nonlinear broadside displacement

Lleida, Fernandez and Masgrau (**Lleida et al,** 1998) show nonlinear 9 microphones array. Each band is composed of 5 microphones, microphones 0,1,4,7 and 8 for band I, microphones 1,2,4,6 and 7 for band II and microphones 2,3,4,5 and 6 for band III, (where band I from 50Hz to 1kHz, band II from 1kHz to 2kHz and band III from 2kHz to 4kHz from total bandwidth of 4kHz) at nonlinearly distances, giving the desired distance between microphones to fulfil the spatial aliasing constraint (Fig. 3-22 (B)).

For the linear displacement, the size of spatial aperture should hold following the relationship to process narrowband signals using a microphone array (**Lleida et al.,** 1998).

$$D << \frac{c}{\pi B} = \frac{2Nc}{\pi f_s} \tag{3-24}$$

$$\text{and } l < \frac{\lambda_{\min}}{2} = \frac{c}{2f_n} \tag{3-25}$$

where D is size of spatial aperture, B is the signal bandwidth, c is sound speed, N is FFT order, f_s is sampling frequency, $\lambda_{\min}$ is minimum wavelength and f_n is Nyquist frequency.

For the speech signal with a sampling frequency of 22050Hz, these conditions impose the following constraint on the microphone array design, $d < 0.015m$ and $N >> 105D = 105(L-1)l$, where L is the number of microphones.

Grenier (**Grenier,** 1992) uses eight microphones for car environments that are located on an arc of a circle, not uniformly spaced over the arc. It has been applied to a front-end application for an automatic speech recognition system by one TMS C30 processor in a real time.

2) Broadside / endfire displacement

Greenberg and Zurek (**Greenberg and Zurek,** 1992) show test results where performance is better with a broadside array with 7 cm spacing between microphones than with a 26 cm broadside (Fig. 3-23 (A)) or a 7 cm endfire configuration in an application of ears, top front glasses frame, shirt pocket and temple of eyeglass frame (Fig. 3-23 (B)). They found that performance degrades in reverberant environments.

Berghe and Wouters (**Berghe and Wouters,** 1998) have implemented with the use of two identical directional microphones mounted in an endfire configuration (with both front-facing) within a single behind-the-ear (BTE) hearing aid.

(A) Broadside displacement (B) Endfire displacement

Fig. 3-23 Example of broadside / endfire displacement

Kompis et al. (**Kompis et al.,** 1998) have introduced a combined method comprising a fixed and adaptive beamformer. They show that for four microphones, two microphones of each behind-the-ear (BTE) in an endfire configuration forming a fixed beamformer (Audio-Zoom) are used and its resulting outputs are then post-processed by an adaptive breamformer. The system has been implemented in real-time on a portable digital signal processor system.

■

6. Summary

Most of spectral subtraction method has been developed independently using one-microphone. However, a spectral subtraction has shown problems because of a musical noise and a limitation in certain environments, hence it is only applicable in a high SNR environment (see p. 19). The other one-microphone method, cepstrum and complex cepstrum method have mainly been developed for echo cancellation and dereverberation with speech analysis and speech enhancement. These methods also have an application limitation because of segmentation and numerical errors (see p. 32). The front-end minimum phase application to a cascaded adaptive filter shows a favourable effect on computational simplicity due to the mirror image property of IIR all-pass components of a non minimum phase system and its structure has been applied to an application for the identification of nonminimum phase (see. p. 97).

Several modified approaches have been developed from the basic structure of main approaches, especially from ANC and beamforming. For a speech enhancement and noise cancellation, a benefit has been found from an application of both ANC using an adaptive filter at the reference input and a beamforming using a speech blocking function for the speech separation. Based on the structure, several modified approaches can be found with the application of TDOA and MSC algorithm, which contributes to a noise reduction and speech enhancement schemes.

However, a misalignment of microphones array and off-axis reflections due to a reverberation easily violate the assumption of speech quality at the microphones and this leads to a speech leakage into the reference and subsequent speech cancellation. Exemplary applications to the problems are summarized as follows.

- The use of longer filter size in a low SNR environment with a delay filter
- The use of sound absorbing materials and a position of reference microphone to noise source location
- The use of directional microphones
- The application of an estimated acoustic transfer function

- The application of a small separation between two microphones
- The application of an intermittent adaptation during the noise period
- The application of a signal separation algorithm, such as SAD, BSS and GSD
- Sub-band application
- The use of a speech directivity algorithm
- The application of a signal blocking algorithm
- The application of a direct speech in front of microphones

With the above described applications and algorithms, several different approaches have been proposed to limit a speech cancellation, where most approaches can be found from the several variants of the standard GSC and ANC implementation.

- Speech controlled adaptation algorithm (**Compernolle and Gerven,** 1992; **Greenberg and Zurek,** 1992; **Nordholm and Claesson,** 1993; **Le Bouquin Jeannes and Faucon,** 1994; **Hoshuyama et al.,** 1999)
- Norm-constrained adaptive filter (**Cox et al.,** 1987)
- Adaptive blocking matrix with coefficient-constrained adaptive filters (**Hoshuyama et al.,** 1999)
- Spatial filter designed blocking matrix (**Nordholm and Claesson,** 1993; **Nordebo et al.,** 1994)
- Blocking matrix with coefficient-constrained sub-band adaptive filters (**Neo and Farhang-Boroujeny,** 2002)

- Incorporating a transfer function with a blocking matrix (**Gannot et al.,** 2001)

- Postfiltering (**Zelinski,** 1988; **Zelinski,** 1990; **Fischer and Simmer,** 1996; **Meyer and Simmer,** 1997; **Bitzer et al.,** 1999; **Brandstein and Ward,** 2001; **Cohen and Berdugo,** 2002; **Cohen et al.,** 2003)

- GSVD-based optimal filtering technique (**Doclo and Moonen,** 2005): a multi-microphone extension of the single microphone signal subspace techniques (**Ephraim and Van Trees,** 1995)

- BSS (blind source separation) (**Parra et al.,** 1998; **Parra and Spence,** 2000) and GSD (generalized sidelobe decorrelator) method (**Fancourt and Parra,** 2001)

With effort over several different approaches, it has been shown that an improved performance in SNR has been found, but with the added complexity of processing and hence few real-time applications.

Chapter 4

Kepstrum approach to speech enhancement

1. The research objective

For speech enhancement and noise cancellation, the ANC approach has the attraction as a noise canceller by the virtue of the following:

- The use of an adaptive filter in the reference input, which minimizes an output power in a MMSE sense

and beamforming using a multiple microphones array gives advantages on speech enhancement by:

- Maximizing a speech directivity
- Signal separation by spatial discrimination

However, it has been found that both methods easily violate the main assumptions (see p. 9) because of:

- Room reverberation
- Microphone misalignment
- Look direction (i.e., the direction of the desired speech source) error
- Speech leakage into reference input
- Multiple noise sources

Various modified methods have been derived from the ANC and beamforming, and provided a solution for improved performance through various approaches. The level of application and algorithms were varied from simple to quite complex, which requires a

computational burden in real-time implementation. The simple approaches give a limited performance but the most complex approaches provided a better performance than the simple approaches (see p. 104-106).

To increase an SNR, much work has been done with simulation tests and the results give an improved performance, but until recently very little was known of their real-time behaviour. For real-time applications in a real environment, it is found that a physical dimension in size of the microphone array and complexity of software computation (which ultimately limits the available bandwidth) are examples to be considered.

In the thesis, a real-time kepstrum approach is proposed for dealing with noise improvements in a reverberant environment. It uses the kepstrum method (**Barrett and Moir,** 1984; **Barrett and Moir,** 1986) based on a modified application from ANC (**Widrow et al.,** 1975) and a beamforming method (**Griffiths and Jim,** 1982). The objective is to provide an improved performance in SNR with an efficient processing technique and fast processing for real-time applications in a real reverberant environment.

The following is a summarized list for the kepstrum approach, where the kepstrum method based on kepstrum analysis is applied to a speech enhancement method.

- Modified application from ANC

    1. A small separation (20-30cm) of microphones and use of a VAD during noise periods (this application satisfies one of assumptions of ANC, an uncorrelated noise). This application gives effects of minimized reverberation. It allows the use of a reduced adaptive filter size for a multiple noise sources.

    2. Directional microphones may also be used for a better performance.

- Modified application from beamforming

    1. Speech directly in front of two microphones and use of sum and subtract function (after phase alignment) for signal blocking. This gives increased speech directivity in the primary input and a refined noise

reference. Therefore, speech distortion can be reduced. This application reduces problems from microphone mismatch or an acoustic path between speech source and microphone.

2. The kepstrum application in front of sum and subtract function (blocking component for signal separation). It can improve a performance in a reverberant environment.

- Kepstrum method

1. System identification of unknown acoustic transfer functions between two microphones and its front-end application in front of signal blocking component. (The unknown acoustic transfer function between channels causes the problem and the signal blocking component does not work properly because of a speech leakage).

2. The application technique is to update noise statistics during the noise periods only and freeze it during speech periods (the adaptive algorithm is treated in the same way).

Fig. 4-1 shows a block diagram of the kepstrum approach, where the kepstrum method estimates the ratio of acoustic transfer functions between two microphones and it is applied to the front-end of a speech enhancement method.

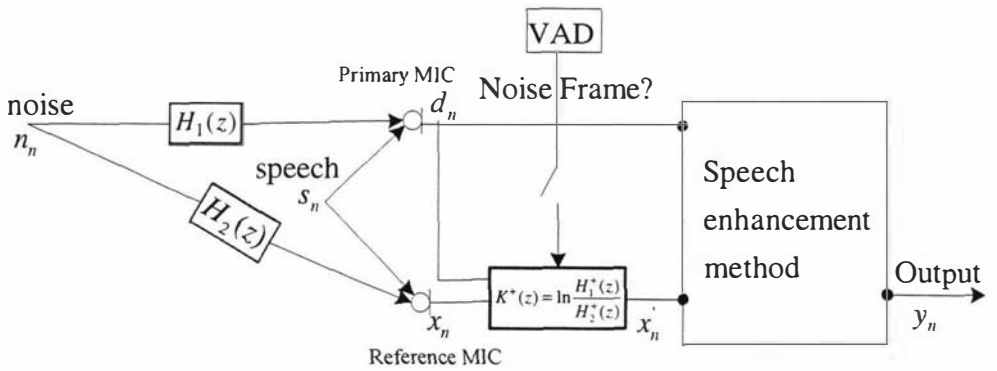


Fig. 4-1 A block diagram of kepstrum approach

By using a kepstrum approach, several favourable effects have been found, which will be investigated. The summarized effects are:

- Invertibility: application flexibility by enabling invertibility, which gives a stable minimum phase transfer function on application of the inverse of a room impulse response.

- Processing robustness: by using it as building blocks. Periodogram estimate applies to kepstrum, TDOA and MSC functions.

- A small number of kepstrum coefficients are required because the minimum phase property shows highly concentrated energy around time zero.

- The use of highly reduced weights in an adaptive algorithm when the front-end minimum phase kepstrum is used with a cascaded adaptive filter.

- Nonminimum phase identification from an application of a front-end minimum phase kepstrum and a cascaded IIR (RLS) filter.

- Improved performance in SNR for speech enhancement.

■

2. Kepstrum approach

Based on a fundamental theory of kepstrum analysis, the kepstrum method provides minimum phase kepstrum estimation and processing technique, and it is applied to speech enhancement methods with a modified application from ANC and beamforming. This is called the kepstrum approach.

2.1 Kepstrum analysis of spectral factor

In addition to its appearance in the logarithmic potential problem (see p. 38), the kepstrum has been used by Szegö (**Szegö,** 1915) and Kolmogorov (**Kolmogorov,** 1939) to solve the problem of factoring the power spectrum $\Phi(w)$ of a random process in order to extract a stable causal system $H_M(z)$, and the problem of spectral factorization is the problem of extracting a causal system $H_M(z)$ whose magnitude spectrum is the square root of the power spectrum.

$$|H(w)| = [\Phi(w)]^{\frac{1}{2}} \tag{4-1}$$

$$K(w) = 2\log|H(w)| = \log\Phi(w) \tag{4-2}$$

Whereas Szegö and Kolmogorov were only concerned with the real part of the function $\log H(z)$ on the unit circle, Robinson (**Robinson,** 1954) was concerned with the imaginary part as well, namely both the real and the imaginary part have physical meaning. Thus, it is important to consider the logarithm of the spectrum $H(w)$.

Kepstrum from power spectrum estimation provides a simple and practical algorithm for obtaining a complex frequency response from a pure magnitude domain, which can be easily implemented. The spectral factors from power spectrum represent a stable minimum phase causal system and non minimum phase anticausal counter part. The *Kolmogorov equation power series* (KEPS) is therefore defined as $K(z) = \sum_{n=-\infty}^{\infty} k_n z^{-n}$ and kepstrum may be identified from two spectral factors found from estimates of the z-transform power spectral density of a random signal plus noise, $\Phi(z)$ accordingly:

$$\Phi(z) = H(z)H(z^{-1}) \tag{4-3}$$

$$\log\Phi(z) = \log H(z) + \log H(z^{-1}) = K^{+}(z) + K^{-}(z) \tag{4-4}$$

Basically, the complex cepstrum has the property that all the information about the minimum phase part of $\Phi(z)$ is contained in the causal part of the kepstrum domain (**Elliott**

and Rafaely, 2000). Kesptrum can be defined as logarithm of power spectrum and it also can be represented as minimum phase spectral factor and non minimum phase counter part.

From the PSD,

$$\Phi(z) = H^+(z)H^-(z) = H^+(z)H^+(z^{-1})$$

Let $z = z^{-1}$, $\Phi(z^{-1}) = H^+(z^{-1})H^+(z)$

It follows that $\Phi(z) = \Phi(z^{-1})$ (4-5)

Defining $z = e^{jw}$,

$$\Phi(w) = \Phi(-w) \quad (4\text{-}6)$$

$$\Phi(w) = \left|H^+(e^{jw})\right|^2 \quad (4\text{-}7)$$

By applying logarithm to both sides,

$$\log\Phi(w) = 2\log\left|H^+(e^{jw})\right| \quad (4\text{-}8)$$

For the N-point discrete form,

$$K\left(\frac{2\pi}{N}k\right) = \log\Phi\left(\frac{2\pi}{N}k\right) = 2\,\mathrm{Re}\left[\log H^+\left(e^{j\frac{2\pi}{N}k}\right)\right] = 2\left|\log H^+\left(e^{j\frac{2\pi}{N}k}\right)\right| \quad (4\text{-}9)$$

$$k_0 = \frac{1}{2N}\sum_{k=0}^{N-1}\log\Phi\left(\frac{2\pi}{N}k\right) \quad (4\text{-}10)$$

$$k_n = \frac{1}{N}\sum_{k=0}^{N-1}\log\Phi\left(\frac{2\pi}{N}k\right)\cos\frac{2\pi}{N}kn \quad (4\text{-}11)$$

$$= \frac{1}{2N}\sum_{k=0}^{N-1}\left(\log\Phi\left(\frac{2\pi}{N}k\right)e^{j\frac{2\pi}{N}kn} + \log\Phi\left(\frac{2\pi}{N}k\right)e^{-j\frac{2\pi}{N}kn}\right)$$

The above (4-10) and (4-11) have the same results from (2-66) and (2-67) by substituting (4-8).

$$k_0 = \frac{1}{N}\sum_{k=0}^{N-1}\log\left|H^+(e^{j\frac{2\pi}{N}k})\right| \quad (4\text{-}12)$$

$$k_n = \frac{2}{N}\sum_{k=0}^{N-1}\log\left|H^+(e^{j\frac{2\pi}{N}k})\right|\cos\frac{2\pi}{N}kn \quad (4\text{-}13)$$

From (4-9),

$$K\left(\frac{2\pi}{N}k\right) = \log\Phi\left(\frac{2\pi}{N}k\right) = 2\left[k_0 + \sum_{n=1}^{N-1}\left(\frac{1}{2}k_{-n}e^{j\frac{2\pi}{N}kn} + \frac{1}{2}k_n e^{-j\frac{2\pi}{N}kn}\right)\right] \quad (4\text{-}14)$$

$$= g_0 + \sum_{n=1}^{N-1}\left(k_{-n} e^{j\frac{2\pi}{N}kn} + k_n e^{-j\frac{2\pi}{N}kn} \right)$$

where $g_0 = 2k_0$.

For the causal kepstrum coefficients,

$$K^+\left(\frac{2\pi}{N}k\right) = \frac{g_0}{2} + \sum_{n=1}^{N-1}\left(k_n e^{-j\frac{2\pi}{N}kn} \right) \tag{4-15}$$

In addition, since k_n are real,

$$K^+\left(\frac{2\pi}{N}k\right) = \frac{g_0}{2} + \sum_{n=1}^{\frac{N}{2}-1} k_n e^{-j\frac{2\pi}{N}kn} \tag{4-16}$$

Therefore, it shows that restoration of complex function can be achieved by using the kepstrum method (**Barrett and Moir,** 1986).

$$K^+\left(\frac{2\pi}{N}k\right) = \log H^+\left(\frac{2\pi}{N}k\right) = \log\left|H^+(e^{j\frac{2\pi}{N}k})\right| + \mathrm{j}\arg H^+(e^{j\frac{2\pi}{N}k}) \tag{4-17}$$

$$= \frac{g_0}{2} + \sum_{n=1}^{\frac{N}{2}-1} k_n e^{-j\frac{2\pi}{N}kn}$$

So its minimum phase kepstrum can be recovered from its positive kepstrum domain by making the zeroth coefficient to half (4-18) and shown in Fig. 4-2.

$$k_n = \begin{cases} k_n & for\ n \geq 1 \\ \frac{1}{2}k_n & for\ n = 0 \\ 0 & for\ n \leq -1 \end{cases} \tag{4-18}$$

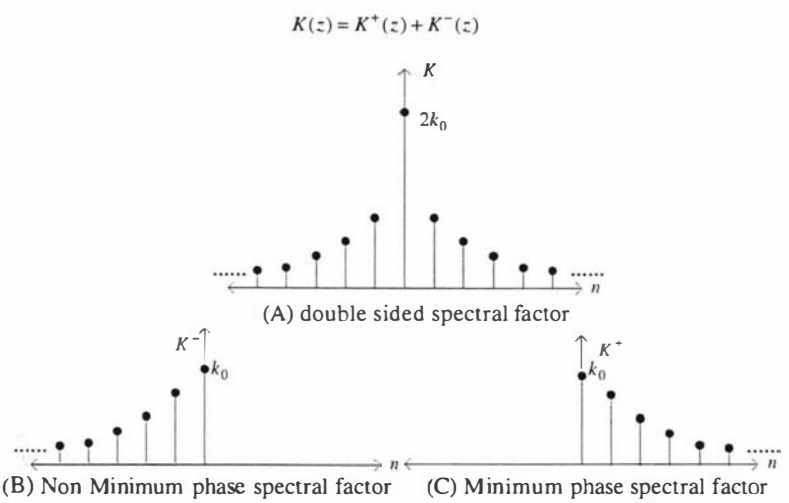


Fig. 4-2 Kepstrum representation from power spectrum ■

2.2 Kepstrum method

Based on a kepstrum analysis, the estimation technique for system identification and its processing technique are introduced and this is called the kepstrum method.

The kepstrum method uses the ratio of acoustic transfer functions from acoustic paths between two microphones during the absence of speech. This method is based on kepstrum analysis, which gives a mathematical construct from the Kolmogorov's fundamental works (**Kolmogorov,** 1941). The similar sounding cepstrum method (**Oppenheim and Schafer,** 1968) originated from a slightly different theoretical framework and although was applied to stochastic signals, the theoretical basis is deterministic. On the other hand, the kepstrum method uses an alternative method which is based on more statistical and mathematical constructs (**Barrett and Moir,** 1984; **Barrett and Moir,** 1986; **Moir and Barrett,** 2003).

The cepstrum method is essentially a practical speech analysis method based on the use of the DFT, using the fast Fourier transform algorithm. However, this method is theoretically based on quantities dependent both on sample length and statistical variation. On the other hand, the kepstrum method could be considered as an alternative method for practical speech enhancement and noise cancellation, which provides a surer theoretical foundation, not subject to statistical variation, by using only the truncated low-time portion of sample length of the kepstrum coefficients.

2.2.1 Kepstrum estimation: system identification of acoustic transfer functions

The kepstrum method based on kepstrum analysis (**Bogert et al.,** 1963; **Schafer,** 1969; **Silvia and Robinson,** 1978) estimates the acoustic transfer functions (the ratio of the two acoustic transfer functions) between two microphone channels during noise periods only.

This is an efficient and fast processing method to identify this acoustic transfer function ratio in a real-time implementation, its benefit coming from the kepstrum processing technique and computational simplicity using the FFT. The technique for estimation of the acoustic transfer function ratio (involving $H_1(z)$ and $H_2(z)$ in Fig. 4-3) uses small (20-30cm) separation between two microphones with the use of VAD when speech is *absent* (Fig. 4-3).

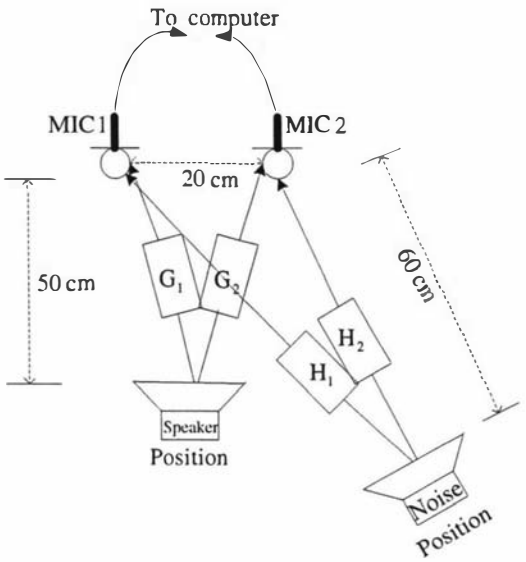


Fig. 4-3 Kepstrum estimation: system identification of acoustic transfer functions during the noise periods only

The estimation procedure for a single acoustic transfer function is illustrated in Fig. 4-4, which shows that periodogram is processed from windowed FFTs as a discrete estimate of continuous power spectral-density.

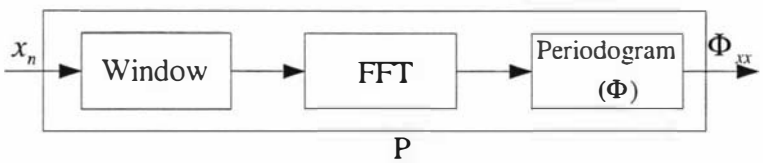


Fig. 4-4 Periodogram estimation procedure

The window size gives a compromising effect between frequency resolution and processing time. The increase of window size gives a better spectral resolution but at the expense of greater processing time and vice versa. The selection of a 2048 window size and 50% overlapping gives a processing time of 46msec, which is a little over speech stationarity (20-40msec) but has excellent frequency resolution (10.76Hz) to differentiate between speech and noise signals.

The periodogram is the most practical estimation tool for power spectrum of a signal as a non parametric spectral estimation method. It has a conceptual simplicity and ease of implementation by FFT.

For the periodogram estimates, the modified WOSA algorithm has been used (see p. 56). The modified WOSA based auto and cross periodograms are processed from 50%

overlapping Hanning windowed 2048 FFTs as a discrete estimate of continuous power spectral density by smoothing methods (2-125, 2-126, and 2-127) with the use of β=0.8 (see p. 58).

2.2.2 Kepstrum processing technique

After taking the log of periodogram, it is found that there exists a bias equal in magnitude to minus Euler's constant γ=0.577215..., so it is added to be unbiased (**Gradshteyn and Ryzhik,** 1979; **Wahba,** 1980)

Therefore, kepstrum coefficients are found from the inverse FFT of the unbiased logarithm of the periodogram (Fig. 4-5). The whole procedure is repeated for each of the two microphones. By subtracting the two sets of kepstrum coefficients we arrive at the kepstrum equivalent of the ratio of the two acoustic transfer functions.

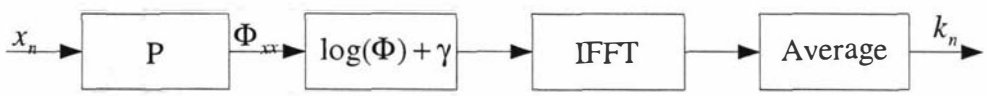


Fig. 4-5 Block diagram for kepstrum processing procedure.
(Window: Hanning, log (Φ): log of periodogram, γ= Euler constant, 0.577215...)

Kepstrum has essential properties which gives an efficient processing technique. For the processing of the ratio of acoustic transfer functions, it just needs a subtraction in kepstrum processing (4-19). For the inverse of the ratio of acoustic transfer functions, it only needs a negative sign (4-20).

$$\ln\frac{H_1(z)}{H_2(z)} \leftrightarrow K_1(z) - K_2(z) \tag{4-19}$$

$$\ln\frac{H_2(z)}{H_1(z)} \leftrightarrow -(K_1(z) - K_2(z)) \tag{4-20}$$

Kepstrum estimation does not provide phase–frequency information. Therefore, methods to recover unknown minimum phase information are considered here by using two methods: A) by adding TDOA delay as phase or B) by restoring phase from the causal kepstrum domain.

I) Kepstrum processing (A): by adding TDOA delay as phase

The whole procedure from Fig. 4-5 is repeated for each of the two microphones. The two set of kepstrum coefficients (k_{1n} and k_{2n}) from the two microphones are then found from the inverse of the natural logarithm of the auto-periodograms. By subtracting the two sets of kepstrum coefficients ($k_{1n} - k_{2n}$), we arrive at the kepstrum (k_n) equivalent to the ratio of the two acoustic transfer functions (since subtraction in logs is division in ordinary algebra) as (4-19). Note that this will be a minimum phase term only as any non-minimum phase information is lost in the periodogram estimates. (adding back the pure time-delay from the estimated TDOA does improve the phase estimate but still misses non-minimum phase zeros).

The difference in kepstrum coefficients from the two channels is truncated in the causal kepstrum domain as explained previously. Then for this first kepstrum processing method, it is converted to an impulse response (h_n) by using the recursive formula (4-21) (**Silvia and Robinson,** 1978).

$$(n+1)h_{n+1} = \sum_{l=0}^{n} (n+1-l)h_l\, k_{n+1-l}\ ,\quad n=0,1,2,3... \tag{4-21}$$

The reference signal (x_n) is then convolved with TDOA delayed impulse response (h_{n-d}) giving a new reference signal ($x_n^{'}$) produced during the noise periods. As described above, the block diagram for the kepstrum processing technique using TDOA delay as phase is shown in Fig. 4-6.

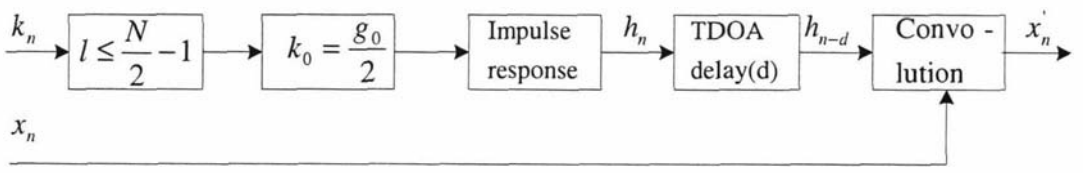


Fig. 4-6 Kepstrum processing (A): by adding TDOA delay as phase

II) Kepstrum processing (B): by restoring phase from the causal kepstrum domain

For the second kepstrum processing method, the kepstrum coefficients are truncated with the processing of the first coefficient to half their previous value and then the kepstrum coefficients are transformed by taking the N point FFT. The magnitude and phase

information are then recovered from the complex output of this FFT. The recovered magnitude and phase information are then multiplied by the FFT of the reference input (x_n) and its output is then inverse FFT transformed back to the time-domain so producing a new refined reference signal ($\dot{x_n}$). This last operation is multiplication in the frequency-domain (or convolution in the time-domain). The block diagram procedure is shown in Fig. 4-7.

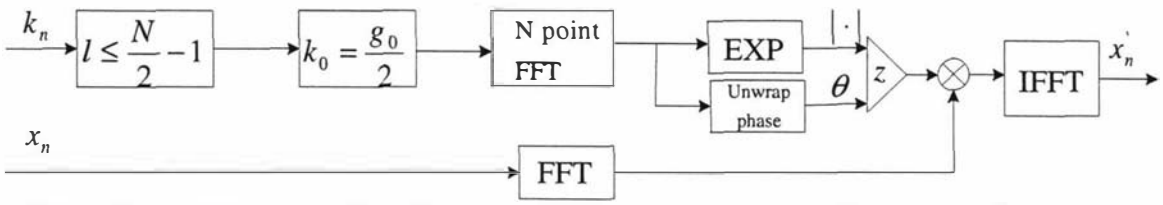


Fig. 4-7 Kepstrum processing (B): by restoring phase from causal kepstrum domain.

Phase unwrapping is unnecessary for the class of minimum phase signals, i.e., a sequence whose z-transform has no poles or zeros outside the unit circle, which implies that $k_n = 0$ for $n < 0$ (**Schafer,** 1968; **Gold and Rader,** 1969; **Oppenheim and Schafer,** 1975).

■

2.3 The effect of a robust processing by building blocks

In addition to use periodogram to kepstrum method, the auto and cross periodograms may also be applied to speech enhancement and noise reduction schemes, such as TDOA (2-155) and MSC (2-129) as well as a VAD application (**Agaiby and Moir,** 1997, a) as illustrated in Fig. 4-8.

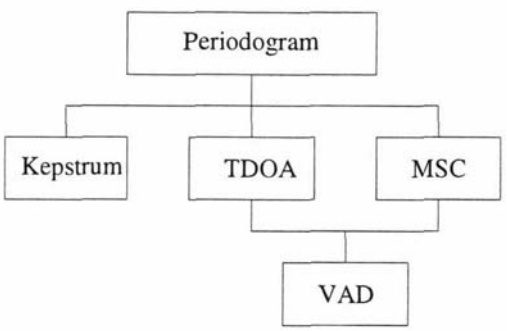


Fig. 4-8 The application of a periodogram estimate

■

2.4 Kepstrum applications

The application technique by the cepstrum method is to use automatic pitch detector (**Noll,** 1967) to select voiced or unvoiced information. On the other hand, the application technique by the kepstrum method is to use a VAD for the noise that is frozen during the speech periods and continuously updated during the noise periods. For the purpose of speech enhancement and noise cancellation, its information may then be applied to the speech enhancement method.

2.4.1 Application to speech enhancement method

The kepstrum approach uses the fact that the kepstrum coefficients are continuously updated during the noise periods and frozen during the speech periods. So, the frozen coefficients are used in the speech period as noise characteristics of the last frame. Its information is applied to the speech enhancement method (the G-J beamformer) as shown in Fig. 4-9. It shows that the kepstrum output (x_n') is added to the primary input (d_n) so producing an enhanced speech signal and it is also subtracted from the primary signal (d_n) so producing a refined noise reference input (Fig. 4-9).

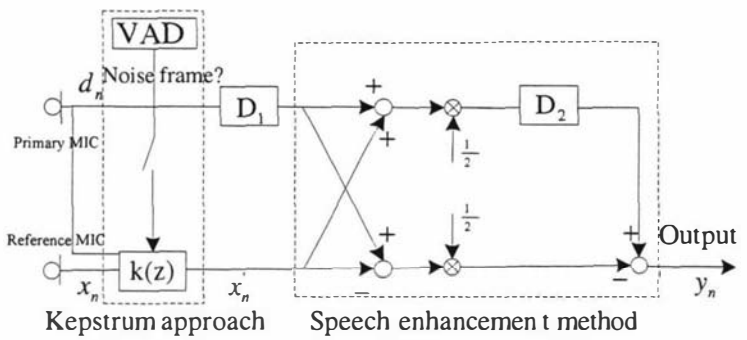


Fig. 4-9 Kepstrum approach to speech enhancement methods (the G-J beamformer)

2.4.2 Application to speech enhancement method with adaptive algorithm

With the use of an adaptive filter as shown in Fig. 4-10, the refined noise reference input is adaptively filtered during the noise periods and its noise statistics are used during speech plus noise periods so it produces a higher performance in SNR in both the stationary and non-stationary noise case.

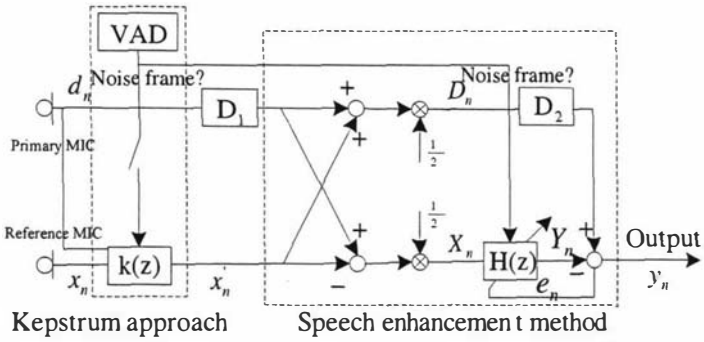


Fig. 4-10 Kepstrum approach to speech enhancement methods (the G-J adaptive beamformer)

■

3. The effect of front-end minimum phase kepstrum application to an adaptive filter

3.1 Invertibility

The real-time application using estimated acoustic transfer functions often suffers from nonminimum phase terms in a real reverberant environment. This results in the identification of a system whose inverse is unstable or at best can be interpreted as uncausal (convergent outside the unit circle). Only minimum phase terms provide invertibility so that their inverse is also minimum phase with stability. The inverse of a transfer function or impulse response gives application flexibility in signal processing, which is sometimes required in real reverberant environments. The use of an adaptive LMS filter gives a simple solution with stability but with a number of non minimum phase terms. This indicates that it requires the use of a large amount of filter weights making reliable real-time processing limited.

■

3.2 Application of front-end minimum phase kepstrum to all-zero FIR filter (LMS or NLMS)

There are two favourable effects to be considered when the minimum phase kepstrum is applied to a cascaded adaptive filter. One is for a highly reduced adaptive filter size and the other one is for a small size of front-end kepstrum coefficient size. Based on theoretical analysis, it will be verified in simulation test (see p. 152) and real-time test (see p. 164).

3.2.1 The effect of highly reduced cascaded adaptive filter size

A stable FIR (finite impulse response) filter can be represented as minimum phase and non minimum terms and it can be factored into a minimum phase filter in cascade with a causal stable all-pass IIR (infinite impulse response) filter (Fig. 4-11(A)). This is sometimes called a phase equalizer because it can be used to compensate or equalize phase response as a complementary filter positioned in cascade of any preceding FIR or IIR filter (**Proakis and Manolakis,** 1992).

Moir and Barrett (**Moir and Barrett,** 2003) have introduced a kepstrum (complex cepstrum) approach to minimum phase Wiener filtering of stationary processes and applied it to speech enhancement with real data.

In this thesis, we use kepstrum analysis in the front to estimate the large minimum phase term and then a cascaded NLMS algorithm to estimate the much smaller residual non

minimum phase zeros which are missed from the kepstrum estimate. The form of the NLMS estimate will be a FIR approximation to an all-pass transfer function (Fig. 4-11(B)).

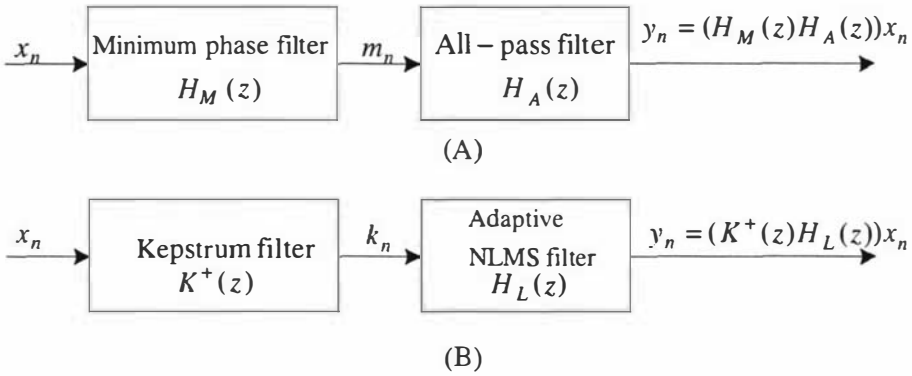


Fig. 4-11 Block diagram of (A) normal factorization represented as minimum phase filter $H_M(z)$ and all-pass filter $H_A(z)$ (B) kepstrum method using kepstrum filter $K^+(z)$ and NLMS algorithm $H_L(z)$

Of course recursive-least squares (RLS) could well be used instead of LMS and this would give rise to a pole-zero i.e., proper all-pass estimate, but RLS is less suitable for high-bandwidth real-time applications due to its computational complexity and often has stability issues.

Throughout this chapter, we use a notation of an n^{th} order causal FIR filter, $H(z)$ by defining as $H(z)=\sum_{n=0}^{\infty} h_n z^{-n}$ convergent within $|z|<1$ and an uncausal filter convergent in $|z|>1$ as $H(z^{-1})=\sum_{n=-1}^{-\infty} h_n z^{-n}$. We also define from an n^{th} degree polynomial $H(z)$ and its corresponding reciprocal polynomial as $z^{-n}H(z^{-1})$. Hence if a polynomial is non-minimum phase then its reciprocal polynomial will be minimum phase and vice-versa.

The general polynomial expression with reciprocal polynomial is shown in Fig. 4-12. Accordingly, the expression using the m roots minimum phase, n roots non minimum phase and its reciprocal polynomial are described in (A) for $H_C(z)=H_M(z)H_A(z)$ showing the minimum phase filter and the cascaded all-pass filter and (B) for $H_C(z)=K^+(z)H_L(z)$ showing the kepstrum filter and the cascaded adaptive NLMS filter.

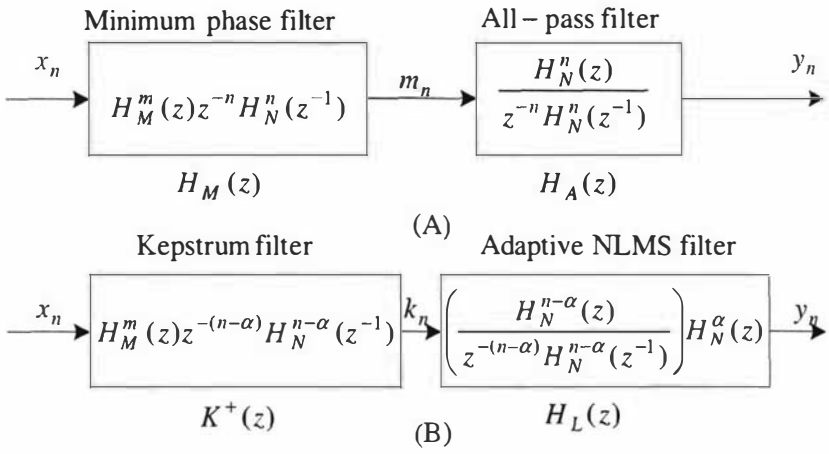


Fig. 4-12 Block diagram of (A) m roots original minimum phase and reflected n roots minimum phase part $H_M(z)$ and all-pass filter $H_A(z)$ (B) kepstrum method using kepstrum filter $K^+(z)$ and NLMS algorithm $H_L(z)$ showing residual α non minimum phase roots with all-pass transfer functions

The favourable effect of front-end minimum phase kepstrum application is investigated. The kepstrum method estimates the minimum phase part of the ratio of the two acoustic transfer functions, and cascaded ordinary LMS (or its variant NLMS) estimates the residual non-minimum phase term. It will be shown that the application of a front-end kepstrum filter is beneficial in that it reduces the number of weights used in the cascaded LMS or NLMS adaptive filter so provides a computational simplicity for real-time processing.

3.2.2 The effect of small amount of kepstrum filter size

According to Parseval's energy theorem (**Papoulis,** 1977), the output energy of an all-pass filter equals any input energy because a magnitude spectrum of an all-pass function is unity, $|H_A(k)|=1$ as described in the N-point discrete form in (4-22) and shown in Fig. 4-13(A).

$$E_y=\sum_{n=0}^{N-1}|y_n|^2=\frac{1}{N}\sum_{k=0}^{N-1}|X(k)H_A(k)|^2$$

$$=\frac{1}{N}\sum_{k=0}^{N-1}|X(k)|^2=\sum_{n=0}^{N-1}|x_n|^2=E_x \qquad (4\text{-}22)$$

Based on (4-22), the application is extended to the combined filter, which is a minimum phase filter in cascaded with an all-pass filter, as shown in Fig. 4-13 (B). The output energy

equals the input energy to an all-pass filter, which is also the output energy of the minimum phase filter.

$$E_y = \sum_{n=0}^{N-1} |y_n|^2 = \sum_{n=0}^{N-1} |m_n|^2 = E_m \tag{4-23}$$

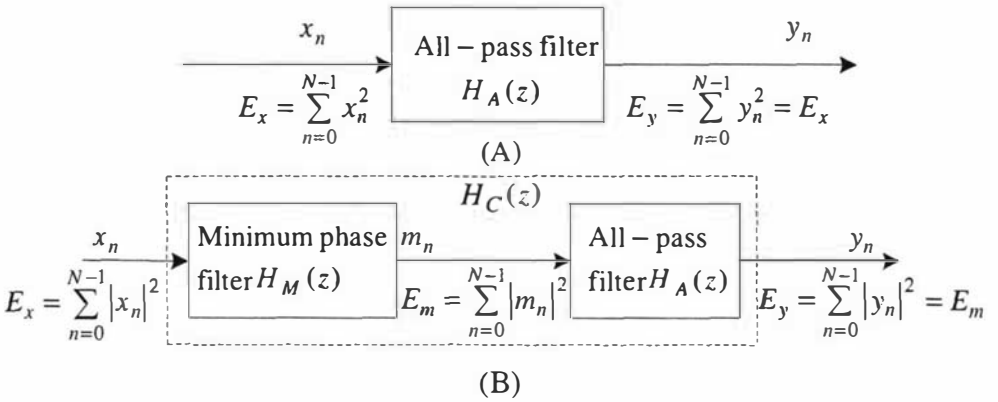


Fig. 4-13 Block diagram of input-output relationship in terms of energy through (A) an all-pass filter (B) a minimum phase filter and an all-pass filter

However, according to the energy concentration property of minimum phase signals, if the systems $H_C(z)$ and $H_M(z)$ have the same magnitude spectrum described as $|H_C(k)| = |H_M(k)|$, then for any n_0, the cumulative partial energy of minimum phase signal shows highly concentrated energy around time zero.

$$E_y^{n_0} = \sum_{n=0}^{n_0} |y_n|^2 \le \sum_{n=0}^{n_0} |m_n|^2 = E_m^{n_0} \tag{4-24}$$

where n_0 is any integer in a time series. $E_y^{n_0}$ and $E_m^{n_0}$ are the output energy and the input energy to an all-pass filter, truncated at time n_0 respectively.

It can also be expressed in the form of energy of an impulse response in (4-25).

$$E_{hn}^{n_0} = \sum_{n=0}^{n_0} |h_N|^2 \le \sum_{n=0}^{n_0} |h_M|^2 = E_{hm}^{n_0} \tag{4-25}$$

where $h_M(n)$ and $h_N(n)$ are minimum phase and non minimum phase impulse responses and $E_{hm}^{n_0}$ and $E_{hn}^{n_0}$ are their energies respectively.

It can be understood that among any signals which have the same magnitude spectrum, only minimum phase signals provide a highly concentrated energy around time zero, which is shown as the fast decaying minimum phase impulse response.

■

3.3 Application to pole-zero IIR filter (RLS)

Any transfer functions can be factored into minimum phase and non minimum phase terms and it can also be decomposed into minimum phase filter and all-pass filter. It will be shown that when the minimum phase kepstrum is positioned in front and it is cascaded with RLS algorithm, non-minimum phase terms can be directly identified from the numerator of all-pass functions for the noise-free case.

However, it cannot be identified by LMS algorithm because it is represented as a combined term of minimum phase and non minimum phase zeros. For the application of an additive white noise, the front-end kepstrum can be used for reducing filter sizes of a cascaded RLS or LMS algorithm because the front-end kepstrum filter absorbs the most of non minimum phase terms from the cascaded RLS or LMS filter, so the remaining non minimum phase terms can be processed in a highly reduced filter size by a cascaded adaptive filter. In addition, the application of the front-end minimum phase kepstrum provides invertibility, which gives a stable minimum phase transfer functions for the inverse of system identification due to the reverberant nature of most rooms environment (Fig. 3-20). Based on theoretical analysis, it will be verified in simulation test (see p. 156).

3.3.1 Identification of unknown system with white noise input

Any noise-free FIR transfer function with m roots of minimum phase zeros and n roots of non minimum phase zeros can be described as polynomial expression of $H(z) = H_M^m(z)H_N^n(z)$ where m roots of minimum phase zeros lie within $|z| < 1$ and n roots of non minimum phase zeros lie outside of the unit circle $|z| > 1$ in the z-plane. It is easy to identify the complete transfer function $H(z)$ using either RLS or LMS adaptive filter as shown in Fig. 4-14.

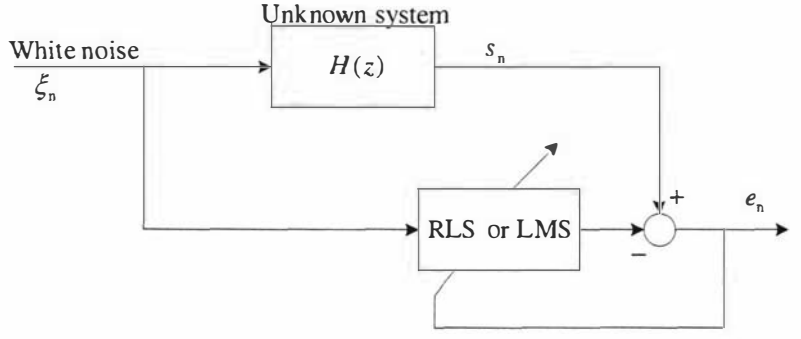


Fig. 4-14 Identification of unknown system with white noise input

Most LMS or its variants are used in signal processing application because non minimum phase terms can occur as an acoustic transfer function between a sound source and a receiving microphone. It can be implemented efficiently in real-time processing due to its simplicity on computation. On the other hand, RLS or its variants are commonly used in control system applications. However, RLS is less suitable for high-bandwidth real-time applications.

While it has always been able to estimate the overall transfer function using methods based on adaptive algorithms such as LMS or RLS, there remains the problem of polynomial factorization to get at the unknown non minimum phase terms.

Let us consider a FIR (finite impulse response) z-transform with the white noise input ξ_n and assume that the white noise input ξ_n and the output s_n can be measured directly. Hence, the system output can be expressed as $s_n = H(z)\xi_n$

where $H(z) = H_M^m(z)H_N^n(z)$. (4-26)

Now, the power spectrum of s_n is

$$\Phi_{ss}(e^{jw}) = \left|H(e^{jw})\right|^2 \sigma_\xi^2 \tag{4-27}$$

where σ_ξ^2 is the variance of zero mean white noise.

Substituting (4-26) into (4-27) gives

$$\Phi_{ss}(e^{jw}) = \left|H_M(e^{jw})H_N(e^{jw})\right|^2 \sigma_\xi^2 \tag{4-28}$$

which can be written as

$$\Phi_{ss}(e^{jw}) = H_M(e^{-jw})H_M(e^{jw})H_N(e^{-jw})H_N(e^{jw})\sigma_\xi^2 \tag{4-29}$$

3.3.2 Identification of unknown non minimum phase system with white noise input

The front-end minimum phase kepstrum filter gives same spectrum as (4-30).

$$s_n = H_M(z)z^{-n}H_N(z^{-1})\xi_n = H_K(z)\xi_n \tag{4-30}$$

where $H_K(z) = H_M(z)z^{-n}H_N(z^{-1})$ indicates that the n zeros of the non minimum phase part have been reflected back within the unit circle via the reciprocal polynomial $z^{-n}H_N(z^{-1})$ so it has original minimum phase zero polynomial cascaded with the reflected non minimum phase zero polynomial.

The essence of this method is to estimate first the transfer function defined as $H_K(z) = H_M(z)z^{-n}H_N(z^{-1})$ by using kepstrum analysis and then to use a cascaded RLS to minimize the remaining residual error (Fig. 4-15).

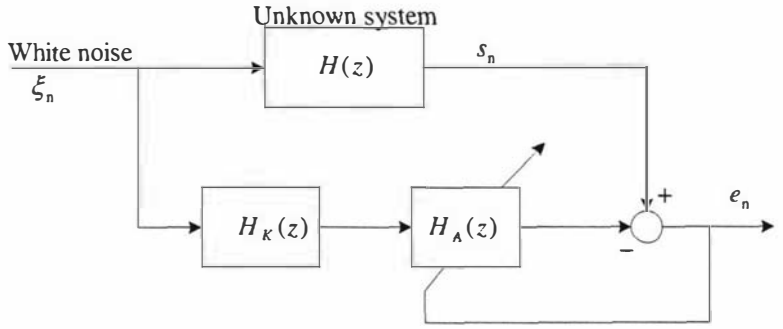


Fig. 4-15 Identification of unknown non minimum phase system with white noise input

This results in an overall estimated transfer function $H(z) = H_M(z)z^{-n}H_N(z^{-1})H_A(z)$ where $H_A(z)$ is an infinite impulse response (IIR) cascade of all-pass z-transform functions. For real non minimum phase zeros, the all-pass transfer function will be first order and for complex (2^{nd} order) non-minimum phase zeros, the all-pass transfer function will be complex 2^{nd} order. The numerator polynomial of $H_A(z)$ will consist of the non-minimum phase zeros only, isolated from the remainder of the transfer function, where $H_A(z)$ gives a pole-zero representation i.e., proper all-pass estimate so non minimum phase zeros can be identified from the numerator of all-pass functions. Therefore, the form of $H_A(z)$ consists of two parts, numerator polynomial of non minimum phase zeros only and denominator polynomial of its reflected minimum phase poles shown as

$$H_A(z) = \frac{H_N(z)}{z^{-n}H_N(z^{-1})} \tag{4-31}$$

where the reciprocal polynomial $z^{-n}H_N(z^{-1})$ has all of its roots within $|z|<1$ making $H_A(z)$ stable. A similar method which relies entirely on recursive estimation is given in Lim and Macleod (**Lim and Macleod,** 1994).

For example, suppose for a non-minimum phase system that

$H(z) = H_M^1(z)H_N^1(z) = (1-0.1z^{-1})(1-2z^{-1})$

Then it shows that it can be easily identified from the original and reflected minimum phase zeros from the front-end kepstrum filter. And non minimum phase zeros can be uniquely

identified from the numerator polynomial of cascaded all-pass transfer function and reflected minimum phase poles can be identified from the denominator polynomial of cascaded all-pass transfer function.

$H_K(z) = H_M^1(z)z^{-1}H_N^1(z^{-1}) = (1-0.1z^{-1})(-2)(1-0.5z^{-1})$ and

$$H_A(z) = \frac{H_N^1(z)}{z^{-1}H_N^1(z^{-1})} = \frac{(1-2z^{-1})}{(-2)(1-0.5z^{-1})}$$

3.3.3 Identification of unknown non minimum phase system with white noise input plus additive white noise

With the application of an additive white noise, it will be shown that it is not able to identify the non minimum phase term directly by RLS algorithm. It indicates that the technique can be only applied to noise-free cases. However, for the case of additive white noise, RLS or LMS can be used as a method of reducing the number of weights when LMS or RLS is cascaded after the kepstrum method. Therefore, the majority of the system can be estimated using the kepstrum method and the remaining non minimum phase terms identified using RLS or LMS. Here Fig. 4-16 shows that the additive white noise is added with a zero-mean additive white measurement noise term v_n and its variance σ_v^2.

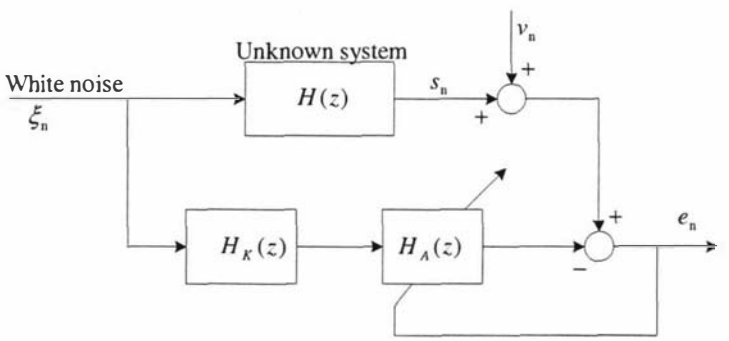


Fig. 4-16 Identification with a white measurement noise

If the minimum phase kepstrum filter $H_K(z)$ is to be identified using the kepstrum method then the kepstrum method will estimate the minimum phase spectral factor $H_K(z)$ found from

$$H_K(z)H_K(z^{-1})\sigma_\varepsilon^2 = H(z)H(z^{-1})\sigma_\xi^2 + \sigma_v^2 \qquad (4\text{-}32)$$

where σ_ϵ^2 is the variance of the white noise innovations process (**Barrett and Moir,** 1987). The error in Fig.4-16 is given by

$$e_n = [H(z) - H_A(\varsigma)H_K(\varsigma)]\xi_n + v_n \tag{4-33}$$

We then minimize the mean-square error σ_e^2 from Fig. 4-16 by completing the square (**Barrett and Moir,** 1987). We assume that the two white noise sources are uncorrelated and use the notation that '*' represents complex conjugate.

$$\sigma_e^2 = \frac{1}{2\pi j}\oint_{|\varsigma|=1}[(H_A(z)H_K(z) - H(z))(H_A(z)H_K(z) - H(z))^*\sigma_\xi^2 + \sigma_v^2\frac{dz}{z}] \tag{4-34}$$

Then (4-34) is a minimum when

$$H_A(z) = \frac{H(z)}{H_K(z)} \tag{4-35}$$

leaving a minimum mean-square error

$$\sigma_{e(\min)}^2 = \frac{\sigma_v^2}{2\pi j}\oint_{|\varsigma|=1}\frac{dz}{z} = \sigma_v^2 \tag{4-36}$$

Clearly (4-35) is no longer an all-pass transfer function like (4-31) and $H_N(z)$ does not appear in the numerator. Instead the non-minimum phase term appears with the overall unknown transfer function $H(z)$. Therefore, for the case of additive white noise, there is no great advantage in using this method since factorization would have to be used.

■

Chapter 5

Experiments

Experiments are implemented in a real-time processing by a software implementation in LabVIEW (laboratory virtual instrument engineering workbench) in real environments, which are a typical office, indoor room with moderate reverberation conditions. This chapter covers a description of real-time processing, an experiment set-up, application using LabVIEW software; the selected VAD and a performance evaluation method; and then test results and discussions through simulation and real-time tests.

1. Real-time processing

1.1 Definition of real-time processing

For the definition of the word 'real-time', many definitions have been found. However, most of them are contradictory and hence controversial. There does not seem to be 100% agreement over the terminology. The following is the definitions found from dictionaries and researchers.

- A real-time system is one in which the correctness of the computations not only depends upon the logical correctness of the computation but also upon the time at which the result is produced. If the timing constraints of the system are not met, system failure is said to have occurred (gillies@ee.ubc.ca).

- The ability of the operating system to provide a required level of service in a bounded response time (POSIX standard 1003.1)

- 1) Time in which the occurrence of an event and the reporting or recording of it are almost simultaneous. 2) The actual time used by a computer in solving a problem the answer to which is immediately available to control effectively a process that is going on at the same time (Webster's dictionary) (**Agnes and Guralnik,** 2000).

- Pertaining to the performance of a computation during the actual time that the related physical process transpires so results of the computation can be used in guiding the physical process (LabVIEW user manual G-11 NI corp.).

- Real-time is a form of transaction processing in which each transaction is executed as soon as complete data becomes available for the transaction. Therefore, real-time processing requires a fast enough data processing to keep up with an outside process (http://www.wordwebonline.com).

- Two types of real-time systems 'soft' and 'hard'. A hard real-time means the type of a typical real-time system which requires a stringent deadline, and soft real-time means a system which has reduced constraints on 'lateness' but still must operate very quickly (http://media. wiley.com).

According to the several definitions, a common basis of a real-time system requires one that must satisfy explicit bounded response time constraints without system failure, with the logical correctness based on both the correctness of the outputs and their timeliness. The response time is called the time between the presentation of a set of inputs and the appearance of all the associated outputs.

A fast response time is not only needed to characterize a real-time system but it simply must have response times that are constrained (or called deadlines) and thus predictable. Most systems can claim to look as if they are real time. Therefore, a refined definition may be to make the classification between 'hard' and 'soft' real-time systems based on their properties, somewhat in terms of the system's tolerance to missed deadlines. Hard real-time systems are regarded as those where failure to meet even one deadline results in total system failure and, in soft real-time systems, missing deadlines leads to performance degradation with reduced constraints on lateness but not failure. However, the definition is controversial, as some users take 'hard' and 'soft' to mean the degree of time constraints.

■

1.2 Performance measurement of real-time processing

A real time system performance is often measured by time-loading or CPU utilization, which is a measure of the percentage of non-idle processing. A system is said to be time-overloaded if it is 100% or more time-loaded. Time-overloading occurs in interrupt-driven systems when higher priority interrupt-driven tasks execute too frequently to allow lower priority tasks to finish on time. Systems that are time-overloaded are unstable and exhibit missed deadlines and unpredictable response times. Table 5-I shows CPU utilization ranges

for a real-time capability. Utilization factors in the 0%-69% range are generally considered as safe. Beyond 70% they have a high risk of missing deadlines, and above 100% is potentially disastrous.

Table 5-I: CPU utilization zones

Utilization %	Zone type	Application types
0 – 25%	Overkill	Various
26 – 50%	Very safe	Various
51 – 68%	Safe	Various
69%	Theoretical limit	Various
70 – 99%	Dangerous	Embedded systems
100% +	Overload	Stressed systems

Fig. 5-1 shows a snapshot of CPU usage on a window task manager during a kepstrum processing task. The snapshot value shows an instant status in the most right ending point with previous values in a waveform, which needs to be averaged for the performance evaluation. This example shows the latest value of processing status, which is 63% according to a classified percentage of CPU utilization, which is based on computer processors of Intel(R) Pentium® 4 CPU 2.8 GHz, ACPI multiprocessor and 1Gb RAM memory.

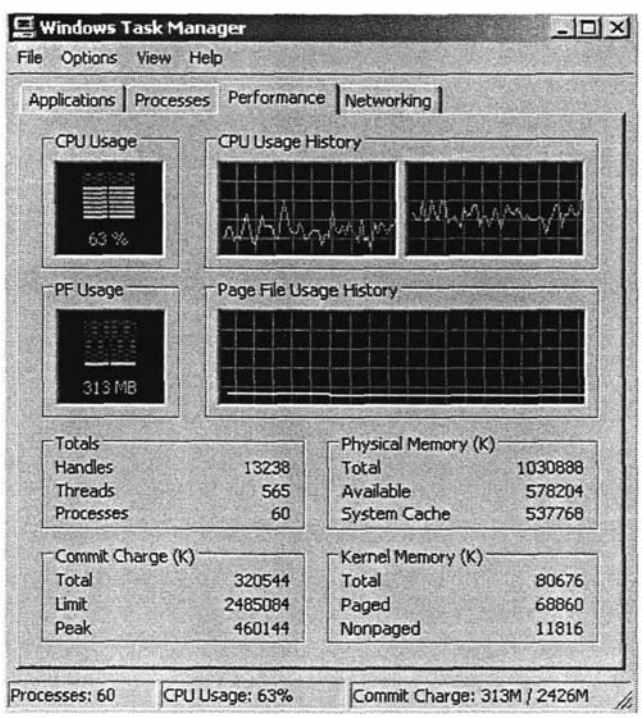


Fig. 5-1 Snapshot of CPU usage on window task manager

Based on this, the average CPU usage of kepstrum and LMS algorithm has been measured. CPU usage between kepstrum processing (A) and kepstrum processing (B) (see p. 117-118) is also compared. The results are listed in Table 5-II, which shows that:

- Kepstrum processing (B) shows an efficient and better real-time CPU usage than LMS adaptive filter according to comparison of kepstrum coefficients and LMS weights. The use of 64 coefficients of kepstrum and 64 weights of LMS gives a better CPU usage than 200 weights of LMS.

- Kepstrum processing (B) shows much better performance than kepstrum processing (A) when the coefficient size is large while the two processing methods show almost the same CPU usage when a small number of coefficients size are used.

Table 5-II Comparison of CPU usage in kepstrum and LMS algorithm

Kepstrum processing	Kepstrum coefficients	LMS weights	Average CPU usage (%)
(B)	64	64	50
—	0	200	60
(B)	200	0	35
(A)	64	—	40
(B)	64	—	39
(A)	1000	—	83
(B)	1000	—	45

For the comparison of computational complexity, a computational complexity for real-time processing is measured by the complexity of multiplication in FLOPS (floating point operations per second). Table 5-III shows the required processing and number of FLOPS for computational complexity of the kepstrum and LMS algorithm.

Table 5-III Comparison of kepstrum and LMS algorithm in FLOPS

Algorithm	Required processing	FLOPS
Kepstrum	WOSA (A)	$N \log_2(5.12/\Delta f)$
	FFT/IFFT (B)	$(N/2)\log_2 N$
	Logarithm/exponential (C)	$N^{1/3}(\log N)^2$
	Total computation	2A + 5B + 3C
NLMS	Real multiplication (D)	$3N^2 + 2N$
	Iterations (E)	20
	Total computation	D·E

For the case that 200 NLMS weights are used, real multiplication is 0.12G ($G = 10^6$) and 2.4G for its iteration to convergence. On the other hand, for kepstrum processing, the total computation is 0.08G per N =2048 samples and a highly reduced processing time can be expected if a small number of 64 kepstrum coefficients are used.

■

1.3 Real-time application

The research work has been done in real-time processing by using LabVIEW® software on a personal computer for an FFT based kepstrum approach. For the application of kepstrum signal processing, a personal computer is used as a system device for software implementation. The system input receives an unprocessed signal from microphones and its output sends out a processed signal in real-time through software implementation on a personal computer. In doing so, there is a certain amount of latency from input to output but for many of the aforementioned applications, this is not restrictive. Therefore, the results can be defined as being real-time since no machine is possible of instantaneous performance other than analogue signal processing. A FFT based real-time bandwidth power spectrum estimation and its overlapping processing techniques ensure a fast processing time by using LabVIEW software.

■

2. Experimental set-up

Experiments have been implemented in two rooms, where the test places are at desk A of both rooms as illustrated in Fig. 5-2. The room dimensions are shown. The background noise level is measured as 48.5dBA for room (A) and 48dBA for room (B) by using a sound level meter (Digitech QM1589). Fig. 5-3 shows a test environment, where desk A is the designated place for the test in each room and all the position for the test are measured from the left upper corner of each room in d(m) x w(m) x h(m). One radio is used for the different scenarios of noise according to three different locations (R1, R2 and R3). However, speaker and computer positions are fixed.

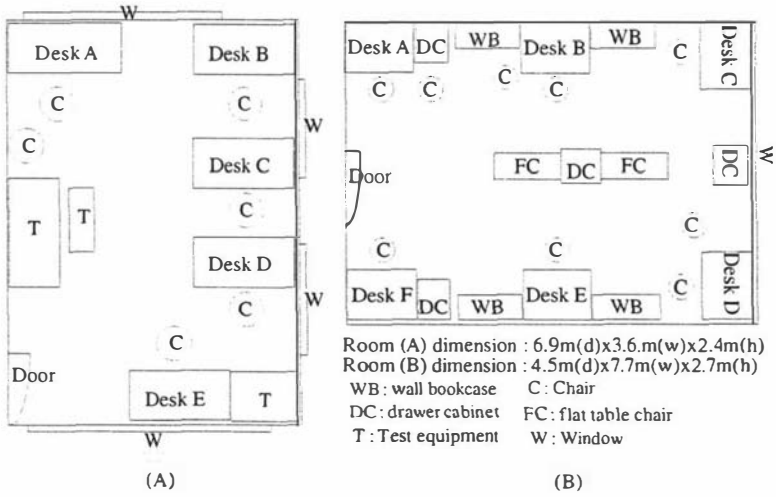


Fig. 5-2 Room environment: room (A) and room (B)

The speech signals are sampled using a standard internal sound card and two preamplifiers with unidirectional/omnidirectional electret condenser microphones. The sampling frequency is chosen to be 22050Hz with 16 bits amplitude resolution per channel, which gives quite a high quality performance as the Nyquist frequency bandwidth is around 11 kHz.

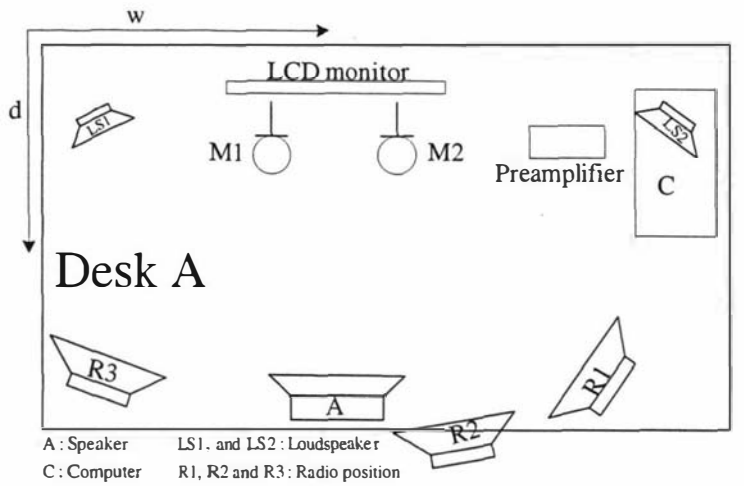


Fig. 5-3 Test environment (desk A)

The room dimension and location of sound sources and equipments are listed in Table 5-IV.

Table 5-IV Room dimension and location of sound sources and equipments

Room dimension	Room(A)	6.90 x 3.60 x 2.40
	Room(B)	4.50 x 7.70 x 2.70
Location of sound source and equipment	Speaker (S)	0.85 x 0.60 x 1.10
	Computer fan (C)	0.10 x 1.10 x 0.70
	Radio1 (R1)	0.70 x 1.10 x 0.70
	Radio2(R2)	1.10 x 0.90 x 0.80
	Radio3(R3)	0.80 x 0.20 x 0.70
	Primary MIC(M1)	0.40 x 0.50 x 0.90
	Reference MIC(M2)	0.40 x 0.70 x 0.90
	Monitor	0.25 x 0.45 x 0.75
	Loudspeaker 1(L1)	0.40 x 0.05 x 0.70
	Loudspeaker 2(L2)	0.30 x 1.15 x 0.85
	Preamplifier	0.25 x 0.90 x 0.70

d(m) x w(m) x h(m)

Room reverberation time (see p. 99) is calculated as 0.79 seconds for room (A) and 1.18 seconds for room (B) from the information of wall materials and absorption coefficients for the frequency 500Hz as shown in Table 5-V. The typical absorption coefficients of various surfaces of wall, floor and ceiling can be found from Donald (**Donald,** 2001). Based on the calculation of reverberation time, it is assumed that both rooms reflect a moderately reverberant situation.

Table 5-V Information of wall material and absorption coefficient according to room environment

Information for room reverberation time calculation		Wall 1	Wall 2	Wall 3	Wall 4	Floor	Ceiling
Room (A)	Material	Glass window	Plaster board	Glass window	Glass window	Carpet on wood	Plaster board
	Absorption coefficient	0.2	0.05	0.2	0.2	0.4	0.05
Room (B)	Material	Plaster board	Plaster board	Plaster board	Heavy plate glass	Carpet on concrete	Plaster board
	Absorption coefficient	0.05	0.05	0.05	0.04	0.21	0.05

■

3. LabVIEW software

The experiment is implemented by using LabVIEW® software of National Instruments (**Johnson,** 1994) on a typical high-performance dual-processor personal computer. The program is called 'g' graphical programming language and this data flow language uses a block diagram for a solution of system modelling and design, which is also suitable for a prototype application as a precursor for real-time hardware implementation using DSP (digital signal processing) processor. By using a real-time bandwidth and overlap processing, the program provides a front-panel graphical output and it acts like a virtual instrument in the industry standard. It enables a real-time monitoring, controlling and measuring of time domain and frequency domain parameters with a data recording and a play back ability.

Fig. 5-4 shows a front-panel of LabVIEW which is programmed by a block diagram. The front panel is comprised of three main parts, (A) control panel, (B) indicator panel and (C) display panel, which are designed for easy monitoring, controlling and measuring during the real-time implementation.

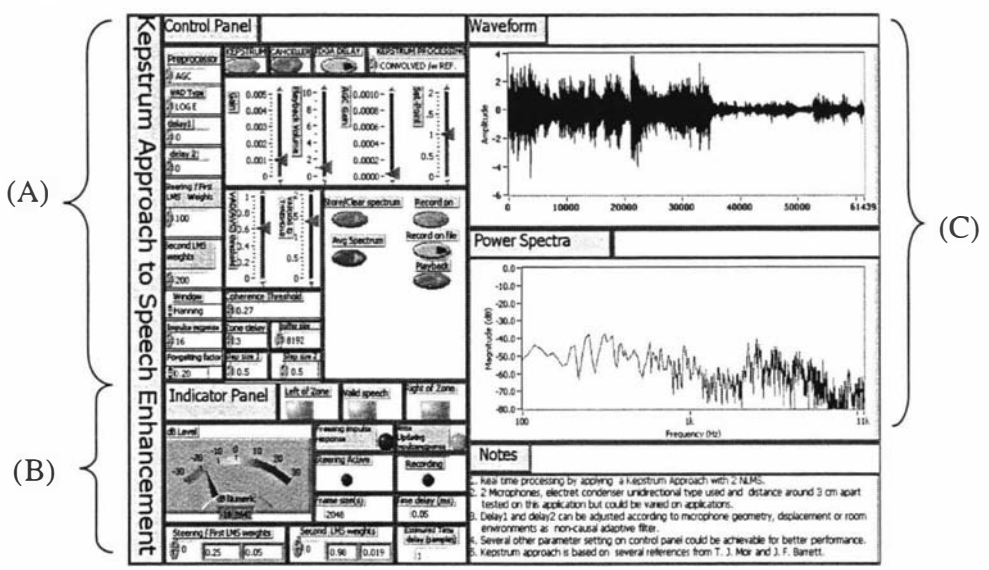


Fig. 5-4 Example of front panel in LabVIEW: (A) control panel (B) indicator panel (C) display panel for time domain waveform and frequency domain power spectra

- **Control panel:** Part (A), a control panel allows a processing control during the real-time processing or simulation test. It is designed to select VAD (MSC and TDOA, log energy, and variance), frame size, window function (hanning and hamming), kepstrum processing of type (A) and (B). The application of adaptive filter, kepstrum filter and TDOA delay is available by an on/off pushbutton. It is possible

to change the data of the delay filter, adaptive filter and kepstrum filter sizes as well as adjust data for the VAD threshold setting.

- **Indicator panel:** From the indicator panel of part (B), VAD operation can be verified from panels of valid speech zone, and left or right of the zone for a noise source indication. Adaptive performance of an adaptive filter can be verified from the panels showing freezing and updating weights. Time delay estimate can be verified, where it is measured in milliseconds or in sample of TDOA function. It is available to verify the performance in dB meter, where it is normally used for measuring overall dB noise reduction (**Moir,** 2006).

- **Display panel:** The on-going performance can be monitored in (C) display panel for the performance in a time domain waveform and frequency domain power spectrum. The data can be stored for further analysis using audio editing tool or the performance can be played back by a command on a control panel.

Table 5-VI shows a specification for the parameter setting and it is set as a default. It can easily be changed during the processing time or off-time. Information for experimental equipments is also shown. Note that default parameters and listed equipment are normally used if not differently specified in the each experiment.

Table 5-VI Specification for default parameter setting and information for experimental equipments

Specifications for parameter setting and equipment information				
Delay1 (D_1)		0	Frame size	2048
Delay2 (D_2)		0	Sampling frequency	22050 Hz, 16 bits/ch
Steering/first NLMS weights (H_1)		100	Nyquist frequency	11025 Hz
Second NLMS weights (H_2)		200	Experimental software	LabVIEW 7.0
Step size of NLMS		0.50	Computer (CPU)	Intel P4 2.8GHz with 1Mb cache, 1 Gb RAM
Window		hanning		
Buffer size		8192	Signal analyzer	Cooledit pro 2.0
Forgetting factor		0.80	Pre-amp	Alto stereo tube
VAD setting	TDOA delay	3 samples	Sound card	Internal, onboard
	MSC	0.40	Microphones	1. Unidirectional electret condenser 2. Omnidirectional electret condenser
Kepstrum coefficients		64		

A block diagram is mainly composed of six parts and its simplified block diagrams are shown in Fig. 5-5. Information for the processing procedure according to each block diagram is listed in Table 5-VII.

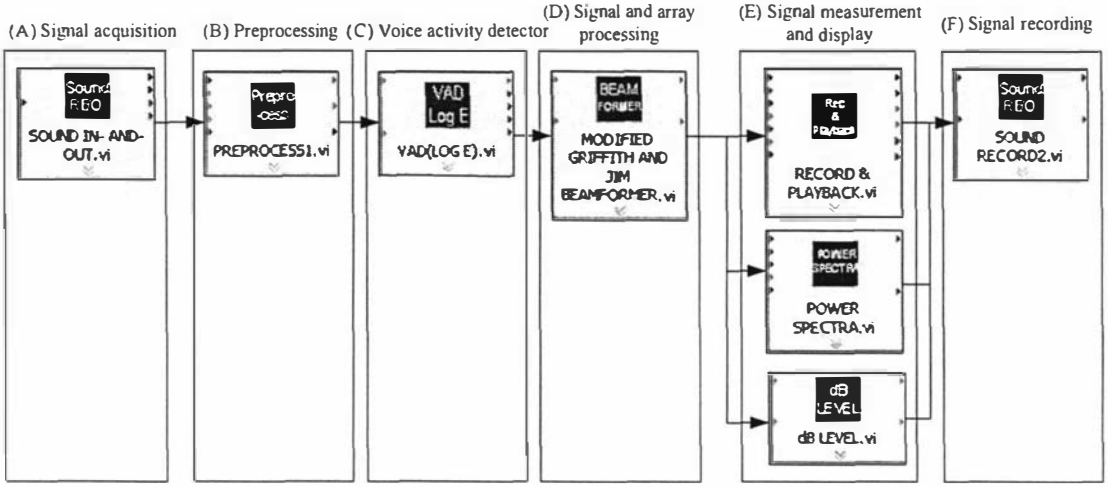


Fig. 5-5 A simplified block diagram in LabVIEW

Table 5-VII Information of simplified block diagram

(A)Signal acquisition	Sound input – stereo, 22050Hz, 16 bits Sound output – mono, 22050Hz, 16 bits
(B) Preprocessing	HPF (High-pass filter) - Cut-off frequency: 50Hz
(C) Voice activity detector	1) TDOA and MSC 2) Variance 3) Log energy
(D) Signal and array processing	Method I- ANC based approach Method II- G-J based approach – with TDOA Method III- G-J based approach – with TDOA and adaptive filter Method IV – G-J based approach – with speech beamforming and adaptive filter
(E) Signal measurement and display	1) Real-time measurement and display in waveforms 2) Real time measurement and display in power spectra 3) Real time measurement and display in dB ratio
(F) Signal recording	Permanent recording

In addition to performance analysis from the front panel of LabVIEW, an audio editing tool, cooledit pro® software has been used for the performance measurement, spectral and spectrogram analysis, which can provide more detailed analyses (Fig. 5-6).

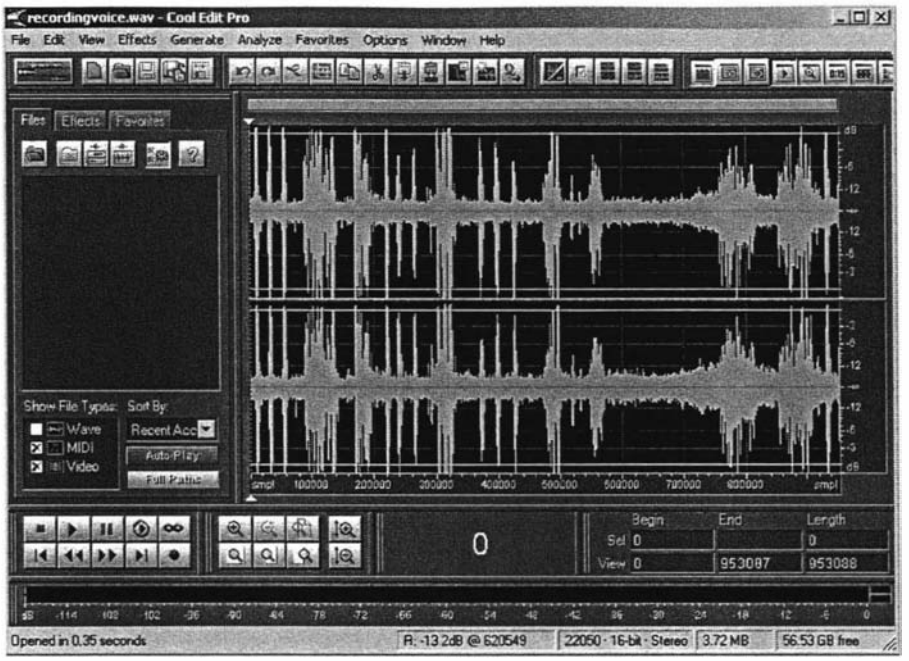


Fig. 5-6 Analysis using audio editing tool, cooledit pro 2.0

4 Performance evaluations

4.1 SNR measure

The SNR was the widely used measure in the past for measuring enhancement algorithms and it is defined as (5-1).

$$\mathrm{SNR(dB)} = 10\log\left(\frac{\sigma_s^2}{\sigma_n^2}\right) \tag{5-1}$$

where σ_s^2 and σ_n^2 are the variance of signal and noise respectively.

NRN (normalized residual noise) method is sometimes used for the performance evaluation of noise cancellation. It can be measured in terms of the ratio of noise signal powers at the output and input in the absence of speech (**Mikhael and Hill,** 1988; **Lu and Clarkson,** 1993), which is defined as (5-2).

$$\mathrm{NRN(dB)} = 10\log\left(\frac{\sigma_{n(A)}^2}{\sigma_{n(B)}^2}\right) \tag{5-2}$$

where A and B represent the noise signals after and before noise cancellation respectively.

The NRN is frequently used to measure the performance of the ANC implemented in diverse situations or with different gradient algorithms. In addition, the SNR improvement can be measured from the subtraction of the output signal SNR and the primary input signal SNR, yielding

$$\text{SNR improvement (dB)} = 10\log\left(\frac{\sigma_{s(A)}^2}{\sigma_{n(A)}^2}\right) - 10\log\left(\frac{\sigma_{s(B)}^2}{\sigma_{n(B)}^2}\right) = 10\log\left(\frac{\sigma_{s(A)}^2}{\sigma_{n(A)}^2}\frac{\sigma_{n(B)}^2}{\sigma_{s(B)}^2}\right) \tag{5-3}$$

The equation (5-3) may be identical as long as there is no signal leakage at the reference input. In this case, (5-3) becomes (5-4).

$$\text{SNR improvement (dB)} = 10\log\left(\frac{\sigma_{n(B)}^2}{\sigma_{n(A)}^2}\right) \tag{5-4}$$

It gives a reciprocal relation between NRN (5-2) and SNR improvement (5-4).

The main benefit of the SNR quality measure is its mathematical simplicity. However, it is found as poor measure of speech quality for a broad range of speech distortions. Another fact indicates that SNR is not well related to any subjective attribute of speech quality.

■

4.2 Segmental SNR measure

For the measurement of SNR in real-time processing, there is a measurement problem with our real-world application because the true signal is unknown (i.e., un-measurable due to noise) and hence its power cannot be calculated on its own. It has been shown that the usual way to calculate SNR for such problems is to measure the power of the noise when the speech signal is absent and then measure the power of the signal plus noise (assuming that the noise has not changed its characteristics). By subtracting the two measures (in dB), 1+SNR (in dB) is found from which SNR can then be calculated as in (5-5)

$$1+\mathrm{SNR}=10\log\left(\frac{\sigma_s^2+\sigma_n^2}{\sigma_n^2}\right) \tag{5-5}$$

where σ_s^2 and σ_n^2 are the variance of signal and noise respectively. In general, variance σ^2 is easily calculated for M samples according to $\sigma^2=\frac{1}{M}\sum_{i=1}^{M}x_i^2$, where x_i, i =1, 2,...M are samples of the signal.

A speech signal has time-varying statistics so a much improved quality measure can be expected if SNR is measured over short frames of quasi-stationarity and the results are averaged. This frame-based measure has been called the segmental SNR.

This assumes of course that the noise power does not change throughout the whole time of measurement, which is in reality a false assumption with nonstationary data. However, it has shown that it is a good approximation and the only method available other than simulation results.

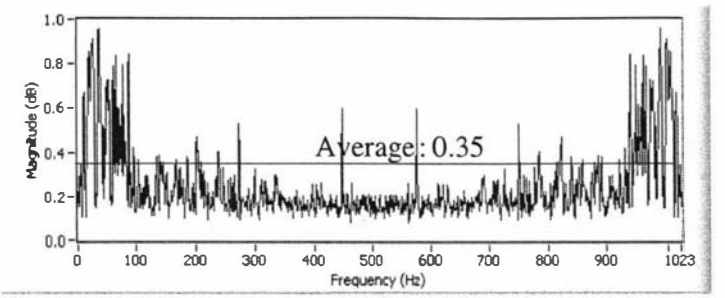
■

5. VAD

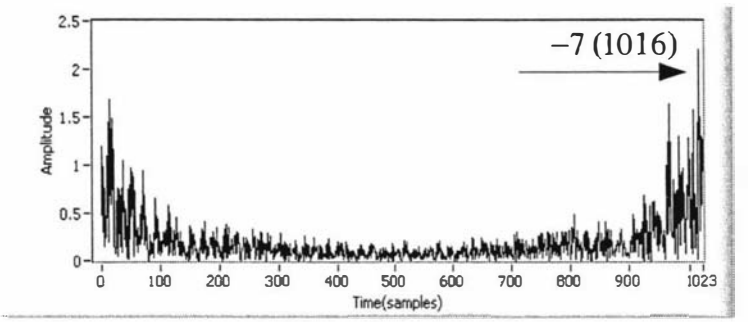
A robust VAD (**Agaiby and Moir,** 1997, a) based on the MSC (2-129) and the TDOA (2-155) function has been used for the test in a real reverberant environment. It is found that GCC based TDOA estimate using HT prefilter gives a better estimate of time-delay than ordinary cross-correlation, particularly in reverberant environments and MSC provides a frequency analysis of the correlation level between the signals of the two microphones (see p. 60-68). This combined method of using MSC and TDOA function gives more constraints, but a better performance than the method using a sole criterion, MSC based VAD method (**Le Bouquin Jeannes and Faucon,** 1994). Therefore, the simple methods based on log energy (**Gerven and Xie,** 1997) or variance (**Moir,** 2001) may also be used with MSC function for the application in reverberant environments.

For the experiment, the output from average MSC>0.40 and TDOA $<|3|$ samples $(136.05\mu s)$ is regarded as a valid speech frame. Speech frames are shown as flagged in the time-domain waveform. Fig. 5-7 shows an example of a VAD decision regarded as a noise frame from the data of average MSC of 0.35 and TDOA of -7 (at sample number 1016 within the frame). Negative delays are shown as maxima which occur past the mid-point of the GCC. Positive delays are assumed to be in the latter first half of the cross-correlation up to $N/2-1$ points. Calibration needs to be done initially to decide which direction is taken to be negative or positive. (i.e., left or right when the microphones are directly in front of the desired speech source).

Fig. 5-8 shows an example of a frame snapshot of threshold value and current average output value from a simple VAD based on log energy (A) and variance function (B). It has been shown that these VAD algorithms (**Gerven and Xie,** 1997; **Moir,** 2001) work well in stationary noise and simple test, but it shows less performance than an algorithm (**Agaiby and Moir,** 1997, a) based on TDOA and MSC function.

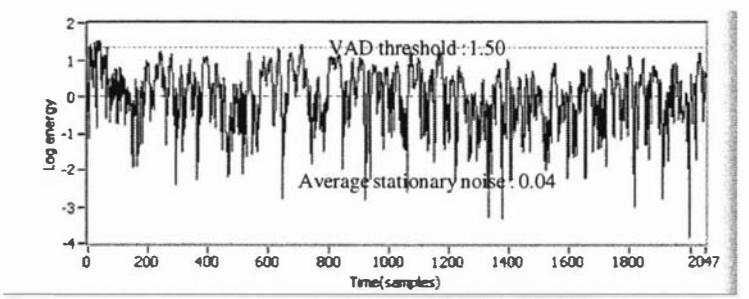


(A)

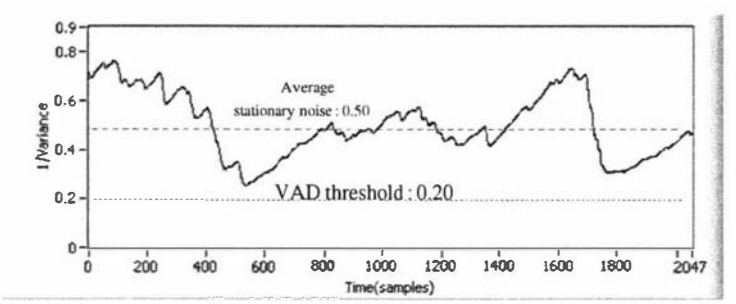


(B)

Fig. 5-7 Example of assumed noise frame found from the VAD: (A) MSC showing average 0.35 and (B) GCC estimate and TDOA showing maximum value at an interpreted -7 samples, which occurs at sample number 1016.



(A)



(B)

Fig. 5-8 Example of VAD based on (A) log energy and (B) variance function showing threshold value and current average output value in frame

■

6. Real-time and simulation tests

6.1 Comparison of omnidirectional and unidirectional microphones

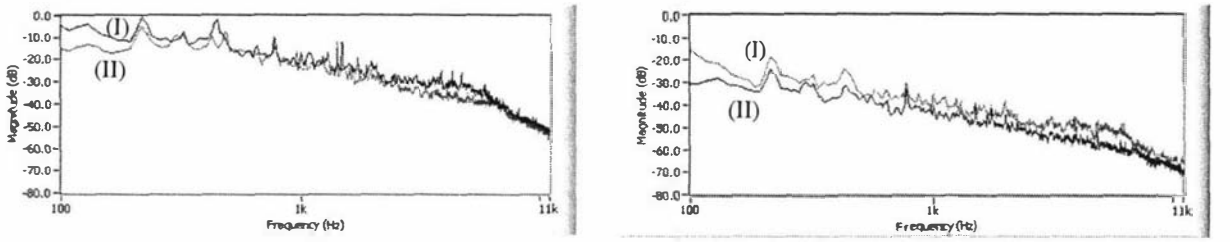
The objective is to provide a performance comparison of a noise reduction between omnidirectional and unidirectional microphones in both stationary and nonstationary noise. Table 5-VIII shows a specification of the two microphone types used. The test has been implemented at desk (A) in a room (A). Fig. 4-9 is used for the test.

Table 5-VIII Specification of microphones

Specification	Microphone type (A)	Microphone type (B)
Sensitivity polar pattern	Omnidirectional	Unidirectional
Physical material	Electret condenser	Electret condenser
Frequency response	20Hz to 16kHz	100Hz to 16kHz
Sensitivity	−65dB +/− 3dB	−68dB +/− 3dB
Impedance	Not specified	500ohms
Size	13(Dia)x30(L)mm	11.5(Dia)x25(L)mm

From Fig. 5-9 and 5-10, it shows that:

- From the application to both G-J beamformer and G-J kepstrum beamformer, the unidirectional microphone gives a better performance in both a stationary computer fan and nonstationary radio music noise.



(A) (B)

Fig. 5-9 Comparison of (I) omnidirectional and (II) unidirectional microphone based on G-J beamformer (A) and G-J kepstrum beamformer (B) in stationary computer fan

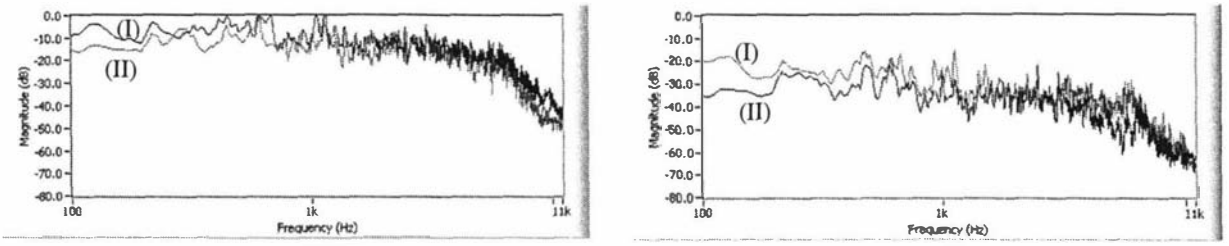


Fig. 5-10 Comparison of (I) omnidirectional and (II) unidirectional microphone based on G-J beamformer (A) and G-J kepstrum beamformer (B) in nonstationary music radio

■

6.2 Comparison of a modified application to ANC and beamforming method

The objective is to provide real-time performance comparisons of a modified application to ANC and G-J beamforming (see p. 107-110). Tests are implemented in room (B) with the use of two unidirectional microphones. With the modified application, performances are to be compared between no front-end TDOA application in ANC (method I) and TDOA function as front-end application in G-J beamformer (method II). It is also compared between TDOA function (method III) and speech beamforming filter (method IV) as a front-end application in G-J adaptive beamformer. In the diagrams of the four methods in Fig. 5-11, D_1 and D_2 refer to a small delays introduced to maintain causality. In some cases, there may be more than one delay. For the performance comparisons under the same condition, both delays are set to zero for all four methods.

1) Method I: ANC based approach

   The objective is to evaluate the performance according to a modified application to ANC, which uses a small separation between two microphones with the use of a VAD during silent periods of speech. The speech appears directly in front of the microphones.

2) Method II: G-J based approach - with TDOA

   The objective is to verify a performance using TDOA delay as a speech directivity function (steering mechanism) with the same application conditions of the modified application to ANC, but not with an adaptive filter.

3) Method III: G-J based approach - with TDOA and adaptive filter

The objective is to test the performance of an adaptive version of method II, which uses a TDOA compensation delay for steering the speech to be in front of the microphones and LMS algorithm to minimize mean-squared error.

4) Method IV: G-J based approach - with speech beamforming and adaptive filter

This method uses two LMS algorithms. The first LMS algorithm is used and updated during periods of speech whilst the second LMS algorithm is used for noise reduction and updated only during periods of noise.

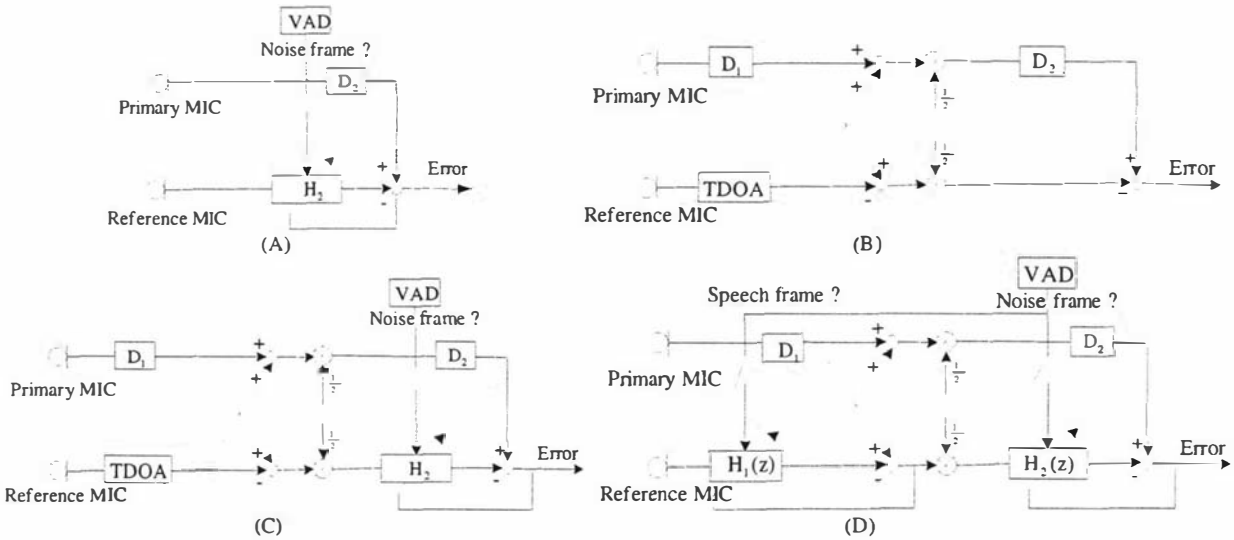


Fig. 5-11 Block diagram of (A) ANC based approach (B) G-J based approach - with TDOA (C) G-J based approach - with TDOA and adaptive filter (D) G-J based approach - with speech beamforming and adaptive filter

Three types of microphone configuration are considered for the potential application to the hearing aids for the impaired as shown in Fig. 5-12.

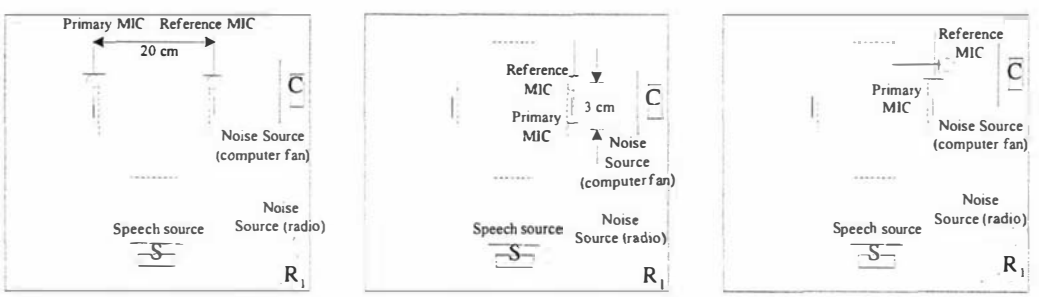


Fig. 5-12 Experimental microphone set-up
(left: broadside, center: endfire and right: endfire variant)

Following is a summary from the test results.

- The modified application to ANC (method I) and G-J beamformer (method II) shows almost the same noise reduction ratios for both stationary and nonstationary noise environments. However, in a speech with noise environment, the performance between the two methods shows a difference of up to 5dB. This indicates that the speech enhancement method enhanced by the directivity (steering) function of TDOA (method II) gives higher performance than the noise cancellation method using the modified application to ANC (method I).

- The modified application to G-J beamforming with TDOA (method III) using benefits of both methods (I) and (II) shows a considerably increased performance. The modified application to G-J beamformer with speech beamforming (method IV) shows the best performance of around 1 or 2 dB better than method (III) in all three tests.

- There appears to be little difference among the three variant methods of testing i.e. broadside, endfire and endfire variant.

Table 5-IX shows test result. Based on the test results, method IV has been tested by using delay filters ($D_1 = 5 \text{ and } D_2 = 50$) and increasing two adaptive filter sizes ($H_1 = 150 \text{ and } H_2 = 500$) from the default setting value. Performance is clearly better at low frequencies of less than about 500Hz but still behaves well up to the Nyquist frequency. Fig. 5-13 shows the corresponding time-domain result with the G-J adaptive beamformer switched on in mid-sentence in a nonstationary radio noise environment and Fig. 5-14 illustrates the average spectrum improvement on a stationary computer fan noise. It can be seen that the noise floor is lowered by over 20dB at 100Hz and around 5-10 dB therein up to 10 kHz.

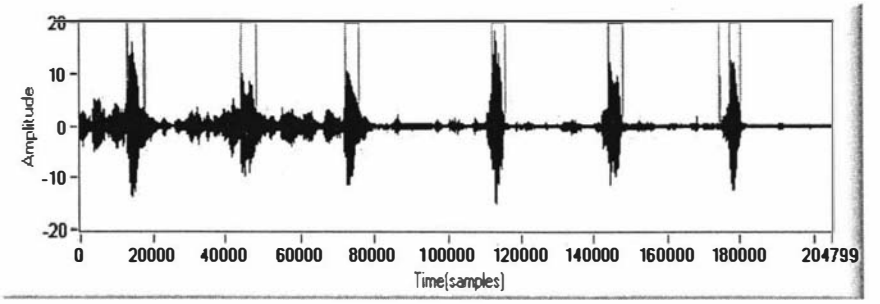


Fig. 5-13 Test waveforms on radio noise and speech
(two LMS adaptive filter is switched on in mid sentence).VAD flag is also shown

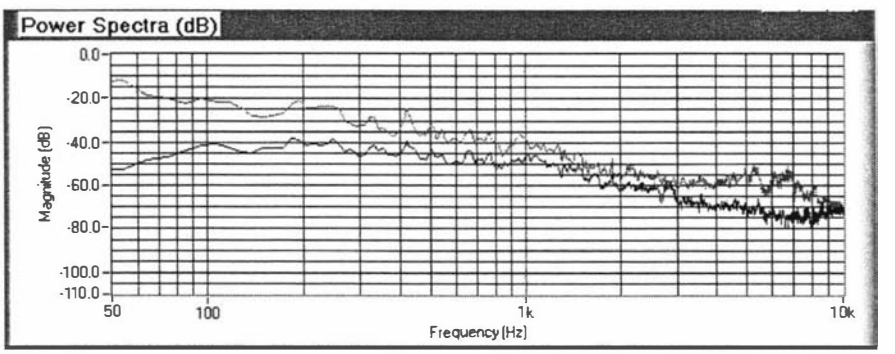


Fig. 5-14 Average power spectra on stationary computer fan noise (top line: ambient noise and bottom line: method IV with delay filters)

Table 5-IX Test results based on stationary (computer fan), nonstationary (radio) noise, with and without speech

Test type	Test type I - based on one ambient noise (computer fan)			Test type II - based on oneambient noise (computer fan) and one additive(radio) noise			Test type III - based on speech with two noises		
Configuration type	Broadside	Endfire	Endfire variant	Broadside	Endfire	Endfire variant	Broadside	Endfire	Endfire variant
Average noise power / Method type	Average noise power (dB)	Average noise power (dB)	Average noise power (dB)	Average noise power (dB)	Average noise power (dB)	Average noise power (dB)	Average noise power (dB)	Average noise power (dB)	Average noise power(dB)
Method I	– 32.04dB	– 33.46dB	– 28.51dB	– 30.18dB	– 30.08dB	– 29.58dB	– 26.04dB	– 28.51dB	– 25.61dB
Method II	– 32.27dB	– 34.45dB	– 33.29dB	– 31.39dB	– 30.87dB	– 30.24dB	– 31.44dB	– 31.50dB	– 31.65dB
Method III	– 42.53dB	– 39.04dB	– 37.81dB	– 36.12dB	– 34.95dB	– 35.77dB	– 33.07dB	– 33.12dB	– 33.25dB
Method IV	– 43.71dB	– 39.99dB	– 38.13dB	– 37.16dB	– 35.44dB	– 36.07dB	– 35.69dB	– 35.05dB	– 34.12dB

The comparison of real-time performances based on the modified application to ANC based approach and G-J based beamforming approach gives an initiative for the kepstrum method to apply to G-J beamforming rather than ANC. Therefore, the kepstrum method is applied in front of the speech enhancement method, G-J beamforming.

■

6.3 The effect of front-end kepstrum application in a reverberant environment

The objective is to test the effect of a front-end kepstrum application to a speech enhancement method for echo or reverberation cancellation. The test is implemented in room (A) in an environment of a stationary computer fan noise. G-J beamformer of Fig. 4-9 and G-J adaptive beamformer of Fig. 4-10 is used for the test. A loudspeaker has been used to produce an echo sound.

1. Fig. 5-15 shows that pitches of echo present in an overall frequency band. G-J beamformer using unidirectional microphones shows better noise reduction than one using omnidirectional microphones, but it is not related to a cancellation of a pitched echo signal.

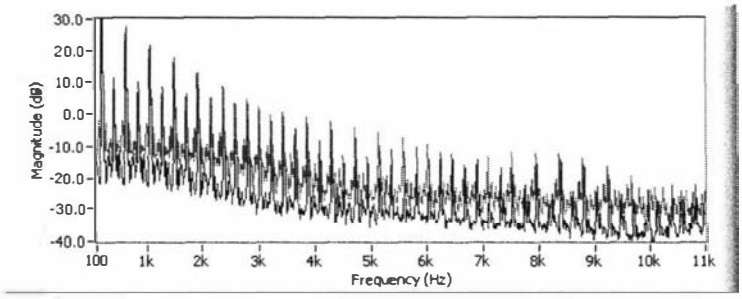


Fig. 5-15 Comparison of omnidirectional (top waveform) and unidirectional (bottom) microphone on echo sound

2. By using unidirectional microphones, the test has been carried out with G-J beamformer (Fig. 4-9), G-J adaptive beamformer (Fig. 4-10) and G-J kepstrum beamformer (i.e., kepstrum applied in G-J beamformer). Fig. 5-16 shows that the kepstrum approach shows a highly reduced noise reduction with echo cancellation, which shows a better performance than G-J adaptive beamformer.

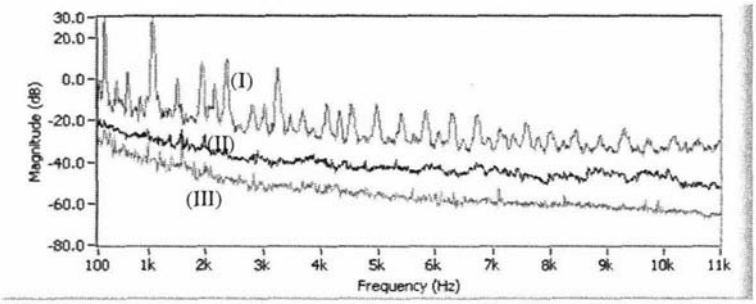


Fig. 5-16 (I) G-J beamformer (II) G-J adaptive beamformer (III) G-J kepstrum beamformer

■

6.4 The effect of minimum phase front-end kepstrum application

The favourable effect of a minimum phase front-end kepstrum application is investigated. It will test invertibility for a stable performance on an inverted room impulse response. Secondly, front-end minimum phase application to a LMS and RLS filter will be tested for the capability of the use of a highly reduced cascaded filter size and the identification of a nonminimum phase term in a cascaded RLS filter respectively.

6.4.1 Invertibility of minimum phase kepstrum application

Invertibility has been tested in room (A) by using unidirectional microphones and kepstrum processing (B) in Fig. 4-9. Fig. 5-17 shows a snapshot of waveforms of a nonstationary radio noise when (A) a kepstrum filter from (4-19) and (C) its inverted kepstrum filter (4-20) have been converted into (B) and (D) impulse responses respectively. It is shown that an inverted impulse response provides a stable performance during both a stationary and nonstationary noise environment.

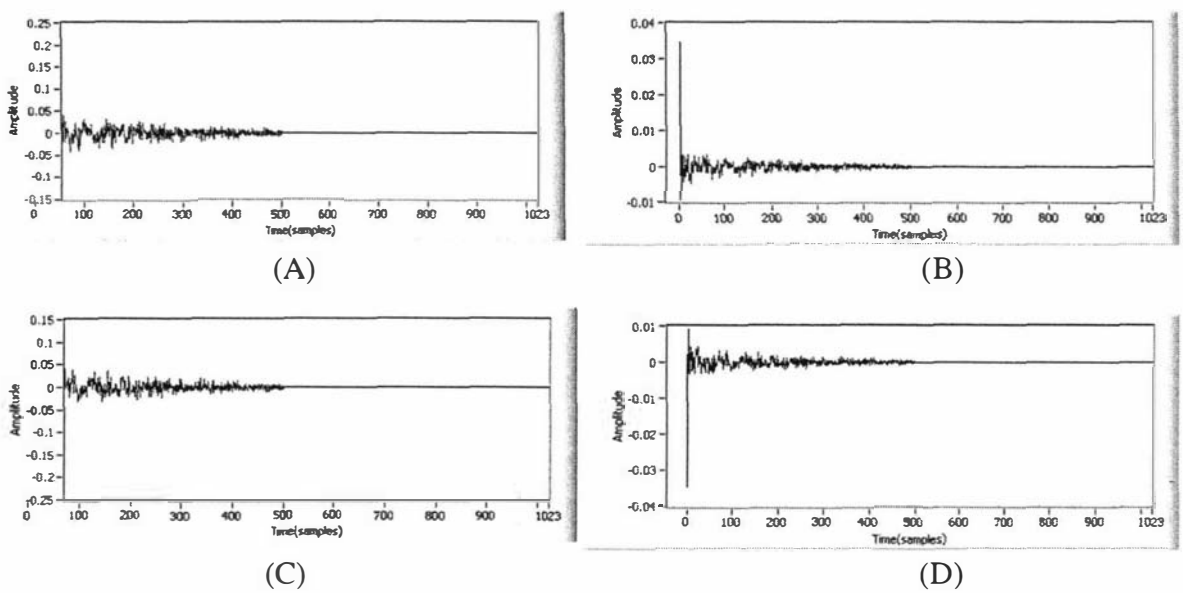


Fig. 5.17 (A) Kepstrum and (B) converted impulse response waveform and (C) inverted kepstsrum and (D) its converted impulse response

6.4.2 Front-end application of minimum phase kepstrum to a cascaded LMS algorithm

A third order FIR filter with one minimum phase zero at $z = 0.1$ and two non minimum phase zeros at $z = 2$ and $z = 5$ is used as a simple example of an unknown system for a simulation test. It can be described by the polynomial expression of $H_S(z) = H_M^1(z)H_N^2(z)$,

where $H_S(z)$ is a third order FIR filter and is taken to be unknown. 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.

$$H_S(z) = H_M^1(z)H_N^2(z) \tag{5-6}$$
$$= (1-0.1z^{-1})(1-2z^{-1})(1-5z^{-1}) = 1-7.1z^{-1}+10.7z^{-2}-z^{-3}$$

In (5-6), the non minimum phase zeros ($z = 2$ and 5) can be reflected back within the unit circle by the reciprocal polynomial $z^{-n}H_N(z^{-1})$ where $n = 2$, so becoming reflected minimum phase zeros ($z = 0.5$ and 0.2). It is known that the magnitude of a transfer function consisting of an all-pass filter (or more generally a cascaded number of all-pass filters) is unity over the entire frequency up to half-sampling. However, the poles and zeros are reciprocals of one another and new reflected poles with stability in the denominator cancel the reflected minimum phase zeros.

For the minimum phase part,

$$H_M(z) = H_M^1(z)z^{-2}H_N^2(z^{-1}) \tag{5-7}$$
$$= (1-0.1z^{-1})(-2)(1-0.5z^{-1})(-5)(1-0.2z^{-1})$$
$$= 10(1-0.8z^{-1}+0.17z^{-2}-0.01z^{-3}) = 10-8z^{-1}+1.7z^{-2}-0.1z^{-3}$$

For the all-pass z- transfer function part,

$$H_A(z) = \frac{H_N^2(z)}{z^{-2}H_N^2(z^{-1})} = \frac{(1-2z^{-1})(1-5z^{-1})}{(-2)(1-0.5z^{-1})(-5)(1-0.2z^{-1})} \tag{5-8}$$

(Here LMS or NLMS will give an estimate which is the power-series expansion of the above.)

$$= 0.1(1-6.3z^{-1}+5.49z^{-2}+.....) = 0.1-0.63z^{-1}+0.549z^{-2}+.....$$

The combined filter $H_C(z)$ from minimum phase part (5-7) and cascaded all-pass filter (5-8) shows the same result as the unknown system (5-6):

$$H_C(z) = H_M(z)H_A(z) \tag{5-9}$$
$$= 10(1-0.8z^{-1}+0.17z^{-2}-0.01z^{-3})(0.1)(1-6.3z^{-1}+5.49z^{-2}+.....)$$
$$= 1-7.1z^{-1}+10.7z^{-2}-z^{-3}+....$$
$$\cong H_M^1(z)H_N^2(z) = H_S(z)$$

1) Simulation test-1

For the simulation test to verify the performance of the kepstrum minimum phase filter and the cascaded adaptive NLMS filter, the kepstrum filter and the NLMS filter are positioned in parallel for separate identification as shown in Fig. 5-18 and in cascade for a combined identification as shown in Fig. 5-20.

With the use of a white noise input, the test result from Fig. 5-20 shows that the kepstrum filter estimates the original minimum phase zeros and reflected minimum phase zeros (from the non minimum phase zeros). On the other hand, the NLMS filter estimates the minimum phase and non minimum phase zeros together. Fig. 5-19 shows waveforms of impulse responses according to the simulation test of $H_k(z)$ and $H_L(z)$ in Fig. 5-18. It can be verified by comparing (5-7) and I-1 in Table 5-X and also (5-9) and I-2 in Table 5-X.

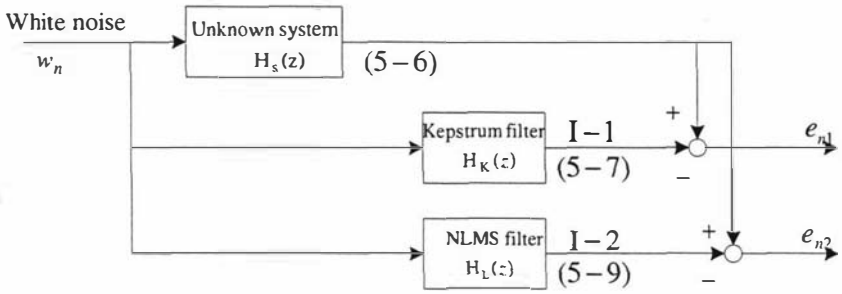
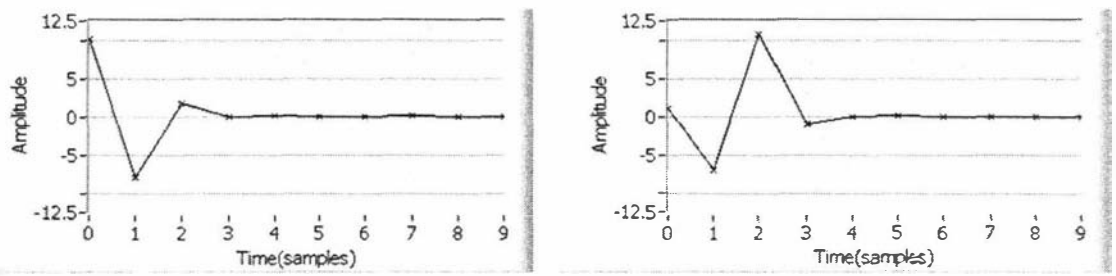


Fig. 5-18 Block diagram for simulation test for kepstrum filter and NLMS filter positioned in parallel



(A) (B)

Fig. 5-19 Waveforms of impulse responses according to simulation test: (A): I-1 and (B): I-2 in Fig. 5-18

Table 5-X Coefficient arrays showing each estimate output from the simulation test based on block diagram in Fig. 5-18

I – 1: Kepstrum estimate of impulse response (coefficients)						
10.009	−7.991	1.700	−0.102	−0.008	−0.002	0.005
I – 2: NLMS estimate of impulse response (weights)						
1.004	−7.100	10.700	−1.002	−0.007	−0.004	−0.002

2) Simulation test-2

Based on the first test result, it can be concluded from the second simulation test in Fig. 5-20 that the front-end kepstrum filter estimates the original and reflected minimum phase zeros and the cascaded NLMS filter estimates the residual non minimum phase terms realized as a cascade of all-pass transfer function estimates. The test results can be verified by comparing (5-7) and II-1in Table 5-XI, also (5-8) and II-2 in Table 5-XI respectively.

Fig. 5-21 shows waveforms of impulse responses according to a simulation test based on Fig. 5-20 and it can be verified by the signals of minimum phase impulse response, Fig. 5-19(A) and Fig. 5-21(A) showing fast decaying waveforms. This can now be compared with non minimum phase system of Fig. 5-19(B), which has the same magnitude spectrum with minimum phase terms of Fig. 5-19(A) and Fig. 5-21(A). The waveforms for the impulse response of the kepstrum minimum phase filter and the cascaded NLMS filter are shown in Fig. 5-21(A) and Fig. 5-21(B) respectively.

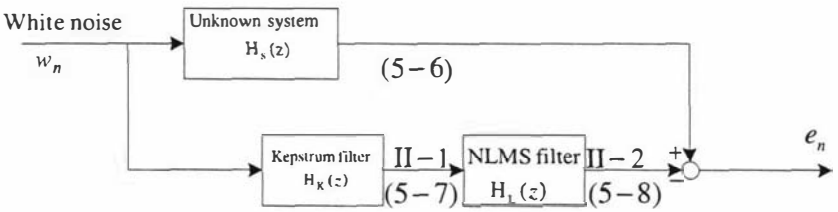


Fig. 5-20 Block diagram for simulation test for kepstrum filter and NLMS filter positioned in cascade

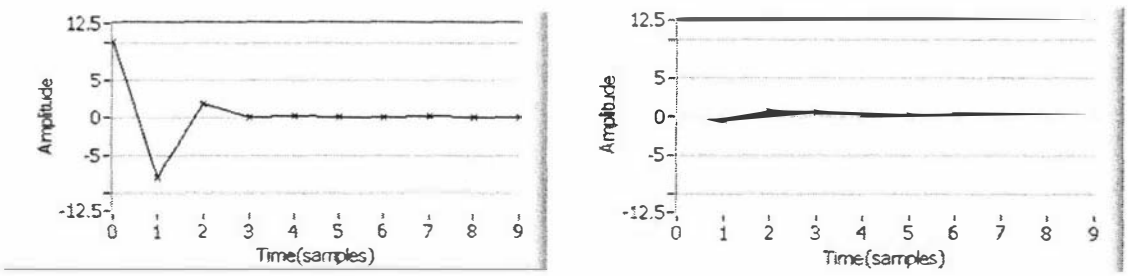


Fig. 5-21 Waveforms of impulse responses according to simulation test: (A): II-1 and (B): II-2 in Fig. 5-20

Table 5-XI Coefficient arrays showing each estimate output from the simulation test based on block diagram in Fig. 5-20

II – 1 : Kepstrum estimate of impulse response (coefficents)						
10.000	– 7.995	1.695	– 0.099	– 0.001	– 0.001	0.000
II – 2 : NLMS estimate of impulse response (weights)						
0.100	– 0.629	0.549	0.447	0.257	0.136	0.069

3) Summary

The waveform of impulse response between NLMS (Fig. 5-19(B)) and cascaded NLMS after kepstrum filter (Fig. 5-21(B)) shows a difference between combined minimum phase and non-minimum phase zeros, and residual non-minimum phase zeros. This shows that the waveform of the residual non-minimum phase term shows a much more flattened shape than the one of combined minimum phase and non-minimum phase terms. It can be induced from simulation tests with white noise input that the front-end kepstrum filter absorbs the minimum phase part so the cascaded NLMS algorithm estimates only a small amount of the residual non minimum phase terms. However, for a real-time processing with the case of additive noise, the cascaded LMS algorithm estimates any residual non-minimum phase terms that the kepstrum method has missed. This therefore results in a greatly reduced number of weights since the majority of the residual transfer function will be of the form of all-pass systems in cascade. That is, an FIR ratio of acoustic transfer functions $H(z)$ which has zeros both outside and inside the unit circle (ignoring pure delays which can be estimated separately as discussed previously using a TDOA method), can be factored as $H(z) = H_M(z)H_N(z)$ where the minimum phase term $H_M(z)$ is estimated using the kepstrum approach and the non-minimum phase cascaded term $H_N(z)$ is estimated using NLMS.

6.4.3 Front-end application of minimum phase kepstrum to a cascaded RLS algorithm

The first simulation test shows pole-zero cancellation about over-parameterization of RLS algorithm when it is used after kepstrum method on a third order noise free FIR system and the second test shows pole-zero identification of speech signal according to the locations of the speech source.

1) Example 1- over parameterization

Consider a third order noise-free FIR system which has one minimum phase zero at $z = 0.5$ and two non-minimum phase zeros at $z = 4$ and $z = 5$. This represents a polynomial of

$$H(z) = H_M^m(z)H_N^n(z) = 1 - 9.5z^{-1} + 24.5z^{-2} - 10z^{-3}$$

where

$H_N^2(z) = (1 - 9z^{-1} + 20z^{-2}) = (1 - 4z^{-1})(1 - 5z^{-1})$ and

$H_M^1(z) = (1 - 0.5z^{-1})$

The kepstrum method identifies $H_K(z) = (1 - 0.9z^{-1} + 0.275z^{-2} - 0.025z^{-3})$ which is easily shown to be $H_K(z) = H_M(z)z^{-2}H_N(z^{-1})$, the original minimum phase term and reflected minimum phase term.

The RLS algorithm converges to

$$H_A(z) = \frac{1 - 9z^{-1} + 20z^{-2}}{1 - 0.45z^{-1} + 0.05z^{-2}}$$

This is easily shown to be a second order all-pass transfer function with zeros at $z = 4$ and $z = 5$, and poles at $z = 0.25$ and $z = 0.2$. The non-minimum phase polynomial is now uniquely identified from the numerator of $H_A(z)$ as shown in Fig. 5-22.

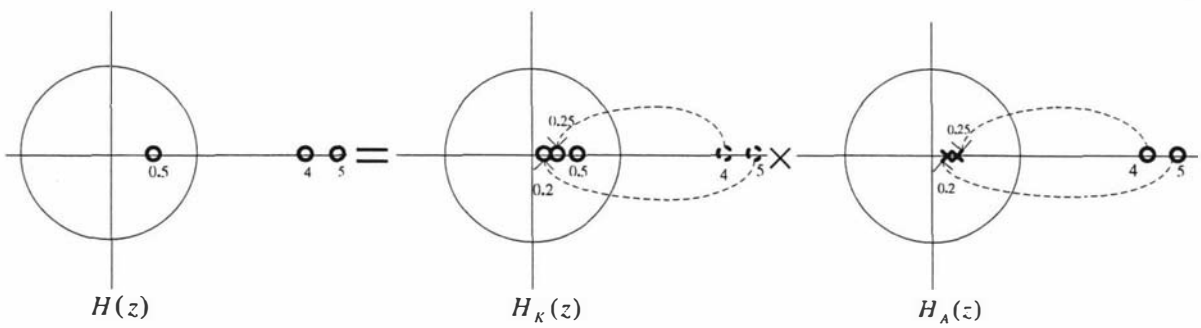


Fig. 5-22 Example of identification of non-minimum phase zeros

Now suppose that the order of the non-minimum phase term is unknown (which is the usual case). Suppose we over-parameterize the RLS algorithm from second order to sixth order. The pole-zero plot of $H_A(z)$ is shown in Fig. 5-23 for this case. It can be seen that there are 4 pole-zero-cancellations and that therefore the filter is still stable since all poles lie within the unit circle. The two zeros at $z = 4$ and $z = 5$ for clarity are not shown. It is also possible to substitute LMS (or RLS with zeros only) instead of pole-zero RLS.

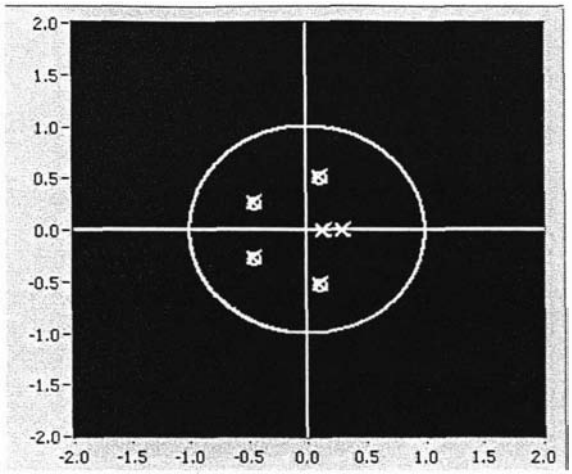


Fig. 5-23 Over parameterization of $H_A(z)$ to order six

However, the all-zero methods (LMS or its variants) identify a combination of minimum and non-minimum phase zeros and polynomial factorization would be needed to separate them. Of course it is not always necessary to obtain a separate non-minimum phase term and for such cases the kepstrum method can be used to reduce the order of the second stage parameter identification method whether that be RLS or LMS.

2) Example 2- pole-zero identification of speech signal

The purpose of this experiment is to estimate the ratio of the two transfer functions. The speech source with negligible ambient noise is located 1 meter away from two microphones approximately 30cm apart (Fig. 5-24).

Each microphone can be considered to be either the input or output of a linear system with transfer function $G_1(z)/G_2(z)$ or $G_2(z)/G_1(z)$, depending on the orientation of the microphones. Of course only one of these ratios of acoustic transfer functions is causal due to the fact that a pure delay (not included in this analysis) will be always presented in each path.

For a period of time, the speech was directly in front of the microphones and at other times, it was at an angle as shown in Fig. 5-24.

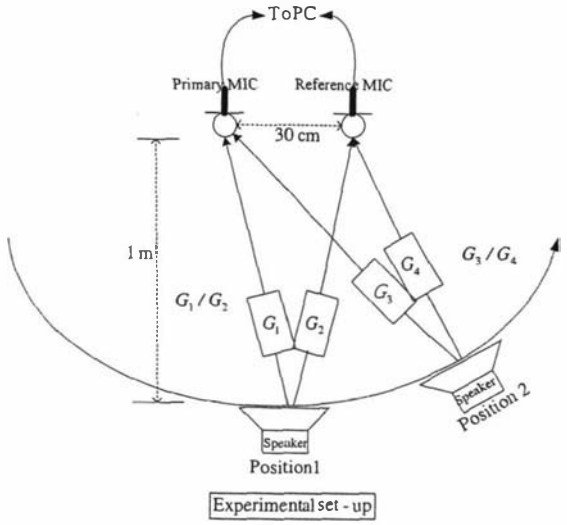


Fig. 5-24 Estimation of the ratio of two acoustic transfer functions according to different position of speech source

The purpose of the experiment was to measure using this new method if the ratio of transfer functions is often non-minimum phase in nature. The speech was recorded in a small office environment with floor space 16m square. The speech was sampled at 22050Hz 16 bits and segmented into FFT blocks of length 4096 points. Although this does not correspond to a statistically stationary period, it nevertheless gives rise to good frequency resolution. Fig. 5-25 shows the snapshot of when the speech was directly in front of the two microphones and a 10^{th} order pole-zero RLS model was used with 32 kepstrum coefficients.

From the Fig. 5-25, it can be seen that for this case the method has identified a minimum phase system and that there are at least three sets of complex near pole-zero cancellations. It should be noted that this case was the exception and that there is nearly always quite a few non-minimum phase zeros.

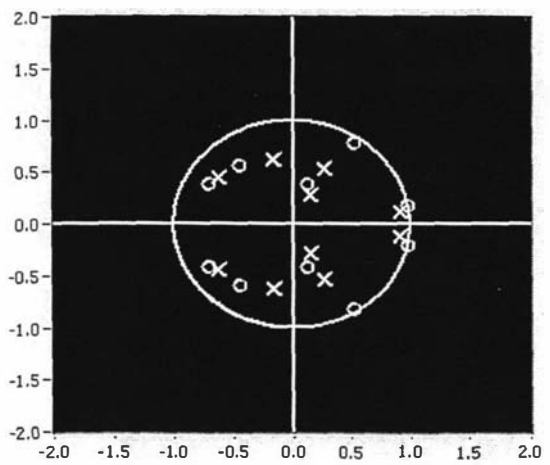


Fig. 5-25 Pole-zero identification when speech is directly in front of the microphones

Fig. 5-26 shows the snapshot when the speech is at the right hand side of the microphones. There are three non-minimum phase zeros, the two shown, which are complex conjugates and a third real zero at $z = -2.5$ outside of the viewing zone. The near pole-zero cancellations are quite apparent making the perceived order of the RLS part third order only.

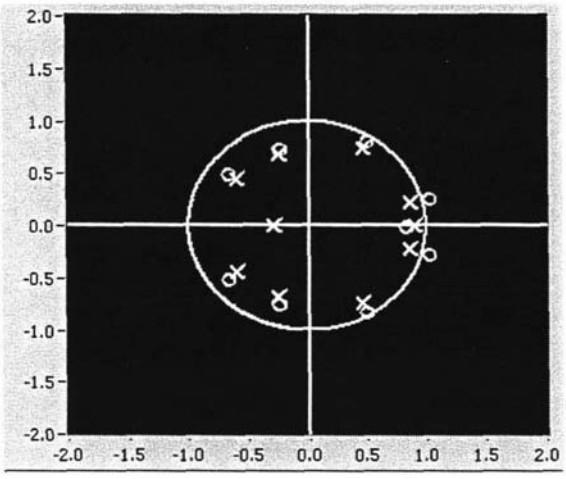


Fig. 5-26 Pole-zero identification when speech is to the right of the microphones

3) Summary

A new method has been shown to identify the non-minimum phase part of a transfer function directly without any form of polynomial factorization. The method uses a hybrid kepstrum and RLS method and is shown to be robust to over-parameterization. The technique only works for the noise-free case but can be used as a method of reducing the number of weights when RLS is cascaded after the kepstrum method. Hence the majority of the system can be estimated using the kepstrum method and the remaining non-minimum phase terms identified using RLS. Hence the most adaptive control or signal processing algorithms can use this method for system identification.

■

6.5 Comparison of kepstrum processing (A) and kepstrum processing (B)

Based on the two kepstrum processing methods (see p. 117-118), the kepstrum approach is tested on both the G-J beamformer and G-J adaptive beamformer by using unidirectional microphones in real-time in a real stationary and non-stationary noise environment in room (B). Both methods show a significant noise reduction ratio, but kepstrum processing (B) shows a better performance in noise reduction than kepstrum processing (A) from the waveforms in Fig. 5-27(A). The test result of average noise power spectra in Fig. 5-27(B)

shows that kepstrum processing (B) (by restoring phase from the causal kepstrum domain) gives a better performance than kepstrum processing (A) (using TDOA delay as phase), by providing differences of average 5-6 dB improvement in stationary noise and 6-8 dB in non-stationary noise.

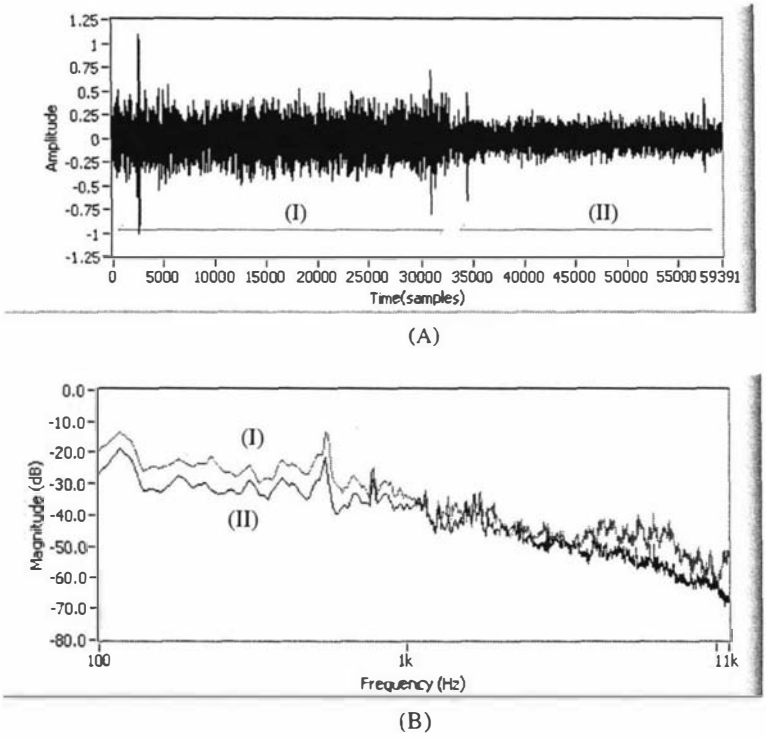


Fig. 5-27 (A) Waveforms and (B) average power spectra showing performance of stationary noise reduction based on (I): kepstrum processing(A) and (II): kepstrum processing (B).

It can be seen from the average noise power spectra in Fig. 5-27(B) that the better performance in kepstrum processing (B) comes from all frequency ranges but it is assumed that higher performance comes from the reduction in the high frequency range. Therefore, it is assumed that kepstrum processing (B) is more effective in a real reverberant environment.

■

6.6 Kepstrum approach to G-J beamforming with a modified application

Based on the result of the analyses between the kepstrum processing (A) and kepstrum processing (B), kepstrum processing (B) is used for the test with the use of unidirectional microphones in room (B). The performance of the kepstrum approach applied to a G-J beamformer (Fig. 4-9) has been compared and it shows an improvement of 12.32dB noise reduction ratio in a stationary noise environment and only 0.49dB speech enhancement ratio

in a speech and a stationary noise environment. Improvements of 15.16dB noise reduction and 3.93dB speech enhancement are obtained in a non-stationary noise environment.

On the other hand, the performance applied to the G-J adaptive beamformer (Fig. 4-10) shows improvements of 11.91dB noise reduction ratio and 4.73dB speech enhancement ratio in a stationary noise environment, and 12.41dB and 6.55dB improvement in a nonstationary noise respectively. Fig. 5-28 shows waveforms at the top showing the performance between the G-J beamformer and G-J kepstrum beamformer and also at the bottom showing the performance between the G-J adaptive beamformer and G-J kepstrum adaptive beamformer, where both are tested in a nonstationary radio noise. The real speech used was 'hello' 'one' 'two' 'three' in the first half (A) and also the same words in the second half (B), where real-time voice has been triggered by the VAD and shown as flagged in the time-domain waveform. The results show more reduced nonstationary noise in the second half (B) in both top and bottom waveforms in Fig. 5-28. The test results are summarized in Table 5-XII.

For the performance comparison in a various conditions, the test has been done in room (A) using omnidirectional microphones in a G-J adaptive beamformer. Male and female speakers have spoken in front of two microphones by using the speech "I am in front of a desk". The test has been done in a different input SNRs (0dB, 10dB and 20dB respectively) at the different radio noise source locations (R_1,R_2 and R_3) and a fixed location of computer fan noise (C). The radio was tuned to real music sound. The test results have shown a highly promising performance for all of the different input SNRs at the different locations of noise sources, and also from both male and female speakers as shown in Table 5-XIII.

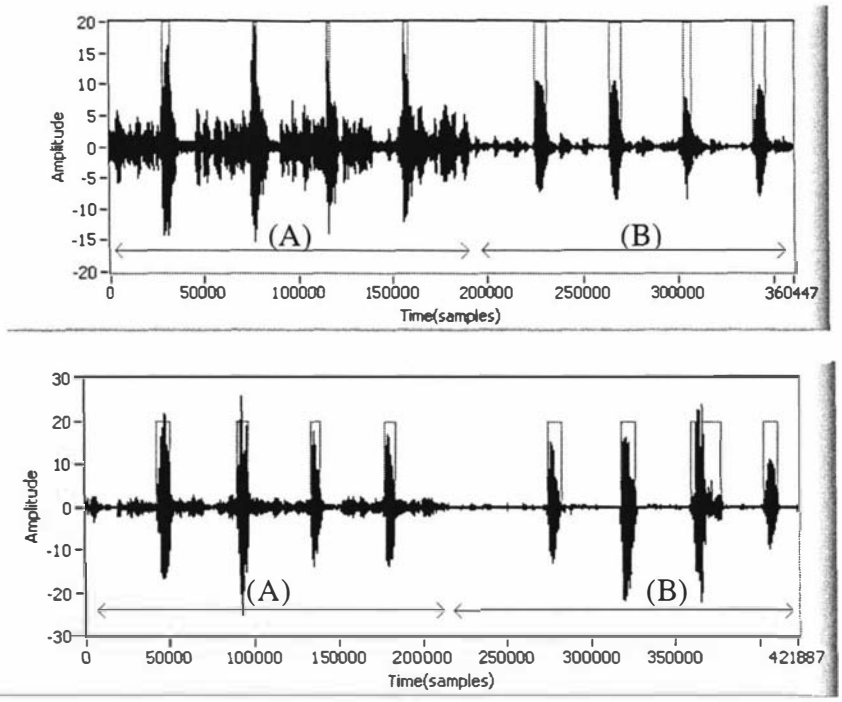


Fig. 5-28 Waveforms in speech with nonstationary (radio) noise showing performance of (top) (A): G-J beamformer, (B): G-J kepstrum beamformer and also (bottom) (A): G-J adaptive beamformer and (B) G-J kepstrum adaptive beamformer. VAD flag is shown in speech periods

Table 5-XII Test results based on stationary (computer fan) and nonstationary (radio) noise

Noise type	Stationary noise						
Processing type	Microphone output	G-J beamformer	G-J Kepstrum(A) beamformer	G-J Kepstrum(B) beamformer	G-J adaptive beamformer	G-J kepstrum(A) adaptive beamformer	G-J kepstrum(B) adaptive beamformer
Measurement / Performance in	Average power (dB)						
Averaged signal plus noise (A)	−14.91	−16.17	−24.60	−28.00	−17.41	−22.23	−24.59
Averaged noise (B)	−37.64	−41.09	−48.73	−53.41	−46.13	−53.78	−58.04
Averaged SNR (A - B)	22.73	24.92	24.13	25.41	28.72	31.55	33.45
SNR comparison		–	−0.79	0.49	3.80/–	6.63/2.83	8.53/4.73
Noise reduction		–	7.64	12.32	5.04/–	12.69/7.65	16.95/11.91

Noise type	Nonstationary noise						
Processing type	Microphone output	G-J beamformer	G-J Kepstrum(A) beamformer	G-J Kepstrum(B) beamformer	G-J adaptive beamformer	G-J kepstrum(A) adaptive beamformer	G-J kepstrum(B) adaptive beamformer
Measurement / Performance in	Average power (dB)						
Averaged signal plus noise (A)	−11.81	−15.14	−24.73	−26.37	−16.28	−22.32	−22.14
Averaged noise (B)	−24.25	−31.62	−40.84	−46.78	−38.49	−44.86	−50.90
Averaged SNR (A - B)	12.44	16.48	16.11	20.41	22.21	22.54	28.76
SNR comparison		–	−0.37	3.93	5.73/–	6.06/0.33	12.28/6.55
Noise reduction		–	9.22	15.16	6.87/–	13.24/6.37	19.28/12.41

Table 5-XIII Performance comparison in various conditions

Test result on kepstrum approach									
Noise sources	$R_1 + C$			$R_2 + C$			$R_3 + C$		
Input SNR	0dB	10dB	20dB	0dB	10dB	20dB	0dB	10dB	20dB
Male	2.08	7.92	13.90	2.12	5.93	13.67	4.58	8.19	15.72
Female	3.78	7.24	13.41	3.33	6.27	13.11	3.07	9.66	13.96

Compared with a recently proposed neural network based algorithms using the two-microphone G-J adaptive beamformer (**Beh et al.,** 2006), the kepstrum approach shows a better performance by 1-6 dB through the entire range of 0dB-20dB input SNR. In addition, the performance comparison with other methods tested in a real room reverberant environment, shows that the kepstrum approach has 2.2dB better SNR performance for a 1.2 dB input SNR than a four-microphone GSD (**Fancourt and Parra,** 2001). GSD is a hybrid method of a combination of geometric beamforming (GSC) and BSS algorithms, and claims a better performance than a four-microphone DS beamformer and GSC (**Griffiths and Jim,** 1982) as shown in Table 5-XIV. However, the two-microphone kepstrum approach gives worse performance than a two-microphone SBAGJ (**Campbell and Shields,** 2003), which uses a sub-band method.

Table 5-XIV Performance comparison with other methods in a real reverberant room environment

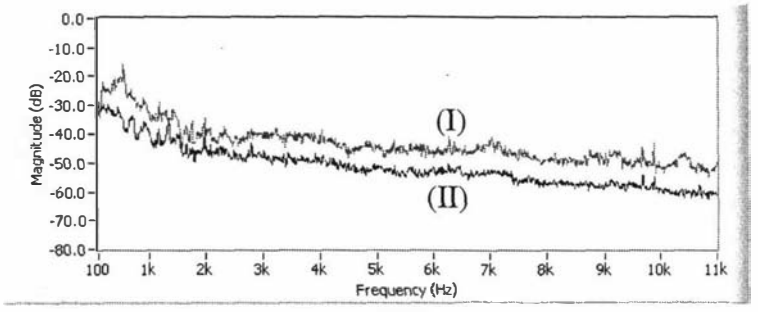
Algorithm	None	DS	GSC	GSD	Kepstrum	SBAGJ
SNR	1.2dB	1.3dB	3.0dB	4.6dB	6.8dB	8.0dB

To verify the effect of a highly reduced adaptive filter size in the application of a front-end kepstrum filter, applied filter size between kepstrum coefficients and NLMS weights has been compared.

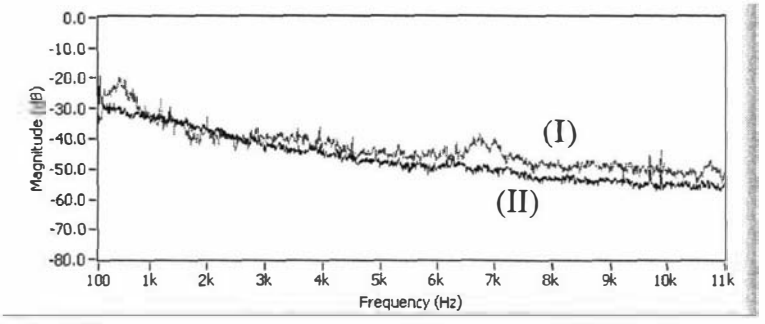
The test results are shown in Table 5-XV and it indicates that with the front-end kepstrum approach, the number of weights can be reduced in size by up to 80%~95% in the adaptive filter of the G-J adaptive beamformer (Fig. 5-29(A) and Fig. 5-29(B)). They also show that the front-end kepstrum application gives favourable results even in a nonstationary noise environment. The comparison of filter size reduction from the method of ANC and G-J adaptive beamforming using speech beamforming can also be found in Jeong and Moir (**Jeong and Moir,** 2005).

Table 5-XV The performance showing the same effect as the kepstrum approach by reducing the adaptive filter size in the G-J adaptive beamformer

Noise type / Adaptive filter size / Processing type	Stationary noise (computer fan)		Nonstationary noise (radio)	
	Filter size H(z)	Reduction ratio (%)	Filter size H(z)	Reduction ratio (%)
Adaptive beamformer	200	−95%	200	−90%
Kepstrum approach (64)	10		20	
Adaptive beamformer	150	−93%	150	−93%
Kepstrum approach (64)	10		10	
Adaptive beamformer	100	−95%	100	−90%
Kepstrum approach (64)	5		10	
Adaptive beamformer	50	−90%	50	−80%
Kepstrum approach (64)	5		10	



(A)



(B)

Fig. 5-29 (A) Average power spectra of stationary noise showing comparison of (A): adaptive beamformer with filter size 200 (I) and kepstrum (64) with adaptive beamformer with filter size 200 (II) and (B): adaptive beamformer with filter size 200 (I) and kepstrum (64) with adaptive beamformer with filter size 10 (II)

■

7. Summary

It has been found that the unidirectional microphone is more appropriate as it shows a better performance than the omnidirectional microphone in a real room reverberant environment. From the performance comparison of kepstrum processing, kepstrum processing (B) (recovering phase from the causal kepstrum domain) has shown a better performance than kepstrum processing (A) (using TDOA delay as phase). It has also found that the kepstrum method is effective in echo cancellation.

With the use of kepstrum processing (B) and frame size of 2048 (or 1024) with a VAD function, it allows a stable real-time processing, but when frame size is increased to 4096 and kepstrum processing (A) is used, it has shown a barely maintained real-time processing, with the system sometimes halting or failing.

The performance of the kepstrum approach is related with other factors from the performance of VAD and the selection of parameters such as frame size and FFT windows. Therefore, it is necessary to set a proper threshold for VAD and other parameters according to environmental conditions. The main limitation is due to the fact that the acoustic noise transfer functions (to the microphones) rapidly change in a real environment and hence during the speech periods, they use the previously frozen transfer functions. Therefore, the more robust the VAD, with more accurate separation ability and a rapidly changing speed between the intermittent utterances, the more possible it is to increase the overall performance of the kepstrum approach.

The kepstrum approach has shown an improved performance in speech enhancement and noise reduction with several favourable effects. Through real-time and simulation tests, it has shown that the front-end kepstrum estimate gives several favourable effects such as 1) invertibility, which gives stable minimum phase transfer function when the inverse of the ratio of acoustic transfer functions between a noise source and input microphones is desired; 2) the use of a highly reduced NLMS filter size by absorbing most of the minimum phase part in the front-end kepstrum estimate; 3) the use of a small number of kepstrum coefficients because of the minimum phase property, which shows highly concentrated energy around time zero; 4) computational simplicity using FFTs and efficiency by real-time kepstrum processing. Based on this, the kepstrum approach could give an application diversity, which may be applied to the area of equalization, noise cancellation, beamforming, line enhancing and deconvolution.

■

Chapter 6

Conclusions

An original and robust kepstrum approach has been proposed. It is based on kepstrum (complex cepstrum) analysis, which is comprised of a kepstrum processing technique using a VAD, and its front-end application to speech enhancement methods with a modified application to a conventional ANC and beamforming.

It has been shown that it can prove to have several favorable effects for real-time processing in a reverberant environment. A front-end application of a minimum phase spectral-factor from an unbiased periodogram estimate gives invertibility, so its inverse may give a stable performance. In addition, it makes use of the small kepstrum filter size with highly reduced cascaded adaptive filter sizes.

For the application to speech enhancement and noise cancellation, the novel application of the kepstrum approach has been shown to give an improvement over the modified G-J adaptive beamformer and more recently proposed algorithms. A higher performance in speech enhancement ratio even in nonstationary noise is found.

It can be concluded that the kepstrum approach is quite generic and can be applicable to the front-end of any speech enhancement method in real-time processing, as it is FFT based and independent of model order. Very few kepstrum coefficients are required and the residual all-pass transfer functions can be estimated with a cascaded recursive estimation method (LMS based) of reduced order.

For future work, the innovations approach using a kepstrum whitening filter will be applied to speech enhancement. The innovations white-noise sequence was first discovered by Kalman and Bucy (**Kalman and Bucy,** 1961) and used among others by Moir and Barrett (**Moir and Barrett,** 2003). It comes from the fact that it is possible to identify the spectral factor of a signal plus noise sequence directly from signal plus noise using the kepstrum method (thus avoiding spectral factorization) and the resulting estimator may be implemented in an innovations based form. The kepstrum method will be applied as a whitening filter to obtain the innovations sequence.

Further to this exercise, there remain many other areas to investigate including window size, VAD type and trade-off between computational complexity and estimation error improvement.

■

References

Agaiby, H. (1999). Word boundary detection for engineering applications, Ph.D. thesis, University of Paisley, Scotland, UK.

Agaiby, H., and **Moir, T. J.,** (1997, a). A robust word boundary detection algorithm with application to speech recognition. Digital signal processing proceedings, 1997, 13th international conference, vol.2, pp.753-755.

Agaiby, H., and **Moir, T. J.,** (1997, b). Knowing the wheat from the weeds in noisy speech. EUROSPEECH '97, 5th European Conference on Speech Communication and Technology, pp.1119-1122, Rhodes, Greece.

Agnes, M. E., and **Guralnik, D.B.,** (2000). "Webster's new world college dictionary, 4th edition."

Allen, J. N., Berkley, D.A., and **Blauert, J.,** (1977). "Multi-microphone signal-processing technique to remove room reverberation from speech signals." J. Acoust. Soc. Am. **62**: pp.912-915.

Audio-technica "A brief guide to microphones." http://www.audio-technica.com/cms/site/9904525cd25e0d8d/index.html: 2005 Audio-Technica U.S., Inc.

Ballow, G. (1991). Handbook for sound engineers: The new audio cyclopedia, Howard W Sam & Company, 2nd edition.

Barrett, J. F., and **Moir, T. J.,** (1984). "Spectrum analysis using kepstrum coefficients." IEE Colloq. Recent Advances in Identification and Signal Processing: IEE, London UK, **Part II**: pp.1-5.

Barrett, J. F., and **Moir, T. J.,** (1986). "The kepstrum method for spectral analysis." International Journal Control **43**(1): pp.29-57.

Barrett, J. F., and **Moir, T. J.,** (1987). "A unified approach to multivariable discrete-time filtering based on the Wiener-theory." Kybernetika **23**(3): pp.177-197.

Bartlett, M. S. (1948). "Smoothing periodograms from time series with continous spectra." Nature (London) **161**: pp.686-687.

Bees, D., Blostein, M, and **Kabal, P.,** (1991). Reverberant speech enhancement using cepstral processing. ICASSP-91, vol.2, pp.977-980.

Beh, J., Baran, R.H., and **Ko, H.,** (2006). "Dual channel based speech enhancement using novelty filter for robust speech recognition in automobile environment." Consumer Electronics, IEEE Transactions **52**(2): pp.583-589.

Bendjima, B., Rivenq, A., Iyad, D., and **Rouvaen, J.M.,** (1999). Speech enhancement for vehicle hands-free mobile telephony. Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th, vol. 4, pp.2168-2172.

Berghe, J. V., and **Wouters, J.,** (1998). "An adaptive noise canceller for hearing aids using two nearby microphones." Journal of The Acoustical Society of America **103**(6): pp.3621-3626.

Berouti, M., Schwartz, R., and **Makhoul, J.,** (1979). Enhancement of speech corrupted by acoustic noise. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79, vol. 4, pp.208-211.

Bitzer, J., Simmer, K.U., and **Kammeyer, K.D.,** (1999). "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement." Proceedings of the ICASSP **5**: pp.2965-2968.

Blackman, R. B., and **Tukey, J.W.,** (1958). The measurement of power spectra. New York, Dover publications, Inc.

Bogert, B. P., Healy, M. J. R., and **Turkey, J. W.,** (1963). The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe

cracking. Proceedings of the symposium on time series analysis, pp.209-243, John Wiley and Sons, NY.

Boll, S., and **Pulsipher, D.,** (1980). "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **28**(6): pp.752-753.

Boll, S. F. (1979). "Suppression of acoustic noise in speech using spectral subtraction." IEEE Trans. Acoust., Speech, Signal Processing **ASSP-27**: pp.113-120.

Brandstein, M. S., and **Ward, D.B.,** (2001). Microphone arrays: signal processing techniques and applications, Springer-Verlag, Berlin, Germany.

Burg, J. (1967). "Maximum entrophy spectral analysis." In 37th meeting society exploration geopysicists.

Campbell, D. R., and **Shields, P.W.,** (2003). "Speech enhancement using sub-band adaptive Griffiths-Jim signal processing." Speech Communication **39**(1-2): pp.97-110.

Capon, J. (1969). "High-resolution frequency-wavenumber spectrum analysis." Proceedings of the IEEE **57**(8): pp.1408-1418.

Carter, G. (1977, a). "Receiver operating characteristics for a linearly thresholded coherence estimation detector." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **25**(1): pp.90-92.

Carter, G. (1981). "Time delay estimation for passive sonar signal processing." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **29**(3): pp.463-470.

Carter, G., Knapp, C., and **Nuttall, A.,** (1973). "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing." Audio and Electroacoustics, IEEE Transactions **21**(4): pp.337-344.

Carter, G. C. (1977, b). "Variance bounds for passively locating an acoustic source with a symmetric line array." The Journal of the Acoustical Society of America **62**(4): pp.922-926.

Carter, G. C. (1987). "Coherence and time delay estimation." Proceedings of the IEEE **75**(2): pp.236-255.

Carter, G. C., and **Robinson, E.R.,** (1993). "Ocean effects on time delay estimation requiring adaptation." Oceanic Engineering, IEEE Journal **18**(4): pp.367-378.

Carter, G. C., Nuttall, A.H., and **Cable, P.G.,** (1972). "The smoothed coherence transfrom(SCOT)." Naval underwater systems center, New London Lab., New London, CT, Tech. Memo TC-159-72, Aug. 8, 1972.

Carter, G. C., Nuttall, A.H., and **Cable, P.G.,** (1973). "The smoothed coherence transform." Proceedings of the IEEE **61**(10): pp.1497-1498.

Chan, Y., Riley, J., and **Plant, J.,** (1979). A parameter estimation approach to time delay estimation. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79, vol. 4, pp.128-131.

Childers, D. G., Skinner, D.P., and **Kemerait, R.C.,** (1977). "The cepstrum: A guide to processing." Proceedings of the IEEE **65**(10): pp.1428-1443.

Cho, Y., and **Ko, H.,** (2004). Speech enhancement for robust speech recognition in car environments using Griffiths-Jim ANC based on two-paired microphones. Consumer Electronics, 2004 IEEE International Symposium, pp.123-127.

Cohen, I., and **Berdugo, B.,** (2002). Microphone array post-filtering for non-stationary noise suppression. Proc. 27th IEEE Int. Conf. Acoustics, Speech, Signal Processing, pp.901-904, Orlando, Fla, USA.

Cohen, I., Gannot, S., and **Berdugo, B.,** (2003). "An integrated real-time beamforming and postfiltering system for nonstationary noise environments." EURASIP Journal on Applied Signal Processing **2003**(11): pp.1064-1073.

Cohen, T. (1970). "Source-depth determinations using spectral, pseudo-autocorrelation and cepstral analysis." Geophys. J. Roy. Astron. Soc., **20**: pp.223-231.

Compernolle, D. V. (1990, a). Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings. Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference, pp.833-836, Albuquerque.

Compernolle, D. V. (1990, b). "Hearing aids using binaural processing principles." Acta Oto-Laryngol. Suppl. **469**: pp.76-84.

Compernolle, D. V., and **Gerven, S. V.,** (1992). Signal separation in a symmetric adaptive noise canceler by output decorrelation. Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference, vol. 4, pp.221-224.

Compernolle, D. V., Ma, W., and **Diest, M.D.,** (1990). "Speech recognition in noisy environments with the aid of microphone arrays." Speech Communication **39**(12): pp.433-443.

Cooley, J. W., and **Tukey, J. W.,** (1965). "An algorithm for the machine calculation of complex Fourier series,." Math. of Comput., **19**: pp.297-301.

Cox, H., Zeskind, R., and **Owen, M.,** (1987). "Robust adaptive beamforming." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **35**(10): pp.1365-1376.

Cron, B. F., and **Sherman, C.H.,** (1962). "Spatial-correlation function for various noise models." J. Acoust. Soc. Am. **34**(11): pp.1732-1736.

Darlington, D. J., and **Campbell, D.R.,** (1996). Subband adaptive filtering applied to speech enhancement. Proceedings of fourth international conference on spoken language ICSLP 1996, vol. 2, pp.921-924.

Doclo, S., and **Moonen, M.,** (2005). "Multimicrophone noise reduction using recursive GSVD-based optimal filtering with ANC postprocessing stage." Speech and Audio Processing, IEEE Transactions **13**(1): pp.53-69.

Donald, E. H. (2001). Musical acoustics, 3rd edition, Brooks/Cole Pub. Co.

Eckart, C. (1952). Optimal rectifier systems for the detection of steady signals, Univ. California, Scripps Inst. Oceanography, Marine Physical Lab. Rep SIO 12692, SIO Ref 52-11.

Elko, G. W. (1996). "Microphone array systems for hands-free telecommunication." Speech Communication **20**: pp.229-240.

Elliott, S. J., and **Rafaely, B.,** (2000). "Frequency-domain adaptation of causal digital filters." IEEE Transactions on Signal Processing **48**(5): pp.1354-1364.

Ephraim, Y., and **Van Trees, H.L.,** (1995). "A signal subspace approach for speech enhancement." Speech and Audio Processing, IEEE Transactions **3**(4): pp.251-266.

Fancourt, C., and **Parra, L.,** (2001). The generalized sidelobe decorrelator. Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop, pp.167-170.

Fay, J. W. (1980). "Coherence bounds for signal-to-noise ratios from magnitude-squared coherence estimates." IEEE Trans. Acoust., Speech, Signal Processing **ASSP-28**: pp.758-760.

Fischer, S., and **Simmer, K.U.,** (1996). "Beamforming microphone arrays for speech acquisition in noisy environments." Speech Communication **20**(3-4): pp.215-227.

Flanagan, J. L. (1972). Speech analysis, synthesis, and perception. 2nd Ed. New York, Springer-Verlag.

Flanagan, J. L., Johnson, J.D., Zahn, R., and **Elko, G.W.,** (1985). "Computer steered microphone arrays for sound transduction in large rooms." J. Acoust. Soc. Am. **78**(5): pp.1508-1518.

Ford, R. D. (1970). Introduction to acoustics, Amsterdam: Elsevier.

Frost, O. L., III, (1972). "An algorithm for linearly constrained adaptive array processing." Proc. IEEE **60**(8): pp.926-935.

Gannot, S., Burshtein, D., and **Weinstein, E.,** (1998). "Iterative and sequential Kalman filter-based speech enhancement algorithms." IEEE Trans. speech and audio pjrocessing **6**: pp.373-385.

Gannot, S., Burshtein, D., and **Weinstein, E.,** (2001). "Signal enhancement using beamforming and nonstationarity with applications to speech." Signal Processing, IEEE Transactions [see also Acoustics, Speech, and Signal Processing, IEEE Transactions] **49**(8): pp.1614-1626.

Gerven, S. V., and **Compernolle, D.V.,** (1995). "Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness." Signal Processing, IEEE Transactions [see also Acoustics, Speech, and Signal Processing, IEEE Transactions] **43**(7): pp.1602-1612.

Gerven, S. V., and Xie, F., (1997). A comparative study of speech detection methods. ESCA. Eurospeech97, pp.1095-1098, Rhodes, Greece.

Gold, B., and **Rader, C.M.,** (1969). Digital processing of signals. New York, McGraw-Hill.

Goubran, R. A., and **Hafez, H.M.,** (1986). Background acoustic noise reduction in mobile telephony. Vehicular Technology Conference, 1986. 36th IEEE, vol. 36, pp.72-76.

Goulding, M. M., and **Bird, J.S.,** (1990). "Speech enhancement for mobile telephony." Vehicular Technology, IEEE Transactions **39**(4): pp.316-326.

Gradshteyn, I. S., and **Ryzhik, I.M.,** (1979). Table of integrals, series, and products. New York: Academic.

Greenberg, J. E., and **Zurek, P. M.,** (1992). "Evaluation of an adaptive beamforming method for hearing aids." J. Acoust. Soc. Am. **91**(3): pp.1662-1676.

Grenier, Y. (1992). A microphone array for car environments. Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference, vol. 1, pp.305-308.

Griffiths, L. J. (1976). An adaptive noise cancelling procedure for multidimensional systems. Proc. Circuits Syst. Conf., Asilomer, CA.

Griffiths, L. J., and **Jim, C. W.,** (1982). "An alternative approach to linearly constrained adaptive beamforming." IEEE Transactions on antennas and propagation **AP-30**(No.1): pp.27-34.

Hahn, W., and **Tretter, S.,** (1973). "Optimum processing for delay-vector estimation in passive signal arrays." Information Theory, IEEE Transactions **19**(5): pp.608-614.

Haigh, J. A., and **Mason, J. S.,** (1993). A voice activity detector based on cepstral analysis. EUROSPEECH, pp.1103-1106, Berlin.

Hannan, E. J., and **Thomson, P.J.,** (1971). "The estimation of coherence and group delay." Biometrika **58**(3): pp.469-481.

Hanson, R. L., and **Lawson, C.L.,** (1969). "Extensions and applications of the Householder algorithm for solving linear least squares problems." Math. Comp. **23**: pp. 917-927.

Harrison, W. A., Lim, J. S. and **Singer, E.,** (1986). "A new application of adaptive noise cancellation." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **34**(1): pp. 21-27.

Haykin, S. (1996). Adaptive filter theory. Upper Saddle River, NJ., Prentice-Hall, Inc.

Haykin, S., Justice, J. H., Owsley, N.L., Yen, J.L., and **Kak, A.C.,** (1985). Array signal processing. New Jersey, Prentice-Hall Inc., Englewood Cliffs,.

Hodgkiss, W. S. (1979). Adaptive array processing: Time vs. frequency domain. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP, vol. 4, pp.282-285.

Homer, J. (2000). "Detection guided NLMS estimation of sparsely parametrized channels." Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions] **47**(12): pp.1437-1442.

Hoshuyama, O., Sugiyama, A., and **Hirano, A.,** (1999). "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters." IEEE Transactions on Signal Processing **47**(10): pp.2677-2684.

Hosoda, K., Yokota, K., Nakano, Y., and **Fukasawa, A.,** (1985). A new adaptive digital echo canceller. Proceedings of the 35th IEEE vehicular technology conference, Boulder, Colorado, pp.310-313.

Hu, Y., and **Loizou, P.C.,** (2002). "A subspace approach for enhancing speech corrupted by colored noise." Signal Processing Letters, IEEE **9**(7): pp.204-206.

Hussain, A., Campbell, D.R., and **Moir, T.J.,** (1997). Multi-sensor sub-band adaptive noise cancellation for speech enhancement in an automobile environment. Adaptive Signal Processing for Mobile Communication Systems (Ref. No. 1997/383), IEE Colloquium, pp.5/1-5/7.

Jenkins, G. M., and **Watts, D.G.,** (1968). Spectral analysis and its applications. San Francisco, CA, Holden-Day.

Jensen, S. H., Hansen, P.C., Hansen, S.D., and **Sorensen, J.A.,** (1995). "Reduction of broad-band noise in speech by truncated QSVD." IEEE Trans. Speech and Audio Processing **3**: pp.439-448.

Jeong, J., and **Moir, T. J.,** (2005). Kepstrum approach to real-time speech enhancement methods using two microphones. International conference on sensing technology, pp. 691-695, Palmerston North, New Zealand.

Johnson, G. W. (1994). LabVIEW graphical programming. New York, McGraw-Hill series.

Juang, B. H. (1991). "Speech recognition in adverse environments." Computer Speech and Language **5**: pp.275-294.

Kalman, R. E., and **Bucy, R. S.,** (1961). "New results in linear filtering and prediction theory." Transactions of the ASME - Journal of Basic Engineering **83**: pp.95-107.

Kamath, S., and **Loizou, P.,** (2002). A multi-band spectral subtraction method for enhancing speech corrupted by colored noise. Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference, vol. 4, pp.IV-4164.

Kates, J. M., and **Weiss, M.R.,** (1996). "A comparison of hearing-aid array-processing techniques." J. Acoust. Soc. Am. **99**: pp.3138-3149.

Kay, S. M. (1993). Fundamentals of statistical signal processing: Estimation theory. Englewood Cliffs, NJ., Prentice-Hall.

Kellermann, W. (1991). A self-steering digital microphone array. ICASSP 91, pp.3581-3584, Toronto, Ont., Canada.

Kemerait, R., and **Childers, D.G.,** (1972). Detection of multiple echoes immersed in noise. Proc. 15th Midwest Symp. Circuit Theory, May 4-5, 1972, pp.1-10.

Knapp, C., and **Carter, G. C.,** (1976). "The generalized correlation method for estimation of time delay." IEEE Transaction on Acoustics, Speech, and Signal Processing **ASSP-24**(4): pp.320-327.

Knecht, W. G., Schenkel, M.E., and Moschytz, G.S., (1995). "Neural network filters for speech enhancement." Speech and Audio Processing, IEEE Transactions on **3**(6): pp.433-438.

Kolmogorov, A. N. (1939). "Sur l'interpolation et extrapolation des suites stationaries." C. R. Acad. Sci. Paris **208**: pp.2043-2045.

Kolmogorov, A. N. (1941). Stationary sequences in Hilbert space, pp1-40(Russian): English translation in T. Kailath (ed) "Linear least squares estimation" Dowden, Hutchinson & Ross, Pennsylvania 1977, pp. 66-89. Bull, Moscow Univ.

Kompis, M., and **Dillier, N.**, (1994). "Noise reduction for hearing aids: combining directional microphones with an adaptive beamformer." J. Acoust. Soc. Am. **96**: pp.1910-1913.

Kompis, M., Feuz, P., Valentini, G., and **Pelizzone, M.,** (1998). A combined fixed/adaptive beamforming noise-reduction system for hearing aids. Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE, vol. 6, pp.3136 -3139.

Le Bouquin Jeannes, R., and **Faucon, G.,** (1994). "Proposal of a voice activity detector for noise reduction." Electronics Letters **30**(12): pp.930-932.

Lehnert, H., and **Blauert, J.,** (1992). "Principles of binaural room simulation." Appl. Acoust. **36**: pp.259-291.

Levitt, H. (2001). "Noise reduction in hearing aids: a review." Journal of rehabilitation research and development **38**(1): pp.111-121.

Lim, J. (1979). "Spectral root homomorphic deconvolution system." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **27**(3): pp.223-233.

Lim, J. S., and **Oppenheim, A.V.,** (1979). "Enhancement and bandwidth compression of noisy speech." Proceedings of the IEEE **67**(12): pp.1586-1604.

Lim, T. J., and **Macleod, M. D.,** (1994). "Adaptive allpass filtering for nonminimum phase identification." IEE Proceedings Vis. Image Processing **141**: pp. 373-379.

Lin, J. H., Sellke, T.M., and Coyle, E.J., (1990). "Adaptive stack filtering under the mean absolute error criterion." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **38**(6): pp.938-954.

Lin, Q., Jan, E., and **Flanagan, J.,** (1994). "Microphone arrays and speaker identification." Speech and Audio Processing, IEEE Transactions **2**(4): pp.622-629.

Liu, Q., Champagne, B., and **Kabal, P.,** (1995). Room speech dereverberation via minimum-phase and all-pass component processing of multi-microphone signals. Communications, Computers, and Signal Processing, 1995. Proceedings. IEEE Pacific Rim Conference, pp.571-574.

Ljung, L., and **Soderstrom, T.,** (1983). Theory and practice of recursive identification, MIT press, Cambridge, MA.

Lleida, E., Fernandez, J., and **Masgrau, E.,** (1998). Robust continuous speech recognition system based on a microphone array. Acoustics, Speech, and Signal Processing, 1998. ICASSP '98. Proceedings of the 1998 IEEE International Conference, vol. 1, pp.241-244.

Lockwood, P., and **Boudy, J.,** (1992). "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars." Speech Communications **11**(2-3): pp.215-228.

Long, G. (1986). Convergence performance comparison of adaptive pole-zero filter and two IIR adaptive filters, Technical report, Carleton university, Canada, February, 1986.

Lu, M. H., and **Clarkson, P. M.,** (1993). "The performance of adaptive noise cancellation systems in reverberant rooms." The Journal of the Acoustical Society of America (J. Acoust. Soc. Am.) **93**(2): pp. 1122-1135.

McCowan, I. A., Pelecanos, J., and **Sridharan, S.,** (2001). Robust speaker recognition using microphone arrays. Proceedings of 2001: A speaker odyssey 2001, Speech research Lab., Queensland university of technology.

Media college "Audio: How microphones work." http://www.mediacollege.com/audio/microphones/how-microphones-work.html: MediaCollege.com, PO Box 128, Te Awamutu, New Zealand.

Meyer, J., and **Simmer, K.U.,** (1997). Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction. Proc. 22nd IEEE Int. Conf. Acoustics Speech, Signal Processing, pp.1167-1170, Munich, Germany.

Mikhael, W. B., and **Hill, P.D.,** (1988). "Performance evaluation of a real-time TMS32010-based adaptive noise canceller (ANC)." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **36**(3): pp.411-412.

Mirchandani, G., Gaus, R., Jr. and **Bechtel, L. K.,** (1986). Performance characteristics of a hardware implementation of the cross-talk resistant adaptive noise canceller. ICASSP 86, pp.93-96, Tokyo.

Miyoshi, M., and **Kaneda, Y.,** (1988). "Inverse filtering of room acoustics." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **36**(2): pp.145-152.

Moir, T. J. (2001). "Automatic variance control and variance estimation loops." Circuits, Systems, and Signal Processing **20**(1): pp.1-10.

Moir, T. J. (2006). Tests on a real-time acoustic beamformer as a virtual instrument. 5th WSEAS International Conference on Artificial Intelligence. pp. 335-340

Moir, T. J., and **Barrett, J. F.,** (2003). "A kepstrum approach to filtering, smoothing and prediction with application to speech enhancement." Proc. R. Soc. Lond. A **2003**(459): pp.2957-2976.

Neely, S. T., and **Allen, J.B.,** (1979). "Invertibility of room impulse responses." J. Acoust. Soc. Am. **66**: pp.165-169.

Neo, W. H., and **Farhang-Boroujeny B.,** (2002). "Robust microphone arrays using subband adaptive filters." IEE Proceedings Vis. Image Signal Processing **149**(1): pp.17-25.

Noll, A. M. (1967). "Cepstrum pitch determination." The Journal of the Acoustical Society of America **41**(2): pp.293-309.

Nordebo, S., Claesson, I, and **Nordholm S.,** (1994). "Adaptive beamforming: spatial filter designed blocking matrix." IEEE Journal of oceanic engineering **19**(4): pp.583-590.

Nordholm, S., and **Claesson, I.,** (1993). "Adaptive array noise suppression of handsfree speaker input in cars." IEEE Transactions on vehicular technology **42**(4): pp.514-518.

Nuttall, A. H., and **Carter, G.C.,** (1982). "Spectral estimation using combined time and lag weighting." Proceedings of the IEEE **70**(9): pp.1115-1125.

Online etymology dictionary http://www.etymonline.com/index.php?term=microphone: November 2001 Douglas Harper.

Oppenheim, A. V. (1969). "Speech analysis-synthesis system based on homomorphic filtering." The Journal of the Acoustical Society of America **45**(2): pp.458-465.

Oppenheim, A. V., and **Schafer, R. W.,** (1968). "Homomorphic analysis of speech." IEEE Trans. Audio and Electroacoust., **AU-16**(2): pp.221-226.

Oppenheim, A. V., and **Schafer, R. W.,** (1975). Digital signal processing. New Jersey, Englewood Cliffs Prentice-Hall.

Oppenheim, A. V., Kopec, G.E., and **Tribolet, J.M.,** (1976). "Signal analysis by homomorphic prediction." IEEE Trans. Acoust., Speech, and Signal Processing **ASSP-24**: pp.327-332.

Oppenheim, A. V., Schafer, R.W., and **Stockham, T.G., Jr.,** (1968). "Nonlinear filtering of multiplied and convolved signals." Proc. IEEE **56**(8): pp.1264-1291.

Orfanidis, S. J. (1996). Introduction to signal processing, Prentice-Hall.

Papoulis, A. (1977). Signal analysis, McGraw-Hill, Inc.

Parra, L., and **Spence, C.,** (2000). "Convolutive blind separation of non-stationary sources." IEEE Transations on speech and audio processing **8**(3): pp.320-327.

Parra, L., Spence, C., and **De Vries B.,** (1998). Convolutive blind source separation based on multiple decorrelation. IEEE workshop on neural networks and signal processing, pp.23-32, Cambridge, U.K.

Per, A. (2001). Teleconferencing, system identification and array processing. Department of Information Technology, Ph.D. thesis, Uppsala University.

Pirz, F. (1979). "Design of a wideband, constant beamwidth, array microphone for use in the near field." AT&T Bell Syst. Tech. J. **58**(8): pp.1839-1850.

Plackett, R. L. (1950). "Some theorems in least squares." Biometrika **37**(1-2): pp.149-157.

Poisson, S. D. (1823). "Sur la distribution de la chaleur dans les corps solides." Ser. I: pp.1-162.

Pollák, P., and **Sovka, P.,** (1995). Cepstral speech/pause detectors. IEEE Workshop on Nonlinear Signal and Image Processing, pp.388-391.

Proakis, J. G., and **Manolakis, D. G.,** (1992). Digital signal processing, principles, algorithms and application, Macmillan Publishing Company, a division of Macmillan, Inc.

Pulsipher, D., Boll, S. F., Rushforth, C., and **Timothy, L.,** (1979). Reduction of nonstationary acoustic noise in speech using LMS adaptive noise cancelling. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79., vol. 4, pp.204-207.

Rabiner, L., Sambur, M., and **Schmidt, C.,** (1975). "Applications of a nonlinear smoothing algorithm to speech processing." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **23**(6): pp.552-557.

Rabiner, L. R., and **Schafer, R. W.,** (1983). Digital processing of speech signals, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Regalia, P. A., Mitra, S.K., and **Vaidyanathan, P.P.,** (1988). "The digital all-pass filter: a versatile signal processing building block." Proceedings of the IEEE **76**(1): pp.19 - 37.

Robinson, E. A. (1954). Predictive decomposition of time series with applications to seismic exploration.(Also in geophysics, 32: pp. 418-484). Cambridge, Mass., MIT.

Rodriquez, J. J. (1987). "Adaptive noise reduction in aircraft communication systems." MIT Lincoln Lab. Lexington, MA, Tech. Rep. 756,.

Roth, P. R. (1971). "Effective measurements using digital signal analysis." IEEE Spectrum **8**: pp.62-70.

Sambur, M. (1978). "Adaptive noise canceling for speech signals." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **26**(5): pp.419-423.

Schafer, R., and **Rabiner, L.,** (1973). "Design and simulation of a speech analysis-synthesis system based on short-time Fourier analysis." Audio and Electroacoustics, IEEE Transactions **21**(3): pp.165-174.

Schafer, R. W. (1968). "Echo removal by discrete generalized linear filtering." Ph.D. Dissertation, M.I.T., Cambridge, MA.

Schafer, R. W. (1969). Echo removal by discrete generalized linear filtering., Res. Lab. Electron. MIT, Tech. Rep., 466.

Schafer, R. W., and **Rabiner, L. R.,** (1970). "System for Automatic Formant Analysis of Voiced Speech." The Journal of the Acoustical Society of America **47**(2(part 2)): pp.634-648.

Schmidt, R. O. (1981). A signal subspace approach to multiple emitter location and spectral estimation, Stanford university, Stanford, CA.: Ph.D. thesis.

Schmidt, R. O. (1986). "Multiple emitter location and signal parameter estimation." Antennas and Propagation, IEEE Transactions [legacy, pre - 1988] **34**(3): pp.276-280.

Schwarz, H. A. (1872). "Zur Integration der partiellen Differentiagleichung." J. Reine Angewandte Math.,: pp.218-254.

Schweppe, F. (1968). "Sensor-array data processing for multiple-signal sources." Information Theory, IEEE Transactions **14**(2): pp.294 - 305.

Shields, P. W., and **Campbell, D.R.,** (2003). "Speech enhancement using sub-band adaptive Griffiths-Jim signal processing." Speech Communication **39**: pp.97-110.

Shynk, J. J. (1992). "Frequency-domain and multirate adaptive filtering." Signal processing magazine, IEEE **9**(1): pp.14-37.

Silvia, M. T., and **Robinson, E. A.,** (1978). "Use of the kepstrum in signal analysis." Geoexploration **16**: pp. 55-73.

Soede, W., Bilsen, F. A., and **Berkhout, A. J.,** (1993). "Assessment of a directional microphone array for hearing-impaired listeners." J. Acoust. Soc. Am. **94**: pp.799-808.

Sondhi, M. M. (1967). "An adaptive echo canceller." Bell syst. Tech. J., **46**: pp.497-511.

Sound on Sound "Choosing a microphone: microphone types and uses." http://www.soundonsound.com/sos/1995_articles/jun95/microphones.html: Media House, Trafalgar Way, Bar Hill, Cambridge, CB23 8SQ, United Kingdom.

Sovka, P., Pollak, P., and **Kybic, J.,** (1996). Extended spectral subtraction. European Signal Processing Conference (EUSIPCO - 96), pp.963-966, Trieste.

Speaks, C. E. (1996). Introduction to sound-acoustics for the hearing and speech sources, Singular publishing group Inc.,.

Stergiopoulos, S., and **Dhanantwari, A.C.,** (1998). Implementation of adaptive processing in integrated active-passive sonars with multi-dimensional arrays. Advances in Digital Filtering and Signal Processing, 1998 IEEE Symposium, 5-6 June, pp.62-66.

Stoffa, P. L., Buhl, P., and **Bryan, G.M.,** (1974). "The application of homomorphic deconvolution to shallow-water marine seismology - Part I: Models." Geophysics **39**(4): pp.401-416.

Stoica, P., and **Sharman, K.C.,** (1990). "Maximum likelihood methods for direction-of-arrival estimation." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **38**(7): pp.1132-1143.

Sweetwater "Live sound microphone: Buying guide." http://www.sweetwater.com/shop/live-sound/microphones/buying-guide.php#3: Sweetwater Sound, Inc. 5501 US Hwy 30 W Fort Wayne, IN 46818.

Szegö, G. (1915). "Ein Grenzwertsatz uber die Toeplitzschen Determinanten einer reellen positiven Funktion." Math. Ann., **76**: pp.490-503.

Toner, E., and **Campbell, D.R.,** (1993). "Speech enhancement using sub-band intermittent adaptation." International Journal of Speech Communication **12**: pp.253-259.

Treicher, J. R. (1977). The spectral line enhancer - The concept, an implementation, and an application. Ph.D dissertation, Dep. Elec. Eng., Stanford Univ., Stanford, CA.,.

Tribolet, J. (1977). Seismic applications of homomorphic signal processing, Thesis. Dept. of Electr. Eng. Comput. Sci.,. Cambridge, Mass., M. I. T.

Tribolet, J., Quatieri, T., and **Oppenheim, A.,** (1977). Short-time homomorphic analysis. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '77., vol. 2, pp.716-722.

Ulrych, T. (1971). "Application of homomorphic deconvolution to seismology." Geophysics **36**(4): pp.650-660.

University of Salford "A guide to microphone specifications." http://www.acoustics.salford.ac.uk/acoustics_world/jd/Microphones/Microphones.htm: Salford, Greater Manchester M5 4WT, UK. T +44 (0)161 295 5000 F. +44 (0) 161 295 5999.

Van Trees, H. L. (1968). Detection, estimation, and modulation. Theory, Part I. New York, Wiley.

Van Veen, B. D., and **Buckley, K. M.,** (1988). "Beamforming: a versatile approach to spatial filtering." ASSP Magazine, IEEE [see also IEEE Signal Processing Magazine] **5**(2): pp.4-24.

Viberg, M., Ottersten, B., and **Kailath, T.,** (1991). "Detection and estimation in sensor arrays using weighted subspace fitting." Signal Processing, IEEE Transactions [see also Acoustics, Speech, and Signal Processing, IEEE Transactions] **39**(11): pp.2436-2449.

Virag, N. (1999). "Single channel speech enhancement based on masking properties of the human auditory system." Speech and Audio Processing, IEEE Transactions **7**(2): pp.126-137.

Wahba, G. (1980). "Automatic smoothing of the log periodogram." J. Amer. Stat. Assoc. **75**: pp.122-132.

Wallace, R. B., and **Goubran, R. A.,** (1992). "Noise cancellation using parallel adaptive filters." IEEE Transaction on circuits and systems-II: Analog and digital signal processing **39**(No.4): pp.239-243.

Weiss, M. (1987). "Use of an adaptive noise canceler as an input preprocessor for a hearing aid." J. Rehabil. Res. **24**(4): pp.93-102.

Welch, P. (1967). "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms." Audio and Electroacoustics, IEEE Transactions **15**(2): pp.70-73.

Welker, D. P., Greenberg, J.E., Desloge, J.G., and **Zurek, P.M.,** (1997). "Microphone-array hearing aids with binaural output - Part II: A two-microphone adaptive system." IEEE Trans. Speech Audio Process **5**(6): pp.543-551.

Wenger, M. P. (2003). Noise rejection, the essence of good speech recognition. Technical report, Emkay Innovative Products. Itasca, IL., USA: pp 1-13.

Widrow, B. (2001). "A microphone array for hearing aids." Circuits and Systems Magazine, IEEE **1**(2): pp.26 -32.

Widrow, B., and **Hoff, M. E.,** (1960). "Adaptive switching circuits." IRE Wescon Convention Record: pp.96-104.

Widrow, B., and **Luo, F.,** (2003). "Microphone arrays for hearing aids: An overview." Speech comnunication **39**(2003): pp.139-146.

Widrow, B., and **Stearns, S. D.,** (1985). Adaptive signal processing, Englewood Cliffs New Jersey: Prentice Hall.

Widrow, B., Glover, J. R. Jr., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E. Jr., and **Goodlin, R. C.,** (1975). "Adaptive noise cancelling: principles and applications." Proceedings of the IEEE **63**(12): pp.1692-1716.

Widrow, B., McCool, J.M., Larimore, M.G., and **Johnson, C.R., Jr** (1976). "Stationary and nonstationary learning characteristics of the LMS adaptive filter." Proceedings of the IEEE **64**(8): pp.1151-1162.

Wiener, N. (1949). Extrapolation, interpolation and smoothing of stationary time series, with engineering applications. New York, Technology Press and Wiley.

Wold, H. (1938). A study in the analysis of stationary time series, reprinted by Almqvist & Wiksell, Stockholm, 1954.

Wood, L. C., and **Treitel, S.,** (1975). "Seismic signal processing." Proc. IEEE **63**: pp.649-661.

Yin, L., Astola, J., and Neuvo, Y., (1993). "A new class of nonlinear filters-neural filters." Signal Processing, IEEE Transactions [see also Acoustics, Speech, and Signal Processing, IEEE Transactions] **41**(3): pp.1201-1222.

Zelinski, R. (1988). A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. Proc. 13th IEEE Int. Conf. Acoustics, Speech, signal Processing, pp.2578-2581, New York, USA.

Zelinski, R. (1990). "Noise reduction based on microphone array with LMS adaptive post filtering." Electronic letters **26**(24): pp. 2036-2037.

Zheng, Y. R., Goubran, R.A., and El-Tanany, M., (2004). "Experimental evaluation of a nested microphone array with adaptive noise cancellers." Instrumentation and Measurement, IEEE Transactions **53**(3): pp.777-786.

Zinser, R., Jr., Mirchandani, G., and Evans, J., (1985). Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85., vol. 10, pp.1253-1256, Tampa, Florida.

Ziskind, I., and **Wax, M.,** (1988). "Maximum likelihood localization of multiple sources by alternating projection." Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions **36**(10): pp.1553-1560.

Zurek, P. M., and **Greenberg, J.E.,** (2000). Two-microphone adaptive array hearing aids with monaural and binaural outputs. Ninth DSP Workshop, First Signal Processing Education Workshop, Hunt, Texas, October 15-18, 2000.

Appendix

Conference proceedings (I)

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

Conference proceedings (II)

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

Conference proceedings (III)

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