System identification based kepstrum analysis and real-time application to noise cancellation

Jinsoo Jeong

Abstract— This paper presents analysis of kepstrum (known elsewhere as complex cepstrum) and its real-time application to noise cancellation. System identification based kepstrum method estimates the ratio of acoustic path transfer functions between two microphones and it is processed in an efficient way for real-time processing. Its front-end application to speech enhancement method will be shown that it effectively cancels echoes and hence reverberation in a noisy environment with computational simplicity on kepstrum processing for a real-time application. Furthermore, from the test based on the three different microphones configuration, it will be shown that kepstrum provides a better noise reduction ratio in endfire microphones configuration to a simple speech enhancement method, G-J beamformer structure.

Keywords— Adaptive noise cancelling, beamforming, cepstrum, complex cepstrum, kepstrum , NLMS, system identification

I. Introduction

THE signal in time domain is usually transformed into the frequency domain to be analyzed. However, the convolved or multiplied signals may not be easily analyzed due to overlapping spectrum densities so that it can not be characterized from the signal information. If the signals can be separated from each other, they can be easily analyzed and may then be processed by a linear system. The cepstrum processing technique gives a solution to the signals which have been convolved or multiplied in time domain because the operation of the nonlinear mapping can be processed by the generalized linear system (homomorphic system).

Historically, it has been found that there is spectral ripple in echo waveform and they come from the computation of logarithm of power spectrum from the observed signal. Furthermore, it has been shown that power spectrum of spectral ripple gives rise to strong peak in a time response of quefrency domain. It is named as cepstrum by Bogert et. al in 1963 [1]. Nowaday cepstrum is defined as inverse Fourier transform of logarithmic spectrum from the observed signal. It provides a statistical variation about the frequency of spectral ripple. The fact that speech signal has also spectral ripple gives motivation to use cepstrum in speech signal processing. For speech signal analysis, the cepstrum method uses a sample length and statistical variation in a time response, where the low time portion information is used for vocal tract transfer function and the high time portion for excitation respectively so that the processed speech signal is synthesized by convolving vocal tract impulse response and excitation parameters between voiced and unvoiced. The applications can be found from research areas, such as pitch detection, formant estimation, speaker verification and speech recognition.

Complex cepstrum has been introduced by Schafer in 1969 [2] by using the information of both the magnitude and phase spectra from the observed signal. The complex cepstrum method is used to recover signals generated by a convolutional process and has been called as homomorphic deconvolution or homomorphic filtering [3]. This method is different with the cepstrum method using the vocal tract transfer function and pitch. The applications can be found from seismic signal, speech and imaging processing.

Kepstrum has been named by Silvia and Robinson in 1978 [4] and used for seismic signal analysis. The method uses the fact that minimum phase spectral factor can be directly obtained from power spectrum estimation and, in the case of signal plus noise, logarithm of each positive- and negative-sided transfer function become the kepstrum spectral factors of the z-transform spectral density and these are represented as a power series expansion. The kepstrum method uses a low time portion of kepstrum information as noise statistics and it has been applied to speech enhancement by Moir and Barrett [5]. The kepstrum and complex cepstrum give almost same results for most purpose. Besides, two methods are similiar in the fact that both come from inverse FFT (fast Fourier transform) of logarithmic power spectrum, but the distinction between them is that kepstrum is characterized by kepstrum coefficients from Kolmogorov power series, which provides theoretical values (true values), whereas complex cepstrum provides empirical values (estimate values) using FFT. Therefore, it could be understood that the kepstrum of the sequence of coefficients in the Kolmogorov series expansion is replaced by the complex cepstrum of an inverse FFT. This gives the relationship between them being that the complex cepstrum is a truncated version of the kepstrum coefficients corresponding to the sample length only, not subject to statistical variation.

In this paper, we analyze kepstrum and investigate kepstrum method, which is comprised of kepstrum estimation and processing technique, for noise cancellation as a front-end application to speech enhancement method. As a result, we will show that kepstrum approach provides an effective solution for practical real-time noise cancellation. Furthermore, kepstrum

Manuscript received March 19, 2009: Revised version received March 20, 2009.

Jinsoo Jeong is with Faculty of Biomedical Engineering and Health Science, Universiti Teknologi Malaysia, 81310, UTM, Skudai, Johor, Malaysia (phone: 607-553-5738; fax: 607-553-5430; e-mail: jinsoojeong1015@ hanmail.net, jeong@utm.my).

has been considered with three different microphones configuration and we have found that kepstrum provides the highest performance in endfire configuration though it works well in all three microphones configuration in different noise source environments with speech.

II. Kepstrum Analysis

A. Kepstrum Analysis of Minimum Phase Transfer Function

The causal transfer function can be expressed from the Schwarz's classical expression [6] as:

$$H_{+}(z) = \frac{1}{2\pi}\int_{0}^{2\pi}\left(\frac{1+z^{-1}e^{j\lambda}}{1-z^{-1}e^{j\lambda}}\right)H_R(\lambda)d\lambda\,,\quad |z|<1 \tag{1}$$

where λ as the integration variable and $z = re^{jw}$.

This equation gives the causal transfer function $H_{+}(z)$ whose real part on the unit circle is $H_R(w)$. Based on this, the phase information can be recovered by Hilbert transform relation. Furthermore, logarithm of minimum phase transfer function $\log H_M(z)$ can be written as:

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

where $\log|H_M(\lambda)|$ is the magnitude of minimum phase logarithmic transfer function.

From (2), it indicates that minimum phase transfer function may be expressed by Schwarz's formula and therefore minimum phase information can be recovered by Hilbert transform relation.

Defining $z = e^{jw}$,

$$\log H_M(z) = \log H_M(e^{jw}) \tag{3}$$
$$= \log|H_M(e^{jw})| + j\arg[\log H_M(e^{jw})]$$

For the N-point discrete form, from (3),

$$\log H_M\left(\frac{2\pi}{N}k\right) = \left[c_0 + \sum_{n=1}^{N-1}\left(\frac{1}{2}c_{-n}e^{j\frac{2\pi}{N}kn} + \frac{1}{2}c_n e^{-j\frac{2\pi}{N}kn}\right)\right]$$
$$+ j\left[\sum_{n=1}^{N-1} - \left(\frac{1}{2}c_{-n}e^{j\frac{2\pi}{N}kn} - \frac{1}{2}c_n e^{-j\frac{2\pi}{N}kn}\right)\right] \tag{4}$$

where magnitude of logarithmic minimum phase transfer function is

$$\log\left|H_M(e^{j\frac{2\pi}{N}k})\right| = \left[c_0 + \sum_{n=1}^{N-1}\left(\frac{1}{2}c_{-n}e^{j\frac{2\pi}{N}kn} + \frac{1}{2}c_n e^{-j\frac{2\pi}{N}kn}\right)\right] \tag{5}$$

and phase of logarithmic minimum phase transfer function is

$$\arg\left[H_M(e^{j\frac{2\pi}{N}k})\right] = j\left[\sum_{n=1}^{N-1} - \left(\frac{1}{2}c_{-n}e^{j\frac{2\pi}{N}kn} - \frac{1}{2}c_n e^{-j\frac{2\pi}{N}kn}\right)\right] \tag{6}$$

Equation (5) shows even function, therefore minimum phase kepstrum (complex cepstrum) coefficients can be processed in the cepstrum domain by multiplying two in positive time series except time series zero.

B. Kepstrum Analysis of Spectral Factor

Kepstrum has originally been used by Kolmogorov [7] to solve the problem of factoring the power spectrum $\Phi(w)$ of a random process to extract a stable causal minimum phase system $H_M(z)$. Since then, the spectral factorization by power spectral estimation has provided an efficient solution of problem of extracting a causal minimum phase system $H_M(z)$ whose magnitude spectrum is the square root of the power spectrum.

$$|H_M(w)| = \sqrt{\Phi(w)} \tag{7}$$

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.

$$K(w) = 2\log|H_M(w)| = \log\Phi(w) \tag{8}$$

The spectral factors from power spectrum represent a stable minimum phase causal system and non minimum phase anticausal mirror image 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:

$$K(z) = \log\Phi(z) = \log H^{+}(z) + \log H^{-}(z) \tag{9}$$
$$= K^{+}(z) + K^{-}(z) \tag{10}$$

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. 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 power spectral density (PSD), $\Phi(z)$, if we let $z = z^{-1}$, then we find that $\Phi(z) = \Phi(z^{-1})$.

Defining $z = e^{jw}$, we have $\Phi(w) = \Phi(-w)$,

so PSD can be written as $\Phi(w) = \left|H^{+}(e^{jw})\right|^2$

By applying logarithm to both sides, we have

$$\log\Phi(w) = 2\log\left|H^{+}(e^{jw})\right| \tag{11}$$

For the N-point discrete form,

$$\log \Phi(2\pi k / N) = 2\left|\log H^{+}(e^{j2\pi k / N})\right| \quad (12)$$

$$= g_0 + \sum_{n=1}^{N-1} (k_{-n} e^{j2\pi kn / N} + k_n e^{-j2\pi kn / N}) \quad (13)$$

where $g_0 = 2k_0$.

For the causal kepstrum coefficients,

$$K^{+}(2\pi k / N) = (g_0 / 2) + \sum_{n=1}^{N-1} k_n e^{-j2\pi kn / N} \quad (14)$$

Since coeffcients of k_n are real,

$$K^{+}(2\pi k / N) = (g_0 / 2) + \sum_{n=1}^{(N/2)-1} k_n e^{-j2\pi kn / N} \quad (15)$$

From (15), it shows that the kepstrum coefficients are processed in the causal kepstrum domain by halving the first zeroth coefficient with the remaining coefficients truncated in size to $(N/2)-1$. It has been found that the use of kepstrum coefficients less than $(N/2)-1$ gives smoothing effect for removing high frequency variations on the spectral factor [8].

III. Cepstrum and Kepstrum Method

In this section, the two methods, cepstrum and kepstrum method are described. Cepstrum method uses the SISO (single input single output) based technique. Hence, from the single input, it uses a sample length and statistic variation for the speech signal analysis. It uses information of a low time portion for vocal tract transfer function and high time portion for excitation so the processed speech signal is synthesized by convolving vocal tract impulse response and excitation parameters between voiced and unvoiced. The technique is to use automatic pitch detector to select voiced or unvoiced information. On the other hand, kepstrum method uses an alternative approach which is based on more theoretical and mathematical construct. It uses the MISO (multiple input single output) based technique and uses information of a low time portion for noise statistics during the noise periods only. The technique is to use a VAD for noise statistics, which are frozen during the noise periods and updated during the speech periods. Its information is then applied to speech enhancement method.

The cepstrum method is essentially a practical approach to signal analysis based on the use of the discrete Fourier transform, using fast Fourier transform algorithm. However, this method is theoretically based on quantity dependent both on sample length and statistical variation.

On the other hand, the kepstrum method could be considered as an alternative approach to supply a surer theoretical foundation, not subject to statistical variation, providing relationship that the complex cepstrum is a truncated version of the kepstrum coefficients corresponding to the truncated low-time portion of sample length.

A. Cepstrum Method

It has been known that there was a spectral ripple in an echo waveform and also 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. 1 [9].

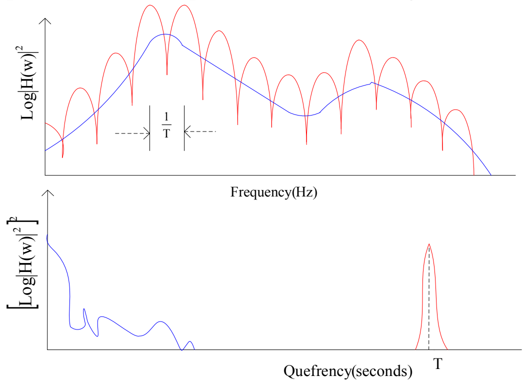


Fig. 1 Example of logarithmic power spectrum (top) and its power spectrum (bottom)

It provides a statistical variation about the frequency of spectral ripple. It also has been found that speech signals also have a spectral ripple [9]. Bogert et al. [1] called this unusual information about frequency of spectral ripple, a cepstrum, and nowadays its processing technique has widely been used to speech signal analysis and application to separate multiplied and convolved signal in time domain.

In a communication system, Stockham [10] have analyzed an audio signal that the signal $s(t)$ can be considered as a product of two components, one, $e(t)$ is slowly varying envelope with a positive value and the other one, $c(t)$ is rapidly varying with constant magnitude with both positive and negative value as shown in Fig 2.

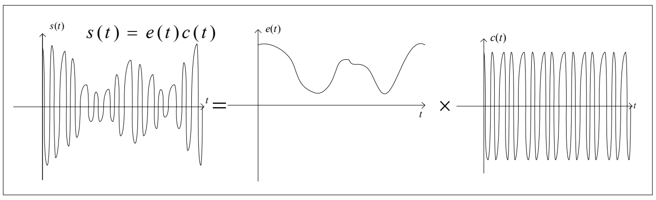


Fig. 2 The signal $s(t)$, envelope $e(t)$ and carrier $c(t)$

The nonlinear system which is obeying a multiplicative superposition can be processed by using homomorphic system (generalized linear system) in a logarithmic time domain (16) (Fig. 3).

$$\log s(t) = \log e(t) + \log|c(t)| \quad (16)$$

If the spectrums of $\log e(t)$ and $\log|c(t)|$ occupy the different frequency band, the linear filtering can be processed so these separated components can be processed independently.

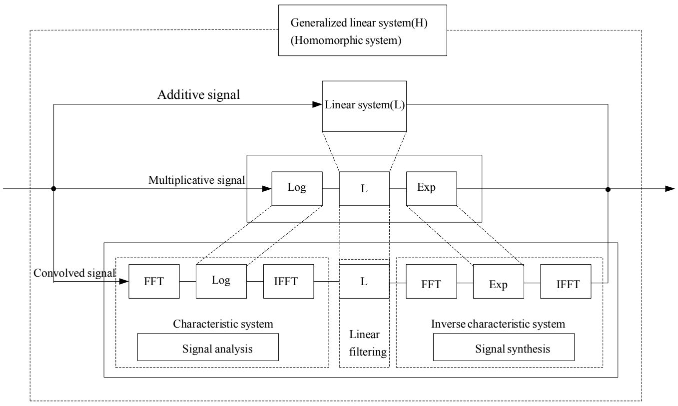


Fig. 3 The generalized linear system

The other example of homomorphic filtering is for the speech analysis and synthesis, introduced by Oppenheim [3] in 1968. Speech can be represented as the output of a linear time varying system whose properties vary slowly with time. This gives the basic principle of speech analysis that if we consider short segments of the speech signal, then each segment can effectively be modeled as having been generated by exciting a linear time invariant system either by a quasi-periodic impulse train or a random noise signal [11].

Fig. 4 shows the example of short segmented signal by 50% overlapping Hanning window with N=2048 point, which has a shape similar to that of half a cycle of a cosine wave. Its defining equation is

$$w(n) = 0.5 - 0.5\cos(2\pi n / N) \qquad (17)$$

where $n = 0, 1, 2, ... N-1$

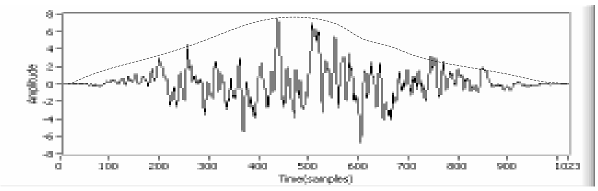


Fig. 4 Example of short segmented (windowed) signal

The problem of speech analysis is to estimate the parameters of the speech model and to measure their variations with time. Since the excitation and impulse response of a linear time invariant system are combined in a convolutional manner, the problem of speech analysis can also viewed as a problem in separating the components of a convolution, which is called deconvolution.

Vocal tract information is gathered in low time portion. On the other hand, sharp peak of pitch period is shown in high time portion (Fig. 5).

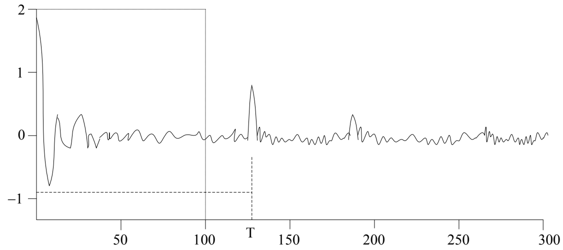


Fig. 5 Vocal tract information in low time potion and pitch information in high time portion

Therefore, speech signal can be synthesized by convolving vocal tract impulse response and excitation function (Fig. 6).

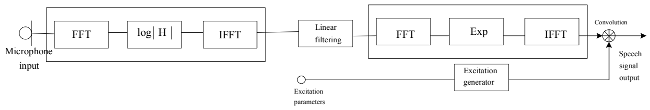


Fig. 6 The system configuration for speech analysis and synthesis

B. Kepstrum Method

For noise cancellation, kepstrum analysis is used for identification of the acoustic transfer functions to get noise statistics using two microphones. It is then processed using kepstrum processing technique.

The estimation procedure for a two acoustic transfer function is illustrated in Fig. 7, which shows that periodogram is processed from windowed FFTs as a discrete estimate of continuous power spectral-density.

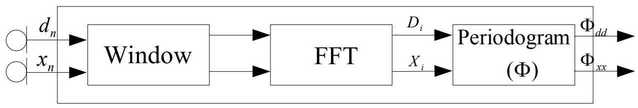


Fig. 7 Periodogram estimation procedure

For the periodogram estimates, the modified WOSA (weighted overlapped segment averaging) algorithm has been used. The modified WOSA based auto periodograms (18 and 19) and cross periodogram (20) are processed from 50% overlapping Hanning windowed 2048 FFTs as a discrete estimate of continuous power spectral density by smoothing methods with the use of β =0.8.

$$\Phi_{dd}(i) = \beta\Phi_{dd}(i-1) + (1-\beta)X_d(i)X_d^*(i) \qquad (18)$$

$$\Phi_{xx}(i) = \beta\Phi_{xx}(i-1) + (1-\beta)X_x(i)X_x^*(i) \qquad (19)$$

$$\Phi_{dx}(i) = \beta\Phi_{dx}(i-1) + (1-\beta)X_d(i)X_x^*(i) \qquad (20)$$

The logarithm of periodogram has found that there exists a bias equal in magnitude to minus Euler's constant γ =0.577215…, so it is added to be unbiased [12]. Therefore, kepstrum coefficients are found from the inverse FFT of the unbiased logarithm of the periodogram (Fig. 8). Whole procedure is repeated for each of the two microphones. By subtracting the two sets of kepstrum coefficients, we have the kepstrum equivalent of the ratio of the two acoustic transfer functions.

Fig. 8 Block diagram for kepstrum processing procedure. (Window: Hanning, log (Φ): log of periodogram, γ = Euler constant, 0.577215…)

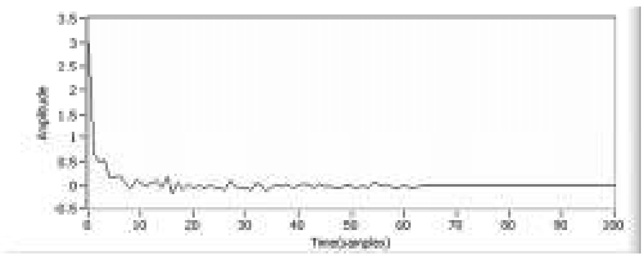
Kepstrum has the property of an efficient processing technique. The processing for the ratio of acoustic transfer functions needs a subtraction in kepstrum processing (21). The inverse of the ratio of acoustic transfer functions only needs a negative sign in kepstrum processing (22). Example of kepstrum processing is shown in Fig. 9.

$$\log\{H_1(z)/H_2(z)\} \leftrightarrow K_1(z)-K_2(z) \quad (21)$$

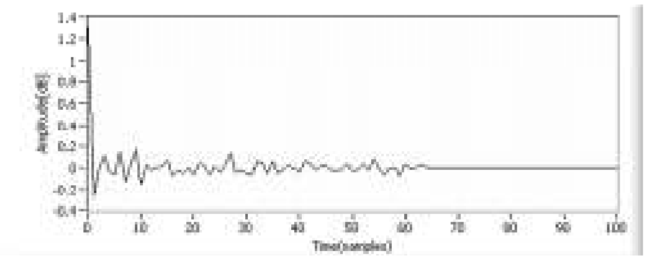
$$\log\{H_2(z)/H_1(z)\} \leftrightarrow -\{K_1(z)-K_2(z)\} \quad (22)$$

where $K_1(z)-K_2(z)=(k_{10}-k_{20})+(k_{11}-k_{21})z^{-1}+.....$

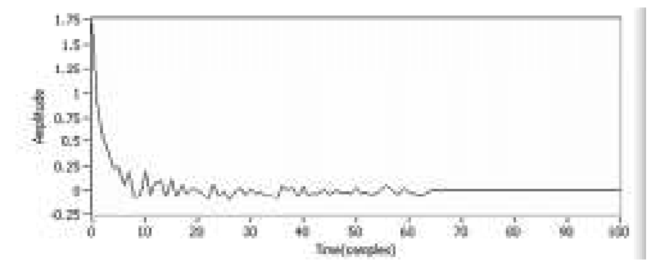
and $-\{K_1(z)-K_2(z)\}=-(k_{10}-k_{20})-(k_{11}-k_{21})z^{-1}-.....$



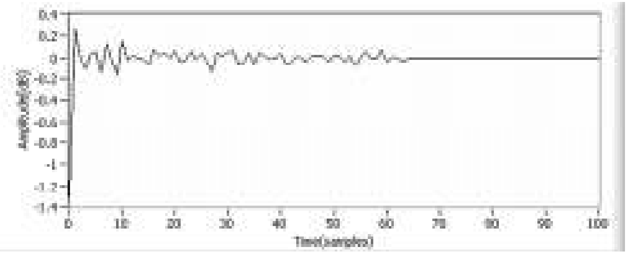
A) Kepstrum coefficients (k_{1n}) of $K_1(z)$



B) Kepstrum coefficients (k_{2n}) of $K_2(z)$



C) Kepstrum coefficients $(k_{1n}-k_{2n})$ of $K_1(z)-K_2(z)$



D) Kepstrum coefficients $(k_{2n}-k_{1n})$ of $K_2(z)-K_1(z)$

Fig. 9 Kepstrum coefficients

Kepstrum estimation does not provide phase–frequency information. Therefore, method to recover unknown minimum phase information is considered by restoring phase from the causal kepstrum domain as shown in Fig. 10. 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 single input (x_n) and its output is then inverse FFT transformed back to the time-domain so producing a new refined reference signal ($x_n^{'}$). This last operation is multiplication in the frequency domain. The block diagram procedure and block diagram of kepstrum approach of the front-end application to speech enhancement method are shown in Fig. 10 and Fig. 12 respectively.

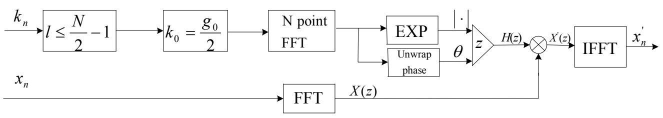
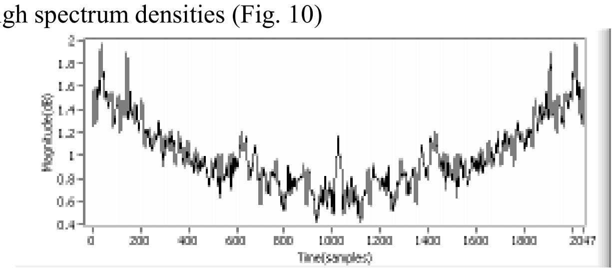


Fig. 10 Kepstrum processing by restoring phase from causal kepstrum domain

The truncated low time portion of kepstrum coefficients gives a smoothing effect which shows removing spectral ripples in high spectrum densities (Fig. 10)



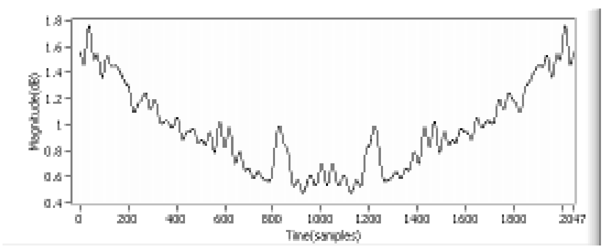


Fig. 11 Smoothing effect (bottom) by truncating kepstrum coefficients

IV. KEPSTRUM APPROACH

This section shows that the front-end kepstrum is applied to the speech enhancement method, G-J (adaptive) beamformer structure [13] as shown in Fig. 12.

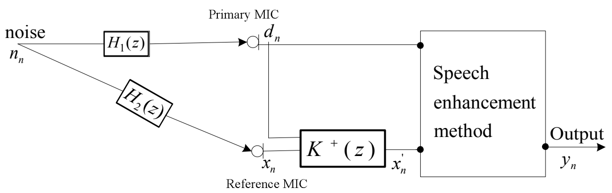


Fig. 12 Block diagram of kepstrum approach

A. G-J Beamformer

The simplest two-microphone G-J beamformer uses sum and difference function as signal separation (Fig. 13).

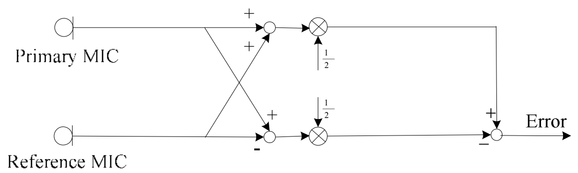


Fig. 13 Block diagram of G-J beamformer

B. G-J (Griffiths and Jim) Aaptive Beamformer

In addition to sum and difference function, NLMS (normalized least mean square) algorithm of adaptive filter [14] is used to minimize the mean square error at the output (Fig. 14). The software implementation of G-J adaptive beamformer can be found from [15, 16], which uses LabVIEW software on PC.

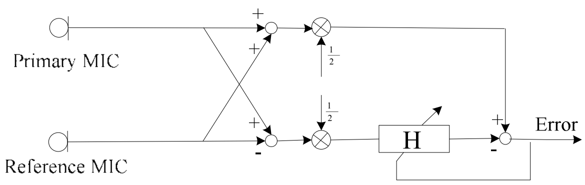


Fig. 14 Block diagram of G-J adaptive beamformer

C. G-J Kepstrum Beamformer (Kepsrum Applied G-J Beamformer)

System identification based analysis [17] shows that the output error is zero if acoustic transfer function is estimated as $H(z)=\{H_1(z)+H_2(z)\}/\{H_1(z)-H_2(z)\}$ in the structure of Fig. 15 (assuming that $H_1(z) \neq H_2(z)$ and that $H_1(z)-H_2(z)$ is minimum phase).

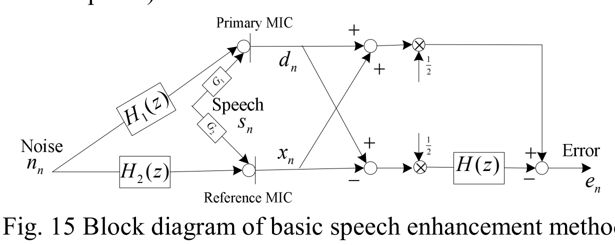


Fig. 15 Block diagram of basic speech enhancement method

It has found that direct estimation of composite transfer functions is difficult in practical application. On the other hand, as illustrated in Fig. 16, if noise signal is cancelled at the reference microphone input, then we may have a desired speech signal only in output error.

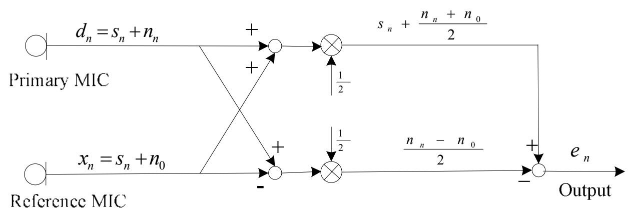


Fig. 16 Block diagram of typical beamforming method

Therefore we apply kepstrum filter to get rid of noise signal in reference input, the output error will then be speech signal only with a negligible noise.

Fig. 17 shows a structure that kepstrum is applied in front of speech enhancement method, G-J beamformer. The objective is to show a favourable effect of kepstrum on the simplest structure of two microphone G-J beamformer and also to compare with G-J adpative beamformer using adaptive filter for the performance of noise cancellation.

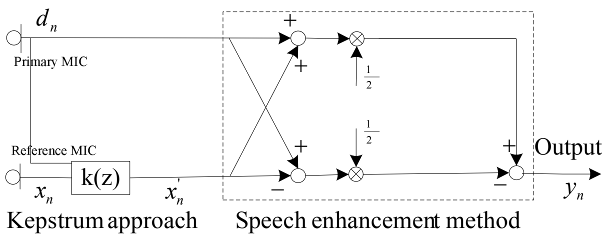


Fig. 17 Block diagram of G-J kepstrum beamformer

D. G-J Kepstrum Adaptive Beamformer (Kepsrum Applied G-J Adaptive Beamformer)

In a same way, kepstrum filter can be applied to front-end of G-J adaptive beamformer (Fig. 18). The method uses adaptive filter, whose weights are continuously updated during noise periods and frozen weights are used during speech periods. Kepstrum coefficients are treated in a same way.

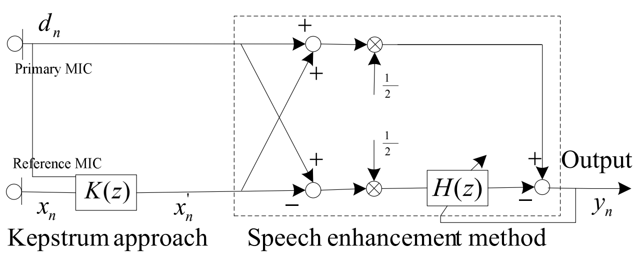


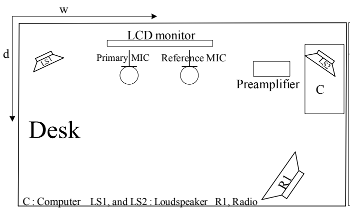
Fig. 18 Block diagram of G-J kepstrum adaptive beamformer

V. EXPERIMENTS

A. Experiment I

A-1Experimental Set-Up

Experiments are implemented in a room, where the test environment and equipment locations in dimension are illustrated in Fig. 19. The background noise level is measured as 48.5 dBA by using a sound level meter (Digitech QM1589).



Dimension and Position for Room and Equipment	d(m) x w(m) x h(m) 6.90 x 3.60 x 2.40
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
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

Fig. 19 Test environment (left) and room dimension, locations of sound source and equipment (right)

The signals are sampled using a standard internal sound card and two preamplifiers (Alto stereo tube type). Two types of omnidirectional / unidirectional electret condenser microphone have been used. The specification of microphones is shown in Table 1.

The sampling frequency is used as 22050Hz with the Nyquist frequency bandwidth of around 11000Hz and sampling resolution is 16 bits per channel.

Room reverberation time is calculated as 0.79 seconds for the frequency 500Hz, which is assumed that rooms reflect a moderately reverberant situation.

Table 1 Specification of microphones

Specification	Microphone type	
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

A-2 Experimental Methodology

The two types of additive noise have been used, computer fan as a stationary noise and music radio tuned to a station as nonstationary noise. G-J beamformer (Fig. 14) and G-J adaptive beamformer (Fig. 7) are used for the test, which are to be compared with G-J kepstrum beamformer (Fig. 17). A loudspeaker has been used to produce an echo sound.

For the performance comparison in a reverberant environment, two types of omnidirectional and unidirectional mirophone have been used.

By using unidirectional microphones, performance has been compared for echo and reverberation cancellation among G-J beamformer, G-J adaptive beamformer and G-J kepstrum beamformer (kepstrum applied in G-J beamformer). The number of kepstrum coefficients and the weights of NLMS is used as 64 and 200 respectively.

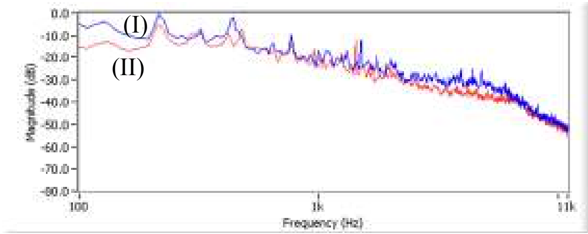
For a real-time application, computational complexity of kepstrum has been measured by the complexity of compution in FLOPS and also in the term of CPU utililization, where the performance is compared with NLMS algorithm.

A-3 Experimental Summary

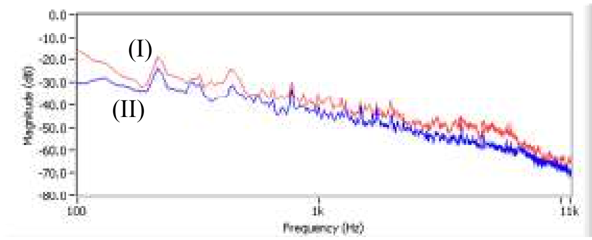
1) Comparison of Omnidirectional Microphone and Unidirectional Microphone

The objective is to provide a performance comparison of a noise reduction between omnidirectional and unidirectional microphones in both stationary and nonstationary noise. Specification of the two microphone types is shown in Table 1.

The application to both G-J beamformer and G-J kepstrum beamformer shows that the unidirectional microphone gives a better performance in both a stationary computer fan and nonstationary radio music noise as shown in Fig. 20 and Fig. 21.

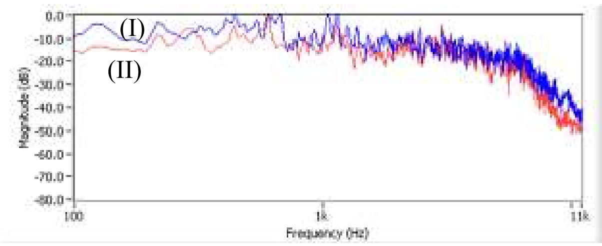


(A)

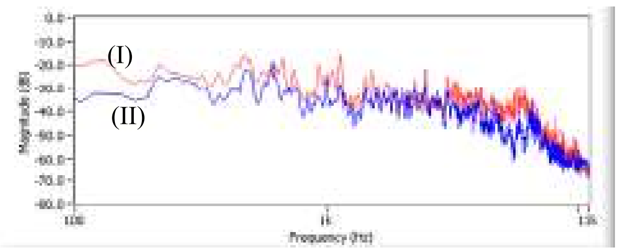


(B)

Fig. 20 Comparison of (I) omnidirectional and (II) unidirectional microphone based on G-J beamformer (A) and G-J kepstrum beamformer (B) in stationary computer fan



(A)



(B)

Fig. 21 Comparison of (I) omnidirectional and (II) unidirectional microphone based on G-J beamformer (A) and G-J kepstrum beamformer (B) in nonstationary music radio

2) The effect of front-end kepstrum application in a reverberant environment

The objective is to test a favorble effect of a front-end kepstrum application to a speech enhancement method for echo or reverberation cancellation.

1) Fig. 22 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.

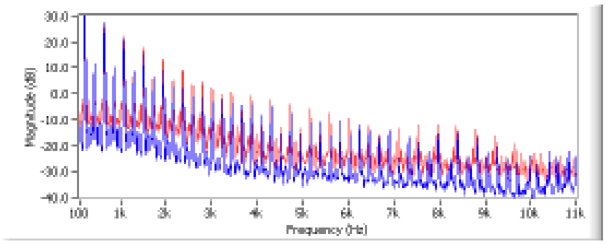


Fig. 22 Comparison of omnidirectional (top waveform) and unidirectional (bottom) microphone on echo sound

2) By using unidirectional microphones, the test has been carried out with G-J beamformer (Fig. 6), G-J adaptive beamformer (Fig. 14) and G-J kepstrum beamformer (Fig. 17) (i.e., kepstrum applied in G-J beamformer). Fig. 23 shows that the kepstrum approach (III) shows a highly reduced noise reduction with echo and additive noises cancellation, which shows a better performance in average power spectra in dB than G-J adaptive beamformer (II).

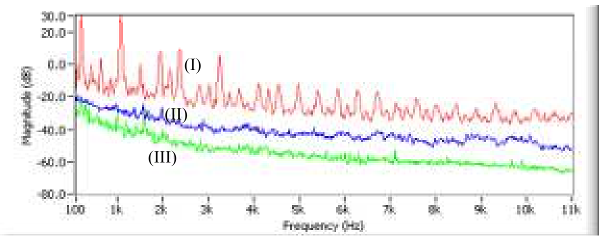
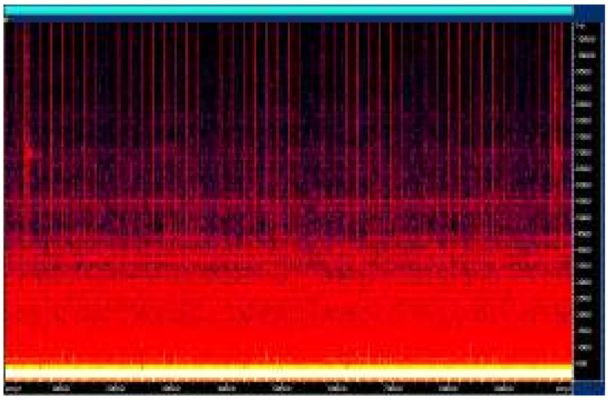
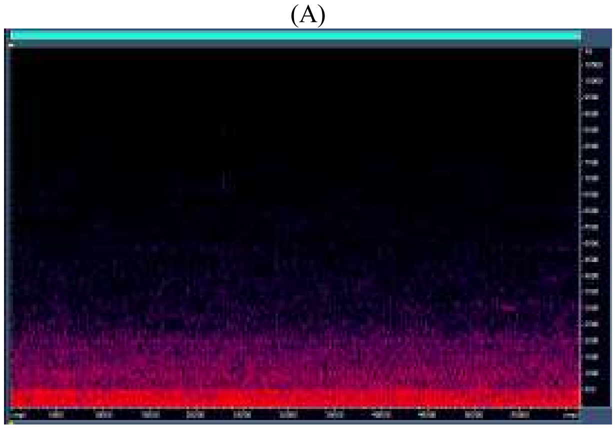


Fig. 23 (I) G-J beamformer (II) G-J adaptive beamformer (III) G-J kepstrum beamformer

3) The performance in echo and additive noise cancellation can also be verified from a spectrogram (frequency vs. time) of G-J beamformer (I) and G-J kepstrum beamformer (III), where it shows that straight lines (e.g., perpendicular to time domain in x-axis) have been removed by using the kepstrum approach as shown in Fig. 24. It shows that kepstrum approach effectively removes echoes as well as additive noise.



(A)



(B)

Fig. 24 Spectrogram analysis between (A) G-J beamformer and (B) G-J kepstrum beamformer

3) Performance measurement of real-time processing

For the real-time processing, comparison of computational complexity is measured by the complexity of multiplication in FLOPS (floating point operations per second). Table 2 shows the required processing and number of FLOPS for computational complexity of the kepstrum and LMS algorithm.

Table 2 Comparison of kepstrum and NLMS algorithm in FLOPS

Algorithm	Required processing	FLOPS
Kepstrum	2 x WOSA 5 x FFT/IFFT 3 x Logarithm/exponential	$2N\log_2(5.12/\Delta f)$ $5(N/2)\log_2 N$ $3N^{1/3}(\log N)^2$
	Total computation	$0.08\ G$
NLMS	Real multiplication (A) Iterations (B)	$3N^2+2N$ 20x (A)
	Total computation	(A) $0.12G$, (B) $2.4G$

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 is used.

Secondly, real-time system performance is measured by CPU utilization, which is based on computer processors of intel(R)

Pentium® 4 CPU 2.8 GHz, 1Gb RAM memory. It shows that kepstrum processing gives better real-time CPU usage than NLMS algorithm of adaptive filter as shown in Table 3.

Table 3 Comparison of CPU usage in kepstrum and NLMS algorithm

Processing type	Kepstrum			NLMS
Coefficients/Weights	64	200	1000	200
Average CPU usage (%)	35	39	45	60

B. Experiment II

Based on the test results of experiment I, the real-time test has been extended on three different microphones configuration as shown in Fig. 25. According to three different microphones set-up, G-J kepstrum beamformer (i.e., kepstrum applied in G-J beamformer) has been compared with G-J beamformer and G-J adaptive beamformer. Test result shows that kepstrum method in G-J beamformer gives the highest performance in endfire microphones configuration from all the test type I (i.e., test based on stationary noise), II (i.e., test based on stationary and nonstationary noise) and III (i.e., test based two noises and speech signal) as shown in Fig. 26 and the test results are listed in Table 4, which is for the performance comparison of kepstrum applied to G-J beamformer and G-J adaptive beamformer. It has been found from the endfire microphones configuration that G-J kepstrum beamformer has almost same performance with G-J keptrum adaptive beamformer in the test type III, which is based on two ambient noises and speech.

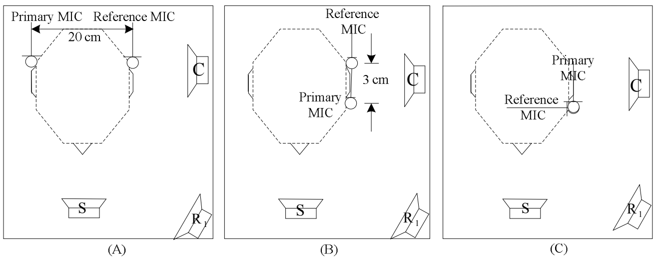


Fig. 25 Experimental microphone set-up (S: speaker, C : computer fan noise, R_1 : radio noise): (A) broadside, (B) endfire and (C) endfire variant

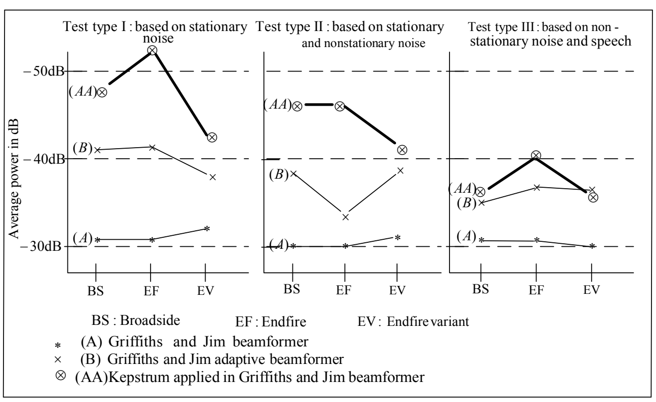


Fig. 26 Comparison of performance based on different microphones set-up

Table 4 Test results based on three different microphones set-up

Test type	Test type I - based on one ambient noise (computer fan)					
Configuration type	Broadside		Endfire		Endfire variant	
Average power in dB / Method type	Average power (dB)	Reduction ratio (dB)	Average power (dB)	Reduction ratio (dB)	Average power (dB)	Reduction ratio (dB)
G - J Beamformer	− 31.55dB	− 16.01dB	− 31.32dB	− 20.57dB	− 32.07dB	− 10.06dB
G - J kepstrum beamformer	− 47.56dB		− 51.89dB		− 42.13dB	
G - J Adaptive beamformer	− 40.92dB	− 16.53dB	− 40.70dB	− 16.63dB	− 38.19dB	− 9.35dB
G - J kepstrum adaptive beamformer	− 57.45dB		− 57.07dB		− 47.54dB	
Test type	Test type II - based on two ambient noises (computer fan and radio)					
G - J Beamformer	− 30.33dB	− 15.58dB	− 30.08dB	− 15.95dB	− 31.16dB	− 9.35dB
G - J kepstrum beamformer	− 45.91dB		− 46.03dB		− 40.51dB	
G - J Adaptive beamformer	− 38.46dB	− 9.79dB	− 33.38dB	− 18.76dB	− 38.65dB	− 13.06dB
G - J kepstrum adaptive beamformer	− 48.25dB		− 52.14dB		− 51.71dB	
Test type	Test type III - based on two ambient noises and speech					
G - J Beamformer	− 30.70dB	− 5.35dB	− 30.97dB	− 9.05dB	− 30.05dB	− 6.02dB
G - J kepstrum beamformer	− 36.05dB		− 40.02dB		− 36.07dB	
G - J Adaptive beamformer	− 35.18dB	− 1.60dB	− 36.79dB	− 4.01dB	− 36.53dB	− 3.94dB
G - J kepstrum adaptive beamformer	− 36.78dB		− 40.80dB		− 40.47dB	

VI. Conclusions

It has been found that the unidirectional microphone is more appropriate as it shows better performance than the omnidirectional microphone in a reverberant environment. By using unidirectional microphone, we have found that the kepstrum method gives the effective solution in echo and additive noise cancellation in a real-time processing. For this test, kepstrum has been applied to the simplest two-microphone G-J beamformer to verify the performance of cancellation of echo and additive noises.

Based on this test result, we have further found that front-end kepstrum application from simple structure of G-J beamformer works well in endfire microphones configuration during speech and noise periods.

For future work, it remains further investigation of kepstrum approach to other technique, such as sub-band method for the improved performance with an efficient processing.

References

[1] B.P. Bogert, M.J.R. Healy, J.W. Turkey, The frequency analysis of time series for echoes: cepstrum, seudo-autocovariance, cross-cepstrum and saphe cracking, Presented at Proceedings of the Symposium on Time Series Analysis, 1963, pp. 209–243.

[2] R.W. Schafer, Echo removal by discrete generalized linear filtering: Res. Lab. Electron. MIT, Tech. Rep., 466, 1969.

[3] A.V. Oppenheim, R.W. Schafer, Homomorphic analysis of speech, IEEE Trans. Audio Electroacoust. AU-16 (1968) 221–226.

[4] M.T. Silvia, E.A. Robinson, Use of the kepstrum in signal analysis, Geoexploration 16 (1978) 55–73.

[5] T. J. Moir, J.F. Barrett, A kepstrum approach to filtering, smoothing and prediction with application to speech enhancement, Proc. R. Soc. Lond. A 2003 (2003) 2957–2976.

[6] H. A. Schwarz, "Zur Integration der partiellen Differentialgleichung." J. Reine Angewandte Math.,: pp.218-254, 1872.

[7] A.N. Kolmogorov, Stationary sequences in Hilbert space, Bull. Moscow Univ. (1941) 1–40 (Russian) [English translation in T. Kailath (Ed.), Linear Least Squares Estimation, Dowden, Hutchinson & Ross, Pennsylvania, 1977, pp. 66–89].

[8] J. F. Barrett, T. J. Moir, "The kepstrum method for spectral analysis" , Int. J. Control, 43 (1986) 29-57.

[9] A. M. Noll, "Cepstrum pitch determination." The Journal of the Acoustical Society of America 41(2): pp.293-309.

[10] T. G. Jr., Stockham, "The application of generalized linearity to automatic gain control." IEEE Trans. Audio and Electroacoust., Au-16 (2): pp 267-270, 1968.

[11] G. Fant, "Acoustic theory of speech production." The Hague, Netherlands: Mouton, 1960.
[12] G.Wahba, Automatic smoothing of the log periodogram, J. Am. Stat. Assoc. 75(1980) 122–132.
[13] L.J. Griffiths, C.W. Jim, An alternative approach to linearly constrained adaptive beamforming, IEEE Trans. Antennas Propag. AP-30 (1982) 27–34.
[14] S. Haykin, Adaptive Filter Theory, third ed., Prentice-Hall, Inc., Upper Saddle River, NJ, 1996
[15]T. J. Moir, "Tests on a real-time adaptive beamformer as a virtual instrument," in *WSEAS 5th International Conference on Signal Processing, Robotics and Automation,*, Madrid,Spain, 15-17 Feb 2006.
[16]T. J. Moir, "Real-time acoustic beamforming on a PC," *WSEAS Transactions on Signal Processing,* vol. 2, pp. 167-174, 2006.
[17]J. Jeong, "Analysis and comparative study on two-microphone noise cancellation and speech enhancement methods for real-time hearing aids application" in *WSEAS 8th International Conference on Signal Processing, Robotics and Automation (ISPRA)* , pp. 143-148, The University of Cambridge, Cambridge, UK, 21-23 Feb. 2009

Jinsoo Jeong was born in Busan, Korea in 1958. He graduated at Dong-A university, Busan, Korea in 1981 with bachelor degree in Electronics Engineering, and then in 1984 Myong-ji university, seoul, Korea with master of engineering degree in Electronics Engineering. In 1986, he further attained master of science degree in Electrical Engineering at Polytechnic University of New York, Brooklyn, New York, USA. He has obtained Ph.D. degree in Information Engineering at Massey University, New Zealand in 2007.

He is currently working at faculty of biomedical engineering and health science, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia as senior lecturer. He has an extensive working experience from Korea and New Zealand at industrial companies as an Engineer. His current research interests are speech enhancement, noise cancellation, system identification, spectral estimations.

Dr. Jeong is currently member of IEEE and contributes as paper reviewer for IEEE conferences.