Analysis of system identification and modified application to two-microphone speech enhancement

Jinsoo Jeong

Abstract— This paper provides analysis of identification of acoustic transfer functions between two microphones and investigation of modified application from two-microphone adaptive noise cancelling and beamforming methods. Based on this, we will perform real-time performance comparisons to obtain the best solution to speech enhancement and noise cancellation. Experiments are processed by software implementation using LabVIEW in a real environment, which is typical indoor office with moderate reverberation condition. The speech performances are analyzed in time and frequency domains using both stationary and nonstationary noises. The analysis on the three type of microphones configuration and computational complexity on NLMS algorithm and TDOA function have also been investigated, which could give rise to fundamental basis for further real-time applications using two-microphone as well as hardware prototype implementation of digital adaptive hearing aids.

Keywords— ANC, beamforming, NLMS, system identification, TDOA, VAD

I. Introduction

FOR speech enhancement and noise cancellation, the ANC (adaptive noise cancelling) approach has the attraction as a noise canceller by using an adaptive filter in the reference input, which minimize an output power in an MMSE (minimize mean square error) sense. Beamforming generally uses a multiple microphones array and gives advantages on speech enhancement by maximizing a speech directivity and signal separation by spatial discrimination.

However, for the real-time application in a realistic environment, signal distortion, complexity on software computation and physical dimension in size of the microphone array should be considered.

Typical ANC approach has problems of signal distortion due to signal leakage into the reference input, uncorrelated noises between two microphones, noncauality and also reverberation. Beamforming approach also has problems of speech distortion due to nature of room reverberation, microphones misalignment, look direction (i.e., the direction of the desired speech source) error, speech leakage into the reference input, and multiple noise sources.

To overcome problems using the two-microphone ANC approach, several applications have been found to improve the performance in SNR (signal-to-noise ratio). Such applications are to reduce a noise in a reverberant environment by using 1) a longer adaptive filter in a low SNR, 2) physical environmental set-up by using sound absorbing materials and locating reference microphone near noise source, 3) directional microphones, 4) estimation of unknown acoustic path transfer function, 5) small separation of distance between two microphones and adpative filter for noise periods only using VAD (voice activity detection) 6) signal separation algorithms, such as CTRANC (crosstalk resistant adaptive noise canceller) and SAD (symmetric adaptive decorrelator) and 7) multiple sub-band processing.

From the conventional beamforming approach, several applications have also been found. Such applications are to enhance a speech signal in a reverberant environment by using 1) speech directivity (delay and sum) fuction, 2) signal blocking (sum and difference) function, 3) speech beamforming (beam-steering filter) or TDOA (time difference of arrival) function, 4) close and direct speech in front of microphones, 5) hybrid with adaptive filtering method with VAD. Generally, microphone arrays based beamforming employs the difference in spatial domain (in location and direction) between the desired speech signal and the noise. This technology may resolve limiting factors of ANC but requires an extended computational complexity and larger physical dimension in size.

In addition to the above methods as described, the various modified methods have been derived from the ANC and beamforming, and provided a solution for improved performance through various approaches by not only using both the benefits, but also remedying application limitations.

In this paper, we analyze the basic structures of noise cancellation and speech enhancement methods using two microphones, and investigate the modified applications for the purpose of 1) preventing a speech distortion, 2) enhancing a speech signal and 3) cancelling the noises effectively from noisy speech signal. It is then applied to the classic ANC [1] and the G-J (Griffiths and Jim) beamformer [2]. As a result, the four methods are compared to find the best solution for the application in a reverberant environment. Secondly, the three different types of microphone configuration are compared for the application of hearing aids. Finally, computational complexity between TDOA function and NLMS (normalized least mean

Manuscript 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).

square) algorithm is compared for the application in a real-time implementation.

II. ANALYSIS

This section provides a theoretical analysis from two fundamental structures for noise cancellation and speech enhancement. The analysis is based on system identification of the ratio of acoustic noise path transfer functions between two microphones and its application to noise cancellation and speech enhancement. During noise alone period, the noise statistics are estimated for the application of noise cancellation and then it is used for the application of speech enhancement during speech plus noise period. This analysis and application may give rise to insight of modified applications for practical real-time processing in a realistic environment.

A. System Identification Based Analysis on Basic Noise Cancellation

Analysis is based on identification of acoustic noise path transfer functions between two microphones and it may be represented as the ratio of acoustic transfer functions, $H_2^{-1}(z)H_1(z)$ as shown in Fig. 1.

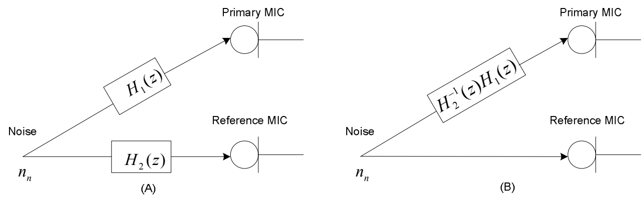


Fig. 1 Acoustic path transfer functions between two microphones showing (A) and (B) as equivalence

Pulsipher et al. [3] have analyzed for the identification of acoustic path transfer functions between two microphones and they investigated it as a data generation model, where the adaptive filter in the correlated noise between two microphones estimates the ratio of acoustic noise path transfer functions.

A simple method to estimate acoustic noise transfer function is to use adaptive filter, which may resolve the nonminimum phase problem in reverberant environment though a large amount of weights are needed.

Typical ANC approach is a noise canceller by using adaptive filter in the reference input, which minimize an output power in an MMSE (minimizing mean square error) sense. This method needs microphones set-up for a practical application such that reference microphone picks up noise signal only and primary microphone picks up the noise signal and the speech signal as well. Therefore, output (y_n) of adaptive filter from the reference input (x_n) approximates recursively from the output error signal (e_n) as shown in Fig. 2.

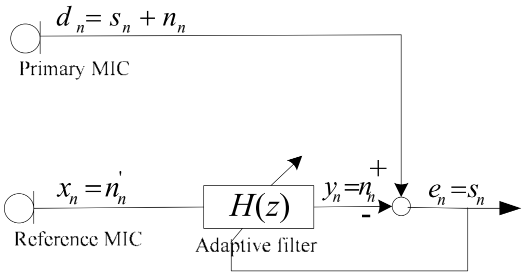


Fig. 2 Block diagram of typical ANC method

Modified application to an adaptive filter is to estimate the ratio of acoustic noise transfer functions during noise periods only and continuously update it until the frozen noise statistics on the last frame is applied to the speech with noise periods, where system output is subtracted from primary input. Therefore, a speech signal is remained alone at the output. This method needs a VAD to differentiate between noise periods and speech with noise periods.

In addition, the application of small separation between two microphones may give favorable effects that reduce significantly filter length required for noise cancellation and minimize the presence of reverberation [4].

Nevertheless, ANC method shows a limitation in maximum cancellation in SNR (signal to noise ratio). The well-known equation [5] shows that the maximum cancellation is related with an MSC (magnitude squared coherence) value from noise between primary microphone and reference microphone.

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

where $\left|\gamma_{xd}(e^{jw})\right|^2$ is MSC function.

According to (1), 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). It also shows that the noise cancellation performance of ANC is highest, when the microphones point in the same direction [5]. Fig. 4 shows example of average MSC showing 0.35 according to their environment.

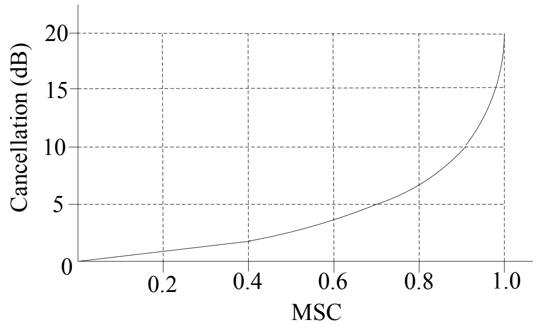


Fig. 3 Theoretical maximum cancellation for ANC as function of MSC

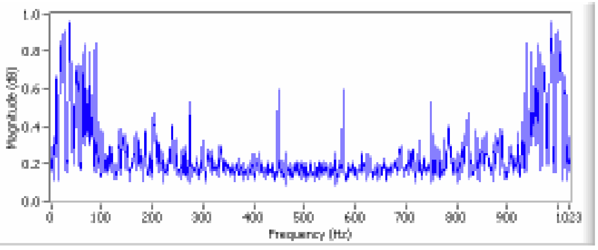


Fig. 4 Example: MSC showing average 0.35

MSC function is widely used in several applications of signal detection [6], time delay estimation [7], SNR estimation [8], a noise reduction scheme [9] and VAD application [10].

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. [1] have analyzed the problem of speech in the reference microphone and has shown that the resulting output SNR is given by (2). This is called 'power inversion', which results in the cancellation of speech at the output.

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

where $SNR_e(z)$ is SNR at error output and $SNR_x(z)$ is SNR at reference microphone input.

In the case that speech signal is not shown in reference microphone input, we have analysis on spectrum of output noise [1] as:

$$\Phi_{ON}(z) \cong \Phi_{nn}(z) |SNR_x(z)| |SNR_d(z)| \tag{3}$$

where $\Phi_{ON}(z)$ is spectrum of output noise, $\Phi_{nn}(z)$ is spectrum of input, $SNR_x(z)$ is SNR at reference microphone input and $SNR_d(z)$ is SNR at primary microphone input.

From the (3), it can be understood that it implies that the output noise spectrum depends on the input noise spectrum and also mplies 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. Finally it 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.

The application of close and direct speech in front of two microphones and the use of adaptive filter using VAD during noise periods only may reduce a speech distortion. Therefore, we now investigate analysis for identification of acoustic path transfer functions between two microphones and its application to basic noise cancellation method.

In the periods of noise alone of Fig. 5, we have equations at primary, reference inputs and output respectively as:

$$d_n = H_1(z)\, n_n \tag{4}$$

$$x_n = H_2(z)\, n_n \tag{5}$$

$$e_n = d_n - y_n = d_n - H(z)\, x_n = \{H_1(z) - H(z) H_2(z)\} n_n \tag{6}$$

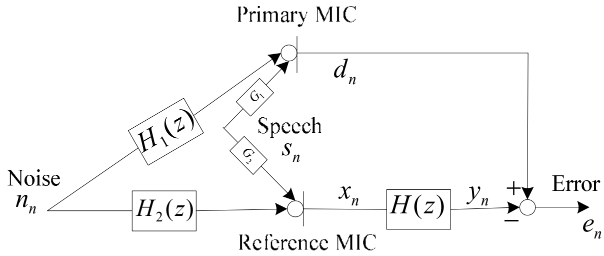


Fig.5 Block diagram of basic noise cancellation method

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 that $H_2(z)$ is minimum phase).

In the periods of speech with noise, also with acoustic transfer function of $H(z) = H_1(z)H_2^{-1}(z)$, it shows that

$$d_n = H_1(z)\, n_n + G_1(z)\, s_n \tag{7}$$

$$x_n = H_2(z)\, n_n + G_2(z)\, s_n \tag{8}$$

$$e_n = d_n - y_n = d_n - H(z)\, x_n = \{G_1(z) - H_1(z) H_2^{-1}(z) G_2(z)\} s_n \tag{9}$$

This indicates that to increase an SNR in speech periods by reducing noise, if we could estimate the ratio of unknown acoustic path transfer functions, $H(z) = H_1(z)H_2^{-1}(z)$, it can effectively cancel noise. Furthermore, 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.

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. 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 both close and direct speech in front of two microphones.

B. Analysis on Basic Speech Enhancement

Typical beamforming approach generally uses a multiple microphones array and gives advantages on speech enhancement by maximizing a speech directivity and speech separation by spatial discrimination.

Analysis shows that in Fig. 6, if we have noise (n_0) with $n_0 \approx 0$ at the reference microphone input, we can have speech signal alone at the output (e_n).

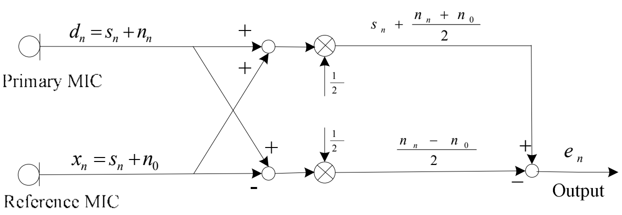


Fig. 6 Block diagram of typical beamforming method

For the identification of acoustic path transfer function between two microphones, in the periods of noise alone of Fig. 7, we have equations at primary, reference inputs and output respectively as:

$$d_n = H_1(z)\, n_n \tag{10}$$

$$x_n = H_2(z)\, n_n \tag{11}$$

$$e_n = 0.5\,(d_n + x_n) - 0.5\, H(z)(d_n - x_n) \tag{12}$$

$$= 0.5\,[\{H_1(z) + H_2(z)\} - H(z)\{H_1(z) - H_2(z)\}]n_n$$

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

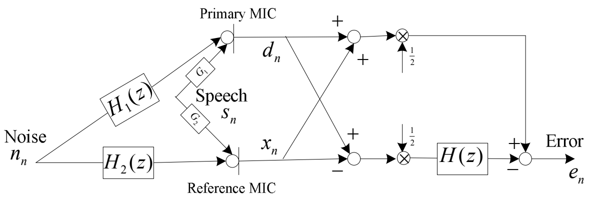


Fig. 7 Block diagram of basic speech enhancement method

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

$$d_n = H_1(z)\, n_n + G_1(z)\, s_n \tag{13}$$

$$x_n = H_2(z)\, n_n + G_2(z)\, s_n \tag{14}$$

$$e_n = 0.5\,(d_n + x_n) - 0.5\, H(z)\,(d_n - x_n) \tag{15}$$

$$= 0.5\,[\{G_1(z) + G_2(z)\} - H(z)\,\{G_1(z) - G_2(z)\}]\, s_n$$

This shows that error is speech signal alone when

$$H(z) = \{H_1(z) + H_2(z)\} / H_1(z) - H_2(z)\} \tag{16}$$

and $G_1(z) = G_2(z) \cong 1$ (17)

Note that it is not filtered error output as noise cancellation case. Therefore, we need to identify $H_1(z) + H_2(z)$ and $H_1(z) - H_2(z)$.

Suppose that $H_1(z) = 4z^{-2}$ and $H_2(z) = 2z^{-3}$

Then, $\dfrac{H_1(z) + H_2(z)}{H_1(z) - H_2(z)}$

$$= \frac{z^{-2}(4 + 2z^{-1})}{z^{-2}(4 - 2z^{-1})} = \frac{4 + 2z^{-1}}{4 - 2z^{-1}}$$

When $s_n = 0$, we know that $d_n' = \left(\dfrac{H_1(z) + H_2(z)}{2}\right) n_n$ and hence we find that $d_n' = \dfrac{1}{2}(4 + 2z^{-1})n_n$. We also find that $x_n' = \dfrac{1}{2}(4 - 2z^{-1})n_n$

Therefore, we can find that $H(z) = \dfrac{4 + 2z^{-1}}{4 - 2z^{-1}}$

For the error output,

$$e_n = d_n' - H(z)x_n' = s_n + \frac{1}{2}(4 + 2z^{-1})n_n - \left(\frac{4 + 2z^{-1}}{4 - 2z^{-1}}\right)$$

$\dfrac{1}{2}(4 - 2z^{-1})n_n = s_n$, which indicates that noise is cancelled and therefore, signal is only remained.

It shows that if we could estimate the ratio of unknown acoustic transfer functions (16), it can effectively enhance a speech and also if the speech can be delivered at an equal distance to both microphones with a minimal attenuation, e.g., $G_1(z) = G_2(z) \cong 1$, the resulting speech distortion will be negligible. For an estimation of unknown acoustic transfer functions, the denominator part of a transfer function in (16) should not be a nonminimum phase for a stable performance.

However, it indicates that for the non-speech distortion, we only require (16). It shows that application of both direct speech in front of the two microphones and a directivity function of sum and difference function can contribute to an increased SNR.

Based on above, for an application in a real environment, due to the nature of room reverberation, speech beamforming or TDOA function may be used to get an enhanced speech signal [11].

C. Summary of Analysis

Ideal direct speech application gives speech leakage in real reverberant environment for both approaches. Therefore, modified applications are:

1) The modified application to an ANC is to use a small separation (say 20~30cm) between two microphones and VAD (to get access to the noise statistics) during the noise periods. It could give benefits of a reduced adaptive filter size and minimized reverberation.

2) The modified application to G-J beamformer uses speech beamforming or TDOA function in front of sum and difference

function for signal blocking. This gives rise to increased speech directivity in the primary input and a refined noise reference in a reference input.

By using modified application 1) and 2), and application based on analysis of Fig. 6, kepstrum approach has been introduced [12,13].

III. Modified Versions of Two-Microphone ANC and Beamforming

A. *Modified Versions of Two-Microphone ANC*

Modified version of classic adaptive noise canceller has been introduced in the need of preventing desired signal from crosstalk due to small separation between microphones. Mirchandani et al. [14] have introduced CTRANC (crosstalk resistant adaptive noise canceller) (Fig. 8) and Compernolle and Gerven [15] uses it to SAD (symmetric adaptive decorrelation) algorithm (Fig. 9).

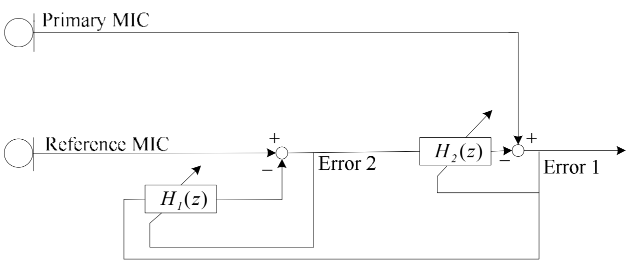


Fig. 8 Crosstalk resistant adaptive noise canceller

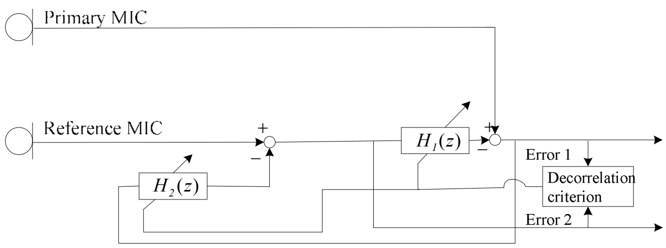


Fig. 9 Symmetric adaptive algorithm

The method using multiple adaptive noise cancellers has been introduced. Wallace and Goubran [16] provide methods using adaptive noise canceller in parallel in multiple reference inputs (Fig. 10) and adding sub-band processing to its algorithm (Fig. 11).

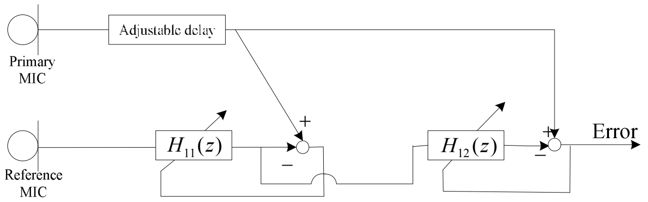


Fig. 10 Two-microphone approach to two stage beamforming multireference adaptive noise canceller

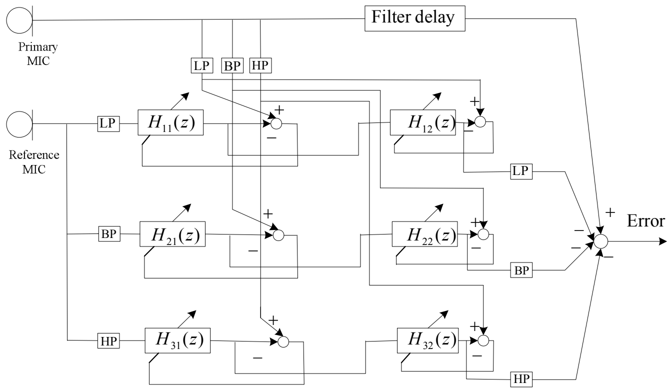


Fig. 11 Two-microphone approach to sub-banded two stage beamforming multi-reference adaptive noise canceller with sub-banded second stage

B. *Modified Versions of Two-Microphone Beamforming*

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 simple methods, delay and sum beamformer (Fig. 12) has been introduced and later, it becomes an important component with adaptive algorithm in modified adaptive beamforming techniques.

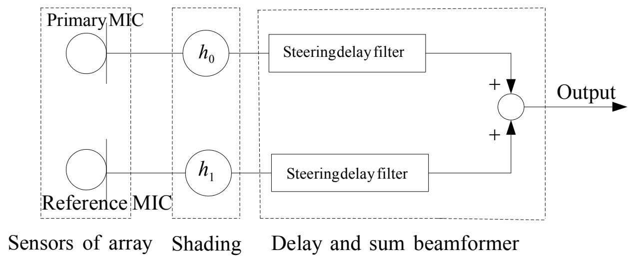


Fig. 12 Two-microphone approach to a weighted delay and sum beamformer

The basic concept of delay and sum beamformer is to provide a spatially enhanced signal by discriminating desired speech signal from unwanted noise signals by DOA (direction of arrival). Delaying appropriately in TDOA (time delay of arrival) and summing all signals arriving from the in-phased angle ultimately produce an increase in SNR.

As a derived version of delay and sum beamformer, the adaptive Frost beamformer with a constrained adaptive algorithm has been proposed [17] (Fig. 13). The purpose is to adapt to preserve desired signals from straight ahead and to minimise noise signals from other direction. So its algorithm is constrained to a chosen frequency response in the look direction, and then iteratively adapts the weights to minimise the noise power at the output.

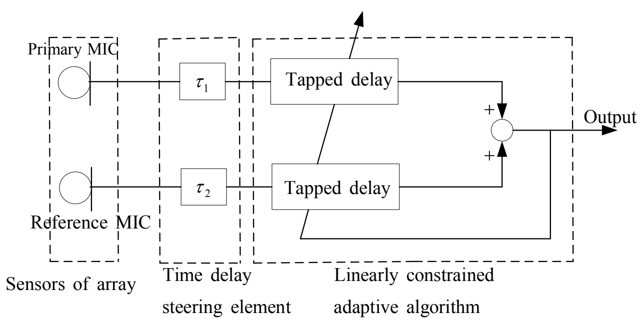


Fig. 13 Two-microphone approach to adaptive Frost beamformer

As an alternative implementation of the adaptive Frost beamformer, G-J beamformer consisting of three building blocks, 1) constrained, fixed beamformer in a primary input 2) blocking matrix to provide a noise reference input and 3) unconstrained adaptive noise canceller has been proposed [2] (Fig. 14).

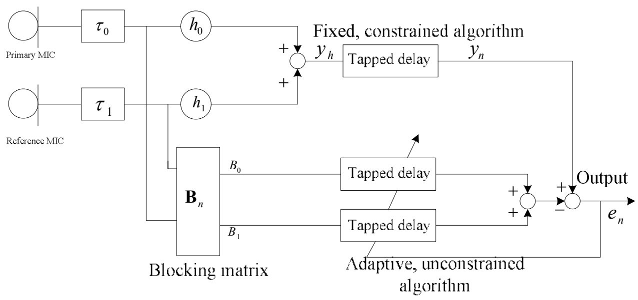


Fig. 14 Two-microphone approach to G-J beamformer

Fixed beamformer is typically designed to maximize directionality and adaptive noise cancellation provides additional benefits in time varying acoustic condition. It alleviates problems of signal cancellation and misadjustment that arise in the presence of strong desired signals.

Structure of G-J beamformer provides concrete basis for the versatile and refined methods to the modified two microphones adaptive beamforming technology. A 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.

Compernolle [11] has introduced a switching two stage adaptive filters using multiple microphones. This method is based on delay and sum beamformer and adaptive noise canceller. 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 switching VAD properly (Fig. 15). Moir [18, 19] has implemented it using LabVIEW software on PC.

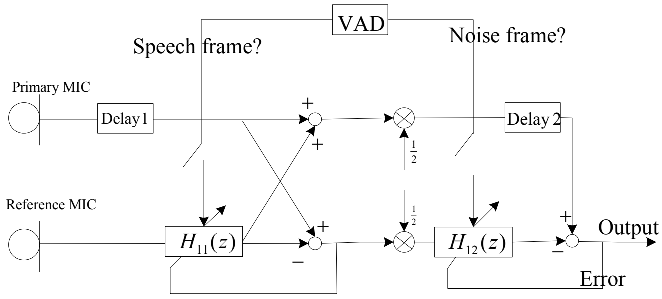


Fig. 15 Two-microphone approach to switching multiple adaptive filters

As a similar method, Berghe and Wouters [20] have proposed two microphones 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 adaptive noise canceller, only allowed to adapt during absence of speech (Fig. 16).

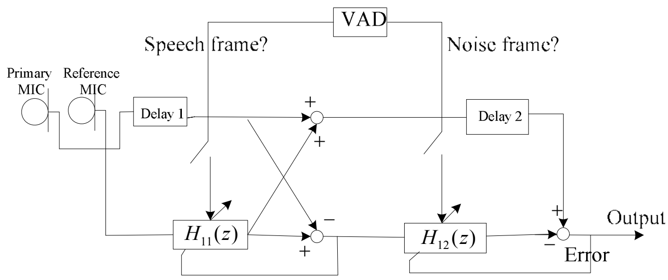


Fig. 16 Adaptive noise canceller for hearing aids using two microphones

The use of a multiple sub-band processing to those methods can add a more improvement in SNR, but at the expense of a significant increase in computation and complexity.

IV. Algorithm for Modified Applications

Based on the analysis, the modified applications utilize such algorithms as NLMS algorithm of adaptive filter, TDOA estimate and VAD function, which are described in this section.

A. Adaptive Filter

It is well known that the performance of ordinary LMS is evaluated in terms of convergence rate and stability. The performance is directly related to the proper selection of step size ($0 < \mu < 1$), which shows a compromising effect between convergence rate and stability. It often gives rise to poor performance due to slow convergence or unstability for a real-time application in nonstationary environment.

However, from the performance analysis of the LMS algorithm [21], it shows that to be convergent or stable in the mean, step size should be $0 < \mu < 2/\lambda_{\max}$. The analysis shows that the maximum value of μ depends on the largest eigenvalue

$\lambda_{\max}$ of the input autocorrelation $\mathbf{R}$ and it is approximated to the trace of autocorrelation, $tr(\mathbf{R})$ and also to input power, $\|\mathbf{x}_n\|^2$ (i.e., $\lambda_{\max} \approx tr(\mathbf{R}) \approx \|\mathbf{x}_n\|^2$). This shows that the input power signal is related to the maximum value of μ. Accordingly, the condition of step size for the stable adaptation should be limited to:

$$0 < \mu < 2/\lambda_{\max} \approx 2/tr(\mathbf{R}) \approx 2/\|\mathbf{x}_n\|^2 \tag{18}$$

Equation (18) is often referred as the condition for convergence in the mean-square.

As described above, the NLMS (19 and 20) uses the input-dependent adaptation step size, which could provide benefits of a faster convergence and better stability than ordinary LMS.

$$h_{n+1} = h_n + \mu_n \mathbf{x}_n e_n \tag{19}$$

$$e_n = d_n - y_n = d_n - h_n \mathbf{x}_n \tag{20}$$

$$\mu_n = \frac{\tilde{\mu}}{\alpha + \|\mathbf{x}_n\|^2}, 0 < \tilde{\mu} < 2 \tag{21}$$

where μ_n is a modified input dependent step size and α of very small positive value is added to prevent the possibility of zero division when we have a very small input value.

B. TDOA Estimate

The estimation procedure for non-parametric power spectrum is illustrated in Fig. 17, which shows example of auto periodogram from reference microphone. It is estimated from windowed FFTs (fast Fourier transforms) as a discrete estimate of continuous power spectral density.

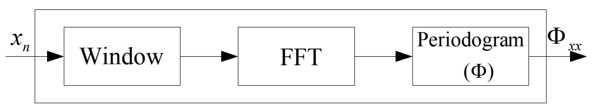


Fig. 17 Periodogram estimation procedure

For the periodogram estimates, the modified WOSA (weighted overlapped segment averaging) algorithm has been used. The smoothing auto periodograms (22 and 23) and cross periodogram (24) are processed from 50% overlapping Hamming windowed 2048 FFTs as a discrete estimate of continuous power spectral density with the use of exponentially forgetting factor, β=0.8 ($0 < \beta < 1$).

$$\Phi_{dd}(i) = \beta\Phi_{dd}(i-1) + (1-\beta)X_d(i)X_d^*(i) \tag{22}$$

$$\Phi_{xx}(i) = \beta\Phi_{xx}(i-1) + (1-\beta)X_x(i)X_x^*(i) \tag{23}$$

$$\Phi_{dx}(i) = \beta\Phi_{dx}(i-1) + (1-\beta)X_d(i)X_x^*(i) \tag{24}$$

The application of 50% overlapping and 2048 window size gives a processing time of 46msec. It is a little over speech stationarity range (20-40msec) but has a dense frequency resolution (10.76Hz) to differentiate between speech and noise signals.

From the estimates of auto- and cross-periodograms, we have TDOA estimate [22] (26) from the maximum value of sampled generalized cross correlation (GCC) (25).

$$R_{dx}(k) = F^{-1}[\psi(i)\Phi_{dx}(i)] \tag{25}$$

$$\text{TDOA } d = \max R_{dx}(k) \tag{26}$$

where F^{-1} denotes inverse FFT and $\psi(i)$ is HT (Hannan-Thompson) weighting function as,

$$\psi(i) = \frac{|\gamma_{dx}(i)|^2}{|\Phi_{dx}(i)|[1-|\gamma_{dx}(i)|^2]} \tag{27}$$

where $|\gamma_{dx}(i)|^2$ is magnitude squared coherence function (MSC) (28).

Fig. 18 shows geometry for TDOA estimate. For example, let us set up the microphones (with distance $(l = 30cm)$ between microphones) that noise source is placed from primary microphone input with distance (d_1) of 30cm and also from reference microphone input with distance (d_2) of 35cm. Now we have a distance difference, $d = d_2 - d_1 = 5\ cm$. For the sampling frequency of 22050Hz, we have a sampling interval of 45.35 μs $(T_s = 1/f_s)$. Therefore, time delay can be calculated by the equation, $\tau = d/c$ $= (0.05m/330m)$ secconds $= 151.5\mu s$. It indicates that 3.34 samples delay (τ/T_s) with bearing of $\theta = 80.40^\circ$ $(\theta = \cos^{-1}(c \cdot \tau / l))$.

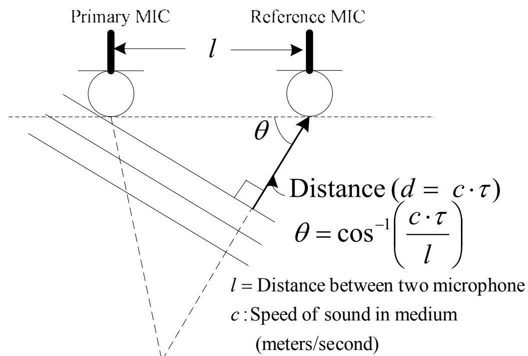


Fig. 18 Geometry used to estimate TDOA of an acoustic source [7]

Fig. 19 shows example of TDOA estimate showing maximum value at an interpreted -7 samples, which occurs at sample number 1016.

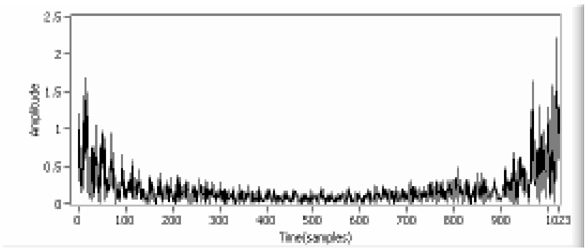


Fig. 19 Example of TDOA

C. VAD

The VAD [23] is used to select periods between noise only and speech plus noise. It uses the values from two functions, the TDOA d samples (26) and the average MSC (28). Both estimates are auto- and cross-periodograms based and used as direction-finder to locate the position of a source and limiter of the viewing zone to avoid problems with reverberations, respectively.

$$|\gamma_{dx}(i)|^2 = \frac{|\Phi_{dx}(i)|^2}{\Phi_{dd}(i)\Phi_{xx}(i)} \tag{28}$$

V. Methods for Modified Applications

With the modified application, performance is to be compared between NLMS algorithm of ANC in noise cancellation method (method I) and TDOA function of G-J beamformer in speech enhancement method (method II). As front-end application in G-J adaptive beamformer, the performance is also to be compared between TDOA function (method III) and speech beamforming filter (method IV).

A. MethodI: ANC Based Approach

Method I uses a modified application to ANC, which uses a small separation between two microphones with the use of a VAD during noise periods. The speech appears directly in front of the microphones.

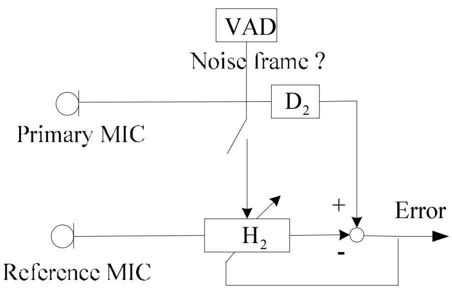


Fig. 20 Block diagram of ANC based approach

B. MethodII: G-J Based Approach – With TDOA

Method II uses TDOA delay as a speech directivity function (steering mechanism) in front of sum and difference function. It applies with the same application conditions of the modified application to ANC, but not with an adaptive filter.

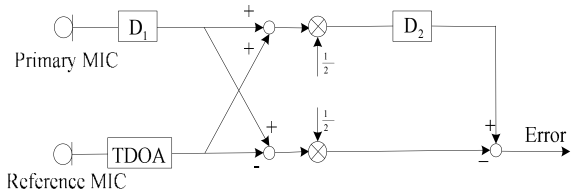


Fig. 21 Block diagram of G-J based approach - with TDOA

C. MethodIII: G-J Based Approach – With TDOA and Adaptive Filter

Method III uses an adaptive version of method II, which uses a TDOA compensation delay for steering the speech to be in front of the microphones and NLMS algorithm to minimize mean-squared error. An NLMS algorithm is used for noise cancellation and updated only during noise periods and frozen in speech periods.

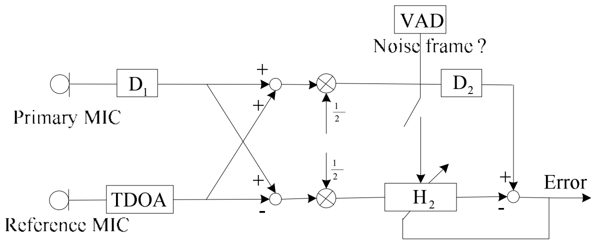


Fig. 22 Block diagram of G-J based approach - with TDOA and adaptive filter

D. MethodIV: G-J Based Approach – With Speech Beamforming and Adaptive Filter

Method IV uses two NLMS algorithms. The weights of first NLMS algorithm are updated during speech periods for speech enhancement whilst the weights of second NLMS algorithm are updated only during noise periods for noise cancellation.

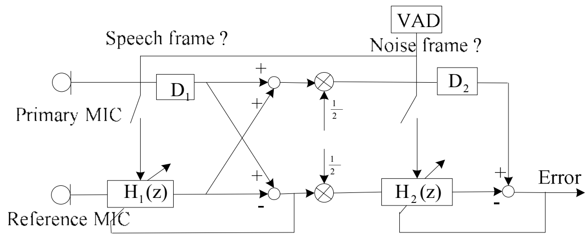
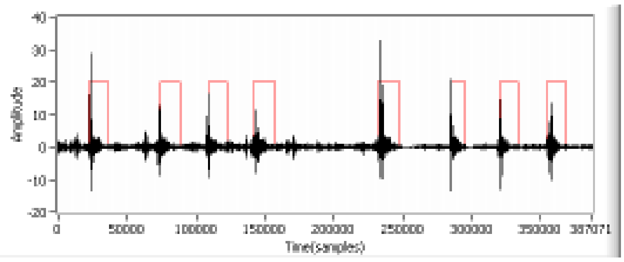
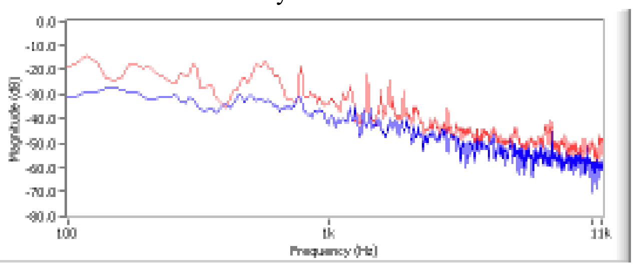


Fig. 23 Block diagram of G-J based approach - with speech beamforming and adaptive filter

Fig. 24 shows performance comparison between method II (G-J beamformer based approach – with TDOA) and method IV (G-J based approach – with speech beamforming and adaptive filter)



(A) Performance in speech with noise period between method II (first half) and method IV (second half): Speech signal is trigged by VAD



(B) Average power spectra between method IV (bottom) and method II (top) during noise period

Fig. 24 Performance of method IV compared with method II

VI. Experiments

A. Experimental Set-Up

Experiments are implemented in a room, where the test places are at desk A as illustrated in Fig. 25. The room dimensions are shown. The background noise level is measured as 48dBA by using a sound level meter (Digitech QM1589).

The speech signals are sampled using a standard internal sound card and two preamplifiers with unidirectional electret condenser microphones. The sampling frequency is chosen to be 22050Hz with 16 bits resolution per channel with the Nyquist frequency bandwidth of around 11 kHz.

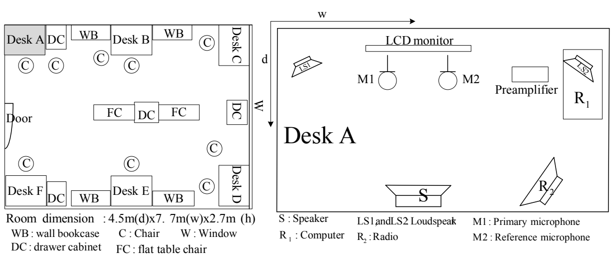


Fig. 25 Experimental environment

Room reverberation time is calculated as 1.18 seconds for the frequency 500Hz with calculation of absorption coefficients of surfaces of wall, floor and ceiling of the room. It is assumed that rooms reflect a moderately reverberant situation.

B. Experimental Methodology

In the diagrams of the four methods in Figs. 20, 21, 22 and 23, 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. Adaptive filter weights, 100 and 200 are used for $H_1(z)$ and $H_2(z)$ respectively.

Three different types of microphone configuration are considered for the application of the hearing aids as shown in Fig. 26.

For the comparison of computational complexity, the required computational complexity for real-time processing of TDOA function and NLMS algorithm is measured by the complexity of multiplication in FLOPS (floating point operations per second).

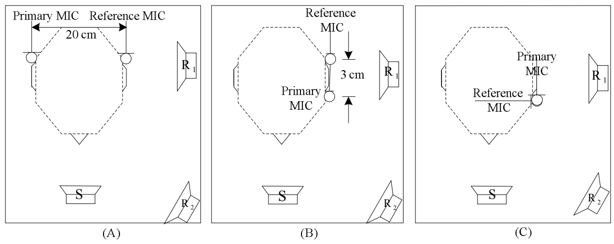


Fig. 26 Experimental microphone set-up (S: speaker, R_1 : computer fan noise, R_2 : radio noise): (A) broadside, (B) endfire and (C) endfire variant

C. Summary

The performance comparisons are described below and summarized in Table 1. Computational complexity has also been measured and illustrated in Table 2.

1) The modified application to ANC (method I) and G-J beamformer (method II) shows almost the same noise reduction ratio in 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 directivity (steering) function of TDOA (method II) of speech enhancement method shows higher performance than NLMS algorithm (method I) of noise cancellation method.

2) For the computational complexity comparison, it shows that in the case when 200 NLMS weights are used, real multiplication is 0.12M ($M = 10^6$) and 2.4M for its iteration for a convergence. On the other hand, for the TDOA function, the total computation is 0.058M per N = 2048 samples. Table 2 shows the required processing and number of FLOPS for computational complexity of the TDOA function and NLMS algorithm.

3) 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. However, we should consider this method with the highest computational complexity and high demand of an accurate performance on VAD, because it will give wrong operation, therefore results in poor performance.

4) It shows little difference among the three different microphones configuration, i.e., broadside, endfire and endfire variant. It is assumed that in a reverberation environment, the

performance does not depend significantly on microphones configuration.

Table 1 Test results based on stationary (computer fan), nonstationary (radio) noise, with and without speech

Test type	Test type I - based on stationary noise (computer fan)			Test type II - based on stationay noise (computer fan) and nonstation ary(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	− 32.0dB	− 33.5dB	− 28.5dB	−30.2dB	−30.1dB	−29.6dB	−26.0dB	−28.5dB	−25.6dB
Method II	− 32.3dB	−34.5dB	−33.3dB	−31.4dB	−30.9dB	−30.2dB	−31.4dB	−31.5dB	−31.7dB
Method III	−42.5dB	− 39.0dB	−37.8dB	−36.1dB	−35.0dB	−35.8dB	−33.1dB	−33.1dB	−33.3dB
Method IV	−43.7dB	− 40.0dB	−38.1dB	−37.2dB	−35.4dB	−36.1dB	−35.7dB	−35.1dB	−34.1dB

Table 2 Comparison of TDOAand NLMS algorithm in FLOPS

Algorithm	Required processing	FLOPS
TDOA	WOSA (A) FFT/IFFT (B)	$N\log_2(5.12/\Delta f)$ $(N/2)\log_2 N$
	Total computation	3A + 2B
NLMS	Real multiplication (C) Iterations (D)	$3N^2+2N$ 20
	Total computation	C · D

VII. CONCLUSIONS

From the theoretical analysis of basic structures of noise cancellation and speech enhancement method, the four different methods has been investigated using modified application from typical ANC and G-J beamformer. The purpose is to find best solution from speech distortion in a reverberant environment, computational complexity in real-time processing and microphones configuration for the application of hearing aids.

It has been shown that speech enhancement method using TDOA as speech directivity function gives better performance than noise cancellation method using NLMS algorithm in adaptive filter, especially in speech with noise period. Furthermore, the front-end application of TDOA function also gives benefit of computational simplicity than rear-end NLMS algorithm in speech enhancement method. In conclusion, the method (III) using TDOA function as speech directivity function from adaptive noise cancelling structure gives the most promising result. Investigation about three microphones configuration shows little difference in this experiment. For future work, the analysis based on identification of acoustic transfer functions between two microphones and the modified application for two-microphone approach could give a fundamental basis to application of hardware prototype implementation of digital adaptive hearing aids.

REFERENCES

[1] B. Widrow, J. R. Glover JR, J. M. Mc Cool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong JR, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications.," *IEEE Proceedings*, vol. 63, pp. 1692-1716, 1975.

[2] L. J. Griffiths and C. W. Jim, "An alternative to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propogation*, vol. 30, pp. 27-34, 1982.

[3] D. Pulsipher, S. F. Boll, C. Rushforth and L. Timothy, "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, 1979

[4] W. A. Harrison, J. S. Lim and E. Singer, "A new application of adaptive noise cancellation," *Acoustics, Speech, and Signal Processing, IEEE Transactions*, vol. 34, pp. 21-27, 1986.

[5] M. M. Goulding, and J. S. Bird, "Speech enhancement for mobile telephony," Vehicular Technology, IEEE Transactions, 39 (4): pp. 316-326, 1990

[6] G. C. Carter, "Receiver operating characteristics for a linearly thresholded coherence estimation detector.", IEEE transactions 25 (1): pp. 90-92, 1977

[7] G.C. Carter, "Coherence and time delay estimation." Proceedings of the IEEE 75 (2): pp.236-255, 1987

[8] J. W. Fay, "Coherence bounds for signal-to-noise ratios from magnitude-squared coherence estimates." IEEE Trans. Acoust., Speech, Signal Processing ASSP-28: pp.758-760, 1980

[9] J. J. Rodriquez, "Adaptive noise reduction in aircraft communication systems." MIT Lincoln Lab. Lexington, MA, Tech. Rep. 756, 1987

[10] R. Le Bouquin Jeannes, and G. Faucon, "Proposal of a voice activity detector for noise reduction." Electronics Letters 30(12): pp.930-932, 1994

[11] D. V. Compernolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," presented at ICASSP-90 pp. 833-836, Albuquerque

[12] J. Jeong, "Kepstrum analysis and real-time application to noise cancellation" in *WSEAS 8th International Conference on Signal Processing, Robotics and Automation (ISPRA)*, pp. 149-154, The University of Cambridge, Cambridge, UK, 21-23 Feb. 2009

[13] J. Jeong and T. J. Moir, "A real-time kepstrum approach to speech enhancement and noise cancellation," *Neurocomputing,* vol. 71, pp. 2635-2649, 2008

[14] G. Mirchandani, R. Jr. Gaus, and L. K. Bechtel, "Performance characteristics of a hardware implementation of the cross-talk resistant adaptive noise canceller". ICASSP 86, pp.93-96, 1986

[15] D. V. Compernolle, and S. V. Gerven, "Signal separation in a symmetric a daptive noise canceler by output decorrelation", Acoustics, Speech, and Signal Processing, IEEE International Conference, vol. 4, pp.221-224, 1992

[16] R. B. Wallace, and R. A. Goubran, "Noise cancellation using parallel adaptive filters." IEEE Transaction on circuits and systems-II: Analog and digital signal processing 39 (4): pp.239-243, 1992

[17] O. L. III Frost, "An algorithm for linearly constrained adaptive array processing." Proc. IEEE 60 (8): pp.926-935, 1972

[18] 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.

[19] T. J. Moir, "Real-time acoustic beamforming on a PC," *WSEAS Transactions on Signal Processing,* vol. 2, pp. 167-174, 2006.

[20] J. V. Berghe, and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones." Journal of the acoustical society of america 103 (6): pp.3621-3626, 1998

[21] S. Haykin, *Adaptive filter theory*, Third ed. Upper saddle river, NJ.: Prentice-Hall, Inc., 1996.

[22] C. Knapp, and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. ASSP-24, pp. 320-327, 1976.

[23] H. Agaiby, and T. J. Moir, "A robust word boundary detection algorithm with application to speech recognition," presented at digital signal processing proceedings, 1997, 13th international conference, vol.2, pp. 753-755, 1997.

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.