Analysis and Comparative Study on Two-Microphone Noise Cancellation and Speech Enhancement Methods for Real-Time Hearing Aids Application

J. JEONG
Faculty of Biomedical Engineering and Health Science
Universiti Teknologi Malaysia
81310 Skudai, Johor
MALAYSIA
jeong@utm.my

Abstract: This paper provides analysis and real-time performance comparisons on modified application from two-microphone classic adaptive noise cancelling and beamforming methods. 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 insight and interests for hardware prototype implementation of digital adaptive hearing aids.

Key-Words: ANC, Beamforming, TDOA, NLMS, Direct Speech, VAD

1 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, noncaualty 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 adaptive 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) function, 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 Griffiths and Jim (G-J) 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 square) algorithm is compared for the application in a real-time implementation.

2 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.

2.1 Analysis on basic noise cancellation
In the periods of noise alone of Fig. 1, we have equations at primary, reference inputs and output respectively as

\[ d_n = H_1(z) n_n \]

\[ x_n = H_2(z) n_n \]

\[ e_n = d_n - y_n = d_n - H(z) x_n = \{H(z) - H(z)H(z)H(z)\} n_n \]

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 \( H(z) = H_1(z)H_2^{-1}(z) \), it shows that

\[ d_n = H_1(z) n_n + G_1(z) s_n \]

\[ x_n = H_2(z) n_n + G_2(z) s_n \]

\[ e_n = d_n - y_n = d_n - H(z) x_n = \{G(z) - H(z)H_1(z)G(z)\} s_n \]

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, \( G_1(z) = G_2(z) \equiv 1 \), 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.

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. 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 [3].
2.2 Analysis on basic speech enhancement

In the periods of noise alone of Fig. 2, we have equations at primary, reference inputs and output respectively as
\[ d_n = H_1(z) n_n \] (7)
\[ x_n = H_2(z) n_n \] (8)
\[ e_n = 0.5 (d_n + x_n) - 0.5 H(z) (d_n - x_n) \] (9)
\[ = 0.5 \left[ \{ H_1(z) + H_2(z) \} - H(z) \{ H_1(z) - H_2(z) \} \right] n_n \]

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. 2 Block diagram of basic speech enhancement method](image)

In the periods of speech with noise using the above expression for \( H(z) \),
\[ d_n = H_1(z) n_n + G_1(z) s_n \] (10)
\[ x_n = H_2(z) n_n + G_2(z) s_n \] (11)
\[ e_n = 0.5 (d_n + x_n) - 0.5 H(z) (d_n - x_n) \] (12)
\[ = 0.5 \left[ \{ G_1(z) + G_2(z) \} - H(z) \{ G_1(z) - G_2(z) \} \right] s_n \]

This shows that error is speech signal alone when \( H(z) = \{ H_1(z) + H_2(z) \}/ \{ H_1(z) - H_2(z) \} \) (13) and \( G_1(z) = G_2(z) \geq 1 \) (14)

It shows that if we could estimate the ratio of unknown acoustic transfer functions (13), 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., \( G_1(z) = G_2(z) \geq 1 \), the resulting speech distortion will be negligible. For an estimation of unknown acoustic transfer functions, the denominator part of a transfer function in (13) should not be a nonminimum phase for a stable performance.

However, it indicates that for the non-speech distortion, we only require (14). 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 [4].

2.3 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 20~30 cm) 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.

3 Algorithms for modified application

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.

3.1 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 [5], it shows that to be convergent or stable in the mean, step size should be \( 0 < \mu < 2/\lambda_\text{max} \).

The analysis shows that the maximum value of \( \mu \) depends on the largest eigenvalue \( \lambda_\text{max} \) of the input autocorrelation \( R \) and it is approximated to the trace of autocorrelation, \( tr(\mathbf{R}) \) and also to input power,
\[ \| \mathbf{s} \|^2 \] (i.e., \( \lambda_\text{max} \approx tr(\mathbf{R}) \approx \| \mathbf{s} \|^2 \)). This shows that the input power signal is related to the maximum value of \( \mu \). Accordingly, the condition of step size for the stable adaptation should be limited to:
\[ 0 < \mu < 2/\lambda_\text{max} \approx 2/\{ tr(\mathbf{R}) \} \approx 2/\| \mathbf{s} \|^2 \] (15)

Eq. (15) is often referred as the condition for convergence in the mean-square.

As described above, the NLMS (16 and 17) uses the input-dependent adaptation step size, which could provide benefits of a faster convergence and better stability than ordinary LMS.
\[ \mathbf{h}_{n+1} = \mathbf{h}_n + \mu_n \mathbf{x}_n e_n \] (16)
\[ e_n = d_n - y_n = d_n - h_n x_n \]  
(17)

\[ \mu_n = \frac{\tilde{\mu} + \alpha }{2}, 0 < \tilde{\mu} < 2 \]  
(18)

where \( \mu_n \) is a modified input dependent step size and \( \alpha \) of very small positive value is added to prevent the possibility of zero division when we have a very small input value.

### 3.2 TDOA estimate

The estimation procedure for non-parametric power spectrum is illustrated in Fig. 3, 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.

\[ \Phi_{xx}(i) = \left| \frac{\Phi_{dd}(i)}{\Phi_{dd}(i) \Phi_{ss}(i)} \right|^2 \]  
(25)

#### 4 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).

#### 4.1 Method I: 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.

\[ R_d(k) = F^{-1}[\psi(i)\Phi_{dd}(i)] \]  
(22)

\[ d = \max R_d(k) \]  
(23)

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

\[ \psi(i) = \frac{|\gamma_{dd}(i)|^2}{\Phi_{dd}(i)[1 - |\gamma_{dd}(i)|^2]} \]  
(24)

\[ |\gamma_{dd}(i)|^2 \] is magnitude squared coherence function (MSC) (25).

#### 4.2 Method II: 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.

\[ |\gamma_{dd}(i)|^2 = \frac{\Phi_{dd}(i)^2}{\Phi_{dd}(i)\Phi_{ss}(i)} \]  
(25)
4.3 Method III: 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. 6 Block diagram of G-J based approach - with TDOA and adaptive filter

4.4 Method IV: G-J based approach - with speech beamforming and adaptive filter
Method IV uses two NLMS algorithms. The weights of first NLMS algorithm is updated during speech periods for speech enhancement whilst the weights of second NLMS algorithm is updated only during noise periods for noise cancellation.

Fig. 7 Block diagram of G-J based approach - with speech beamforming and adaptive filter

5 Experiments
5.1 Experimental set-up
Experiments are implemented in a room, where the test places are at desk A as illustrated in Fig. 8. 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. 8 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.

5.2 Experimental methodology
In the diagrams of the four methods in Figs. 4, 5, 6 and 7, D1 and D2 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. 9.

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

5.3 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.

6 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, these remains further investigation to application of hardware prototype implementation of digital adaptive hearing aids.

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

<table>
<thead>
<tr>
<th>Test type</th>
<th>Test type I: Based on stationary noise (computer fan)</th>
<th>Test type II: Based on stationary noise (radio)</th>
<th>Test type III: Based on stationary noise (computer fan)</th>
<th>Test type IV: Based on stationary noise (radio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power noise</td>
<td>Average</td>
<td>Power noise</td>
<td>Average</td>
<td>Power noise</td>
</tr>
<tr>
<td>Type</td>
<td>Broadside</td>
<td>Endfire</td>
<td>Variant</td>
<td>Broadside</td>
</tr>
<tr>
<td>Method I</td>
<td>34.5dB</td>
<td>34.0dB</td>
<td>33.5dB</td>
<td>33.0dB</td>
</tr>
<tr>
<td>Method II</td>
<td>38.5dB</td>
<td>38.0dB</td>
<td>37.5dB</td>
<td>37.0dB</td>
</tr>
<tr>
<td>Method III</td>
<td>42.5dB</td>
<td>42.0dB</td>
<td>41.5dB</td>
<td>41.0dB</td>
</tr>
<tr>
<td>Method IV</td>
<td>46.5dB</td>
<td>46.0dB</td>
<td>45.5dB</td>
<td>45.0dB</td>
</tr>
</tbody>
</table>

Table 2 Comparison of TDOA and NLMS algorithm in FLOPS

Algorithm | Required processing | FLOPS |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TDOA</td>
<td>WOLA (A)</td>
<td>$N \log_2 \left( \frac{M}{2} \right)$</td>
</tr>
<tr>
<td></td>
<td>FFT/IFFT (B)</td>
<td>$\frac{N^2 - 2}{2} \log_2 N$</td>
</tr>
<tr>
<td></td>
<td>Total computation</td>
<td>$\Delta + \Delta$</td>
</tr>
<tr>
<td>NLMS</td>
<td>Real multiplication (C)</td>
<td>$\frac{3N^2 + 2N}{2}$</td>
</tr>
<tr>
<td></td>
<td>Iterations (D)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Total computation</td>
<td>C - D</td>
</tr>
</tbody>
</table>

References:


