Adaptive Microphone Array-based Filter in the Speech Enhancement

Chih-Chia Yao
Department of Computer Science and Information Engineering
Chaoyang University of Technology,

Abstract - This study applied microphone array technology to develop an adaptive hybrid filter for enhancing speech signals. In the past decade, numerous studies have confirmed the excellent performance of microphone array technology in speech source localization and noise reduction. The proposed hybrid filter applies adaptive wavelet filter and spectral subtraction method for signal processing before filtering the signals by microphone array. Both the spectral subtraction method and adaptive wavelet thresholding method are highly effective for signal denoising. However, spectral subtraction method is ineffective for low-SNR signals whereas adaptive wavelet thresholding method is ineffective for high-frequency signals. The filtering behavior between adaptive wavelet filter and spectral subtraction method is controlled by a feed-forward fuzzy neural network. Experimental results reveal that the proposed filter outperforms other methods of speech signal enhancement.

Keywords: speech enhancement; microphone array; wavelet filter; spectral subtraction method;

I. INTRODUCTION

Speech enhancement is a continuing challenge in the field of speech and signal processing, particularly in applications such as mobile communications, speech recognition and hearing aids [1]. Although speech enhancement algorithms have been studied intensively in the last two decades, unresolved issues include distortion of the original speech signal and residual noise in the form of musical tones created by the enhancement algorithms [2].

Microphone arrays are now used to enhance speech by analyzing the spatial diversity indicated by sensors at different spatial locations [3]. Array signal processing techniques combine the outputs of the various sensors in a process, termed beamforming, aimed at cleaning the received signals from the contaminating interference and noise. Studies indicate that adaptive processing of array signals significantly improves target-to-jammer ratio, depending on acoustic conditions and the degree of reverberation.

Classical beamforming theory dates back to the mid-to-late twentieth century during which major advances were made [4]. Beamforming is the filtering and discriminating of active speech sources from various noise sources based on location. Proposed beamformers include delay-and-sum beamformers, superdirective beamformers, differential microphone arrays, and frequency-in-variant beamformers [5, 6, 7, 8]. The beamformer output can also be enhanced by applying a post-filter [9]. Because the typical beamformer does not provide adequate noise reduction, various post-filtering techniques are needed to further filter residual noise out of the beamformer output.

Many filtering techniques have been developed to augment the basic recognizer with additional components to cope with environmental interference such as additive noise. Proposed solutions can be roughly classified as digital signal processing and statistical analysis. The former usually remove an estimate of the distortion from the contaminated signals, e.g. spectral subtractive [10]. The latter uses statistical modeling to predict structures and patterns in the signal process [11].

The principle of a spectral subtractive algorithm is to obtain the best possible estimations of the short time spectra of a speech signal from a given contaminated speech. This approach estimates the power spectral density of a clean speech signal by subtracting the power spectral density of the noise from that of the contaminated signal. The main attractions of a spectral subtractive algorithm are its simply implementation and its capability to accommodate varying subtraction parameters.

However, spectral subtractive algorithms have several deficiencies such as imprecise estimation of signal and noise parameters and mismatched probability distribution models of speech or noise. Most noise suppression methods result in some loss of speech information [11]. Subtractive-type denoising methods introduce musical residual noise that perceptibly degrades processed speech. Various methods of suppressing musical noise have been proposed [12, 13].

Wavelet thresholding method is also widely used to shrink the signal and remove the noise [14]. In wavelet transform the signals are decomposed into a sparse, multi-scale representation. Because of its flexible time-frequency resolution, the wavelet transform is an appropriate tool for analyzing signals consisting of short high-frequency bursts and long quasi-stationary segments. By adequately choosing adaptive wavelet threshold, the corrupting white Gaussian noise can be efficiently removed by subtracting a threshold from noisy wavelet coefficients [15].

This study proposes an adaptive microphone array-based filter to reduce the effect of additive noise. The pre-filter and post-filter in the proposed framework cooperate to filter out signals. The pre-filter removes contaminated signals by adaptive wavelet thresholding method and spectral subtraction method [15, 16]. Both the spectral subtraction method and adaptive wavelet thresholding method provide excellent signal denoising. However, spectral subtraction method is ineffective for low-SNR signals, and adaptive wavelet thresholding method is ineffective for high-frequency signals. Hence, a feed-forward fuzzy neural network was adopted to reach the near-optimal signal mixed ratio of spectral subtraction to adaptive wavelet thresholding so that their disadvantages would offset each other. The post-filter with microphone array can then effectively compensate for noise and reverberation.

The remainder of this paper is organized as follows. In Section 2 we review the basic concept of microphone array,
adaptive wavelet filter and spectral subtraction method. In Section 3 the scheme of the adaptive microphone array-based filter is introduced. In Section 4 a performance evaluation of the proposed system is presented and some comparisons with other protocols are made. Our conclusions are made in Section 5.

II. BASIC CONCEPTS

A. Microphone Array

Consider the linear microphone array consisting of \( N \) microphones and with \( d_n \) the distance between the \( n \)th microphone and the reference point, arbitrarily chosen here as the center of the microphone array. Let \( \sigma_n(\omega, \phi, \theta) \) denotes the noise, where \( \omega \) represents the frequency, \( \phi \) and \( \theta \) represent the azimuthal and the elevation angle \((0 \leq \phi \leq 2\pi, 0 \leq \theta \leq \pi)\), respectively. Besides, the speech source \( S(\omega) \) is located at an angle \((\phi_s, \theta_s)\) in the microphone array.

The \( n \)th microphone signal \( Y_n(\omega) \) is expressed as

\[
Y_n(\omega) = g_s(\omega, \phi_s, \theta_s)S(\omega) + V_n(\omega)
\]  

(1)

with \( S(\omega) \) be the speech component of the reference signal received at the reference point, \( V_n(\omega) \) be the noise component of the \( n \)th microphone signal, and

\[
g_s(\omega, \phi_s, \theta_s) = a_s(\omega, \phi, \theta)e^{-j\omega(\phi_s\theta_s)}e^{-j\omega\theta_s}
\]  

(2)

where the delay \( \tau(\phi, \theta) \) in number of samples is equal to \( f_s|\frac{r-r_s}{c}| \), with \( c \) is the sound speed in air, usually around \( c=340\text{m/s} \), \( f_s \) is the sampling frequency and \( r_m \), \( m=1,2,\ldots,M \), are the position vectors of the microphones to the source point \( r \). The vector form of microphone signals can be written as

\[
Y(\omega) = g_s(\omega)S(\omega) + V(\omega)
\]  

(3)

The output signal \( Z(\omega) \) is equal to

\[
Z(\omega) = \sum_{j=1}^{N} W_n(\omega)Y_j(\omega) = W^T(\omega)Y(\omega)
\]  

(4)

Where \( W(\omega) = [W_1(\omega) W_2(\omega) \ldots W_N(\omega)] \) be the vector of weights.

B. Adaptive Wavelet Filter

Wavelet transformation is widely used in signal processing because processing signals in the frequency domain is often easier to implement [15]. Let \( x(t) \) be clean speech with finite length. The continuous wavelet transform (CWT) of a signal \( x(t) \) is given as follows:

\[
X_{\psi_a}(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-\tau}{a}\right)dt
\]  

(5)

where \( \tau \) and \( a \) represent the time shift and scale variables, respectively, and \( \psi(\cdot) \) is the mother wavelet chosen for the transform. In the discrete version (DWT), the scale and translation parameters of the discrete wavelet family are given by \( a = 2^n \) and \( \tau = n2^n \).

Let \( n \) be noise signal, contaminated speech \( y(t) \) can be expressed as

\[
y(t) = x(t) + n(t)
\]  

(6)

If \( W \) denotes the wavelet transform matrix, Eq. (6) can be written in the wavelet domain as

\[
Y(t) = X(t) + N(t)
\]  

(7)

where \( Y(t) = W \cdot y(t) \), \( X(t) = W \cdot x(t) \) and \( N(t) = W \cdot n(t) \). The estimated speech signal \( \hat{X}(t) \) can be obtained by using the thresholding function

\[
\hat{X}(t) = F_t(Y, T)
\]  

(8)

where \( F_t(Y, T) \) denotes the thresholding function and \( T \) is the threshold. The standard thresholding function includes the soft thresholding function which is defined as

\[
F_t(Y, T) = \begin{cases} 
\text{sign}(Y)|Y| - T, & |Y| \geq T \\
0, & |Y| < T
\end{cases}
\]  

(9)

The proper value for the threshold can be determined in many ways. In 2006, Ghanbari et al. proposed an adaptive thresholding method which provided excellent performance.

C. Spectral Subtraction Method

Spectral subtraction, proposed by Boll in 1979 [13], is a signal processing method in frequency domain that is applied widely. The noisy speech \( y(t) \) is assumed to consist of the clean speech \( x(t) \) additively degraded by uncorrelated random noise \( n(t) \), as follows:

\[
\text{Time domain: } y(t) = x(t) + n(t)
\]  

(10)

\[
\text{Frequency domain: } Y(\omega) = X(\omega) + N(\omega)
\]  

(11)

or

\[
Y e^{-j\phi_s} = X e^{-j\phi_s} + N e^{-j\phi_s}
\]  

where \( Y(\omega) \), \( X(\omega) \), and \( N(\omega) \) are discrete Fourier transforms, with amplitudes \( Y \), \( X \), and \( N \), and phases \( \phi_s \), \( \phi_x \), and \( \phi_n \), respectively, at frequency or frequency channel.

The short-time power spectrum of the noisy speech can be approximated by

\[
|Y e^{j\phi_s}|^2 \approx |X e^{j\phi_s}|^2 + |N e^{j\phi_s}|^2
\]  

(11)

The term \( |N e^{j\phi_s}|^2 \) cannot be obtained directly and is approximated as \( E\{|N e^{j\phi_s}|^2\} \), where \( E\{\cdot\} \) denotes the expectation operator. Typically, \( E\{|N e^{j\phi_s}|^2\} \) is estimated during non-speech activity, and is denoted by \( \tilde{N} e^{j\phi_s} \). Thus,

\[
|X e^{j\phi_s}|^2 = |Y e^{j\phi_s}|^2 - \tilde{N} e^{j\phi_s}
\]  

(12)

Berouti et al. proposed an important variation of spectral subtraction for reduction of residual musical noise [17]. An overestimate of the noise power spectrum is subtracted and the resulted spectrum is limited from going below a preset minimum level. The proposed algorithm could be expressed as
\[ \mathbf{R}_{e} e^{j\omega} = \begin{cases} \mathbf{R}_{e} e^{j\omega} - \alpha \mathbf{R}_{e} e^{j\omega}, & \text{if } \| \mathbf{R}_{e} e^{j\omega} \| > \beta \| \mathbf{R}_{e} e^{j\omega} \|, \\ \beta \| \mathbf{R}_{e} e^{j\omega} \|, & \text{otherwise} \end{cases} \]  (13)

where \(\alpha\) is the subtraction factor and \(\beta\) is the spectral parameter.

III. ADAPTIVE MICROPHONE ARRAY-BASED FILTER
A. Architecture of Adaptive Microphone Array-based Filter

This section proposes a novel hybrid filter for speech enhancement called Adaptive Microphone Array-based Filter. The schematic diagram of the adaptive microphone array-based filter is shown in Fig. 1. The adaptive microphone array-base filter comprises hybrid Wavelet-Spectral filters (H_WS filter), a superdirective beamformer and an analysis compensation system. First, H_WS filters which were composed by adaptive wavelet filters and spectral subtraction methods were served as former processing. A feed-forward fuzzy neural network then acts as a controller by providing the signal mixed ratio between wavelet filter and spectral subtraction method. Figure 2 shows a schematic diagram of the H_WS filter.

After H_WS filters, the signals were filtered by microphone array. By using superdirective beamformer, the microphone array can suppress interfering signals arriving from directions other than the look-direction. Besides, an analysis compensation system was proposed to eliminate unwanted spur and compensate the filtered signals.

B. Hybrid Wavelet-Spectral Filter (H_WS Filter)

A microphone array that uses beamforming technology not only suppresses interfering signals from directions other than the look-direction, it also reduces the reverberation. In practice, however, this microphone array does not provide adequate noise reduction in a high-noise environment.

Additive noise is conventionally classified as stationary or non-stationary. Spectral subtraction method and adaptive wavelet filter reportedly perform well in filtering stationary noise and non-stationary noise, respectively [12]. However, spectral subtraction method performs poorly on low SNR signals. And Residual noises could be produced. Adaptive wavelet thresholding method is ineffective for high-frequency signals and can result in distortion. The proposed filters combine adaptive wavelet filter and spectral subtractive method to remove contaminated signals. Ideally, the filtering algorithm should vary from frame to frame based on the local information. However, setting the conditions under which a certain filter should be selected is extremely difficult, if not impossible, since the local conditions can be evaluated only vaguely in some portions of the contaminated signals. After training with a set of input signals and desired signals, the neural controller acquires the desired classifier function.

**Definition 3.1:** The output of the H_WS filter is defined by
\[ y_\mu(n) = y_u(n)\alpha + y_s(n)(1-\alpha), \quad \alpha \in [0,1] \]  (14)

where \(n\) is the frame index, \(y_\mu(n)\), \(y_u(n)\) and \(y_s(n)\) are the output signal of H_WS filter, spectral subtraction method and adaptive wavelet filter, respectively.

C. Superdirective Beamformer

Let \(\omega\) represents the frequency, \(W\) represents the vector of the filter’s weight and \(V(\omega)\) represents the noise signal. The superdirective technique was used to calculate the channel filters maximizing the array gain, while maintaining a minimum constraint on the noise gain. The optimal filters are calculated as:
\[ W_{opt} = \left( \Phi_V \right)^{-1} N \cdot g_s \]  (15)

where \(\Phi_V(\omega)\) is the propagation vector between the source and each microphone. \(\Phi_V(\omega)\) is the noise coherence matrix and \(\Phi_V(\omega) = \Phi_V(\omega)/\Phi(\omega)\).

Once the optimal filters \(\Phi(\omega)\) have been calculated, the superdirective beamformer output is calculated as
\[ Z(\omega) = W_{opt} Y_N(\omega) \]  (16)

where \(Y_N(\omega)\) is the N-channel output column vector from the H_WS filter \(Y_N(\omega) = [y_{w1}(w) y_{w2}(w) \cdots y_{wN}(w)]\).

D. Feature Selection

Clearly, choosing appropriate critical features is essential for effective speech analysis in noisy environment. Some researchers have focused on energy, zero crossing rate and time duration to discriminate between the speech signal and background noise. Other proposed parameters include linear prediction coefficient, linear prediction error energy, pitch information and time-frequency parameter. However, adapting these parameters to variable-level background noise is still problematic, even when complex decision strategies are used. This study incorporated four parameters for use as linguistic variables to analyze the presence of speech in a noisy environment. The four parameters were energy, zero crossing rate, entropy and Mel frequency cepstral coefficient.

- **Energy**
  Energy is a good measure of noise when SNR exceeds 0dB. The formula for Energy is
  \[ E_{j,k} = \sum_{n} S_j^2(n), \quad j = 0,1,2,\ldots,N \]  (17)
  where \(k\) represented the frame, and \(S_j(n)\) represented the \(j\)th energy value.

- **Zero Crossing Rate**
  Zero crossing rate (ZCR) is a basic and easily computed acoustic feature. The ZCR is the number of zero crossing of a waveform within a given frame. The ZCR of both unvoiced sounds and environment noise generally exceed that of voiced sounds. A Zero Crossing Rate can be calculated by the mathematical formula:
  \[ ZCR = \frac{1}{2} \sum_{n=1}^{N} \left| \text{sgn}[S(n)] - \text{sgn}[S(n-1)] \right| \]  (18)
  where \(\text{sgn}[S(n)] = 1\) if \(S(n) \geq 0\); otherwise \(\text{sgn}[S(n)] = -1\).

- **Entropy**
  Entropy is a measure of the variation in energy distribution in a system. When the noise model is stationary or
slightly non-stationary, the corresponding entropy remains stable or is only slightly changed. The formula for entropy is

\[ E(S) = \sum p(x_i) \log p(x_i) \]  

(19)

- Mel Frequency Cepstral Coefficient

Mel frequency cepstral is an effective way to generate representative features from signals. Each coefficient has a value for each frame of the sound. Mel frequency cepstral coefficients (MFCCs) are derived from a cepstral representation of the audio clip. The difference between the cepstrum and the mel frequency cepstrum is that the frequency bands in the MFC are equally spaced on the mel scale, which is a better approximation of the human auditory response compared to the linearly-spaced frequency bands used in the normal cepstrum. The energy value of each band was calculated by

\[ Y(m) = \log \left( \sum_{k \in \text{band}} \left| X(k) \right|^2 B_m(k) \right) \]  

(20)

where \( B_m(k) \) was the triangular bandpass filter of \( m \) th band. Taking the Discrete Cosine Transform on the derived energy value from Eq. (19), a Mel Frequency Cepstral Coefficient was then obtained, which can be expressed as follows:

\[ \text{MFCC}(n) = \frac{1}{M} \sum_{m=1}^{M} Y(m) \cos \left( \frac{\pi n (m - 1)}{2} \right) \]  

(21)

where \( M \) was the number of bands.

E. Feed-forward Fuzzy Neural Network

A feed-forward fuzzy neural network in the hybrid wavelet-spectral filter provided a signal mixed ratio for wavelet filter and spectral subtraction method.

The four layers of the feed-forward fuzzy neural network are Input Layer, Rule Layer, Hidden Layer, and Output Layer. Each layer is described below.

Layer 1 is the input layer, which directly transmits the input variable to the next layer. Each node corresponds to an input variable.

Layer 2 is the fuzzification layer, which sets the fuzzy sets of the corresponding input variables (linguistic variable). The activation values of the nodes represent the membership degrees of the input variables. Table 1 shows the fuzzy sets for each input variable.

Layer 3 is the fuzzy rule layer. The number of nodes in this layer is equal to the number of fuzzy rules. A node in this layer represents a fuzzy rule. For each node, \( n \) fixed links before layer 3 represent the IF-part of the fuzzy rule, and the links after layer 3 represent the THEN-part of the fuzzy rule. The output of this layer is normalized firing strength with product inference.

Layer 4 was Output Layer. The node acts as a defuzzifier and computes the output value. The output of the FNN system with singleton fuzzification, product inference and center average defuzzification can be expressed as

According to the universal approximation theorem, the above FNN can uniformly approximate any well-defined nonlinear function over a compact set \( U \) to the desired accuracy.

IV. EXPERIMENTAL RESULTS

This section demonstrated the effectiveness of our proposed system. The experimental results that pertained to our proposed speech enhancement system were compared to spectral subtraction method, wavelet filter, Time-Scale adapted, Enhancement WT and Phase-Error filter.

In the experiment, four microphones were used to receive audio signal. Figure 3 shows the geometric description of the experimental apparatus. The distance between speech source and noise source is 18 cm. The interval between two microphones is 9 cm.

Tests for the enhancement task were performed using the Aurora 2 database, which is currently a standard databases for noisy speech enhancement and recognition [18]. The Aurora 2 database provides not only the uncontaminated speech of 4004 sentences, but also the speech of 48048 sentences contaminated with eight noise types (Subway, Babble, Car, Exhibition, Restaurant, Street, Airport and Station), which are mixed with -10, -5, 0, 5, 10, 15 and 20dB.

Tenfold cross-validation was performed on the dataset to evaluate how well the algorithm generalizes to future data. The tenfold cross-validation method extracts a certain proportion, typically 10%, of the training set as the tuning set, which is a surrogate of the testing set. In each training session, the proposed method was applied to the rest of the training data to obtain a filter. The tuning set correctness of this filter was then computed.

The comparisons of SNR also confirm the effectiveness of the proposed system. Tables 2 and 3 show the SNR before and after signal denoising with the proposed system. The experimental results show that the proposed system effectively filters noise.

Figure 4 compares the average SNR value obtained using the proposed method with those obtained by other methods, including spectral subtraction, wavelet filter, TSA (time-scale adapted) [19], enhancement wavelet filter and phase-error filter [20]. The experimental results confirm that the proposed method outperforms other methods.

V. DISCUSSIONS

The proposed adaptive microphone array-based filter reduces additive noise. In the proposed framework, contaminated signals were filtered by adaptive wavelet thresholding method and spectrum subtraction method, respectively. Microphone arrays were then used to provide an effective means of compensating for the effects of noise and reverberation. Between the adaptive wavelet thresholding method and spectral subtraction method, a feed-forward fuzzy neural network optimized the mixed signal ratio in order to compensate for other disadvantages. Moreover, an analytic compensation system was also used to compensate the filtering signal and to suppress spur noise. Experimental results
confirm that the proposed method is more effective for removing noise compared to conventional methods.

Figure 1: The schematic diagram of adaptive microphone array-based filter.

Figure 2: The schematic diagram of hybrid wavelet-spectral filter.

Table 1: Fuzzy set for each input variable

<table>
<thead>
<tr>
<th>Linguistic variable</th>
<th>Fuzzy sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>ZCR</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Entropy</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>High</td>
</tr>
<tr>
<td>MFCC(n)</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>

Table 2. The SNR of the signals in Aurora 2 database.

<table>
<thead>
<tr>
<th>dB</th>
<th>Subway</th>
<th>Babble</th>
<th>Car</th>
<th>Exh.</th>
<th>Rest</th>
<th>Street</th>
<th>Airport</th>
<th>Sta.</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>20dB</td>
<td>18.5</td>
<td>18.3</td>
<td>18.4</td>
<td>18.6</td>
<td>18.38</td>
<td>18.32</td>
<td>18.45</td>
<td>18.38</td>
<td>18.42</td>
</tr>
<tr>
<td>15dB</td>
<td>13.5</td>
<td>13.6</td>
<td>13.4</td>
<td>13.6</td>
<td>13.45</td>
<td>14.07</td>
<td>14.28</td>
<td>13.55</td>
<td>13.68</td>
</tr>
<tr>
<td>10dB</td>
<td>8.85</td>
<td>8.9</td>
<td>8.95</td>
<td>8.92</td>
<td>9.31</td>
<td>9.07</td>
<td>9.28</td>
<td>8.93</td>
<td>9.03</td>
</tr>
<tr>
<td>5dB</td>
<td>4.95</td>
<td>4.9</td>
<td>4.98</td>
<td>4.96</td>
<td>4.97</td>
<td>4.95</td>
<td>4.92</td>
<td>4.94</td>
<td>4.95</td>
</tr>
<tr>
<td>0dB</td>
<td>0.84</td>
<td>0.62</td>
<td>0.08</td>
<td>0.12</td>
<td>0.11</td>
<td>0.3</td>
<td>0.26</td>
<td>0.09</td>
<td>0.3</td>
</tr>
<tr>
<td>-5dB</td>
<td>-4.81</td>
<td>-4.26</td>
<td>-4.3</td>
<td>-4.13</td>
<td>-5.01</td>
<td>-4.86</td>
<td>-4.97</td>
<td>-5.06</td>
<td>-4.68</td>
</tr>
</tbody>
</table>

Table 3. The SNR of the signals after enhancing by our proposed algorithm.

<table>
<thead>
<tr>
<th>dB</th>
<th>Subway</th>
<th>Babble</th>
<th>Car</th>
<th>Exh.</th>
<th>Rest</th>
<th>Street</th>
<th>Airport</th>
<th>Sta.</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>20dB</td>
<td>24.43</td>
<td>22.38</td>
<td>22.55</td>
<td>26.73</td>
<td>25.57</td>
<td>26.49</td>
<td>22.41</td>
<td>25.58</td>
<td>24.52</td>
</tr>
<tr>
<td>15dB</td>
<td>21.03</td>
<td>18.97</td>
<td>20.52</td>
<td>22.97</td>
<td>18.84</td>
<td>23.25</td>
<td>19.66</td>
<td>21.26</td>
<td>20.81</td>
</tr>
<tr>
<td>10dB</td>
<td>17.17</td>
<td>13.26</td>
<td>14.25</td>
<td>17.69</td>
<td>17.78</td>
<td>15.31</td>
<td>15.49</td>
<td>17.71</td>
<td>16.08</td>
</tr>
<tr>
<td>5dB</td>
<td>10.95</td>
<td>8.98</td>
<td>10.78</td>
<td>12.78</td>
<td>11.54</td>
<td>13.52</td>
<td>10.17</td>
<td>13.01</td>
<td>11.47</td>
</tr>
<tr>
<td>0dB</td>
<td>4.53</td>
<td>3.54</td>
<td>7.66</td>
<td>7.28</td>
<td>7.74</td>
<td>6.98</td>
<td>7.03</td>
<td>1.95</td>
<td>5.84</td>
</tr>
<tr>
<td>-5dB</td>
<td>2.85</td>
<td>-1.62</td>
<td>3.57</td>
<td>3.24</td>
<td>3.45</td>
<td>2.73</td>
<td>3.02</td>
<td>1.1</td>
<td>2.29</td>
</tr>
<tr>
<td>-10dB</td>
<td>-0.02</td>
<td>-3.33</td>
<td>0.85</td>
<td>0.12</td>
<td>0.08</td>
<td>0.27</td>
<td>0.89</td>
<td>0.1</td>
<td>-0.13</td>
</tr>
</tbody>
</table>

Figure 3: Geometric description of our used microphone array

Figure 4: Comparison of our proposed method with other methods

References


