ABSTRACT

In this paper, we will compare and evaluate the various noise cancellation schemes available in what we will term a dual microphone system. A dual microphone system (DMS) is a composite directional audio-capturing device which consists of two microphones, each microphone having possibly different directional characteristics, e.g., omnidirectional, bidirectional or cardioid. By recasting the various combinations of two microphones for a DMS into a coherent and familiar framework of Generalized Sidelobe Canceller (GSC), we will subsequently derive the expected noise reduction of various structures under incoherent, coherent and diffuse noise fields, followed by a series of experiments in a typical office environment. These results are indicative of the achievable reduction of noise in real applications. The relationship and differences between the various methods are also discussed.

1. INTRODUCTION

In recent years, dual microphone systems (DMS) have become a promising solution for directional audio-capturing (and, indirectly, noise reduction) for speech communications in particular, where mobility for the user is necessary, see [1–5]. Equipped with adaptive noise cancellation [6], the performance of a DMS can be optimized by adapting to the changing and noisy environments. Compared with using a single directional microphone having fixed directional characteristics, the DMS provides more flexibility and can lead to a much higher improvement in the signal-to-noise ratio (SNR). Furthermore, a DMS is compact and, for example, can be incorporated into a hearing aid [2,4]. Configurations using two closely spaced microphones also eliminate the need for near-field compensation [7].

The DMS can be broadly divided into two groups according to the types of microphones used. The first group, as was partially proposed in [1], consists of two intrinsically omnidirectional, cardioid or bidirectional microphones (see Fig. 1 for the different combinations), followed by an adaptive noise canceller [6], see Fig. 2. Spatial information of the incoming signals is obtained via the inherent directionality of the individual microphones, rather than through the time differences caused by the inter-microphone separation. Consequently, the two microphones can be placed as close as physically allowable.

The second group of DMS consists of two omnidirectional microphones that are separated by a small distance d. By making use of the delays and signal gains in combining the outputs from these two microphones, different intermediate polar patterns can be generated which are then used to form the various combinations in Fig. 1. This configuration is flexible because the looking direction of the system can be easily adjusted [2–5].

In this paper, we will limit ourselves to the evaluation of DMS with very closely-spaced microphones. We will consider the noise-reducing capability of each configuration under typical noise conditions as well as in a real environment. The conclusions reached can then be used to predict the likely achievable performance of a given DMS in real-world environments. Some guidelines for real system design will also be provided.

The paper is organized as follows. In Section 2, we will review available algorithms for DMS. The various configurations are then recast into a consistent and unified representation that is used to derive the theoretical noise reduction under different noise fields, given in Section 2.2. Detailed experimental results are given in Section 3 followed by some conclusions in Section 4.

2. DUAL-MICROPHONE SYSTEMS

In this section we will briefly review the components of a DMS.

2.1. Adaptive Noise Canceller

Fig. 2. Block diagram of an adaptive noise cancellation system.

Let us designate each of the two microphones in a DMS as being ‘primary’ and ‘reference’ respectively. The core technology
in a DMS is the adaptive noise canceller [6]. In this canceller, the ‘primary’ microphone picks up both the desired signal \( s(t) \) and the noise \( n(t) \), whereas the ‘reference’ microphone picks up mostly a noise \( n'(t) \) component (with some leakage of the desired signal in practice), see Fig. 2. The reference signal is then adaptively subtracted from the primary signal to produce a desired and hopefully noise-free signal\(^1\). Based on this principle, alternative noise cancellation systems can be constructed by using combinations of two microphones, each having different directionalities and/or orientations as shown in Fig. 1.

A noise canceller can also be constructed by using two omnidirectional microphones which are separated by a small distance \( d \), and used in either broadside or endfire orientation. By using the delay-differential method [3], ‘virtual’ cardioid or bidirectional microphones can be formed from the outputs of the two omnidirectional microphones. That is, the various configurations in Fig. 1 which use combinations of intrinsically directional or omnidirectional microphones can be similarly realized in a virtual manner. As an example, configuration in Fig. 1b can be achieved using two differential cardioid microphones placed back-to-back, see, for example, [2,3].

To collectively represent the various DMS involving two omnidirectional microphones, we will recast them into the form of a General Side-lobe Canceller [8], as shown in Fig. 3. Denote the inputs of the two microphones as \( x_1(t) \) and \( x_2(t) \) respectively. The upper branch in Fig. 3 enhances signals coming from a particular direction by using delays and weighting coefficients, while the aim of the bottom branch is to block the desired signal. The two branches are then summed and, through an adaptive filter \( W \), the reference signal \( y_2(t) \) is then subtracted from \( y_1(t) \) to obtain the desired output \( y(t) \). The various signals of interest are

\[
\begin{align*}
y_1(t) &= a_1x_1(t - \tau_{11}) + a_2x_2(t - \tau_{12}) \\
y_2(t) &= x_1(t - \tau_{21}) - x_2(t - \tau_{22}) \\
y(t) &= y_1(t - \tau_c) - W(t)Y_2(t) \\
Y_2(t) &= \begin{bmatrix} y_2(t) & y_2(t-1) & \cdots & y_2(t-L+1) \end{bmatrix}^T \\
W(t) &= \begin{bmatrix} w_0(t) & w_1(t) & \cdots & w_{L-1}(t) \end{bmatrix}^T
\end{align*}
\]

where \( \tau_c \) is the time delay for causality, \( Y_2(t) \) the reference vector at time \( t \) and \( L \) the filter length of \( W \). The Norm-Constrained NLMS [6] algorithm will be used in the updating of \( W \):

\[
W'(t + 1) = \begin{cases} W(t) + \mu y(t) Y_2(t) / \| Y_2(t) \|^2 & \text{for } \Omega > K \\ W' & \text{otherwise} \end{cases}
\]

where \( \mu \) is the step size, \( W' \) denotes the temporal filter, and \( \Omega \) and \( K \) are the squared-norm of \( W(t) \) and a threshold respectively. If \( \Omega > K, W(t + 1) \) will be restrained by scaling.

The parameters for various configurations which make use of two omnidirectional microphones have been summarized in Table 1, following the notation of Fig. 3. We will assume the desired signal (and hence the desired direction) is from the left, therefore for broadside configurations, the desired direction is 0°.

<table>
<thead>
<tr>
<th>No.</th>
<th>( \tau_{11} )</th>
<th>( \tau_{12} )</th>
<th>( \tau_{21} )</th>
<th>( \tau_{22} )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>Taps</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>(B)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>(C)</td>
<td>( d/c )</td>
<td>( d/c )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>(D)</td>
<td>0</td>
<td>0</td>
<td>( d/c )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>L</td>
</tr>
<tr>
<td>(E)</td>
<td>0</td>
<td>( d/c )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>L</td>
</tr>
<tr>
<td>(F)</td>
<td>0</td>
<td>( d/c )</td>
<td>( d/c )</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>L</td>
</tr>
<tr>
<td>(G)</td>
<td>0</td>
<td>( d/c )</td>
<td>( d/c )</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(H)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A_1 )</td>
<td>( A_2 )</td>
<td>-</td>
</tr>
</tbody>
</table>

\^1\text{We will further assume that after adaptation } W'n'(t) \approx n(t) \text{ where } W \text{ is a vector of adaptation weights, which is satisfied in many practical situations.}

Table 1. Parameters of different configurations for DMS. (Note: \( d \) = inter-microphone distance, \( c \) = speed of sound).

2.2. Theoretical Performance

In this section, we will compare the performances of the schemes in Table 1 under incoherent, coherent and diffuse noise fields. The analysis will be based on the approach in [10] and uses the complex coherence function. Due to the space limitations, we will directly present the results, see [11] for the details.

In an incoherent noise field, e.g. due to the microphone self noise, all methods have the same noise reduction (NR) of 2 (equivalent to 3dB). However, in a coherent noise field, e.g. single point source in free field, the NR of methods (B)-(G) are infinite in all directions while the NR for method (H) is infinite in the back hemisphere only. Method (A) has a fixed NR which depends on the...
direction of arrival $\theta$, given by $\text{NR} = \frac{2}{[1 + \cos(\omega d \cos \theta/c)]}$, as shown in [10].

Diffuse noise fields can be regarded as a good approximation of a highly reverberant room, hence this provides a truer reflection of many real-world indoor environments. The resulting NRs under a diffuse noise field are shown in Table 2. To better understand the performance of Fig. 3, the NR of the upper branch $(\text{NR}_U)$, the adaptive lower branch $(\text{NR}_L)$ and the overall NR $= \text{NR}_U \times \text{NR}_L$ have been separately tabulated. From Table 2, it can be seen that (A) and (B) have the same performance. In fact, (B) produces no noise reduction in the adaptive branch as $\text{NR}_L = 1$. Furthermore, although the $\text{NR}_U$ and $\text{NR}_L$ of (C)-(H) are different, the overall NR are all identical to method (H), which is the optimal design under this condition [9].

When $\gamma = \omega d/c$ is small, which corresponds to low frequencies and a small $d$, the performances of (A) and (B) are very poor as shown in Fig. 4, while the NR of (C)-(H) can be up to 6dB. On the other hand, when $\gamma$ increases, the noises received by the two microphones gradually become incoherent and the NRs of (A)-(H) reduce to 3dB.

<table>
<thead>
<tr>
<th>No.</th>
<th>$\text{NR}_U$</th>
<th>$\text{NR}_L$</th>
<th>$\text{NR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>$\frac{1}{1 + \sin(\gamma)}$</td>
<td>non-adaptive</td>
<td>$\frac{1}{1 + \sin(\gamma)}$</td>
</tr>
<tr>
<td>(B)</td>
<td>$\frac{1}{1 + \sin(\gamma)}$</td>
<td>1</td>
<td>$\frac{1}{1 + \sin(\gamma)}$</td>
</tr>
<tr>
<td>(C)</td>
<td>$\frac{1}{1 + \cos \gamma \sin(\gamma)}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2(1 + \sin \gamma \cos \gamma)}$</td>
</tr>
<tr>
<td>(D)</td>
<td>$\frac{1}{1 + \cos \gamma}$</td>
<td>1</td>
<td>$\frac{1}{1 + \sin \gamma \cos \gamma}$</td>
</tr>
<tr>
<td>(E)</td>
<td>$\frac{1}{1 + \sin \gamma}$</td>
<td>$\frac{1}{\sin \gamma}$</td>
<td>$\frac{1}{2(1 + \sin \gamma \cos \gamma)}$</td>
</tr>
<tr>
<td>(F)</td>
<td>$\frac{1}{1 + \cos 2\gamma}$</td>
<td>$\frac{1}{1 + \sin \gamma \cos \gamma}$</td>
<td>$\frac{1}{2(1 + \sin \gamma \cos \gamma)}$</td>
</tr>
<tr>
<td>(G)</td>
<td>$\frac{1}{1 + \sin \gamma \cos \gamma}$</td>
<td>$\frac{1}{2(1 + \sin \gamma \cos \gamma)}$</td>
<td>$\frac{1}{2(1 + \sin \gamma \cos \gamma)}$</td>
</tr>
<tr>
<td>(H)</td>
<td>non-adaptive</td>
<td>non-adaptive</td>
<td>non-adaptive</td>
</tr>
</tbody>
</table>

Fig. 4. Theoretical noise reduction for methods (A)-(H) as function of $\gamma$, where $\gamma = \omega d/c$.

### 3. EXPERIMENTAL RESULTS AND DISCUSSIONS

A number of experiments were conducted in an office as shown in Fig. 5, with $T_s \sim 0.10s$ for frequency range 250Hz-4kHz. The volume of the test signal was set at 75 dB SPL, using a GENELEC 1029A loudspeaker. The model of the omnidirectional, cardioid and bidirectional microphones used were Panasonic WM034CY195M, HORN EM9752U-474 and Panasonic WM-55D103, respectively. The two input channels were pre-amplified and recorded simultaneously at $F_s = 22050$ Hz using the soundcard of a Pentium 4 PC. The inter-microphone distance for the DMS consisting of two omnidirectional microphones was $d = 0.0156m$. For combinations of omnidirectional, cardioid and bidirectional microphones, the two microphones were placed together. The loudspeaker was placed at three different distances, namely 0.3m, 0.6m, and 1m from the microphones. A stepper motor was used to rotate the DMS 360° at increments of 9°. All equipments were placed 1m above the floor.

![Fig. 5. Configuration of the experiment environment.](image)

![Fig. 6. Experimentally measured polar patterns for different combinations of two microphones, where $R = 0.3m - 0.6m - 1.0m$.
](image)

The measured polar patterns for each of the configurations have been tabulated in Fig. 6. It can be seen that the performances depend highly on the distance between the sound source and the microphones. Except for Fig. 6(d) and 6(f), the noise cancel-
ination levels of all other methods are similar for signals coming from 90°-270°. As the distance between the loudspeaker and the microphones was increased from to 0.60m and then to 1.00m, the maximum noise cancellation was also reduced from about 18dB to 15dB and then to only 10dB. This is likely due to a lack of coherence between the two microphones, which is the result of strong nonstationary, multipath, nonlinear and the time-varying acoustic environments. In addition, inconsistency between the two microphones may have also contributed to this performance loss.

Since the cardioid-bidirectional configuration in Fig. 1(d) cannot be constructed using only two omnidirectional microphones, it was studied only experimentally. As shown in Fig. 6(i), this configuration produces a sharp polar pattern. Comparing against the configuration in Fig. 6(d), there is no directional ambiguity because the primary microphone in Fig. 6(i) is a directional microphone, i.e. cardioid.

In terms of the frequency characteristics, we note that only method (D) has a uniform frequency response for all directions since the primary signal for this configuration is obtained using an omnidirectional microphone. Methods (A)-(C) and (H) have uniform response at 0° only. On the other hand, the responses of methods (E)-(G) are frequency-dependent for all directions and therefore such configurations require low pass filters [3]. Additional improvements to the methods here, especially in conditions having multiple interferences can be achieved by using subband adaptive filtering [5], at the price of increased complexity.

3.1. Remarks On Design Considerations

Based upon the previous analysis and experimental results, we can provide the following guidelines when designing a DMS. Since the extent of noise reduction depends critically on the types of noises as indicated in Section 2.2, the characteristics of the noise field of the application environment should be estimated, and then the configuration chosen accordingly.

In case the direction selectivity is crucial, method (B) should be used. However, its poor performance in reverberant environments, e.g. Fig. 6(d) as the distance between the signal source and DMS is increased, means that this configuration will be ineffective in most offices. In this scenario, the scheme in Fig. 1(d) is a good option as shown in Fig. 6(i). On the other hand, if the environment contains largely diffuse noise, to save cost, a simple hypercardioid combination scheme, i.e. method (H), is ideal. Finally, if the main purpose is to reduce the noise coming from the back of the device, the efficient methods in [2] can be used. To make the system smaller, combinations of intrinsic types of microphones as opposed to using two omnidirectional microphones can be used, but in most cases the look direction is not easily adjustable.

4. CONCLUSIONS

In this paper, we have characterized the performance of DMS based on different configurations. While the first group of DMS which uses two microphones that have intrinsically directional properties can be made compact, the DMS using two omnidirectional microphones provides more flexibility. Within the second group, methods (B)-(H) show good noise reduction under coherent noise fields. In diffuse noise field, methods (C)-(F) have better noise reduction than methods (A)-(B), especially for lower frequencies. At the same time, the first group also provides similar performance with their equivalent schemes.

In addition, by using bidirectional-type reference signal, the dual microphone system has a good spatial selectivity; on the other hand, by using a cardioid-type reference signal, the system becomes more insensitive to the errors in look direction. In a real environment, the performance of the DMS will depend highly on the distance between the sound sources and the microphones. These considerations will aid the designers of DMS in adopting the appropriate configurations.

5. REFERENCES