THEORETICAL COMPARISONS OF DUAL MICROPHONE SYSTEMS

Jianfeng Chen, Louis Shue, Koksoon Phua, Hanwu Sun

Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
Email: {jfchen,lshue,ksphua,hwsun}@i2r.a-star.edu.sg

ABSTRACT

In this paper, we will compare and evaluate the various noise cancellation schemes available in what we will term the 'dual microphone systems' (DMS). A DMS is a directional audio-capturing device consisting of two microphones with possibly different directional characteristics: omnidirectional, bidirectional or cardioid. A general structure is proposed to coherently represent the different schemes for the DMS. This is followed by a theoretical derivation of performance of various DMS configurations under incoherent, coherent and diffuse noise fields. The relationships between the different configurations and some guidelines for design of DMS are also presented.

1. INTRODUCTION

In recent years, dual microphone systems (DMS) have become a promising solution for directional audio capture (and, indirectly, noise reduction) in speech communications where mobility for the user is necessary [1–5]. Equipped with adaptive noise cancellation [6], a DMS optimizes its performance in changing and noisy environments. Compared with using a single directional microphone with fixed polar pattern, the DMS provides more flexibility and can lead to much higher gain in signal-to-noise ratio (SNR). Furthermore, a DMS is compact and can even be incorporated into a hearing aid [2,4]. The schemes 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 microphone used. The first group, as partially proposed in [1], consists of a combination of omnidirectional, cardioid and bidirectional microphones, see Fig. 1, followed by an adaptive noise canceller [6], as shown in Fig. 2. The primary microphone picks up the desired signal $s(t)$ with some noise while the signal from the reference microphone contains almost entirely of noise; an estimate of $s(t)$, namely $y(t)$, is formed after the adaptive canceller. Any spatial information of the signal is obtained directly through the inherent directionality of the individual microphone rather than through the differences caused by the interspacing distance of the two microphones, as in the second group.

The second group consists of two omnidirectional microphones that are separated by a small distance d. By making use of delays and gains, see Fig. 3, in combining the outputs from these two omnidirectional microphones, different polar patterns can be generated that can result in the different combinations in Fig. 1. This configuration is flexible because the looking direction can be easily adjusted and efficient algorithms are also available, e.g., [2–5].

In this paper, we will limit our discussions to the DMS with very closely-spaced microphones. We will consider the potential benefits in noise reduction under typical noise conditions, and thereby deduce useful design guidelines for DMS. The rest of the paper is organized as follows. In Section 2, we will review current algorithms in DMS and propose to recast these schemes into a consistent framework. In Section 3 we will present the theoretical performance under three different noise conditions, followed by detailed simulation results[1] in Section 4. Some concluding remarks are given in Section 5.

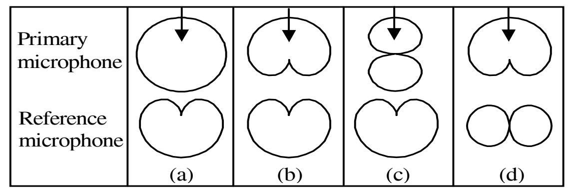

Fig. 1. Various configurations for dual microphone system: (a) omni-cardioid (b) cardioid-cardioid (c) bidirectional-cardioid (d) cardioid-bidirectional (Note: the arrows denote the desired signal.)

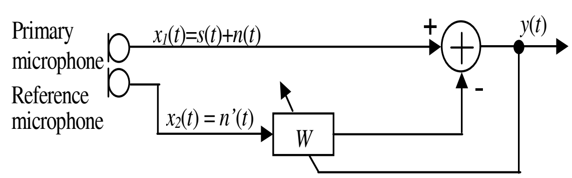

Fig. 2. Block diagram of an adaptive noise cancellation system.

2. DUAL-MICROPHONE SYSTEMS

In this section we will recast the basic configuration of Fig. 3 into the familiar GSC [9] framework for microphone arrays.

Let us designate the two microphones in Fig. 2 as being 'primary' and 'reference'. In an adaptive noise canceller [6]. The 'primary' microphone (MIC#1 in Fig. 3) picks up both the desired signal $s(t)$ and the noise $n(t)$, whereas the 'reference' microphone (MIC#2 in Fig. 3) picks up mostly only a noise $n'(t)$ component. The reference signal is then adaptively subtracted from the primary signal to produce an estimate of the desired signal[2].

A noise canceller can also be constructed using two omni-

[1] Due to space limitations, the experimental results are presented in a companion paper [8].

[2] We will further assume that after adaptation $Wn'(t) \approx n(t)$, which is satisfied in many practical situations.

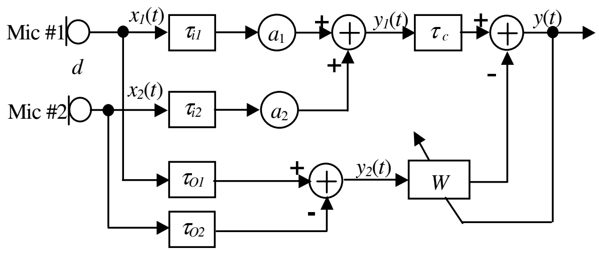

Fig. 3. Coherent structure for adaptive dual omni-microphone array

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

No.	τ_{i1}	τ_{i2}	τ_{O1}	τ_{O2}	a_1	a_2	Taps
(A)	0	0	-	-	1	1	-
(B)	0	0	0	0	1	1	L
(C)	d/c	0	d/c	0	1	1	L
(D)	0	0	d/c	0	1	0	L
(E)	0	0	d/c	0	1	-1	L
(F)	0	d/c	d/c	0	1	-1	L
(G)	0	d/c	d/c	0	1	-1	1
(H)	0	0	0	0	A_1	A_2	-

directional microphones, separated by a small distance d, and used in either broadside or endfire orientation. Virtual cardioid or bidirectional microphones can be formed from the two microphones via the delay-differential method [3]. For example, configuration in Fig. 1b can be achieved using two differential and virtual cardioid microphones placed back-to-back, see also [2,3].

To collectively represent the various dual omni-microphone schemes, let us denote the outputs of the two microphones as $x_1(t)$ and $x_2(t)$ respectively. The upper branch enhances signal from the desired direction by using delays and weighting coefficients, while the bottom branch blocks the signal from the desired direction. 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

$$y_1(t) = a_1 x_1(t - \tau_{i1}) + a_2 x_2(t - \tau_{i2}) \quad (1)$$

$$y_2(t) = x_1(t - \tau_{o1}) - x_2(t - \tau_{o2}) \quad (2)$$

$$y(t) = y_1(t - \tau_c) - W(t) Y_2(t) \quad (3)$$

$$Y_2(t) \triangleq \begin{bmatrix} y_2(t) & y_2(t-1) & \cdots & y_2(t-L+1) \end{bmatrix}^T \quad (4)$$

where τ_c is the time delay for causality, $Y_2(t)$ the 'reference' vector at time t and L the filter length of W. A commonly used adaptive algorithm, the Norm-Constrained LMS [6] will be used:

$$W' = W(t) + 2\mu y(t) \frac{Y_2(t)}{\|Y_2(t)\|^2} \quad (5)$$

$$W(t+1) = \begin{cases} \sqrt{K/\Omega} \cdot W', \Omega \triangleq \|W'\|^2 & \text{for } \Omega > K \\ W' & \text{otherwise} \end{cases} \quad (6)$$

where μ is the step size, W' denotes the temporal filter, and Ω 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 of the various noise cancelling schemes using two omni-microphones have been summarized in Table 1:

(A) Conventional delay-and-sum beamforming.

(B) GSC (broadside) [9]: generalised sidelobe canceller using two omni-microphones in broadside orientation.

(C) GSC (endfire): two omni-microphones in endfire orientation.

(D) Omni-cardioid (Fig. 1(a)): omni-directional microphone is MIC#1 in Fig. 3; differential cardioid microphone formed by combining MIC#1 and MIC#2 as in Fig. 3.

(E) Bidirectional-cardioid (Fig. 1(c)): two differential microphones formed by combining MIC#1 and MIC#2 as in Fig. 3.

(F) Cardioid-cardioid A (Fig. 1(b)): two differential cardioid microphones formed by combining MIC#1 and MIC#2 as in Fig. 3.

(G) Cardioid-cardioid B (Fig. 1(b)): adaptive null-forming scheme with filter length of 1 tap [2,3]. The main difference with (F) is the lack of an adaptive filter W as shown in Fig. 3.

(H) Superdirective [10]: endfire array where A_1 and A_2 are filters.

3. THEORETICAL PERFORMANCE ANALYSIS

In this section, we will derive the theoretical performance for the schemes in Table 1 under incoherent, coherent and diffuse noise. The analysis is based on the complex coherence function [11].

Denote outputs of the two microphones in frequency domain as $X_1(\omega)$ and $X_2(\omega)$. The complex coherence $\Gamma_{X_1X_2}(\omega)$ is defined:

$$\Gamma_{X_1X_2}(\omega) = \frac{P_{X_1X_2}(\omega)}{\sqrt{P_{X_1X_1}(\omega) P_{X_2X_2}(\omega)}} \quad (7)$$

where $P_{X_1X_2}$, $P_{X_1X_1}$ and $P_{X_2X_2}$ are the cross- and auto-power densities of $x_1(t)$ and $x_2(t)$ respectively.

In an incoherent noise field (e.g., microphone self-noise), the coherence function $\Gamma_{X_1X_2}(\omega)$ is independent of the inter-microphone distance d and is 0 for all frequencies ω

$$\Gamma_{X_1X_2}(\omega) = 0, \forall \omega. \quad (8)$$

For the case of coherent noise (e.g., sound source is in the far field and absence of reverberation) the microphone outputs are completely identical except for a time delay and the complex coherence function $\Gamma_{X_1X_2}(\omega)$ is then given by

$$\Gamma_{X_1X_2}(\omega) = e^{j\omega d \cos\theta / c} \quad (9)$$

where θ is the angle between the sound source and the array.

In a diffuse noise field the coherence function is real-valued and given by [11]

$$\Gamma_{X_1X_2}(\omega) = \mathrm{sinc}(\omega d / c) \quad (10)$$

where $\mathrm{sinc}\,\gamma = (\sin\gamma)/\gamma$. The diffuse noise field is often regarded as a good approximation of a reverberant room, hence the analysis under this noise field provides a good indication for many realistic environments. If we assume that no desired signal is present and that the noise is $P_{NN} = P_{X_1X_1} = P_{X_2X_2}$, then by rewriting (1) and (2) in frequency domain, the auto- and cross-spectral densities of $Y_1(\omega)$ and $Y_2(\omega)$ are

$$P_{Y_1Y_1}(\omega) = P_{NN}(a_1^2 + a_2^2 + 2a_1a_2\,\mathrm{Re}\{\Gamma_{X_1X_2}(\omega)e^{-j\omega(\tau_{i1}-\tau_{i2})}\}) \tag{11}$$

$$P_{Y_2Y_2}(\omega) = P_{NN}(2 - 2\,\mathrm{Re}\{\Gamma_{X_1X_2}(\omega)e^{-j\omega(\tau_{O1}-\tau_{O2})}\}) \tag{12}$$

$$\begin{aligned} P_{Y_2Y_1}(\omega) = P_{NN}(&a_1e^{-j\omega(\tau_{O1}-\tau_{i1})} - a_2e^{-j\omega(\tau_{O2}-\tau_{i2})} \\ &- a_1e^{-j\omega(\tau_{O2}-\tau_{i1})}\Gamma_{X_1X_2}(\omega)^* \\ &+ a_2e^{-j\omega(\tau_{O1}-\tau_{i2})}\Gamma_{X_1X_2}(\omega)) \end{aligned} \tag{13}$$

where * is the conjugate operator. The noise reduction (NR) of the fixed upper branch (NR$_U$) and the adaptive bottom branch (NR$_L$) of Fig. 3, such that NR $=$ NR$_U$ $\times$ NR$_L$, are then defined as

$$\mathrm{NR}_U = \frac{P_{NN}(\omega)}{P_{Y_2Y_1}(\omega)} \tag{14}$$

$$\mathrm{NR}_L = \frac{P_{Y_1Y_1}(\omega)}{P_{Y_1Y_1}(\omega) - |H_{opt}(\omega)|^2 P_{Y_2Y_2}(\omega)} \tag{15}$$

Minimizing the output power results in [11]

$$H_{opt}(\omega) = \frac{P_{Y_2Y_1}(\omega)}{P_{Y_2Y_2}(\omega)}. \tag{16}$$

By applying the parameters in Table 1 to (11)-(16) the following observations can be made:

1. In an incoherent noise field, all methods have a noise reduction NR=2, i.e., 3dB.
2. In a coherent noise field, (A) has a fixed NR which depends on the direction of arrival θ, given by NR $= 2/(1+\cos(\gamma\cos\theta))$, where $\gamma = \omega d/c$. NR of methods (B)-(F), (H) are infinite in all directions while NR for method (G) is infinite in the back hemisphere only.
3. For diffuse noise fields, (A) and (B) have the same performance, see Table 2. Furthermore, while NR$_U$ and NR$_L$ of (C)-(H) are different, the overall NR are all identical to method (H), which is optimal for diffuse noise fields [10].
4. For small γ, which corresponds to low frequency and a small d, the performances of (A) and (B) are very poor, while the NR of (C)-(H) can be up to 6 dB. On the other hand, when γ increases, the noises received by the two microphones gradually become incoherent and the NRs of (A)-(H) meet at 3dB.

4. SIMULATIONS AND DISCUSSIONS

In this section, we will study the performance of the various schemes in a simulated office environment $(7 \times 3.50 \times 2.85)$m using two omnidirectional microphones with $d = 0.0156m$ - corresponding to one-sample delay for sampling rate of 22050Hz. Simulations were performed to mimic coherent and reverberant noise fields for different reverberation times T_r. A band-limited (250-4000Hz) white noise was used as the test signal. The imaging method described in [12] was used to simulate the transfer function between the loudspeaker and each microphone.

For methods (B)-(F), the Norm-Constrained NLMS algorithm with $\mu = 0.1$ and $L = 64$ was used[3]. The technique proposed in [2] was implemented for (G), and standard superdirective weightings optimized for diffuse noise field was used for (H). For the simulations, the array was rotated 360° at increments of 9°. Since it was observed that the results of (C)-(F) were identical, we will only report results for methods (B) and (D)-(H).

Table 2. Noise reduction of various DMS under diffuse noise fields, $\gamma = \omega d/c$, $\mathrm{sinc}\,\gamma = (\sin\gamma)/\gamma$.

No.	NR$_U$	NR$_L$	NR
(A)	$\frac{2}{1+\mathrm{sinc}(\gamma)}$	non-adaptive	$\frac{2}{1+\mathrm{sinc}(\gamma)}$
(B)	$\frac{2}{1+\mathrm{sinc}(\gamma)}$	1	$\frac{2}{1+\mathrm{sinc}(\gamma)}$
(C)	$\frac{2}{1+\cos\gamma\,\mathrm{sinc}(\gamma)}$	$\frac{1-\mathrm{sinc}^2\gamma\cos^2\gamma}{1-\mathrm{sinc}^2\gamma}$	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$
(D)	1	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$
(E)	$\frac{1-\cos\gamma}{1-\mathrm{sinc}\,\gamma}$	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{(1+\mathrm{sinc}\,\gamma)(1-\cos\gamma)}$	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$
(F)	$\frac{1-\cos 2\gamma)}{1-\mathrm{sinc}\,\gamma\cos\gamma}$	$\frac{(1-\mathrm{sinc}\,\gamma\cos\gamma)^2}{(1-\mathrm{sinc}^2\gamma)(1-\cos^2\gamma)}$	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$
(G)	$\frac{1-\cos 2\gamma)}{1-\mathrm{sinc}\,\gamma\cos\gamma}$	$\frac{(1-\mathrm{sinc}\,\gamma\cos\gamma)^2}{(1-\mathrm{sinc}^2\gamma)(1-\cos^2\gamma)}$	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$
(H)	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$	non-adaptive	$\frac{2(1-\mathrm{sinc}\,\gamma\cos\gamma)}{1-\mathrm{sinc}^2\gamma}$

4.1. Coherent Noise Field

As shown in Figs. 4a-4d, the norm constraints K in (6) can be used as a beamwidth controller. Increasing K results in sharper beam patterns and vice versa. As alluded to in Section 3, all of the methods show good noise cancellation under coherent noise.

As can be seen from Figs. 4a-4c, the polar patterns of (D)-(F) are exactly the same. Within the directions of interest at about 0°±10°, their responses are all relatively flat. This may be due to the flat notch of the reference channel in these configurations, which is equivalent to a cardioid microphone. On the other hand, the polar pattern of (B) (Fig. 4d) is very sharp, because the reference channel of (B) is equivalent to a bidirectional microphone and its null is much sharper than that of the cardioid counterpart. This seems to indicate that the bidirectional reference signal leads to good spatial selectivity while the cardioid reference signal makes the system robust against errors in target signal direction. However, (B) has an inherent drawback of direction ambiguity.

Method (G) shown in Fig. 4e, due to the constraint in the adaptive filter W, results in a polar pattern with a wide frontal hemisphere [2]. In Fig. 4f, (H) has a fixed hypercardioid pattern. If certain prior knowledge is available, such as the directions of interferences and target signals, (H) can be optimized, and in principle the noise reduction for a single coherent noise is infinite [10].

4.2. Reverberant Environment

The effects of different reverberation times T_r on the polar pattern are shown in Fig. 5. We can see that the noise reduction performance worsens significantly as T_r increases. For (D)-(F) (Figs. 5a- 5c), with T_r increasing from 0.03s to 0.10s and then to 0.30s, the sidelobes increase from 5dB to 8dB and then to 10dB. At the same time, the mainlobes also become wider. On the other hand, although (H) (Figs. 5f) has similar performance to (D)-(F) in reverberant conditions, its mainlobe does not change much. It is also observed that there are two nulls at 90° and 270° for (D)-(G). These nulls actually fall into the plane which crosses the mid-

[3]As the performance of (A) is rather clear, it will not be discussed further.

point of the two microphones and is orthogonal to the line joining the two microphones. In this plane, the transfer functions of the propagation paths from the loudspeaker to the two microphones are almost similar and a reverberant environment results in least degradation in noise reduction. For method (H), the two nulls are at 109° and 251° [10].

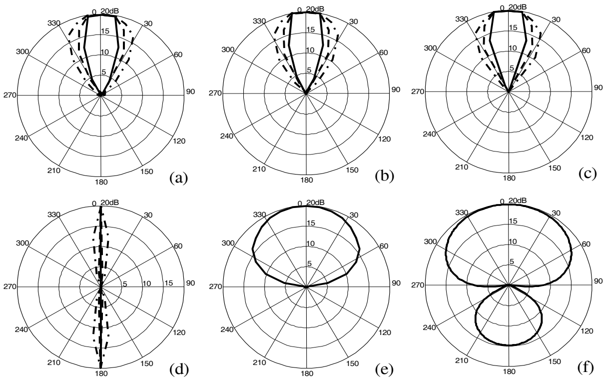

Fig. 4. Simulated results for different combinations of 2 omni-microphones, for K=10 —, 5 - -, 2.5 -.-.. Concentric circles on the polar plots denote 0, 5, 15 and 20dB respectively. (a) Method (D); (b) method (F); (c) Method (E); (d) Method (B); (e) Method (G) [2,3]; (f) Method (H) [10].

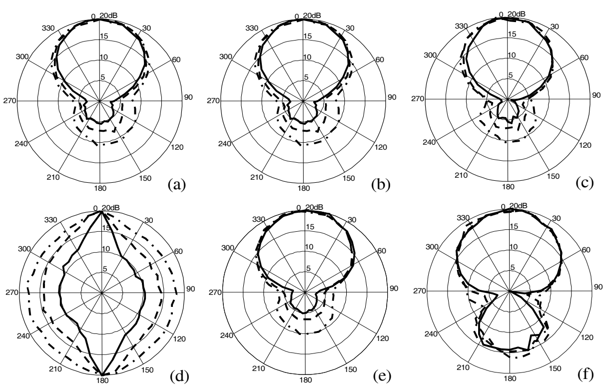

Fig. 5. Simulated results for different combinations of 2 omni-microphones, for $K = 50$, T_r =0.03ms —, 0.1ms - -, 0.3ms -.-.. Concentric circles on the polar plots denote 0, 5, 15 and 20dB respectively. (a) Method (D); (b) method (F); (c) Method (E); (d) Method (B); (e) Method (G) [2,3]; (f) Method (H) [10].

4.3. Design Considerations

In terms of the frequency characteristic, we note that only the frequency response of (D) is uniform for all directions since the primary signal is obtained using an omni-microphone. In contrast, the responses of (A)-(C) and (H) are uniform at 0°only. On the other hand, the responses of (E)-(G) are frequency-dependent for all directions and this leads to the use of a low pass filter as in [3]. Additional improvements to the performance of these methods, especially in multiple interference scenarios can be achieved by using subband adaptive filtering [5] with an increase in complexity.

5. CONCLUSIONS

In this paper, the performance of two groups of DMS is studied. While the first group using two different types of microphones can be made more compact since a inter-microphone distance is not required, the second group using two omni-microphones provides more flexibility. Under the second group, methods (B)-(H) show good noise reduction performance under coherent noise field. In diffuse noise field, methods (C)-(F) have better noise reduction ability than methods (A)-(B), especially in low frequency band. At the same time, the performance of the second group is similar that of their equivalent schemes in the first group. In addition, by using bidirectional-type reference signal, the dual microphone system has a good spatial selectivity, while by using cardioid-type reference signal, the DMS is more robust against errors in the looking-direction. In real environments, it was observed that the performance of the DMS depends highly on the distance between the sound sources and the microphones.

6. REFERENCES

[1] T. Sasaki and K. Gyotoku, “Microphone apparatus,” US Patent 5 471 538, 1995.

[2] F.-L. Luo, J. Yang, C. Pavlovic, and A. Nehorai, “Adaptive null-forming scheme in digital hearing aids,” *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1583–1590, July 2002.

[3] G. W. Elko and A.-T. N. Pong, “A simple adaptive first-order differential microphone,” in *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 1995, pp. 169–172.

[4] J. V. Berghe, “An adaptive noise canceller for hearing aids using two nearby microphones,” *J. Acoust. Soc. Am.*, vol. 103, no. 6, pp. 3621–3626, June 1998.

[5] C. Liu, B. C. Wheeler, W. D. O’Brien, C. R. Lansing, R. C. Bilger, D. Jones, and A. S. Feng, “A two-microphone dual delay-line approach for extraction of a speech sound in the presence of multiple interferers,” *J. Acoust. Soc. Am.*, vol. 110, no. 6, pp. 3218–3231, Dec. 2001.

[6] J. Bernard Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. James R. Zeidler, E. Dong, and R. C. Goodlin, “Adaptive noise canceling: principles and applications,” *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[7] T. D. Abhayapala, R. A. Kennedy, R. C. Williamson, and D. B. Ward, “Nearfield broadband adaptive beamforming,” in *Proc. 5th Int. Symp. Signal Processing and Applications*, Aug. 1999, pp. 839–842.

[8] J. Chen, L. Shue, K. Phua, and H. Sun, “Experimental study of dual microphone systems,” 2003, submitted for publication.

[9] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propagat.*, vol. 30, no. 1, pp. 27–34, 1982.

[10] H. Cox, R. M. Zeskind, and M. M. Owen, “Robust adaptive beamforming,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1365–1375, Oct. 1987.

[11] J. Bitzer and K. U. Simmer, “Multichannel noise reduction: algorithms and theoretical limits,” in *Proc. EURASIP European Signal Processing*, Sept. 1998, pp. 105–108.

[12] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.