SOURCE SEPARATION AND SPEECH DEREVERBERATION BASED ON BLIND MULTICHANNEL IDENTIFICATION IN REVERBERANT ENVIRONMENTS

Yiteng (Arden) Huang†, *Jacob Benesty*‡, *and Jingdong Chen*†

† Bell Laboratories, Lucent Technologies
600 Mountain Avenue
Murray Hill, New Jersey 07974, USA
{arden, jingdong}@research.bell-labs.com

‡ Université du Québec, INRS-EMT
800 de la Gauchetière Ouest, Suite 6900
Montréal, Québec, H5A 1K6, Canada
benesty@inrs-emt.uquebec.ca

ABSTRACT

Separating independent speech sources from their convolutive mixtures in a reverberant acoustic environment is a challenging problem because of two difficulties: (a) very little is known about the source signals or the way they are mixed, and (b) both spatial interference from competing sources and temporal echoes due to room reverberation are observed in the mixtures. In this paper, after blindly identifying the acoustic MIMO system, we deal with spatial interference and temporal echoes in two different steps by converting an $M \times N$ MIMO system into M SIMO systems. The performance is evaluated by simulations with measurements obtained in the varechoic chamber at Bell Labs.

1. INTRODUCTION

Source separation and speech dereverberation have many applications in hands-free communication, multi-modal collaborative, and robust intelligent systems. In this paper, we study this problem in a reverberant acoustic environment where there are M *independent* speech sources and N microphones with $M < N$. At the nth microphone and the kth sample time, we have:

$$\begin{aligned} x_n(k) &= \sum_{m=1}^{M} \mathbf{h}_{nm}^T \mathbf{s}_m(k, L_h) + b_n(k), \qquad (1) \\ & k = 1, 2, \cdots, K, \;\; n = 1, 2, \cdots, N, \end{aligned}$$

where $(\cdot)^T$ denotes the transpose of a matrix or a vector,

$$\begin{aligned} \mathbf{h}_{nm} &= \begin{bmatrix} h_{nm,0} & h_{nm,1} & \cdots & h_{nm,L_h-1} \end{bmatrix}^T, \\ & n = 1, 2, \cdots, N, \;\; m = 1, 2, \cdots, M, \end{aligned}$$

is the impulse response (of length L_h, $\forall m, n$) between source m and microphone n,

$$\mathbf{s}_m(k, L_h) = \begin{bmatrix} s_m(k) & s_m(k-1) & \cdots & s_m(k - L_h + 1) \end{bmatrix}^T$$

is a vector containing the last L_h samples of the mth source signal s_m, and $b_n(k)$ is a zero-mean additive white Gaussian noise (AWGN) with variance $\sigma_b^2, \forall n$.

Using the z transform, the signal model of the MIMO system (1) is expressed as

$$X_n(z) = \sum_{m=1}^{M} H_{nm}(z) S_m(z) + B_n(z), \;\; n = 1, 2, \cdots, N, \quad (2)$$

where $H_{nm}(z) = \sum_{l=0}^{L_h - 1} h_{nm,l} z^{-l}$.

Traditional blind source separation (BSS) is accomplished by independent component analysis (ICA), which applies a group of de-mixing filters to the microphone signals and reinforces mutual independence among the outputs, regarded as the estimates of the source signals. Existing ICA algorithms differ in the way that the dependence of the separated speech signals is defined. The well-known independence criteria include second-order statistics (SOS), higher (than second) order statistics, and information-theory-based measures (please refer to the books [1], [2] and references therein for a more detailed discussion on various ICA methods). Alternatively, we will propose in this paper a new method for source separation and speech dereverberation based on blind channel identification. The rigorous derivation of this method would help us better understand the procedure and limitation of BSS techniques. More insightful performance merits in addition to signal-to-interference ratio (SIR) will be suggested and used in the simulations.

2. BLIND IDENTIFICATION OF A MIMO SYSTEM

In this paper, we intend to separate competing speech sources after blindly identifying the MIMO FIR system. Blind MIMO identification is difficult even for communication systems with short channel impulse responses. It becomes dramatically complicated when an acoustic system is the target as the case studied in this paper. Trying to solve it all at once involves a huge number of parameters to estimate and the current research in this area remains at the stage of feasibility investigations. Moreover, scaling and permutation ambiguities are similar to what have been observed in the BSS problem. Therefore we choose to decompose the problem into several subproblems in which SIMO systems are blindly identified. We assume that from time to time each speaker occupies at least one exclusive interval alone and when they start talking simultaneously the room acoustics have not yet significantly varied. Then in each single-talk interval a SIMO system will be blindly identified and its channel impulse responses will be saved for later use in source separation and speech deconvolution during double or multiple talk periods. In this paper, we assume that the SIMO systems under investigation are all blindly identifiable (for blind identifiability please see [3]) and we will employ the unconstrained normalized multichannel frequency-domain LMS (UNMCFLMS) algorithm [4].

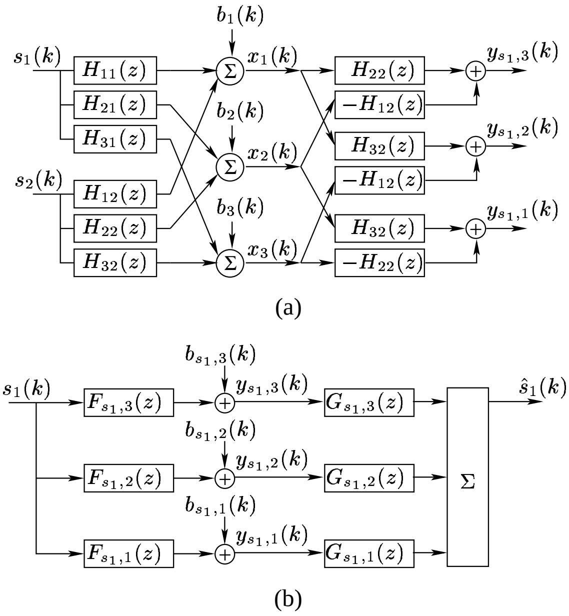

Figure 1: Illustration of the two-stage procedure for source separation and speech dereverberation with respect to s_1 in a 2×3 MIMO system. (a) Cancellation of spatial interference from s_2 and (b) dereverberation in light of the Bezout theorem.

3. SEPARATING SPATIAL INTERFERENCE AND TEMPORAL ECHOES

3.1. Example: Conversion of a 2×3 MIMO System to Two SIMO Systems

For a 2×3 MIMO system, the spatial interference is cancelled by using two microphone signals at a time, as illustrated in Fig. 1 (a). As a results, the 2×3 MIMO system is converted to two SIMO systems with the two speech sources as the inputs. Then for s_1, we have

$$\begin{aligned} Y_{s_1,p}(z) &= H_{s_1,p1}(z)X_1(z) + H_{s_1,p2}(z)X_2(z) + \\ &\quad H_{s_1,p3}(z)X_3(z) \\ &= \sum_{q=1}^{3} H_{s_1,pq}(z)X_q(z), \quad p = 1, 2, 3, \end{aligned} \tag{3}$$

where $H_{s_1,pp}(z) = 0,\ \forall p$. The polynomials $H_{s_1,pq}(z)$, $p, q = 1, 2, 3$, $p \neq q$, are chosen such that:

$$Y_{s_1,p}(z) = F_{s_1,p}(z)S_1(z) + B_{s_1,p}(z), \quad p = 1, 2, 3. \tag{4}$$

As shown in Fig. 1 (a), one possibility is to choose:

$$\begin{array}{rclrcl} H_{s_1,12}(z) &=& H_{32}(z), & H_{s_1,13}(z) &=& -H_{22}(z), \\ H_{s_1,21}(z) &=& H_{32}(z), & H_{s_1,23}(z) &=& -H_{12}(z), \\ H_{s_1,31}(z) &=& H_{22}(z), & H_{s_1,32}(z) &=& -H_{12}(z). \end{array} \tag{5}$$

In this case, we find that:

$$\begin{aligned} F_{s_1,1}(z) &= H_{32}(z)H_{21}(z) - H_{22}(z)H_{31}(z), \\ F_{s_1,2}(z) &= H_{32}(z)H_{11}(z) - H_{12}(z)H_{31}(z), \\ F_{s_1,3}(z) &= H_{22}(z)H_{11}(z) - H_{12}(z)H_{21}(z). \end{aligned} \tag{6}$$

Since $\deg\left[H_{nm}(z)\right] = L_h - 1$, where $\deg[\cdot]$ is the degree of a polynomial, therefore $\deg\left[F_{s_1,p}(z)\right] \leq 2L_h - 2$. Since neither $\{H_{12}(z), H_{22}(z), H_{32}(z)\}$ nor $\{H_{11}(z), H_{21}(z), H_{31}(z)\}$ share common zeros as assumed for blind identifiability of these SIMO systems, $\{F_{s_1,1}(z), F_{s_1,2}(z), F_{s_1,3}(z)\}$ will not share any common zeros.

The second SIMO system corresponding to the source s_2 can be derived in a similar way, but the procedure is omitted due to space limitations.

3.2. Generalization

The approach to separating spatial interference and temporal echoes explained in the previous subsection on a simple example will be generalized here to an $M \times N$ MIMO system ($M < N$). We begin with writing (2) into vector/matrix form

$$\vec{\mathbf{X}}(z) = \mathbf{H}(z)\vec{\mathbf{S}}(z) + \vec{\mathbf{B}}(z), \tag{7}$$

where

$$\begin{aligned} \vec{\mathbf{X}}(z) &= \left[X_1(z)\ X_2(z)\ \cdots\ X_N(z)\right]^T, \\ \mathbf{H}(z) &= \begin{bmatrix} H_{11}(z) & H_{12}(z) & \cdots & H_{1M}(z) \\ H_{21}(z) & H_{22}(z) & \cdots & H_{2M}(z) \\ \vdots & \vdots & \vdots & \vdots \\ H_{N1}(z) & H_{N2}(z) & \cdots & H_{NM}(z) \end{bmatrix}, \\ \vec{\mathbf{S}}(z) &= \left[S_1(z)\ S_2(z)\ \cdots\ S_M(z)\right]^T, \\ \vec{\mathbf{B}}(z) &= \left[B_1(z)\ B_2(z)\ \cdots\ B_N(z)\right]^T. \end{aligned}$$

Let us choose M from N microphone outputs and we have $P = C_N^M$ different ways of doing so. For the pth ($p = 1, 2, \cdots, P$) combination, we denote the index of the M selected microphone signals as p_m, $m = 1, 2, \cdots, M$, and get an $M \times M$ MIMO subsystem. For this subsystem, we consider the following equation:

$$\vec{\mathbf{Y}}_p(z) = \mathbf{H}_{s,p}(z)\vec{\mathbf{X}}_p(z), \quad p = 1, 2, \cdots, P, \tag{8}$$

where

$$\begin{aligned} \vec{\mathbf{Y}}_p(z) &= \left[Y_{s_1,p}(z)\ Y_{s_2,p}(z)\ \cdots\ Y_{s_M,p}(z)\right]^T, \\ \mathbf{H}_{s,p}(z) &= \begin{bmatrix} H_{s_1,p1}(z) & H_{s_1,p2}(z) & \cdots & H_{s_1,pM}(z) \\ H_{s_2,p1}(z) & H_{s_2,p2}(z) & \cdots & H_{s_2,pM}(z) \\ \vdots & \vdots & \vdots & \vdots \\ H_{s_M,p1}(z) & H_{s_M,p2}(z) & \cdots & H_{s_M,pM}(z) \end{bmatrix}, \\ \vec{\mathbf{X}}_p(z) &= \left[X_{p_1}(z)\ X_{p_2}(z)\ \cdots\ X_{p_M}(z)\right]^T. \end{aligned}$$

Let $\mathbf{H}_p(z)$ be the $M \times M$ matrix obtained from the system's channel matrix $\mathbf{H}(z)$ by keeping its rows corresponding to the M selected microphone signals. Then similar to (7), we have

$$\vec{\mathbf{X}}_p(z) = \mathbf{H}_p(z)\vec{\mathbf{S}}(z) + \vec{\mathbf{B}}_p(z), \tag{9}$$

where $\vec{\mathbf{B}}_p(z) = \left[B_{p_1}(z)\ B_{p_2}(z)\ \cdots\ B_{p_M}(z)\right]^T$. Substituting (9) into (8) yields

$$\vec{\mathbf{Y}}_p(z) = \mathbf{H}_{s,p}(z)\mathbf{H}_p(z)\vec{\mathbf{S}}(z) + \mathbf{H}_{s,p}(z)\vec{\mathbf{B}}_p(z). \tag{10}$$

In order to remove the spatial interference, the objective here is to find the matrix $\mathbf{H}_{s,p}(z)$ whose components are linear combinations of $H_{nm}(z)$ such that the product

$$\mathbf{\Phi}_p(z) \triangleq \mathbf{H}_{s,p}(z)\mathbf{H}_p(z) \tag{11}$$

would be a diagonal matrix and

$$Y_{s_m,p}(z) = F_{s_m,p}(z)S_m(z) + B_{s_m,p}(z), \quad m = 1, 2, \cdots, M,\ p = 1, 2, \cdots, P. \tag{12}$$

Obviously a good choice for $\mathbf{H}_{s,p}(z)$ is the adjoint of matrix $\mathbf{H}_p(z)$, i.e., the (i, j)th element of $\mathbf{H}_{s,p}(z)$ is the (j, i)th cofactor of $\mathbf{H}_p(z)$. Consequently, the polynomial $F_{s_m,p}(z)$ would be the determinant of $\mathbf{H}_p(z)$ and

$$F_{s_m,p}(z) = \sum_{q=1}^{M} H_{s_m,pq}(z)H_{p_qm}(z).$$

Again, it can be shown that $\{F_{s_m,p}(z),\ p = 1, 2, \cdots, P\}$ do not share common zeros and the length of the FIR filter $f_{s_m,p}$ would be $L_f \le M(L_h - 1) + 1$.

4. SPEECH DEREVERBERATION FOR SIMO SYSTEMS

For the SIMO system with respect to source s_m ($m = 1, 2, \cdots, M$), we apply the polynomials $G_{s_m,p}(z)$ (p = 1, 2, $\cdots$, P) to its outputs, as shown in Fig. 1 (b), and add the results to get

$$\widehat{S}_m(z) = \sum_{p=1}^{P} G_{s_m,p}(z)Y_{s_m,p}(z). \tag{13}$$

Since $\{F_{s_m,p}(z),\ p = 1, 2, \cdots, P\}$ do not share common zeros, we know from the Bezout theorem that there exists a set of $G_{s_m,p}(z)$ such that

$$\sum_{p=1}^{P} F_{s_m,p}(z)G_{s_m,p}(z) = 1, \tag{14}$$

and $\widehat{S}_m(z) = S_m(z)$ in the absence of noise. The idea of using the Bezout theorem for dereverberation of an acoustic SIMO system was first proposed in [5] in the context of room acoustics, where the method is more widely referred to as the MINT (multichannel inverse theorem) technique.

To find the dereverberation filters, we write the Bezout equation (14) in the time domain as:

$$\mathbf{F}^{\mathrm{c}}_{s_m}\mathbf{g}_{s_m} = \sum_{p=1}^{P} \mathbf{F}^{\mathrm{c}}_{s_m,p}\mathbf{g}_{s_m,p} = \mathbf{e}_1, \tag{15}$$

where

$$\begin{aligned}
\mathbf{F}^{\mathrm{c}}_{s_m} &= \begin{bmatrix} \mathbf{F}^{\mathrm{c}}_{s_m,1} & \mathbf{F}^{\mathrm{c}}_{s_m,2} & \cdots & \mathbf{F}^{\mathrm{c}}_{s_m,P} \end{bmatrix}, \\
\mathbf{g}_{s_m} &= \begin{bmatrix} \mathbf{g}^T_{s_m,1} & \mathbf{g}^T_{s_m,2} & \cdots & \mathbf{g}^T_{s_m,P} \end{bmatrix}^T, \\
\mathbf{g}_{s_m,p} &= \begin{bmatrix} g_{s_m,p,0} & g_{s_m,p,1} & \cdots & g_{s_m,p,L_g-1} \end{bmatrix}^T,
\end{aligned}$$

L_g is the length of the FIR filter $g_{s_m,p}$,

$$\mathbf{F}^{\mathrm{c}}_{s_m,p} = \begin{bmatrix} f_{s_m,p,0} & 0 & \cdots & 0 \\ \vdots & f_{s_m,p,0} & \cdots & 0 \\ f_{s_m,p,L_f-1} & \vdots & \ddots & \vdots \\ 0 & f_{s_m,p,L_f-1} & \ddots & f_{s_m,p,0} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & f_{s_m,p,L_f-1} \end{bmatrix}$$

is an $(L_f + L_g - 1) \times L_g$ matrix, and $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ is an $(L_f + L_g - 1) \times 1$ vector. In order to have a unique solution for (15), L_g must be chosen in such a way that $\mathbf{F}^{\mathrm{c}}_{s_m}$ is a square matrix. In this case, we have:

$$L_g = \frac{L_f - 1}{P - 1} \le \frac{M(L_h - 1)}{P - 1}. \tag{16}$$

The shortest dereverberation filters are idea but impractical. In our implementation, we allow a larger L_g than necessary and solve (15) in the least squares sense: $\mathbf{g}_{s_m,\mathrm{LS}} = \mathbf{F}^{\mathrm{c}\,\dagger}_{s_m}\mathbf{e}_1$, where $\mathbf{F}^{\mathrm{c}\,\dagger}_{s_m} = \left(\mathbf{F}^{\mathrm{c}\,T}_{s_m}\mathbf{F}^{\mathrm{c}}_{s_m}\right)^{-1}\mathbf{F}^{\mathrm{c}\,T}_{s_m}$ is the pseudo-inverse of the matrix $\mathbf{F}^{\mathrm{c}}_{s_m}$. If a decision delay d is taken into account, then the dereverberation filters turn out to be

$$\mathbf{g}_{s_m,\mathrm{LS}} = \mathbf{F}^{\mathrm{c}\,\dagger}_{s_m}\mathbf{e}_d, \tag{17}$$

where $\mathbf{e}_d = \begin{bmatrix} \underbrace{0 \ \cdots \ 0}_{d} & 1 & \underbrace{0 \ \cdots \ 0}_{L_f+L_g-d-2} \end{bmatrix}^T$.

5. SIMULATIONS

In this section, we will evaluate the performance of the proposed blind source separation and speech dereverberation algorithm via simulations in realistic acoustic environments.

5.1. Performance Measures

Similar to what was adopted in our earlier study [4], we will use the normalized projection misalignment (NPM) to evaluate the performance of a BCI algorithm. To assess the performance of source separation and speech dereverberation, two measures, namely signal-to-interference ratio (SIR) and speech spectral distortion, are used in the simulations. For speech spectral distortion, we employed the Itakura-Saito (IS) distortion measure [6]: d_{IS,s_m}. For the SIR, we referred to the notion given in [7] but defined the measure in a different manner since their definition is applicable only for an $M \times M$ MIMO system. In this paper, our interest is in the more general $M \times N$ MIMO systems with $M < N$.

We first define the average input SIR at microphone n as:

$$\mathrm{SIR}^{\mathrm{in}}_n \triangleq \frac{1}{M}\sum_{m=1}^{M}\left(\frac{E\left\{[h_{nm} * s_m(k)]^2\right\}}{\sum_{i=1,i\neq m}^{M} E\left\{[h_{ni} * s_i(k)]^2\right\}}\right), \tag{18}$$

where $E\{\cdot\}$ and $*$ denote mathematical expectation and linear convolution, respectively. Then the overall average input SIR is given by:

$$\mathrm{SIR}^{\mathrm{in}} \triangleq \frac{1}{N}\sum_{n=1}^{N}\mathrm{SIR}^{\mathrm{in}}_n. \tag{19}$$

The output SIR is defined using the same principle but the expression will be more complicated. For a concise presentation, we denote $\phi_{p,ji}$ ($p = 1, 2, \cdots, P,\ i, j = 1, 2, \cdots, M$) as the impulse response of the equivalent channel from the ith input to the jth output for the pth $M \times M$ separation subsystem. From (10) and (11), we know that $\phi_{p,ji}$ corresponds to the (j, i)th element of $\mathbf{\Phi}_p(z)$ and $\psi_{p,mm} = f_{s_m,p}$. Then the average output SIR for the pth subsystem is:

$$\mathrm{SIR}^{\mathrm{out}}_p \triangleq \frac{\sum_{m=1}^{M} E\left\{[\phi_{p,ii} * s_i(k)]^2\right\}}{\sum_{j=1}^{M}\sum_{i=1,i\neq j}^{M} E\left\{[\phi_{p,ji} * s_i(k)]^2\right\}}, \tag{20}$$

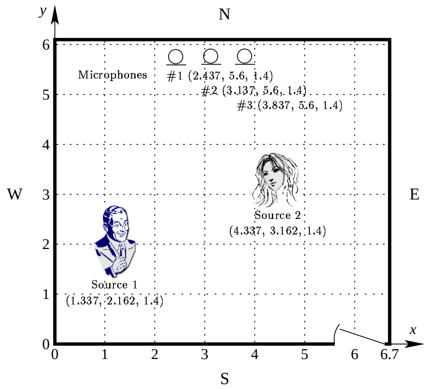

Figure 2: Floor plan of the varechoic chamber at Bell Labs (coordinate values measured in meters).

and the overall average output SIR is found as:

$$\mathrm{SIR}^{\mathrm{out}} \triangleq \frac{1}{P} \sum_{p=1}^{P} \mathrm{SIR}_p^{\mathrm{out}}. \tag{21}$$

5.2. Experimental Setup and Results

The simulations were conducted with the impulse responses measured in the varechoic chamber at Bell Labs [8]. A diagram of the floor plan layout is given in Fig. 2, which shows the positions of the three microphones and two speech sources (one male and the other female). Speech signals were sampled at 8 kHz. The wall panels in the chamber were adjusted with four room acoustic conditions being formed. The microphone outputs are computed by convolving the speech signals and corresponding channel impulse responses. At each microphone, additive noise was inserted at 75 dB signal-to-noise ratio (SNR). For BCI, both adaptive and batch algorithms were investigated. For speech dereverberation, the delay d was fixed as $L_h - 1$.

Table 1 summarizes the experimental results for all four different room acoustics. We see that as room reverberation and channel length increase, channel estimates become less accurate. But although the distortions in the final recovered speech signals grow too, they remain at an acceptable level (the IS measure is less than 0.2 even in the worst case), implying some promise of successful use of the proposed algorithm in prospect speech processing systems.

6. CONCLUSIONS

Blind separation of speech sources from their convolutive mixtures is a very difficult problem in a real reverberant environment. Existing blind source separation methods focus primarily on the signal-to-interference ratio and observe high distortion in their separated signals when room reverberation is significant. A source separation and speech dereverberation was proposed based on blind channel identification techniques. After the mixing system is blindly identified, we showed that spatial interference and temporal echoes could be separated and we then dealt with them in two sequential steps. We conducted experiments using various real impulse responses measured in the varechoic chamber at Bell Labs. The results demonstrated the success and robustness of the proposed algorithm in highly reverberant acoustic environments.

Table 1: Performance of the source separation and speech dereverberation algorithm based on BCI techniques in the varechoic chamber at Bell Labs with different panel configurations.

	NPM (dB)		$\mathrm{SIR}^{\mathrm{in}}$	$\mathrm{SIR}^{\mathrm{out}}$	After S.S.		After S.D.	
BCI	H_{s_1}	H_{s_2}	(dB)	(dB)	$d_{\mathrm{IS},s_1}^{\mathrm{SS}}$	$d_{\mathrm{IS},s_2}^{\mathrm{SS}}$	$d_{\mathrm{IS},s_1}^{\mathrm{SD}}$	$d_{\mathrm{IS},s_2}^{\mathrm{SD}}$
	89% panels open, $T_{60} = 240$ ms, $L_h = 256$							
(a)	-17.472	-17.801	0.231	52.016	1.9476	2.1368	0.0476	0.0258
(b)	-43.556	-35.374	0.231	74.970	2.1041	2.2104	0.0002	0.0006
	75% panels open, $T_{60} = 310$ ms, $L_h = 256$							
(a)	-18.737	-18.057	0.391	52.899	2.6878	2.8424	0.0307	0.0205
(b)	-50.533	-42.994	0.391	74.966	2.8021	3.0904	0.0032	0.0031
	30% panels open, $T_{60} = 380$ ms, $L_h = 512$							
(a)	-13.366	-11.581	0.249	44.780	2.4084	3.9952	0.0755	0.1773
(c)	-38.711	-29.868	0.249	73.932	2.8960	4.2943	0.0019	0.0044
	Panels all closed, $T_{60} = 580$ ms, $L_h = 512$							
(a)	-13.818	-14.470	0.475	44.999	2.6414	4.8217	0.1592	0.1988
(c)	-50.533	-42.994	0.475	73.873	2.6321	4.4309	0.0004	0.0007

NOTES:

H_{s_m} represents the SIMO system corresponding to source s_m.
T_{60} denotes 60-dB reverberation time in the 20-4000 Hz band.
S.S. and S.D. stand for source separation and speech dereverberation, respectively.
(a) the adaptive UNMCFLMS BCI algorithm [4]. (b) the Batch (SVD) BCI with 2500 samples. (c) the Batch (SVD) BCI with 3000 samples.

7. REFERENCES

[1] S. Haykin, Ed., *Unsupervised Adaptive Filtering, Volume I: Blind Source Separation*, New York: Wiley, 2000.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, New York: Wiley, 2002.

[3] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

[4] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[5] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 145–152, Feb. 1988.

[6] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2000, vol. 2, pp. 1041–1044.

[8] A. Härmä, "Acoustic measurement data from the varechoic chamber," Technical Memorandum, Agere Systems, Nov. 2001.