ADAPTIVE DECORRELATION FILTERING ALGORITHM FOR SPEECH SOURCE SEPARATION IN UNCORRELATED NOISES

Rong Hu and Yunxin Zhao

Department of Computer Science
University of Missouri, Columbia, MO 65211 USA
rhq2c@mizzou.edu   zhaoy@missouri.edu

ABSTRACT

A time-domain vector formulation is presented on analysis and solution of the adaptive decorrelation filtering (ADF) system for blind speech source separation in additive background noises. The formulation leads to a derivation of a gradient descent algorithm and offers insights into the impact of uncorrelated white noises on ADF system. A new noise-adapted ADF algorithm is then derived by modifying the decorrelation criterion function to exclude noise impact on cross-correlation information of ADF outputs. Speech separation simulations were based on convolutive mixtures of TIMIT speech data generated with long impulse responses measured in real reverberant acoustic experiment. The proposed algorithm significantly improved convergence rate, gains in target-to-interference ratio at ADF outputs, and phone accuracy of separated target speech source.

1. INTRODUCTION

Blind separation of simultaneous speech signals in real acoustic environment is a difficult and important topic, with potential applications in automatic hands-free speech recognition (ASR) and assistive speech communication. Adaptive decorrelation filtering (ADF) algorithm [1] and certain ICA techniques for convolutive mixtures [2] offer possible solutions for this task.

Weinstein et al [1] established decorrelation as an estimation criterion for blind separation of convolutive mixtures of speech signals. For a general convolutive mixing system, decorrelation cannot guarantee unique solution and an FIR constraint needs to be imposed on the form of the mixing model. For real acoustic paths, the constraint can be met as long as FIR taps are long enough.

Recently, effective preprocessing and post filtering techniques have been developed for ADF to work in reverberant acoustic conditions [3]. However, diffusive noise adds another level of difficulty to the blind source separation task since both filtering and adaptation processes of ADF model are affected by noise. Previously, frequency domain analysis of noise effects on ADF was made in [4] and a subspace based noise-reduction front end was used to improve the working conditions for ADF system. However, there were no effective methods proposed to make ADF robust in the presence of noise.

This paper formulates the ADF model in a time-domain vector form to derive a gradient descent algorithm and provide analysis of noise effects. A noise-adapted ADF algorithm is then proposed to improve the performance of ADF de-coupling filter estimation. Speech source separation and phone recognition simulations were conducted for the new method.

This work is supported in part by National Institutes of Health under the grant NIH 1 R01 dc04340-01A2 and the National Science Foundation under the grant NSF EIA 9911095

2. ADF SYSTEM MODEL

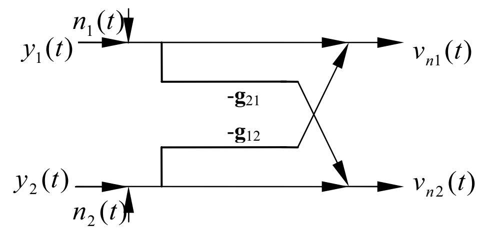

Figure 1. ADF separation system model in background noise

For a two-speaker-two-microphone system, the noise-free signal mixing process ($n_j(t)$=0) is modeled as

$$\begin{bmatrix} Y_1(z) \\ Y_2(z) \end{bmatrix} = \begin{bmatrix} 1 & G_{12}(z) \\ G_{21}(z) & 1 \end{bmatrix} \begin{bmatrix} H_{11}S_1(z) \\ H_{22}S_2(z) \end{bmatrix} \qquad (1)$$

where $G_{ij}(z) = H_{ij}(z)/H_{jj}(z)$ are the cross-coupling filters from the jth speaker to the ith microphone. The ADF source separation model is shown in Fig.1, where the cross-coupling filters $\mathbf{g}_{ij}=[g(0), \ldots, g(N\text{-}1)]^T$, with $(\cdot)^T$ for transpose, are to be identified. The following notations will be used in the rest parts of this paper: variables in bold lower case for vectors, bold capital for matrices, '*' for convolution, E{·} for expectation, and the correlation vector formed by a signal sample $a(t)$ and a signal vector $\mathbf{b}(t)$ is denoted by $\mathbf{r}_{a\mathbf{b}} = E\{a(t)\mathbf{b}(t)\}$.

3. ADF FORMULATION IN NOISE

3.1. Vector Formulation of ADF Algorithm

The input-output (I/O) relation of ADF source separation system for clean (noise-free) speech-mixtures can be put in a vector form of $\mathbf{v} = \mathbf{G} \cdot \tilde{\mathbf{y}}$, with

$$\mathbf{G} = \begin{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0}_{N\times(N-1)} \end{bmatrix} & -\mathbf{G}_{12} \\ -\mathbf{G}_{21} & \begin{bmatrix} \mathbf{I} & \mathbf{0}_{N\times(N-1)} \end{bmatrix} \end{bmatrix}, \quad \tilde{\mathbf{y}} = \begin{bmatrix} \tilde{\mathbf{y}}_1(t) \\ \tilde{\mathbf{y}}_2(t) \end{bmatrix}, \text{ and } \mathbf{v} = \begin{bmatrix} \mathbf{v}_1(t) \\ \mathbf{v}_2(t) \end{bmatrix}$$

where $\tilde{\mathbf{y}}_j(t)$, $\mathbf{v}_i(t)$, and $\mathbf{G}_{ij}$ are (2N-1)×1 input vector, N×1 output vector, and N×(2N-1) system matrix, respectively with $i,j=1,2 \quad i \neq j$, and $\mathbf{I}$ denotes the N×N identity matrix. Specifically,

$\mathbf{v}_i(t)=[v_i(t),\cdots,v_i(t-N+1)]^T$,

$\tilde{\mathbf{y}}_j(t)=[y_j(t),\cdots,y_j(t-N+1),y_j(t-N),\cdots y_j(t-2N+2)]^T$,

$$\mathbf{G}_{ij}=\begin{bmatrix} g_{ij}(0) & g_{ij}(1) & \cdots & g_{ij}(N-1) & 0 & \cdots & 0 \\ 0 & g_{ij}(0) & g_{ij}(1) & \cdots & g_{ij}(N-1) & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & g_{ij}(0) & g_{ij}(1) & \cdots & g_{ij}(N-1) \end{bmatrix}.$$

The I/O relation of input and output correlation matrices is $\mathbf{R}_{\mathbf{vv}}=\mathbf{G}\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}\mathbf{G}^T$, with

$$\mathbf{R}_{\mathbf{vv}}=E\{\mathbf{v}(t)\mathbf{v}^T(t)\}=\begin{bmatrix}\mathbf{R}_{\mathbf{v}_1\mathbf{v}_1} & \mathbf{R}_{\mathbf{v}_1\mathbf{v}_2}\\ \mathbf{R}_{\mathbf{v}_2\mathbf{v}_1} & \mathbf{R}_{\mathbf{v}_2\mathbf{v}_2}\end{bmatrix}$$

where the off-diagonal and diagonal blocks are

$$\mathbf{R}_{\mathbf{v}_i\mathbf{v}_j}=\mathbf{R}_{\mathbf{y}_i\mathbf{y}_j}-\mathbf{R}_{\mathbf{y}_i\tilde{\mathbf{y}}_i}\mathbf{G}_{ji}^T-\mathbf{G}_{ij}\mathbf{R}_{\tilde{\mathbf{y}}_j\mathbf{y}_j}+\mathbf{G}_{ij}\mathbf{R}_{\tilde{\mathbf{y}}_j\tilde{\mathbf{y}}_i}\mathbf{G}_{ij}^T \quad (2)$$

$$\mathbf{R}_{\mathbf{v}_i\mathbf{v}_i}=\mathbf{R}_{\mathbf{y}_i\mathbf{y}_i}-\mathbf{R}_{\mathbf{y}_i\tilde{\mathbf{y}}_j}\mathbf{G}_{ij}^T-\mathbf{G}_{ij}\mathbf{R}_{\tilde{\mathbf{y}}_j\mathbf{y}_i}+\mathbf{G}_{ij}\mathbf{R}_{\tilde{\mathbf{y}}_j\tilde{\mathbf{y}}_j}\mathbf{G}_{ij}^T \quad (3)$$

The system output correlation matrix is also related to system input-output cross-correlation matrix as

$$\mathbf{R}_{\mathbf{vv}}=\mathbf{G}\mathbf{R}_{\tilde{\mathbf{y}}\mathbf{v}}=\mathbf{R}_{\mathbf{yv}}-\begin{bmatrix}\mathbf{0}_{N\times(2N-1)} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{0}_{N\times(2N-1)}\end{bmatrix}\cdot\mathbf{R}_{\tilde{\mathbf{y}}\mathbf{v}} \quad (4)$$

where $\mathbf{y}_i(t)=[y_i(t),\cdots,y_i(t-N+1)]^T$.

Imposing decorrelation conditions on (4) to force the off-diagonal blocks of $\mathbf{R}_{\mathbf{vv}}$ to zero, we obtain

$$\begin{bmatrix}\mathbf{0}_{N\times(2N-1)} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{0}_{N\times(2N-1)}\end{bmatrix}\cdot\mathbf{R}_{\tilde{\mathbf{y}}\mathbf{v}}=\mathbf{R}_{\mathbf{yv}}-\begin{bmatrix}\mathbf{R}_{\mathbf{v}_1\mathbf{v}_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{\mathbf{v}_2\mathbf{v}_2}\end{bmatrix} \quad (5)$$

which is an over-determined system of equations since the solution of ADF coefficients need 2N constraints only. Instead of solving $\mathbf{g}_{ij}$'s by the block-diagonalization of (4) directly, (as actually did by minimizing a criterion in [5]) we can derive solutions by appropriately selecting a subset of constraints from the off-diagonal blocks in (5). Choosing N constraints from the 1st row and N constraints from the (N+1)-th rows of (5) respectively, the equations for ADF system solutions are obtained as

$$\begin{bmatrix}\mathbf{0} & \mathbf{R}_{\mathbf{y}_1\mathbf{v}_1}^T\\ \mathbf{R}_{\mathbf{y}_2\mathbf{v}_2}^T & \mathbf{0}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{g}_{12}\\ \mathbf{g}_{21}\end{bmatrix}=\begin{bmatrix}\mathbf{r}_{y_2\mathbf{v}_1}\\ \mathbf{r}_{y_1\mathbf{v}_2}\end{bmatrix} \quad (6)$$

The least-square solution of Eq. (6) coincides with the least-cross-correlation of ADF outputs because the error vector of (6) actually coincides with $\mathbf{r}_{v_i\mathbf{v}_j}$, i.e., $\mathbf{r}_{v_i\mathbf{v}_j}=\mathbf{r}_{y_i\mathbf{v}_j}-\mathbf{R}_{\mathbf{y}_j\mathbf{v}_j}^T\mathbf{g}_{ij}$.

By alternating between the following two cross-correlation minimization steps

$$\mathbf{g}_{12}^{opt}=\arg\min J_{12}=\tfrac{1}{2}\left(\mathbf{r}_{v_1\mathbf{v}_2}^T\mathbf{r}_{v_1\mathbf{v}_2}\right)$$
$$\mathbf{g}_{21}^{opt}=\arg\min J_{21}=\tfrac{1}{2}\left(\mathbf{r}_{v_2\mathbf{v}_1}^T\mathbf{r}_{v_2\mathbf{v}_1}\right)$$

the parameters of ADF separation filters can be searched by the gradient descent procedure $\mathbf{g}_{ij}^{(t)}=\mathbf{g}_{ij}^{(t-1)}-\mu(t)\cdot\nabla_{\mathbf{g}_{ij}}J_{ij}$. Assuming independence of $\mathbf{g}_{12}$ and $\mathbf{g}_{21}$, the gradient vectors $\nabla_{g_{ij}}J_{ij}$ are derived from the 1st row of (2) as

$$\nabla_{\mathbf{g}_{ij}}J_{ij}=\tfrac{\partial}{\partial\mathbf{g}_{ij}}\tfrac{1}{2}\left(\mathbf{r}_{v_i\mathbf{v}_j}^T\mathbf{r}_{v_i\mathbf{v}_j}\right)=\tfrac{\partial}{\partial\mathbf{g}_{ij}}\left(\mathbf{r}_{v_i\mathbf{v}_j}^T\right)\cdot\mathbf{r}_{v_i\mathbf{v}_j}=-\mathbf{R}_{\mathbf{y}_j\mathbf{v}_j}\mathbf{r}_{v_i\mathbf{v}_j} \quad (7)$$

Alternative methods could be used to solve (6) with varying performances. In fact, the RLS-like algorithm proposed in [1] can be derived from the alternating solution of (6) with Newton's method that uses Hessian matrices implied by (7).

Under the assumption that the real parts of the eigen-values of $\mathbf{R}_{\mathbf{y}_j\mathbf{v}_j}$ remain positive [4], the adaptation direction in (7) can be simplified to $\Delta\mathbf{g}_{ij}\approx\mathbf{r}_{v_i\mathbf{v}_j}$, whose instantaneous implementation coincides with the basic ADF algorithm in [1], derived from the zero-searching problem using the method of Robbins-Monro stochastic approximation [8].

3.2. Analysis of Noise Effects

Assume that spatially uncorrelated white additive noises are present at the inputs, as shown in Fig.1, and denote the ADF output in noise by $\mathbf{v}_n$. The I/O relation becomes $\mathbf{v}_n=\mathbf{G}\cdot(\tilde{\mathbf{y}}+\tilde{\mathbf{n}})$, where $\tilde{\mathbf{n}}=[\tilde{\mathbf{n}}_1^T(t) \ \ \tilde{\mathbf{n}}_2^T(t)]^T$ is a (4N-2)×1 noise vector, uncorrelated with speech mixtures, with

$\tilde{\mathbf{n}}_j(t)=[n_j(t),\cdots,n_j(t-N+1),n_j(t-N),\cdots n_j(t-2N+2)]^T$

The noise correlation matrix is

$$\mathbf{R}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}(t)=blkdiag\left(\sigma_1^2\mathbf{I}_{(2N-1)\times(2N-1)},\ \ \sigma_2^2\mathbf{I}_{(2N-1)\times(2N-1)}\right) \quad (8)$$

where σ_1^2 and σ_2^2 are power of $\mathbf{n}_1$ and $\mathbf{n}_2$, respectively.

Under input noise, the new I/O relation of correlation matrices becomes

$$\mathbf{R}_{\mathbf{v}_n\mathbf{v}_n}=\mathbf{G}\left(\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}+\mathbf{R}_{\tilde{\mathbf{n}}\tilde{\mathbf{n}}}\right)\mathbf{G}^T \quad (9)$$

From (8) and (9), the noisy ADF output correlation $\mathbf{R}_{\mathbf{v}_n\mathbf{v}_n}$ can be determined by the noise-free output correlation $\mathbf{R}_{\mathbf{vv}}$, the noise power σ_1^2 and σ_2^2, and the cross-coupling filters $\mathbf{g}_{ij}$'s as follows:

$$\mathbf{R}_{\mathbf{v}_n\mathbf{v}_n}=\mathbf{R}_{\mathbf{vv}}+$$
$$\begin{bmatrix}\sigma_1^2\mathbf{I}+\sigma_2^2\mathbf{G}_{12}\mathbf{G}_{12}^T & -\sigma_1^2[\mathbf{I}\ \ \mathbf{0}]\mathbf{G}_{21}^T-\sigma_2^2\mathbf{G}_{12}[\mathbf{I}\ \ \mathbf{0}]^T\\ -\sigma_1^2\mathbf{G}_{21}[\mathbf{I}\ \ \mathbf{0}]^T-\sigma_2^2[\mathbf{I}\ \ \mathbf{0}]\mathbf{G}_{12}^T & \sigma_2^2\mathbf{I}+\sigma_1^2\mathbf{G}_{21}\mathbf{G}_{21}^T\end{bmatrix}$$

The effects of noises on the filtering process of ADF can be analyzed from the auto-correlation blocks of $\mathbf{R}_{\mathbf{v}_n\mathbf{v}_n}$ by

$$\mathbf{R}_{\mathbf{v}_{ni}\mathbf{v}_{ni}}=\mathbf{R}_{\mathbf{v}_i\mathbf{v}_i}+\sigma_i^2\mathbf{I}+\sigma_j^2\mathbf{G}_{ij}\mathbf{G}_{ij}^T, \quad (10)$$

which shows that the effects of input noises on ADF outputs could be classified into two types: those propagated by direct-paths and those propagated by cross-channel paths. The direct-path noise remains to be white with the same power as input, while the cross-channel noise is colorized by the de-coupling filters $\mathbf{g}_{ij}$'s.

The adaptation of ADF model parameters cannot reduce output noise. However, the influence of background noises on outputs depends on ADF coefficients. At the system output, input noises have the least effects in the trivial cases of $\mathbf{g}_{ij}=0$ where only direct-path noises are present; input noises have the strongest effects when $\mathbf{g}_{ij}$'s are close to the ideal de-coupling filters, under the source mixing condition that the direct paths are close to the cross-coupling paths [4] such that the magnitudes of $\left|G_{ij}(f)\right|$'s are close to 1. In the latter cases, output level of noise energy nearly doubles that of the input noise energy.

The performance of the parameter estimation process of ADF is deteriorated by the presence of noise. It is desirable to analyze such noise effects on adaptation, and to counter these effects for more accurate estimation of cross-coupling filters. The cross-correlation matrices between noisy output vectors are

$$\mathbf{R}_{\mathbf{v}_{ni}\mathbf{v}_{nj}} = \mathbf{R}_{\mathbf{v}_i\mathbf{v}_j} - \sigma_i^2[\mathbf{I} \quad \mathbf{0}]\mathbf{G}_{ji}^T - \sigma_j^2\mathbf{G}_{ij}[\mathbf{I} \quad \mathbf{0}]^T \qquad (11)$$

where clean $\mathbf{v}$ is unavailable and only noisy $\mathbf{v}_n$ is observable. Due to the noise effects shown in the RHS of (11), directly using the criterion of decorrelation between ADF outputs ($\mathbf{v}_n$) for the derivation of filter adaptation procedure is unsuitable.

3.3. Noise-Adapted ADF Algorithm

To improve the performance of ADF adaptations, noise effects need to be excluded from the objective functions for source separation. From Eq. (11), it is obvious that the noise-free output cross-correlation vectors can be estimated by

$$\mathbf{r}_{v_i\mathbf{v}_j} = \mathbf{r}_{v_{ni}\mathbf{v}_{nj}} + \sigma_i^2 g_{ji}(0)\mathbf{e}_1 + \sigma_j^2\mathbf{g}_{ij} \qquad (12)$$

where $\mathbf{e}_1 = [1 \quad 0 \quad \cdots \quad 0]^T$. Therefore, we form a noise-adapted decorrelation criterion function from (12) as follows:

$$J'_{ij} = \tfrac{1}{2}\left(\mathbf{r}_{v_{ni}\mathbf{v}_{nj}} + \sigma_i^2 g_{ji}(0)\mathbf{e}_1 + \sigma_j^2\mathbf{g}_{ij}\right)^T\left(\mathbf{r}_{v_{ni}\mathbf{v}_{nj}} + \sigma_i^2 g_{ji}(0)\mathbf{e}_1 + \sigma_j^2\mathbf{g}_{ij}\right)$$

and similar to (7), the gradient vectors are given by

$$\begin{aligned}\nabla_{\mathbf{g}_{ij}} J'_{ij} &= -\mathbf{R}_{\mathbf{y}_j\mathbf{v}_j} \cdot \left(\mathbf{r}_{v_{ni}\mathbf{v}_{nj}} + \sigma_i^2 g_{ji}(0)\mathbf{e}_1 + \sigma_j^2\mathbf{g}_{ij}\right) \\ &= -\left(\mathbf{R}_{(\mathbf{y}_j+\mathbf{n}_j)\mathbf{v}_{n2}} - \sigma_j^2\mathbf{I}\right)\cdot\left(\mathbf{r}_{v_{ni}\mathbf{v}_{nj}} + \sigma_i^2 g_{ji}(0)\mathbf{e}_1 + \sigma_j^2\mathbf{g}_{ij}\right) \qquad (13)\end{aligned}$$

3.4. Implementation Considerations

The computation of gradient directions in (13) requires matrix-vector multiplications. To reduce complexity, (13) can be approximated by omitting the multiplying correlation matrix $\mathbf{R}_{\mathbf{y}_j\mathbf{v}_j}$. Another simplification comes from the observation that $g_{ij}(0) \approx 0$ because lengths of the cross-channel acoustic path and the direct acoustic path are usually different. The vectors $\mathbf{r}_{v_{ni}\mathbf{v}_{nj}}$'s can be further replaced by their instantaneous estimates $v_{ni}(t)\mathbf{v}_{nj}(t)$. Therefore, the simplified instantaneous implementation of the noise-adapted ADF algorithm becomes

$$\mathbf{g}_{ij}^{(t)} = \mathbf{g}_{ij}^{(t-1)} + \mu(t)\left(v_{ni}(t)\mathbf{v}_{nj}(t) + \sigma_j^2\mathbf{g}_{ij}^{(t-1)}\right) \qquad (14)$$

where the adaptation gain is normalized by the short-time energy estimates of inputs as in [6], i.e.,

$$\mu(t) = 2\gamma \big/ N\left(\sigma_{y_{n1}}^2(t) + \sigma_{y_{n2}}^2(t)\right) \qquad (15)$$

where γ is the adaptation step-size.

4. EXPERIMENTS

4.1. Speech and Acoustic Path Data

The source speech signals were taken from the TIMIT database. Target $s_1(t)$ and jammer $s_2(t)$ were convolutively mixed using impulse responses of acoustic paths measured in real acoustic environment (RWCP [7]). The target speech contained 40 sentences of 4 speakers (faks0, felc0, mdab0, mreb0) and the jammer speech were randomly selected TIMIT sentences excluding those of target speakers. The acoustic paths corresponded to the two speaker locations of 130° and 50° that were approximately 2 meters away from the 15th and 3rd microphones of a circular microphone array (15 cm in radius), in a recording room with the reverberation time of $T_{[60]}$=0.3sec [7]. The mixed speech data were contaminated by varying levels of white Gaussian noises, as illustrated in Fig.1. In all simulations, filter lengths were set to N = 400 and adaptation step-size was chosen to be γ =0.01. Input noise level was measured with respect to the energy of mixture speech in SNRs. The noise power σ_1^2 and σ_2^2 in (14) were assumed to be known, where in practice they can be measured during speech inactive periods.

4.2. Comparison of Convergence Rates

The performance of the proposed algorithm in adaptive estimation of de-coupling filters was evaluated by convergence rate, which was measured in terms of normalized ADF filter errors at SNR levels of 25dB, 15dB, and 5dB, shown in Fig.2 (a)-(c). The results show that the noise-adapted ADF algorithm outperformed baseline ADF algorithm, and as SNR decreased the convergence rate improvement became more significant.

4.3. Target-to-Interference Ratio

Associated with error reduction in adaptive filter estimation, the noise-adapted ADF algorithm provided improvement to gains of TIR (target-to-interference ratio) over the baseline algorithm. Noise energy was not included in the calculation of TIRs to focus on the performance of speech separation only. The output-to-input TIR gains from baseline and noise-adapted ADFs under different input noise conditions are shown in Table 1. The initial TIR values at ADF inputs were 0.53dB and -0.55dB, respectively.

***Table 1**. Comparison of target-to-interference ratios (dB)*

SNR	Baseline ADF		Noise-Modifed ADF	
	TIR Gain $_{v1}$	TIR Gain $_{v2}$	TIR Gain$_{v1}$	TIR Gain$_{v2}$
0dB	3.72	3.36	7.02	6.89
5dB	5.01	4.58	7.52	7.35
10dB	6.13	5.78	7.76	7.56
15dB	7.00	6.73	7.90	7.71
20dB	7.56	7.34	7.97	7.80
25dB	7.84	7.66	8.00	7.84
30dB	7.96	7.79	8.02	7.86

4.4. Phone Recognition Accuracy

Finally, the effectiveness of the proposed noise-adapted ADF algorithm was compared with baseline algorithm on the task of phone recognition. Speech feature vector contained 13 cepstral coefficients and their first and second-order time derivatives. Acoustic model had 39 context-independent phone units, each unit modeled by 3 emission states of HMM, and each state had a size-8 Gaussian mixture density. Phone bigram was used as "language model." Both training and test data were processed by cepstral mean subtraction.

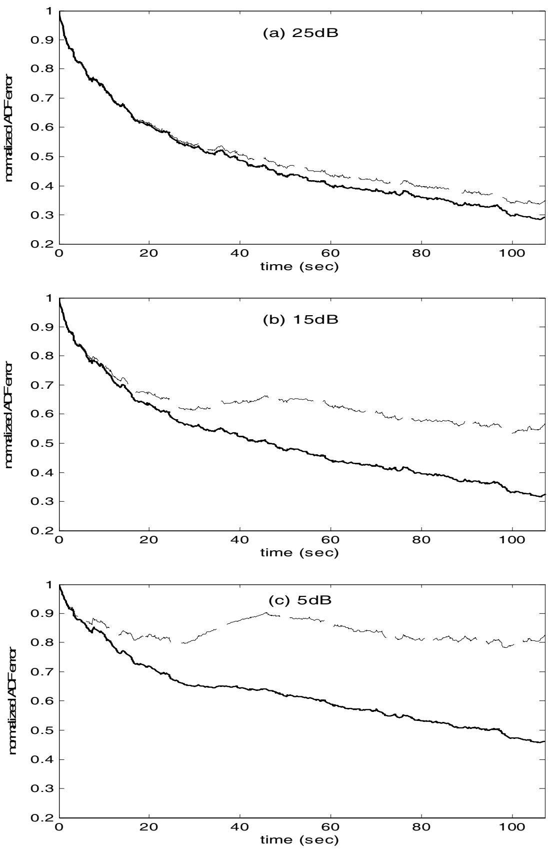

Figure2. Comparison of Convergence Rates (solid: noise-adapted ADF; dashed: baseline ADF)

For purpose of reference, phone recognition accuracies were first evaluated for various noise corrupted cases, including speech mixture ($y_{n1}(t)$), speech mixture $y_{n1}(t)$ separated by ideal de-coupling filters, target source ($s_{n1}(t)$), and directly passed target source $s_1(t)*h_{11}(t)+n(t)$. The results are shown in Table 2. Since noisy speech mixtures contained two sources of comparable energies, while noisy sources contained only a single speech source, the SNR of the latter case was set to be 3dB lower than that of the former. The accuracies of ideally separated noisy mixtures (column 3) were lower than those of directly passed noisy target source (column 6) in most cases, consistent with the analysis of noise effects in Section 3.2.

The performance of blind speech source separation was next evaluated and the ADF de-coupling filters estimated at different levels of input noises were used to separate speech mixtures without additive noise. Phone accuracy results obtained from noise-adapted ADF algorithms were consistently higher than those from baseline algorithm, as shown in Table 3, and the improvement of the proposed method over the baseline ADF was significant at low SNRs.

***Table 2**. Phone accuracy (%) under reference conditions*

SNR	Mixture $y_{n1}(t)$		SNR	Source $s_1(t)$	
	$y_{n1}(t)$	Ideally Separated		$s_{n1}(t)$	$s_1(t)*h_{11}(t)+n(t)$
0dB	10.9	9.9	-3dB	11.3	9.2
5dB	17.0	17.0	2dB	19.7	16.4
10dB	19.4	21.6	7dB	25.3	23.1
15dB	20.6	26.2	12dB	32.8	27.4
20dB	23.9	32.3	17dB	41.0	33.1
25dB	26.3	37.4	22dB	50.4	42.3
30dB	28.4	42.3	27dB	57.9	50.0
clean	30.4	52.0	clean	68.0	59.7

***Table 3**. Comparison of phone accuracy (%) on clean speech mixtures separated by ADF de-coupling filters obtained from baseline and noise-adapted ADF*

SNR	Baseline (%)	Noise-Adapted (%)
0dB	32.9	39.6
5dB	35.8	40.6
10dB	37.3	41.9
15dB	39.3	43.5
20dB	42.3	43.5
25dB	42.4	43.8
30dB	43.5	44.0

5. CONCLUSION

The vector formulation of ADF provides a clear insight into the effects of noises on ADF systems and leads to the derivation of a new noise-adapted ADF algorithm with a simple form. The proposed algorithm significantly improved the performance of ADF de-coupling filter estimation, and achieved enhanced separation effects for speech sources. Although noise impacts on de-coupling filter adaptation were reduced by the new method, de-noising techniques for ADF outputs need to be explored based on current noise analysis, which will be a topic of future study.

REFERENCES

[1] E. Weinstein, M. Feder, and A. V. Oppenheim, “Multi-channel signal separation by decorrelation,” *IEEE Trans. on SAP*, Vol. 1, pp. 405-413, Oct. 1993.
[2] A. Hyvarinen, J. Karhunen, and E. Oja, “Independent Component Analysis,” John Wiley, pp.361-370, 2001.
[3] Y. Zhao and R. Hu, “Fast convergence speech source separation in reverberant acoustic environment,” Proc. of ICASSP 2004, Vol. III, pp.897-900, 2004
[4] K. Yen Y. and Y. Zhao, “Co-channel speech separation in the presence of correlated and uncorrelated noises,” *ESCA Eurospeech* 1999, pp. 2587-2589.
[5] H. Sahlin and H. Broman, “MIMO signal separation for FIR channels: a criterion and performance analysis,” *IEEE Trans. on SP*, Vol.48, No.3, Mar. 2000.
[6] K. Yen Y. and Y. Zhao, “Adaptive co-channel speech separation and recognition,” *IEEE Trans. on SAP*, Vol.7, No.2, pp.138-151, 1999.
[7] RWCP Sound Scene Database in Real Acoustic Environments, ATR Spoken Language Translation Research Lab, Japan 2001.
[8] H. Robbins and S. Monro, “A stochastic approximation method,” Annals of Math.l Stat., Vol. 22, pp. 400-407, 1951.