# Adaptive Speech Enhancement for Speech Separation in Diffuse Noise

*Rong Hu and Yunxin Zhao*

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

## Abstract

An adaptive enhancement method is proposed to improve recognition accuracy on the outputs of blind speech separation (BSS) system based on adaptive decorrelation filtering (ADF) in diffuse noise. A divide and conquer strategy is taken to deal with the noise effects on both system adaptation and ADF outputs. First, fast noise compensation (NC) is performed for filter adaptation, forcing ADF to focus on the task of separation; then, output noises are reduced by conventional speech enhancement, such as spectral subtraction or subspace methods. To make stationary-noise reduction techniques fit for output noises with time-varying properties caused by ADF adaptations, a fast adaptive procedure is developed to map known stationary input noise statistics to output. Separation and recognition experiments were conducted for both real and simulated diffuse noises, based on TIMIT speech data and impulse response data from a room with reverberation time $T_{60} = 0.3sec$. The proposed techniques significantly improved phone recognition accuracy of ADF results.

**Index Terms**: speech enhancement, blind speech source separation, diffuse noise.

## 1. Introduction

Combating the adverse circumstances brought by speaker interferences and environmental noises have been challenging tasks for decades in hands-free automatic speech recognition (ASR) and speech communication. A variety of blind source separation (BSS) and independent component analysis (ICA) [1] algorithms have been proposed for the separation of interfering speech. For the reduction of noise effects, a vast number of speech enhancement algorithms already exist.

For practical speech applications of BSS methods in noise, the difficulties are two folds: 1) the working conditions of BSS algorithms may be affected by the presence of noise, resulting in degraded separation performances; 2) a BSS algorithm itself, aiming mainly at source separation, is limited in ability to suppress diffuse noise. For the first problem, the general approach to improve separation performance in noisy BSS is doing "bias removal" [1]. How the separation performances are affected by noises depends on specific algorithms, and some noise compensation (NC) algorithms, e.g., [2] and [3], were proposed for corresponding separation models. For the second problem of output noise suppression, the mechanism similarities between BSS and adaptive null beamformer were established in [4] and [5]. According to [6], these "spatial inverse" type of approaches are only suitable for directional interferences, not for omni-directional ambient noises. Therefore, efforts should also be devoted to the reduction of output noise after separation. However, conventional speech enhancement algorithms for stationary noises cannot be applied directly. This is because the adaptation of separation parameters makes the output noise properties time varying. Such variation becomes severer when the mixing acoustic paths changes, for example, when speaker moves.

In previous studies [7, 8], we significantly improved convergence rate and separation performance for the adaptive decorrelation filtering (ADF) [9] separation model in noise-free speech application, and algorithms of a fast noise compensation (NC) and fast ADF (FADF) adaption were developed in [3] to improve speech separation performances in diffuse noises. However, as indicated by the above analysis, it is difficult for ADF to remove both the output noise and the adaptation bias together, all by itself. Therefore, we propose a divide and conquer strategy for application of ADF in noise, treating problems of adaptation compensation and speech enhancement separately. One potential solution is to remove the noise from speech inputs, such as the subspace processing [10] performed prior to ADF separation; but such a noise reduction can not improve the condition for subsequent source separation, due to the distortions introduced by speech enhancement pre-processing. In this paper we propose a post-processing adaptive enhancement technique. First, the effective block-wise NC-FADF [3] algorithm is applied to perform speech separation. Then, as separation filters change over time, output noise properties are tracked by a fast adaptive mapping procedure. Finally, the adaptively estimated output noise statistics are used for speech enhancement. Speech separation and phone recognition experiments were conducted to evaluate the proposed separation and enhancement techniques.

## 2. ADF model and noise compensation

### 2.1. Noisy ADF separation model

The following notations are used in our discussions: vector variables are in bold lower case, matrices are in bold upper case, superscript $^T$ is for transposition, $\mathbf{I}$ is identity matrix, $E\{\}$ is for expectation, and * for convolution, $N$ is filter length and block length, $m$ is block index. Speech and noise signal vectors contain $N$ consecutive samples up to current time $t$; their $(2N-1)$-point counterparts are marked with tilde. The cross-correlation vector between a scalar $a$ and a vector $\mathbf{b}$ is denoted as $\mathbf{r}_{a\mathbf{b}} = E\{a\mathbf{b}\}$, and the correlation matrix formed by vectors $\mathbf{a}$ and $\mathbf{b}$ is defined as $\mathbf{R}_{\mathbf{ab}} = E\{\mathbf{a}\mathbf{b}^T\}$.

Figure 1 shows the ADF noisy speech separation model with filters, $\mathbf{g}_{ij} = [g_{ij}(0), ..., g_{ij}(N-1)]^T$, $i, j = 1, 2$, $i \neq j$. By formulating the system parameters into $2N \times (4N-2)$ filter matrix

$$\mathbf{G} = \begin{bmatrix} [\mathbf{I}_N \ \mathbf{0}_{N\times(N-1)}] & -\mathbf{G}_{12} \\ -\mathbf{G}_{21} & [\mathbf{I}_N \ \mathbf{0}_{N\times(N-1)}] \end{bmatrix}, \quad (1)$$

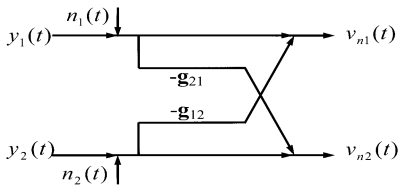


Figure 1: *Noisy ADF separation system*

the ADF system I/O relations [11] can be described as

$$\mathbf{v}_n = \mathbf{G}(\tilde{\mathbf{y}} + \tilde{\mathbf{n}}), \tag{2}$$

where $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1^T(t), \tilde{\mathbf{y}}_2^T(t)]^T$ and $\tilde{\mathbf{n}} = [\tilde{\mathbf{n}}_1^T(t), \tilde{\mathbf{n}}_2^T(t)]^T$ are $(4N-2)\times 1$ vectors of clean speech mixture and noise, respectively, with $\tilde{\mathbf{y}}_i = [y_i(t), \ldots, y_i(t-2N+2)]^T$, $\tilde{\mathbf{n}}_i = [n_i(t), \ldots, n_i(t-2N+2)]^T$, $i = 1, 2$. The $k$-th row of the $N\times(2N-1)$ Toeplitz matrix $\mathbf{G}_{ij}$ is $[\mathbf{0}_{1\times(k-1)}, \mathbf{g}_{ij}^T, \mathbf{0}_{1\times(N-k)}]$, $k = 1, ..., N$.

### 2.2. ADF adaptation and output noise components

The basic ADF adaptation given in [9] for clean mixtures is

$$\mathbf{g}_{ij}(t+1) = \mathbf{g}_{ij}(t) + \mu_{ij}(t) v_i(t) \mathbf{v}_j(t), \tag{3}$$

where $\mu_{ij}(t)$ is the step size. It was also derived by minimizing the cross-correlation objective functions $J_{ij} = \frac{1}{2}\mathbf{r}_{v_i\mathbf{v}_j}^T \mathbf{r}_{v_i\mathbf{v}_j}$ [11] under some approximate assumptions. From (2), the noisy output correlation matrix contains speech-only and noise-only components, i.e., $\mathbf{R}_{\mathbf{v}_n\mathbf{v}_n}$=$\mathbf{R}_{\mathbf{vv}} + \mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}$, where the clean speech $\mathbf{v} = [\mathbf{v}_1^T(t), \mathbf{v}_2^T(t)]^T$ and the noise component $\boldsymbol{\eta} = [\boldsymbol{\eta}_1^T(t), \boldsymbol{\eta}_2^T(t)]^T$, and $\boldsymbol{\eta}$ satisfies the I/O relations of correlation vectors

$$\mathbf{r}_{\eta_i\boldsymbol{\eta}_j} = \mathbf{r}_{n_i\mathbf{n}_j} - \mathbf{G}_{ji}\mathbf{r}_{n_i\tilde{\mathbf{n}}_i} - \mathbf{R}_{\mathbf{n}_j\mathbf{n}_j}\mathbf{g}_{ij} + \mathbf{G}_{ji}\mathbf{R}_{\tilde{\mathbf{n}}_i\mathbf{n}_j}\mathbf{g}_{ij}, \tag{4}$$

$$\mathbf{r}_{\eta_i\boldsymbol{\eta}_i} = \mathbf{r}_{n_i\mathbf{n}_i} - \mathbf{G}_{ij}\mathbf{r}_{n_i\tilde{\mathbf{n}}_j} - \mathbf{R}_{\mathbf{n}_i\mathbf{n}_j}\mathbf{g}_{ij} + \mathbf{G}_{ij}\mathbf{R}_{\tilde{\mathbf{n}}_j\mathbf{n}_j}\mathbf{g}_{ij}. \tag{5}$$

It can be seen that as filters $\mathbf{g}_{ij}$ evolve, the noise properties at ADF output vary. The cross-correlation term (4) causes a bias in filter adaptation and it should be compensated. The auto-correlation (5) represents ADF output noise statistics and it needs to be removed to enhance the separated speech.

### 2.3. Noise compensated fast ADF

The compensation problem is treated first. The time varying noise bias of (4) can be subtracted from the noisy objective function $J_{n_{ij}} = \frac{1}{2}\mathbf{r}_{v_{n_i}\mathbf{v}_{\mathbf{n}_j}}^T \mathbf{r}_{v_{n_i}\mathbf{v}_{\mathbf{n}_j}}$, leading to noise compensated adaptation. We use the block-wise NC-FADF derived in [3]. Let the start time for the $m$-th block be $t_m$, the update of filters from current block ($m$-th) to the next block ($(m+1)$-th) is given as

$$\mathbf{g}_{ij}^{m+1} = \mathbf{g}_{ij}^m + \mu_{ij}^m N(\hat{\mathbf{r}}_{v_{n_i}\mathbf{v}_{\mathbf{n}_j}}^m - \hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_j}^m), \tag{6}$$

$$\hat{\mathbf{r}}_{v_{n_i}\mathbf{v}_{\mathbf{n}_j}}^m = \frac{1}{N}\sum_{k=0}^{N-1} v_{n_i}(t_m+k)\mathbf{v}_{n_j}(t_m+k), \tag{7}$$

$$\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_j}^m = \hat{\mathbf{r}}_{n_i\mathbf{n}_j} - \mathbf{a}_{ij}^m - \mathbf{b}_{ij}^m + \mathbf{c}_{ij}^m, \tag{8}$$

where both (7) and (8) are implemented with FFT-based fast algorithms [12]. The computations of the output bias terms in (8), $\mathbf{a}_{ij}^m = \mathbf{G}_{ji}^m\hat{\mathbf{r}}_{n_i\tilde{\mathbf{n}}_i}$, $\mathbf{b}_{ij}^m = \hat{\mathbf{R}}_{\mathbf{n}_j\mathbf{n}_j}\mathbf{g}_{ij}^m$, $\mathbf{c}_{ij}^m = \mathbf{G}_{ji}^m\mathbf{d}_{ij}^m$, and $\mathbf{d}_{ij}^m = \hat{\mathbf{R}}_{\tilde{\mathbf{n}}_i\mathbf{n}_j}\mathbf{g}_{ij}^m$, have corresponding convolution representations. Their vector components are listed below:

$$\begin{aligned}
a_{ij}^m(k) &= g_{ji}^m(n) * \xi_{ij}^a(n)|_{n=2N-2-k},\\
\xi_{ij}^a(n) &= \hat{r}_{n_i\tilde{n}_i}(2N-2-n),\\
c_{ij}^m(k) &= g_{ji}^m(n) * \xi_{ij}^c(n)|_{n=2N-2-k},\\
\xi_{ij}^c(n) &= d_{ij}^m(2N-2-n).\\
b_{ij}^m(k) &= g_{ij}^m(n) * \xi_{ij}^b(n)|_{n=k+N-1}\\
\xi_{ij}^b(n) &= \hat{r}_{n_j\tilde{n}_j}(|n-N+1|),\\
d_{ij}^m(k) &= g_{ij}^m(n) * \xi_{ij}^d(n)|_{n=k+N-1}\\
\xi_{ij}^d(n) &= \hat{r}_{n_i\tilde{n}_j}(N-1-n).
\end{aligned}$$

The block step-size $\mu_{ij}^m$ in (6) is given by

$$\mu_{ij}^m = \mu^m \cdot \hat{\sigma}_{v_j}^2(m)/\hat{\sigma}_{av}^2(m), \tag{9}$$

$$\mu^m = \gamma/\big(N(\sigma_{y_{n1}}^2(m) + \sigma_{y_{n2}}^2(m))\big), \tag{10}$$

$$\hat{\sigma}_{av}^2(m) = \frac{1}{2}\left(\hat{\sigma}_{v_1}^2(m) + \hat{\sigma}_{v_2}^2(m)\right), \tag{11}$$

where the constant gain factor $\gamma$ ($0 < \gamma < 1$) controls convergence speed, $\sigma_{y_{ni}}^2(m)$'s are short-term input powers, $\hat{\sigma}_{av}^2(m)$ is the estimated average output speech power, and the output speech power $\hat{\sigma}_{v_j}^2(m)$ is estimated by

$$\hat{\sigma}_{v_j}^2(m) \approx \mathbf{v}_{n_j}^T\mathbf{v}_{n_j}/N - \hat{r}_{n_j}(0) + 2\mathbf{g}_{ji}^{mT}\hat{\mathbf{r}}_{n_j\mathbf{n}_i} - \mathbf{g}_{ji}^{mT}\mathbf{b}_{ji}^m. \tag{12}$$

The block-wise computation of ADF outputs are implemented with overlap-add based fast filtering [12]. The algorithm is referred to as FADF if NC-FADF omits the compensation term $\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_j}^m$ in (6) and keeps the first term in R.H.S. of (12) only. For more details of FFT-based computation of $\hat{\mathbf{r}}_{v_{n_i}\mathbf{v}_{\mathbf{n}_j}}^m$, $\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_j}^m$, and step-size $\mu_{ij}^m$ please refer to [3].

## 3. Adaptive enhancement of separated speech

### 3.1. Tracking of ADF Output Noise Auto-Correlations

Although NC-FADF improves the separation performance of ADF, the separation results $\mathbf{v}_{\mathbf{n}_i}$'s are still contaminated by noise. Additional speech enhancement processing should be integrated with ADF at each output to reduce noise. Usually, speech enhancement algorithms require the knowledge of properties of the noise that needs to be removed. For online separation applications, we need to track the time-varying output noise properties as filters evolve from block to block. The idea is based on fast computation of (5). Similar to the derivations of (8), we obtain auto-correlation of ADF output noise for the $m$-th block

$$\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_i}^m = \hat{\mathbf{r}}_{n_i\mathbf{n}_i} - \mathbf{a}_{ii}^m - \mathbf{b}_{ii}^m + \mathbf{c}_{ii}^m, \tag{13}$$

where $\mathbf{a}_{ii}^m = \mathbf{G}_{ij}^m\hat{\mathbf{r}}_{n_i\tilde{\mathbf{n}}_j}$, $\mathbf{b}_{ii}^m = \hat{\mathbf{R}}_{\mathbf{n}_i\mathbf{n}_j}\mathbf{g}_{ij}^m$, $\mathbf{c}_{ii}^m = \mathbf{G}_{ij}^m\mathbf{d}_{ii}^m$, and $\mathbf{d}_{ii}^m = \hat{\mathbf{R}}_{\tilde{\mathbf{n}}_j\mathbf{n}_j}\mathbf{g}_{ij}^m$. Since input noise is stationary, its auto and cross correlations can be assumed to be known, or measured *a priori*. The fast mappings from input correlations to output, depending only on current system parameters $\mathbf{g}_{ij}^m$'s and $\mathbf{G}_{ji}^m$'s, are implemented as the fast convolutions of the following signal sequences:

$$a_{ii}^m(k) = g_{ij}^m(n) * \xi_{ii}^a(n)|_{n=2N-2-k}, \tag{14}$$

$$\xi_{ii}^a(n) = \hat{r}_{n_i\tilde{n}_j}(2N-2-n), \tag{15}$$

$$c_{ii}^m(k) = g_{ij}^m(n) * \xi_{ii}^c(n)|_{n=2N-2-k}, \quad (16)$$

$$\xi_{ii}^c(n) = d_{ij}^m(2N-2-n), \quad (17)$$

$$b_{ii}^m(k) = g_{ij}^m(n) * \xi_{ii}^b(n)|_{n=k+N-1}, \quad (18)$$

$$\xi_{ii}^b(n) = \hat{r}_{n_i\tilde{n}_j}(N-1-n), \quad (19)$$

$$d_{ii}^m(k) = g_{ij}^m(n) * \xi_{ii}^d(n)|_{n=k+N-1}, \quad (20)$$

$$\xi_{ii}^d(n) = \hat{r}_{n_j\tilde{n}_j}(|N-1-n|). \quad (21)$$

### 3.2. Enhancement of separated speech

Utilizing the adaptively estimated noise statistics $\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_i}^m$, a lot of speech enhancement algorithms can be considered for post enhancing ADF outputs. Two representative examples of single channel speech enhancement methods, spectral subtraction and generalized subspace method are compared in the current work for the reduction of ADF output noises in each block. The basic spectral subtraction approach is to be tested for two reasons: 1) it is simple to implement and suitable for fast algorithm; 2) it may, to some extent, provide us with a performance lower-bound, among speech enhancement algorithms with higher complexities.

#### 3.2.1. Spectral subtraction

The spectral subtraction algorithm [13] is taken in the basic form. For block $m$, the estimate of clean speech amplitude is given by

$$\left|\hat{V}_i^m(f)\right| = \begin{cases} \left(\left|V_{n_i}^m(f)\right|^2 - E\left\{|\Phi_i^m(f)|^2\right\}\right)^{\frac{1}{2}}, if \frac{E\left\{|\Phi_i^m(f)|\right\}^2}{\left|V_{n_i}^m(f)\right|^2} \le 1 \\ 0, \qquad otherwise, \end{cases} \quad (22)$$

and the phase of $\hat{V}_i^m(f)$ equals that of $V_{n_i}^m(f)$. The noise power spectral density required in (22) at each block $m$ is directly transformed from the adaptive estimate of vector (13) as

$$E\left\{|\Phi_i^m(f)|^2\right\} = \boldsymbol{FFT}\left(E\left\{\boldsymbol{\phi}_{\boldsymbol{\eta}_i}^m\right\}\right), \quad (23)$$

where, $\boldsymbol{\phi}_{\boldsymbol{\eta}_i}^m$ is the short-term correlation vector. According to the definition in [13], for the signal vector of length $N$ supported only in $m$-th block,

$$\begin{aligned} \phi_{\boldsymbol{\eta}_i}^m(n) &= \textstyle\sum_{t=t_m}^{t_m+N-1} \eta_i(t)\eta_i(t-n) \\ &= N \cdot \tfrac{1}{N} \textstyle\sum_{t=t_m}^{t_m+N-1} \eta_i(t)\eta_i(t-n). \end{aligned}$$

Therefore, the time summation vector $\boldsymbol{\phi}_{\boldsymbol{\eta}_i}^m$ is related to the average vector $\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_i}^m$ by

$$E\left\{\boldsymbol{\phi}_{\boldsymbol{\eta}_i}^m\right\} = N\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_i}^m. \quad (24)$$

No further processing or modification is taken to suppress musical noise, because it is usually achieved at the cost of increased residual noise and because we care more about machine recognition rather than human perception.

#### 3.2.2. Generalized subspace appraoch

For subspace method, we choose the time domain constrained (TDC) type of generalized subspace (GSub) approach proposed by Hu and Loizou [14], because of its ability to handle colored noise. TDC-GSub processing is applied to every block of ADF outputs. This method requires the noise auto-correlation matrix $\mathbf{R}_{\boldsymbol{\eta}_i\boldsymbol{\eta}_i}^m$. It can be constructed by forming a symmetric Toeplitz matrix from the output auto-correlation vector obtained in (13). In fact, $\hat{\mathbf{r}}_{\eta_i\boldsymbol{\eta}_i}^m$ constitutes the first column and the first row of $\mathbf{R}_{\boldsymbol{\eta}_i\boldsymbol{\eta}_i}^m$.

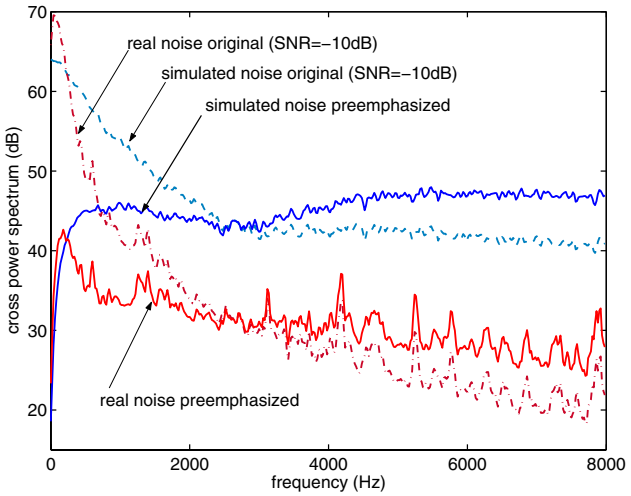


Figure 2: *Noise cross power spectra.*

Another type of information the TDC-GSub algorithm takes is the auto-correlation matrix of noisy ADF output, $\mathbf{R}_{\mathbf{v}_{n_i}\mathbf{v}_{n_i}}^m$, which is estimated from ADF outputs of the current block. The SNR thresholds used for eigen-domain filtering are chosen to be the same as in [14].

## 4. Simulation experiments and results

### 4.1. Experimental setup

Speech sentences in TIMIT database were used as clean sources. Speech mixtures were generated by convolutively mixing sources using real acoustic impulse responses measured in a room with reverberation time $T_{[60]}$=0.3$sec$ [15]. Two microphones (#13 and #15) were mounted in a circular microphone array of radius 15cm . The target speech had 40 sentences from 4 speakers (faks0, felc0, mdab0, mreb0) approximately $2m$ away from the microphones.

Noise data included both cases of simulated and real recorded diffuse noises. For the simulated case, noise were designed to be speech-shaped by the following procedure:

$$n_1(t) = \beta_1 \sum_{k=1}^{P_1} a_k^{(1)} n_1(t-k) + (1-\beta_1)n_2(t) + \varepsilon_1(t), \quad (25)$$

$$n_2(t) = \beta_2 \sum_{k=1}^{P_2} a_k^{(2)} n_2(t-k) + (1-\beta_2)n_1(t) + \varepsilon_2(t), \quad (26)$$

where $\varepsilon_i(t)$'s are white Gaussian excitations, $\beta_1$=0.65, $\beta_2$=0.6, $P_1$=2, $P_2$=3, and $a_k^{(i)}$'s are linear prediction coefficients (LPC) estimated from clean TIMIT data. Real diffuse noises were recorded with a pair of omnidirectional microphones placed 21$cm$ apart on a conference table in the middle of a computer lab, where an air-conditioning and ventilation system and 8 desktop workstations were working simultaneously. With stationary assumption on the input diffuse noise, we estimated its correlation properties from the 5-second segment of noise-only data preceding the 1st speech sentence. The estimates of cross power spectra for both types of noises are illustrated in Figure 2.

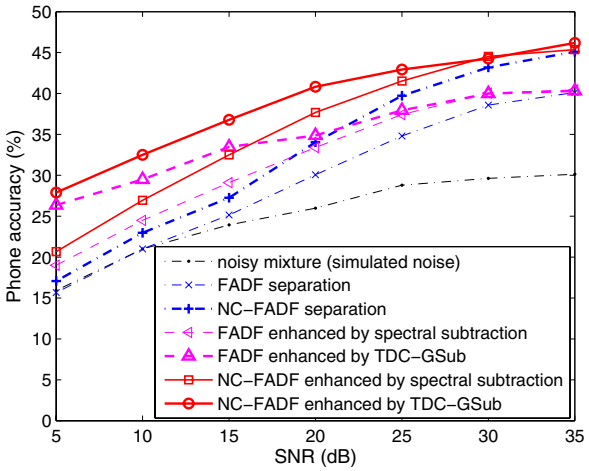


Figure 3: *Phone accuracies (simulated noise)*

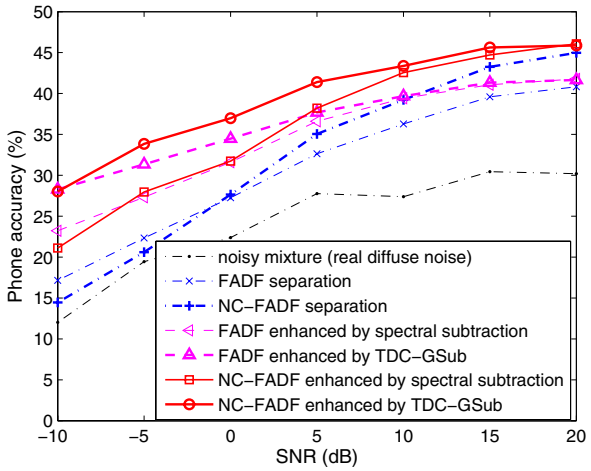


Figure 4: *Phone accuracies (real diffuse noise)*

### 4.2. Speech enhanced ADF and phone recognition

Speech separation experiments were conducted to evaluate the proposed method, using both NC-FADF and FADF, with and without adaptive speech enhancements. In all cases, preemphasis (1-$z^{-1}$) was applied to mixtures to flatten the long-term spectrum of speech for faster convergence [8]. Since SNR was altered by preemphasis differently for simulated (decreased by 3dB) and real diffuse (increased by 12dB) noises, the range of initial SNR's were chosen differently for these two cases so that the target speech was in the same SNR's after preemphasis. Block-length was $N = 400$ and adaptation step-size was set to be $\gamma = 0.01$; FFT length was 1024. After adaptive online noise reduction, a stable deemphasis $1/(1 - 0.98z^{-1})$ was applied to the enhanced speech. Phone recognition were performed for noisy mixture, noisy separated speech, and enhanced separated speech. Phone accuracy results in both noise cases are shown in Figures 3 and 4, respectively.

## 5. Conclusion

The proposed adaptive enhancement techniques significantly improved the phone recognition accuracy of the ADF separation outputs. The combination of NC-FADF with TDC-GSub achieved highest performance. At low SNR's, the gains of phone accuracy are mainly provided by speech enhancement; at high SNR's, the improvement of accuracy comes mainly from better noise compensated speech separation.

## 6. References

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley, 2001.

[2] R. Aichner and H. Buchner and W. Kellermann, "Convolutive blind source separation for noisy mixtures," in *Proc. CFA/DAGA*, 2004, *http://www.lnt.de/LMS/publications/web/lnt2004_2.pdf*.

[3] R. Hu and Y. Zhao, "Noise-compensated fast adaptive decorrelation filtering for competing speech separation," *Submitted to IEEE Sig. Proc. Letters*.

[4] S. Araki, S. Makino, R. Mukai, and H. Saruwatai, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," in *Proc. Eurospeech*, Sept. 2001, vol. 4, pp. 2595–2598.

[5] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. SAP*, vol. 11, pp. 109–116, Mar. 2003.

[6] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. SAP*, vol. 8, no. 5, pp. 497–507, Sept. 2000.

[7] R. Hu and Y. Zhao, "Variable step size adaptive decorrelation filtering for competing speech separation," in *Eurospeech'05*, Sept. 2005, vol. I, pp. 2297–2300.

[8] Y. Zhao, R. Hu, and X. Li, "Speedup convergence and reduce noise for enhanced speech separation and recognition," *IEEE Trans. SAP, to appear, July, 2006*.

[9] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multichannel signal separation by decorrelation," *IEEE Trans. SP*, vol. 43, pp. 405–413, Oct. 1993.

[10] K. Yen, J. Huang, and Y. Zhao, "Co-channel speech separation in the presence of correlated and uncorrelated noises," in *ESCA Eurospeech'99*, 1999, pp. 2587–2589.

[11] R. Hu and Y. Zhao, "Adaptive decorrelation filtering algorithm for speech source separation in uncorrelated noises," in *ICASSP*, 2005, vol. I, pp. 1113–1116.

[12] A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.

[13] J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. of the IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[14] Y. Hu and C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. SAP*, vol. 11, no. 4, pp. 334–341, July 2003.

[15] "RWCP sound scene database in real acoustic environments," ATR Spoken Language Translation Research Lab, Japan, 2001.