F-SEONS: A Second-Order Frequency-Domain Algorithm for Noisy Convolutive Source Separation

Inseon Jang, Kyeongok Kang
Radio and Broadcasting Laboratory
ETRI, Korea
Email: {jinsn,kokang}@etri.re.kr

Sangki Kim, Seungjin Choi
Department of Computer Science
POSTECH, Korea
Email: {ihistory,seungjin}@postech.ac.kr

***Abstract*— We present a frequency-domain method of *noisy* convolutive source separation, where we extends the SEONS algorithm [2] that jointly exploits the nonstationarity and temporal structure of sources. Thus, the method is called *F-SEONS*, implying Frequency-domain SEONS. Unlike most of existing methods of convolutive source separation, we consider the case of noisy data and show that our F-SEONS algorithm identify multivariate FIR channels in a robust way. In addition, we employ an H_∞ filtering method in order to further suppress the sensor noise. Numerical experiments with comparing to other methods, confirm the high performance of our proposed method.**

I. Introduction

The goal of source separation is to recover unknown sources without resorting to any knowledge of the propagation channel characteristics, given only sensor signals that are modelled as convolutive mixtures of mutually independent sources. Source separation has been recognized as a promising technique to segregate speech signals such that automatic speech recognition systems are able to decode each speech signal in the presence of interfering sounds or noise sources.

A number of methods for source separation of convolved mixtures, have been developed. Methods are categorized into two main approaches: (1) time-domain methods; (2) frequency-domain methods. An exemplary time-domain method is to use a dynamic recurrent network for which several associated learning algorithms were developed (for instance, see [3], [12] and references therein). A basic idea in frequency-domain methods, is to use the short-time Fourier transform (STFT) in order to transform the convolution into multiplication so that convolutive source separation involves with instantaneous mixtures at each frequency bin [4], [10].

Most of existing methods neglected the sensor noise in the model of convolutive mixtures, because taking the additive noise into account, makes the problem much harder. SEONS [2] is an algebraic method of source separation which exploits the nonstationarity and temporal structure of sources, that was originally developed for the case of instantaneous mixtures. SEONS includes SOBI [1] as its special case and it was shown in [2] that SEONS was insensitive to additive white noise, since only time-delayed correlation matrices were exploited. In this paper, we extend SEONS to the case of convolutive mixtures, which leads to the frequency-domain SEONS, *F-SEONS*. As in [8], F-SEOSN exploits the nonstationarity in the frequency-domain. In addition, F-SEONS also exploit the temporal structure of sources, in the same way as SEONS, but in the frequency-domain. We show that F-SEONS is superior to most of existing methods, in the presence of additive white noise.

Even though F-SEONS algorithm identify the multivariate FIR channel in a robust way, restored speech signals are still contaminated by noise, because the reconstruction is done by a linear transform. In order to further suppress the noise effect in segregated speech signals, we employ an H_∞ filtering method [9] which was shown to be better than traditional Wiener or Kalman filtering techniques.

II. Problem Formulation

Let $\boldsymbol{x}(t) \in \mathbb{R}^n$ be a vector whose elements $x_i(t)$ are the signals measured at an array of microphones. Denote the n-dimensional original source vector by $\boldsymbol{s}(t)$. Taking the delay and multipath effect into account, the ith microphone signal, $x_i(t)$, is modelled by

$$x_i(t) = \sum_{j=1}^{n} \sum_{\tau=0}^{P} h_{ij}(\tau) s_j(t-\tau) + v_i(t), \tag{1}$$

where $h_{ij}(t)$ is the room impulse response between the jth source and the ith microphone and $v_i(t)$ is the additive sensor noise.

Let us define $H_{ij}(z^{-1}) = \sum_{\tau=0}^{P} h_{ij}(\tau) z^{-\tau}$ where z^{-1} is the time-shift operator, i.e., $z^{-1}s_i(t) = s_i(t-1)$. Then we can rewrite (1) as

$$x_i(t) = \sum_{j=1}^{n} H_{ij}(z^{-1}) s_j(t) + v_i(t). \tag{2}$$

In a compact form, we have

$$\boldsymbol{x}(t) = \boldsymbol{H}(z^{-1})\boldsymbol{s}(t) + \boldsymbol{v}(t), \tag{3}$$

where $\boldsymbol{H}(z^{-1})$ is a polynomial matrix whose (i,j)-element is $H_{ij}(z^{-1})$.

Convolutive source separation aims at recovering unknown sources, $\{s_i(t)\}$, without knowing the multivariate impulse response, $\boldsymbol{H}(z^{-1})$, given only measurement data, $\{\boldsymbol{x}(t)\}$. A frequency-domain approach to the convolutive source separation, is to transform the model (1) into the frequency domain, in order to solve simultaneously a separation problem at each frequency bin. A linear convolution can be approximated by a circular convolution if the frame size T of the DFT is much larger than the channel length P. Then we can write (3) approximately

$$\boldsymbol{x}(\omega, t) \approx \boldsymbol{H}(\omega)\boldsymbol{s}(\omega, t) + \boldsymbol{v}(\omega, t). \quad (4)$$

Note that the solution for each frequency would have an arbitrary permutation, due to the indeterminacy of source separation. Hence, a special care is required, in order to take care of the permutation problem [6], [8].

The goal of convolutive source separation is to estimate the mixing filter $\boldsymbol{H}(\omega)$ or its inverse $\boldsymbol{W}(\omega) = \boldsymbol{H}^{-1}(\omega)$ such that $\boldsymbol{G}(\omega) = \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)$ becomes a generalized permutation matrix.

III. F-SEONS

A main idea of SEONS [2] is to find an unitary joint approximate diagonalizer of a set of time-varying time-delayed correlation matrices. Since time-delayed correlations of white noise are zeros, SEONS allows us to estimate the mixing matrix in a robust manner in the presence of white noise. Now we bring this idea into the convolutive source separation problem, in order to develop the F-SEONS algorithm. In the frequency-domain, we define a time-delayed cross-spectral density matrix and identify the mixing filter $\boldsymbol{H}(\omega)$ through a joint approximate diagonalizer of time-varying time-delayed cross-spectral density matrices.

We first partition the data into K non-overlapping blocks of size T and estimate time-delayed cross-spectral density matrices at each data block, which are defined by

$$\boldsymbol{R}_x(\omega, k; \tau) = \frac{1}{N_s} \sum_{i=0}^{N_s-1} \boldsymbol{x}(\omega, k, i; 0)\boldsymbol{x}^H(\omega, k, i; \tau), \quad (5)$$

where N_s is the number of overlapped segments in each data block and

$$\boldsymbol{x}(\omega, k, i; \tau) = \sum_t \zeta(t - iT_s - (k-1)T - \tau)\boldsymbol{x}(t)\mathrm{e}^{-\frac{j2\pi\omega t}{T_b}}, \quad (6)$$

where $\zeta(t)$ is a windowing function whose value is one for $t \in [0, T_b)$, otherwise zero. T_s represents a time-shift between adjacent overlapped segments, T_b is the frame size for DFT, and τ is the time-lag which runs from 1 to J.

The cross-spectral density matrices of measurement signals $\boldsymbol{x}(t)$ satisfy

$$\boldsymbol{R}_x(\omega, k; \tau) = \boldsymbol{H}(\omega)\boldsymbol{R}_s(\omega, k; \tau)\boldsymbol{H}^H(\omega), \quad (7)$$

for $\tau = 1, \ldots, J$ and $k = 1, \ldots, K$, where the cross-spectral density matrices of white noise disappear, due to the temporal independence. Note that the time-delayed cross-spectral density matrices of sources are diagonal for all $\omega \in [0, 2\pi)$, because they are assumed to be mutually independent. At each frequency bin, ω, the mixing filter $\boldsymbol{H}(\omega)$ is estimated using a joint approximate diagonalization method that was also used in SOBI [1], [2].

At each frequency bin, we estimate $\boldsymbol{H}(\omega)$ as in SEONS. Please refer to [2] for detailed illustration of SEONS. The F-SEONS algorithm is summarized below.

Algorithm Outline: F-SEONS

1) Do the robust whitening.
   a) Compute
   $$\boldsymbol{M}_x(\omega, k; \tau) = \frac{1}{2}\left\{\boldsymbol{R}_x(\omega, k; \tau) + \boldsymbol{R}_x^H(\omega, k; \tau)\right\},$$
   for $\tau = 1, \ldots, J$ and $k = 1, \ldots, K$.
   b) Find
   $$\boldsymbol{C}(\omega) = \sum_{\tau=1}^{J}\sum_{k=1}^{K} \alpha_p \boldsymbol{M}_x(\omega, k; \tau),$$
   which is positive definite.
   c) The robust whitening matrix $\boldsymbol{Q}(\omega)$ is computed by
   $$\boldsymbol{Q}(\omega) = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\omega)\boldsymbol{U}^H(\omega),$$
   where $\boldsymbol{U}(\omega)$ and $\boldsymbol{\Sigma}(\omega)$ are eigenvector and eigenvalue matrices of $\boldsymbol{C}(\omega)$.
2) Calculate
$$\boldsymbol{M}_z(\omega, k; \tau) = \boldsymbol{Q}(\omega)\boldsymbol{M}_x(\omega, k; \tau)\boldsymbol{Q}^H(\omega),$$
for $k = 1, \cdots, K$ and $\tau = 1, \cdots, J$.
3) Find an unitary joint approximate diagonalizer $\boldsymbol{V}(\omega)$ of $\{\boldsymbol{M}_z(\omega, k; \tau)\}$, which satisfies
$$\boldsymbol{V}^H(\omega)\boldsymbol{M}_z(\omega, k; \tau)\boldsymbol{V}(\omega) = \boldsymbol{\Lambda}_{k\tau},$$
where $\boldsymbol{\Lambda}_{k\tau}$ is a set of diagonal matrices.
4) The demixing filter is computed as
$$\boldsymbol{W}(\omega) = \boldsymbol{V}^H(\omega)\boldsymbol{Q}(\omega).$$

Arbitrary permutations for each frequency ω are not distinguishable, which cause a serious problem in a frequency-domain source separation. Although a variety of research is still going on to solve this problem, in this paper we adopt a method of imposing a smoothness constraint on the demixing filter as in [8].

IV. H_∞ Filtering for Denoising

Restored speech signals $s_i(t)$ are still contaminated by noise, since the reconstruction is carried out by a linear transform. In order to suppress the noise in segregated speech signals, here we apply an H_∞ filtering technique. To this end, we assume that the restored speech signal, $s_i(t)$, satisfies an autoregressive (AR) model (all pole model) that has the form

$$s_i(t) = \sum_{k=1}^{q} a_k s_i(t-k) + \epsilon_i(t), \quad (8)$$

where $\{a_k\}$ are the AR model coefficients and $\epsilon_i(t)$ is an innovation. The observed noisy speech signal $y_i(t)$ is described by

$$\xi_i(t) = s_i(t) + \nu_i(t), \quad (9)$$

where $\nu_i(t)$ is the measurement noise.

We can write (8) and (9) in a state space model that has the form

$$\vec{\boldsymbol{s}}_i(t) = \boldsymbol{A}\vec{\boldsymbol{s}}_i(t-1) + \vec{\boldsymbol{b}}\epsilon_i(t) \qquad \text{(state equation)}, \quad (10)$$

$$\xi_i(t) = \vec{\boldsymbol{c}}^T \vec{\boldsymbol{s}}_i(t) + \nu_i(t) \qquad \text{(measurement equation)}, (11)$$

where

$$\vec{\boldsymbol{s}}_i(t) = \left[s_i(t-q+1), s_i(t-q+2), \ldots, s_i(t)\right]^T,$$

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ a_q & a_{q-1} & a_{q-2} & \cdots & a_2 & a_1 \end{bmatrix},$$

$$\vec{\boldsymbol{b}} = \vec{\boldsymbol{c}} = [0, \ldots, 0, 1]^T. \quad (12)$$

Speech enhancement methods based on Wiener/Kalman filters assume that both $\epsilon_i(t)$ and $\nu_i(t)$ are white or color Gaussian processes [5], [7]. However, neither the speech nor the noise may be Gaussian. The Gaussian assumptions may lead to an estimate that is vulnerable to statistical outliers. In contrast, the H_∞ filtering-based speech enhancement method [9] makes no assumption on $\epsilon_i(t)$ and $\nu_i(t)$ and is interested not only in the estimation of $\vec{\boldsymbol{s}}_i(t)$ but also in the estimation of some arbitrary linear combination of $\vec{\boldsymbol{s}}_i(t)$, i.e., $z_i(t) = \vec{\boldsymbol{\beta}}^T \vec{\boldsymbol{s}}_i(t)$ where $\vec{\boldsymbol{\beta}} \in \mathbb{R}^q$. The discrete H_∞ filtering method is interpreted as a *minimax* problem, which leads to the performance criterion that has the form

$$\min_{\hat{z}_i(t)} \max_{(\nu_i(t), \epsilon_i(t), \vec{\boldsymbol{s}}_i(0))} \mathcal{J}_i, \quad (13)$$

where

$$\mathcal{J}_i = -\frac{1}{2}\gamma^2 \left\| \vec{\boldsymbol{s}}_i(0) - \hat{\vec{\boldsymbol{s}}}_i(0) \right\|^2_{\boldsymbol{P}_0^{-1}} + \frac{1}{2}\sum_{t=0}^{N} \Big[\rho^2 \left\| z_i(t) - \hat{z}_i(t) \right\|^2 - \gamma^2 \left(\| |\epsilon_i(t) \|^2_{\boldsymbol{\Sigma}_\epsilon^{-1}} + \sigma_{\nu_i}^{-2} \left\| \nu_i(t) \right\|^2 \right) \Big], \quad (14)$$

where $\hat{\vec{\boldsymbol{s}}}_i(t)$ is the estimate of $\vec{\boldsymbol{s}}_i(t)$ and $\hat{z}_i(t) = \vec{\boldsymbol{\beta}}^T \hat{\vec{\boldsymbol{s}}}_i(t)$ and $\{(\vec{\boldsymbol{s}}_i(0) - \hat{\vec{\boldsymbol{s}}}_i(0)), \epsilon_i(t), \nu_i(t)\} \neq 0$.

Algorithm Outline: H_∞ filtering for Denoising

1) Initialize γ^2, ρ^2, $\sigma^2_{\epsilon_i}$, $\sigma^2_{\nu_i}$ and $\boldsymbol{P}(0) > 0$.
2) Update $\boldsymbol{P}(t)$ by

$$\boldsymbol{P}(t+1) = \boldsymbol{A}\boldsymbol{P}(t) \ \ \boldsymbol{I} - \gamma^{-2}\boldsymbol{\Psi}\boldsymbol{P}(t) + \sigma_{\nu_i}^{-2}\vec{\boldsymbol{c}}\vec{\boldsymbol{c}}^T\boldsymbol{P}(t) \ \ ^{-1}\boldsymbol{A}^T + \sigma^2_{\epsilon_i}\vec{\boldsymbol{b}}\vec{\boldsymbol{b}}^T, \quad \boldsymbol{\Psi} = \rho^2 \vec{\boldsymbol{\beta}}\vec{\boldsymbol{\beta}}^T.$$

3) Estimate the gain of the H_∞ filter, $\boldsymbol{\kappa}(t)$,

$$\boldsymbol{\kappa}(t) = \boldsymbol{A}\boldsymbol{P}(t) \ \ \boldsymbol{I} - \gamma^{-2}\boldsymbol{\Psi}\boldsymbol{P}(t) + \sigma_{\nu_i}^{-2}\vec{\boldsymbol{c}}\vec{\boldsymbol{c}}^T\boldsymbol{P}(t) \ \ ^{-1}\vec{\boldsymbol{c}}\sigma_{\nu_i}^{-2}.$$

4) Estimate the enhanced speech signal $\hat{\vec{\boldsymbol{s}}}_i(t)$ by

$$\hat{\vec{\boldsymbol{s}}}_i(t) = \boldsymbol{A}\hat{\vec{\boldsymbol{s}}}_i(t-1) + \boldsymbol{\kappa}(t) \ \ \xi_i(t) - \vec{\boldsymbol{c}}^T\boldsymbol{A}\hat{\vec{\boldsymbol{s}}}_i(t-1) \ \ .$$

V. Numerical Experiments

We performed numerical experiments for the case where two speech source signals were convolved with a multivariate FIR filter whose characteristics is described by

$$H_{11} \ \ z^{-1} \ = 1 + 0.67z^{-1} + 0.55z^{-2} - 0.45z^{-3} + 0.37z^{-4},$$

$$H_{12} \ \ z^{-1} \ = 0.6 - 0.40z^{-1} - 0.33z^{-2} + 0.27z^{-3} - 0.22z^{-4},$$

$$H_{21} \ \ z^{-1} \ = 0.5 + 0.34z^{-1} + 0.27z^{-2} - 0.22z^{-3} + 0.18z^{-4},$$

$$H_{22} \ \ z^{-1} \ = 1 - 0.57z^{-1} + 0.49z^{-2} + 0.40z^{-3} + 0.26z^{-4}.$$

Original clean speech signals were taken from OGI NB-95 DB which was also used to construct ST-NB 95 DB for cocktail party speech recognition [11]. We compared our F-SEONS algorithm with the Parra-Spence algorithm [8] where the nonstationarity was also exploited and the least squares method was employed. In our experiments, we used $T_b = 512$, $K = 5$, $J = 3$, $T_s = 4$.

The first experiment involves with the case of noisy data where several levels of additive white Gaussian noise (AWGN) were added. We used two signal to interference (SIR) measures which were defined by

$$\mathrm{SIR}_I = \frac{\sum_\omega \sum_i E\left\{|H_{ii}(\omega)|^2 |s_i(\omega)|^2\right\}}{\sum_\omega \sum_{i \neq j} \sum_j E\left\{|H_{ij}(\omega)|^2 |s_j(\omega)|^2\right\}},$$

$$\mathrm{SIR}_O = \frac{\sum_\omega \sum_i E\left\{|G_{ii}(\omega)|^2 |s_i(\omega)|^2\right\}}{\sum_\omega \sum_{i \neq j} \sum_j E\left\{|G_{ij}(\omega)|^2 |s_j(\omega)|^2\right\}}, \quad (15)$$

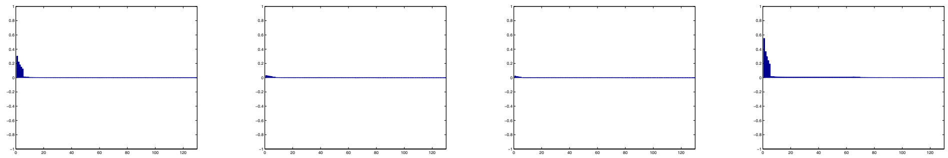
where $\boldsymbol{G}(\omega) = \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)$. At each SNR, we carried out 50 independent runs and calculated averaged SIRs. A comparison is summarized in Table I. In the case of no noise, our F-SEONS algorithm and the Parra-Spence algorithm work well, although our method produces better performance. In the presence of AWGN, our F-SEONS algorithm outperforms the Parra-Spence algorithm (see Fig. 1).

TABLE I

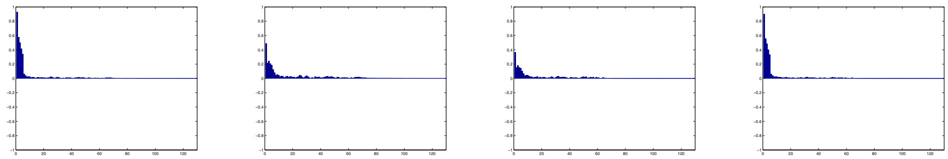
COMPARISON OF THE F-SEONS ALGORITHM AND THE PARRA-SPENCE ALGORITHM FOR THE CASE OF NOISY CONVOLVED MIXTURES, IN TERMS OF AVERAGED SIR_O (IN THESE EXPERIMENTS $\mathrm{SIR}_I = 4.6045$).

AWGN (dB)	SIR_O	
	F-SEONS	Parra-Spence
No noise	42.8173	32.7438
30	39.3525	25.0661
20	29.0861	17.3735
10	17.2102	11.2024
5	15.4790	8.9272

In the second experiment, we applied the H_∞ filtering to segregated speech signals, in order to further reduce the noise. AR coefficients $\{a_k\}$ of order 5, were estimated within each segment of length 128 samples (corresponding to the duration of 16ms). Initialization for the H_∞ filter was done using the following values, $\sigma_{\epsilon_i} = 0.1$, $\sigma_{\nu_i} = 0.1$, $\vec{\boldsymbol{s}}_i(0) = \mathbf{1}$, $\gamma^2 = 1.02$, $\rho^2 = 0.001$.



(a)



(b)

Fig. 1. Impulse responses of global system $\boldsymbol{G}(\omega) = \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)$ for the case of: (a) F-SEONS (SNR=10dB); (b) Parra-Spence (SNR=10dB). At each row, from left to right, $G_{11}(\omega)$, $G_{12}(\omega)$, $G_{21}(\omega)$, $G_{22}(\omega)$. Even in the case of SNR=10dB, F-SEONS produces zero G_{12} and G_{21} in (a), whereas, the Parra-Spence algorithm shows a large amount of cross-talking in (b).

In order to measure the performance of denoising, we also evaluated the denoising performance of the $\boldsymbol{H}_\infty$ filter in terms of the SNR between the segregated speech signal and the denoised output of the $\boldsymbol{H}_\infty$ filter. At each SNR, we carried out 50 independent runs and calculated averaged SNR between outputs (see Table II). In Table II, SNR$_I$ is the SNR between $\boldsymbol{y}_{st}(t) = \sum_{\tau=0}^{P} \boldsymbol{G}(\tau)\boldsymbol{s}(t-\tau)$ $\boldsymbol{y}(t) = \sum_{\tau=0}^{P} \boldsymbol{W}(\tau)\boldsymbol{x}(t-\tau)$, and SNR$_O$ is the SNR between the the separated speech which is not considering separated noise, i.e. $\boldsymbol{y}_{st}(t)$ and output speech of the H_∞ filter $\boldsymbol{y}_{de}(t)$. These quantities are defined by

$$\text{SNR}_I = \frac{\sum_{i=1}^{N} E[|\boldsymbol{y}_{st}(t)|^2]}{\sum_{i=1}^{N} E[|\boldsymbol{y}(t) - \boldsymbol{y}_{st}(t)|]^2},$$

$$\text{SNR}_O = \frac{\sum_{i=1}^{N} E[|\boldsymbol{y}_{st}(t)|^2]}{\sum_{i=1}^{N} E[|\boldsymbol{y}_{de}(t) - \boldsymbol{y}_{st}(t)|]^2}.$$

TABLE II

THE RESULTS OF H_∞ FILTERING FOR SEGREGATED SPEECH SIGNALS.

AWGN (dB)	SNR$_I$(dB)	SNR$_O$(dB)
30	26.2095	32.0824
20	16.3984	23.8823
10	6.4392	17.9489
5	3.5086	8.0485

VI. CONCLUSION

In this paper, we have proposed a new frequency-domain method for noisy convolutive source separation, which is not sensitive to additive white noise. The method, F-SEONS, successfully generalized our original SEONS to the case of convolutive mixtures. We showed that F-SEONS algorithm were able to estimate the demixing filter correctly, in the presence of white noise. In addition, we also used the H_∞ filter in order to further suppress the remaining noise in segregated speech signals, and showed that this postprocessing indeed led to denoised signals.

ACKNOWLEDGMENT

This work was supported by KISTEP Brain Neuroinformatics Program, ETRI, and POSTECH Basic Research Fund.

REFERENCES

[1] A. Belouchrani, K. Abed-Merain, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.

[2] S. Choi, A. Cichocki, and A. Belouchrani, "Second order nonstationary source separation," *Journal of VLSI Signal Processing*, vol. 32, pp. 93–104, 2002.

[3] S. Choi, H. Hong, H. Glotin, and F. Berthommier, "Multichannel signal separation for cocktail party speech recognition: A dynamic recurrent network," *Neurocomputing*, vol. 49, no. 1-4, pp. 299–314, Dec. 2002.

[4] R. H. Lambert, "Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures," Ph.D. dissertation, University of Southern California, May 1996.

[5] K. Y. Lee and K. Shirai, "Efficient recursive estimation for speech enhancement in colored noise," *IEEE Signal Processing Letters*, vol. 3, pp. 196–199, 1996.

[6] N. Mitianoudis and M. E. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 5, pp. 489–497, 2003.

[7] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, 1987, pp. 177–180.

[8] L. C. Parra and C. Spence, "Convolutive blind source separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, pp. 320–327, May 2000.

[9] X. Shen and L. Deng, "A dynamic system approach to speech enhancement using the H infinity filtering algorithm," *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 4, pp. 391–399, 1999.

[10] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[11] E. Tessier, F. Berthommier, H. Glotin, and S. Choi, "A model of source segregation using the localization cue for robust cocktail-party speech recognition," in *Proc. Int'l Conf. Speech Processing*, 1999, pp. 97–102.

[12] K. Torkkola, "Blind separation of convolved sources based on information maximization," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, 1996, pp. 423–432.