Comparison of Time-Domain ICA, Frequency-Domain ICA and Multistage ICA for Blind Source Separation

Tsuyoki NISHIKAWA, Hiroshi SARUWATARI, and Kiyohiro SHIKANO

Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, JAPAN

E-mail: tsuyo-ni@is.aist-nara.ac.jp

ABSTRACT

We propose a new algorithm for blind source separation (BSS), in which frequency-domain independent component analysis (FDICA) and time-domain ICA (TDICA) are combined to achieve a superior source-separation performance under reverberant conditions. Generally speaking, conventional TDICA fails to separate source signals under heavily reverberant conditions because of the low convergence in the iterative learning of the inverse of the mixing system. On the other hand, the separation performance of conventional FDICA also degrades significantly because the independence assumption of narrow-band signals collapses when the number of subbands increases. In the proposed method, the separated signals of FDICA are regarded as the input signals for TDICA, and we can remove the residual crosstalk components of FDICA by using TDICA. The experimental results obtained under the reverberant condition reveal that the separation performance of the proposed method is superior to that of conventional ICA-based BSS methods.

1 Introduction

Blind source separation (BSS) is the approach taken to estimate the original source signals using only the information of the mixed signals observed in each input channel. This technique is applicable to the realization of high-quality hands-free telecommunication systems. The BSS methods based on independent component analysis (ICA) [1] can be classified into two groups in terms of the processing domain, i.e., frequency-domain ICA (FDICA) in which the complex-valued inverse of the mixing matrix is calculated in the frequency domain [2, 3, 4], and time-domain ICA (TDICA) in which the inverse system of the mixing FIR-filter matrix is calculated in the time domain [1, 5]. The recently developed BSS techniques can achieve a good source-separation performance under artificial or short reverberant conditions. However, the performances of these methods under heavily reverberant conditions significantly degrade because of the following problems. (1) In conventional FDICA, the separation performance is saturated before reaching a sufficient performance because we transform the fullband signals into the narrow-band signals and the independence assumption collapses in each narrow-band [6]. (2) In conventional TDICA, the convergence degrades because the iterative learning rule becomes more complicated as the reverberation increases.

In order to resolve the problems, we propose a new BSS algorithm called multistage ICA (MSICA), in which FDICA and TDICA are combined. By using the proposed method, we can achieve a superior separation performance even under heavily reverberant conditions. The results of the signal separation experiments reveal that the separation performance of the proposed algorithm is superior to those of the conventional ICA-based BSS methods.

2 Sound Mixing Model of Microphone Array

In this study, a straight-line array is assumed. The number of array elements (microphones) is K and the number of multiple sound sources is L, and we deal with the case of $K = L = 2$.

In the frequency domain, the observed signals in which multiple source signals are mixed linearly are given by

$$\boldsymbol{X}(f) = \boldsymbol{A}(f)\boldsymbol{S}(f), \qquad (1)$$

where $\boldsymbol{X}(f) = [X_1(f), \cdots, X_K(f)]^{\mathrm{T}}$ is the observed signal vector, and $\boldsymbol{S}(f) = [S_1(f), \cdots, S_L(f)]^{\mathrm{T}}$ is the source signal vector (see Fig. 1). $\boldsymbol{A}(f)$ is the mixing matrix which is assumed to be complex-valued because we introduce a model to deal with the arrival lags among the elements of the microphone array and the room reverberations.

3 Conventional ICA and Its Problems

3.1 Frequency-Domain ICA

The conventional BSS based on FDICA is conducted with the following steps: (1) transform the observed fullband signals into the narrow-band signals, (2) optimize the inverse of the mixing matrix $\boldsymbol{A}(f)$ in each subbband, and (3) reconstruct the fullband separated signal from the narrow-band separated signals. FDICA has the following advantages and disadvantages.

Advantages:

(F1) We can simplify the convolutive mixture down to simultaneous mixtures by the frequency transform.

(F2) It is easy to converge the separation filter in iterative ICA learning with high stability.

Disadvantages:

(F3) The separation performance is saturated before reaching a sufficient performance because the independence assumption collapses in each narrow-band [6] (see, e.g., Sect. 5.2).

(F4) Permutation among source signals and indeterminacy of each source gain in each subband.

As for disadvantage (F4), various solutions have already been proposed [2, 7, 8]. However, the collapse of the independence assumption, (F3), is a serious and inherent problem, and this prevents us from applying FDICA in a real acoustic environment with a long reverberation.
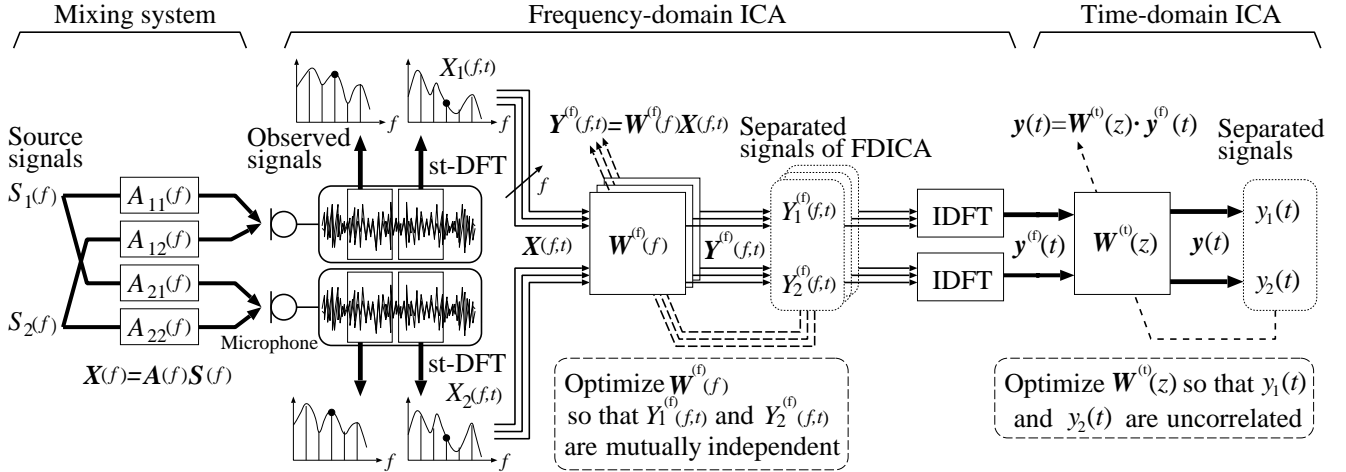
Figure 1: Blind source separation procedure performed in multistage ICA.

3.2 Time-Domain ICA

In the conventional BSS based on TDICA, each element of the mixing matrix is represented as a FIR filter. We can optimize its inverse, i.e., form an inverse filter system, by using the fullband observed signals themselves. TDICA has the following advantages and disadvantages.

Advantages:

(T1) We can treat the fullband speech signals where the independence assumption of sources usually holds.

(T2) High-convergence possibility near the optimal point.

Disadvantages:

(T3) The iterative rule for FIR-filter learning is complicated.

(T4) The convergence degrades under reverberant conditions.

It is known that TDICA works only in the case of mixtures with a short-tap FIR filter, i.e., less than 100 taps. Also, TDICA fails to separate source signals under real acoustic environments because of disadvantages (T3) and (T4).

4 Proposed Method: Multistage ICA

As described above, the conventional ICA methods have some disadvantages. However, note that the advantages and disadvantages of FDICA and TDICA are mutually complementary, i.e., (F3) can be resolved by (T1) and (T2), and (T3) and (T4) can be resolved by (F1) and (F2). Hence, in order to resolve the disadvantages, we propose a new algorithm, MSICA, in which FDICA and TDICA are combined (see Fig. 1).

MSICA is conducted with the following steps. In the first stage, we perform FDICA to separate the source signals to some extent with the high-stability advantages of FDICA, (F1) and (F2). In the second stage, we regard the separated signals of FDICA as the input signals for TDICA, and we remove the residual crosstalk components of FDICA by using TDICA. Finally, we regard the output signals of TDICA as the resultant separated signals. MSICA can achieve a high stability and a separation performance superior to that of conventional FDICA and TDICA. In the following sections, we describe details of the ICA-learning rules for each stage.

4.1 First-Stage ICA: Frequency-Domain ICA

In the first-stage ICA, we introduce the fast-convergence FDICA proposed by one of the authors [4]. We perform the signal separation procedure as described below (see FDICA in Fig. 1).

In FDICA, first, the short-time analysis of observed signals is conducted by frame-by-frame discrete Fourier transform (DFT). By plotting the spectral values in a frequency bin of each microphone input frame by frame, we consider them as a time series. Hereafter, we designate the time series as $\boldsymbol{X}(f,t) = [X_1(f,t), \cdots, X_K(f,t)]^T$. Next, we perform signal separation using the complex-valued inverse of the mixing matrix, $\boldsymbol{W}^{(f)}(f)$, so that the L time-series output $\boldsymbol{Y}^{(f)}(f,t) = [Y_1^{(f)}(f,t), \cdots, Y_L^{(f)}(f,t)]^T$ becomes mutually independent; this procedure can be given as

$$\boldsymbol{Y}^{(f)}(f,t) \;=\; \boldsymbol{W}^{(f)}(f)\boldsymbol{X}(f,t). \qquad (2)$$

We perform this procedure with respect to all frequency bins. Finally, by applying the inverse DFT and the overlap-add technique to the separated time series $\boldsymbol{Y}^{(f)}(f,t)$, we reconstruct the resultant source signals in the time domain, $\boldsymbol{y}^{(f)}(t)$.

In conventional FDICA , the optimal $\boldsymbol{W}^{(f)}(f)$ is obtained by the following iterative equation [2]:

$$\boldsymbol{W}_{i+1}^{(f)}(f) = \boldsymbol{W}_i^{(f)}(f) + \eta \Big[\mathrm{diag}\Big(\big\langle \boldsymbol{\Phi}(\boldsymbol{Y}^{(f)}(f,t))\boldsymbol{Y}^{(f)}(f,t)^{\mathrm{H}} \big\rangle_t \Big)$$
$$- \big\langle \boldsymbol{\Phi}(\boldsymbol{Y}^{(f)}(f,t))\boldsymbol{Y}^{(f)}(f,t)^{\mathrm{H}} \big\rangle_t \Big] \boldsymbol{W}_i^{(f)}(f), \qquad (3)$$

where $\langle \cdot \rangle_t$ denotes the time-averaging operator, i is used to express the value of the i-th step in the iterations, and η is the step-size parameter. Also, we define the nonlinear vector function $\boldsymbol{\Phi}(\cdot)$ as

$$\boldsymbol{\Phi}(\boldsymbol{Y}^{(f)}(f,t)) \equiv \big[\Phi(Y_1^{(f)}(f,t)), \cdots, \Phi(Y_L^{(f)}(f,t)) \big]^T, \quad (4)$$
$$\Phi(Y_l^{(f)}(f,t)) \equiv \big[1 + \exp(-\mathrm{Re}[Y_l^{(f)}(f,t)]) \big]^{-1}$$
$$+ j \cdot \big[1 + \exp(-\mathrm{Im}[Y_l^{(f)}(f,t)]) \big]^{-1}, \quad (5)$$

where $\mathrm{Re}[Y_l^{(f)}(f,t)]$ and $\mathrm{Im}[Y_l^{(f)}(f,t)]$ are the real and imaginary parts of $Y_l^{(f)}(f,t)$, respectively.

4.2 Second-Stage ICA: Time-Domain ICA

In the second-stage ICA, we introduce the TDICA which uses nonstationarity of the source signals (see TDICA in Fig. 1). We separate the sources by minimizing the non-negative cost function which takes the minimum value only when the second-order cross-correlation becomes zero if the source signals are nonstationary. The cost function can be given as [5]

$$Q\big(\boldsymbol{W}^{(\mathrm{t})}(z)\big) = \frac{1}{2B} \sum_{b=1}^{B} \left\{ \log \frac{\det \mathrm{diag}\boldsymbol{R}_y^{(b)}(0)}{\det \boldsymbol{R}_y^{(b)}(0)} \right\}, \qquad (6)$$

where B is the number of local analysis blocks. $\boldsymbol{R}_y^{(b)}(n)$ is the correlation matrix of the separated signals, i.e., $\boldsymbol{R}_y^{(b)}(n) = \langle \boldsymbol{y}(t)\boldsymbol{y}(t-n)^{\mathrm{T}}\rangle_t^{(b)}$, where $\langle \cdot \rangle_t^{(b)}$ denotes the time-averaging operator for the b-th local analysis block, $\boldsymbol{y}(t)$ is the resultant separated signal vector, and $\boldsymbol{W}^{(\mathrm{t})}(z)$ is the z-transform of the separation filter coefficient $\boldsymbol{w}^{(\mathrm{t})}(n)$ ($n = 0, \cdots, N-1$); these are given as

$$\boldsymbol{y}(t) = [y_1(t), \cdots, y_L(t)]^{\mathrm{T}} = \boldsymbol{W}^{(\mathrm{t})}(z) \cdot \boldsymbol{y}^{(\mathrm{f})}(t), \quad (7)$$

$$\boldsymbol{W}^{(\mathrm{t})}(z) = \sum_{n=0}^{N-1} \boldsymbol{w}^{(\mathrm{t})}(n) z^{-n}, \qquad (8)$$

where z^{-1} is used as the unit-delay operator for convenience, i.e., $z^{-n} \cdot x(t) = x(t-n)$, and $\boldsymbol{y}^{(\mathrm{f})}(t)$ is the time-domain output of FDICA.

Equation (6) becomes zero only when $y_i(t)$ and $y_j(t)$ are uncorrelated for all of the local analysis blocks . Calculating the natural gradient of $Q\big(\boldsymbol{W}^{(\mathrm{t})}(z)\big)$, we obtain the iterative equation of the separation filter $\boldsymbol{w}^{(\mathrm{t})}(n)$ to minimize Eq. (6) as

$$\boldsymbol{w}_{i+1}^{(\mathrm{t})}(n) = \boldsymbol{w}_i^{(\mathrm{t})}(n) - \frac{\alpha}{B} \frac{\partial Q\big(\boldsymbol{W}_i^{(\mathrm{t})}(z)\big)}{\partial \boldsymbol{w}_i^{(\mathrm{t})}(n)} \boldsymbol{W}_i^{(\mathrm{t})}(z)^{\mathrm{T}} \boldsymbol{W}_i^{(\mathrm{t})}(z)$$

$$= \boldsymbol{w}_i^{(\mathrm{t})}(n) + \frac{\alpha}{B} \sum_{b=1}^{B} \Big\{ \boldsymbol{R}_y^{(b)}(0)^{-1} \boldsymbol{R}_y^{(b)}(n)$$

$$- \big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\big)^{-1} \boldsymbol{R}_y^{(b)}(n) \Big\} \boldsymbol{W}_i^{(\mathrm{t})}(z), \qquad (9)$$

where α is the step-size parameter. Equation (9) evaluates only off-diagonal of $\boldsymbol{R}_y^{(b)}(0)$. Therefore we expand Eq. (9) to the following equation to evaluate the off-diagonal of $\boldsymbol{R}_y^{(b)}(n)$ for all time delays n:

$$\boldsymbol{w}_{i+1}^{(\mathrm{t})}(n) = \boldsymbol{w}_i^{(\mathrm{t})}(n) + \frac{\alpha}{B} \sum_{b=1}^{B} \Big\{ \big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\big)^{-1} \mathrm{diag}\boldsymbol{R}_y^{(b)}(n)$$

$$- \big(\mathrm{diag}\boldsymbol{R}_y^{(b)}(0)\big)^{-1} \boldsymbol{R}_y^{(b)}(n) \Big\} \boldsymbol{W}_i^{(\mathrm{t})}(z). \qquad (10)$$

5 Experiments and Results

5.1 Experimental Setup

A two-element array with the interelement spacing of 4 cm is assumed. The speech signals are assumed to arrive from two directions, $-30°$ and $40°$ (direction normal to the array is set to be $0°$). The distance between the microphone array and the loudspeakers is 1.15 m. Two kinds of sentences, spoken by two male and two female speakers selected from the ASJ continuous speech corpus for research, are used as the original speech samples. The sampling frequency is 8 kHz and the length of speech is limited to within 3 seconds. Using these sentences, we obtain 12 combinations with respect to speakers and source directions. As for the mixing system, we use the impulse responses recorded in a real room with the reverberation time of 300 ms. In order to evaluate the performance, we used the *noise reduction rate* (NRR), defined as the output SNR in dB minus input SNR in dB.
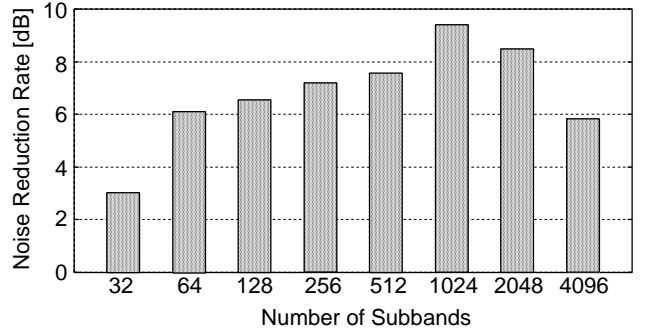

Figure 2: Relation between separation performances and the number of subbands in conventional FDICA.

5.2 Relation between Separation Performance and Number of Subbands in FDICA

In order to confirm the low-independence problem of sub-band signals in FDICA ((F3) described in Sect. 3.1), we carried out the preliminary experiment under the following analysis conditions. The number of subbands (frame length in DFT) is set to be from 32 to 4096, the frame shift is 16 taps, the window function is a Hamming window, the number of iterations in ICA is 30, and the step-size parameter η for iterations is set to be 1.0×10^{-5}.

Figure 2 shows the NRR results for different numbers of subbands in FDICA. As shown in Fig. 2, the NRR of FDICA obviously degrades when the number of subbands becomes too large, and the separation performance is saturated before reaching a sufficient performance. This is because we transform the fullband signals into the narrow-band signals and the independence assumption collapses in each frequency band, particularly when the number of subbands is large. On the basis of this result, we should cascade another signal processing analysis, e.g., TDICA, with FDICA to obtain the further separation performances.

5.3 Relation between Separation Performance and Filter Length in TDICA

We carried out the experiments using TDICA and MSICA to evaluate the contribution of increments of separation-filter length for improving the separation performances under reverberant conditions. The analysis conditions of these experiments are as follows: the filter length N is set to be from 10 to 2000 taps, the maximum number of iterations is 500, and the step-size parameter α for iterations is set to be $1/N$. As for the local analysis block, we divided the signals equally into B parts ($B = 1 \sim 10$). We chose the optimal B and number of iterations for each filter length because the convergence is different for every filter length. As for the FDICA part in MSICA, the analysis conditions are the same as those given in Sect. 5.2, except for the number of subbands (which is fixed at 1024 bands).

Figures 3(a) and (b) show the NRR results in the conventional TDICA and MSICA for different filter lengths. As shown in Fig. 3(a), when the separation filter is lengthened, the separation performance of the conventional TDICA degrades. This also implies that the simple TDICA separates only the direct components of arriving signals. On the other hand, in Fig. 3(b), the separation performance of MSICA is improved when the filter length is longer. This reveals that the TDICA part in MSICA can separate the source signals even with the reverberation components, and the TDICA is still useful near the optimal point.
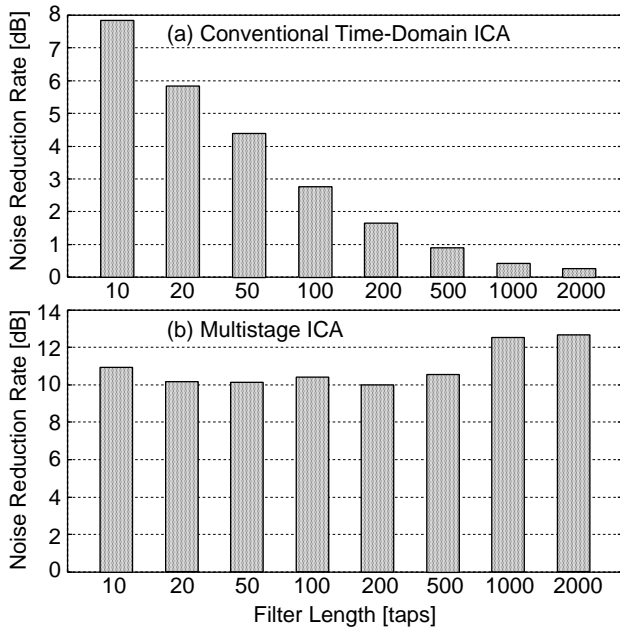
Figure 3: Relation between separation performance and filter length in (a) TDICA and (b) TDICA part in MSICA.

Figure 4: Comparison of noise reduction rates obtained by MSICA, conventional FDICA and the conventional TDICA.

5.4 Comparison between Conventional ICA and MSICA

We compared the performance of the proposed MSICA with that of the conventional ICA under the reverberant condition. As for FDICA, the analysis conditions are the same as those given in Sect. 5.2, except for the number of subbands (which is fixed at 1024 bands). As for the conventional TDICA, the number of local analysis blocks, B, is fixed at 3 blocks, the number of iterations is 400, and the filter length is 10 taps. As for the TDICA part in MSICA, the number of local analysis blocks, B, is fixed at 9 blocks, the number of iterations is 400, and the filter length is 1000 taps.

Figure 4 shows the NRRs of the conventional FDICA, conventional TDICA, and MSICA. In this figure, we separately plot the NRRs for different combination of speakers, and the averages of their NRRs. The results reveal that the separation performances of the proposed MSICA are superior to those of the conventional FDICA and TDICA with every combination. Specifically, compared with the conventional ICA, the proposed method can improve the NRR by about 2.7 dB over that of FDICA and by about 6.2 dB over that of TDICA, for an average of 12 combinations.

As described in Sect. 5.2, the FDICA in this study showed the saturation of NRR when we used the 1024-subband analysis. As described in Sect. 5.3, the simple TDICA could not separate the source signals accurately under the reverberant condition. These findings indicate the practical limitations of the separation performances of conventional ICA-based BSS methods. From the results of Fig. 4, however, we can confirm that the proposed MSICA can inherently remove these limitations, and is effective for improving the separation performance and convergence under reverberant conditions.

6 Conclusion

In this paper, we propose a new algorithm for BSS, in which FDICA and TDICA are combined to achieve a superior source-separation performance under reverberant conditions. The results of the signal separation experiments reveal that
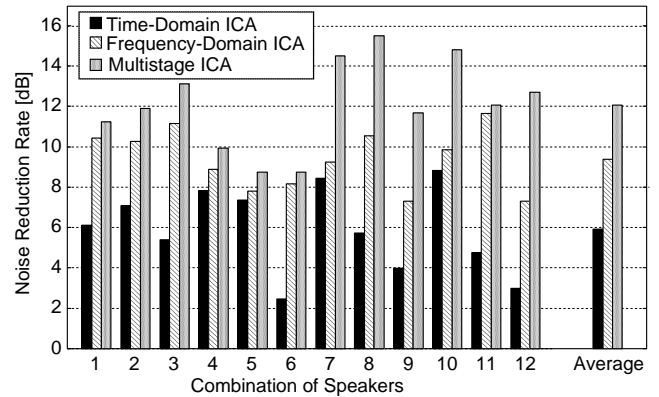
the separation performance of the proposed algorithm is superior to that of conventional ICA-based BSS methods, and the combination of FDICA and TDICA is inherently effective for improving the separation performance. Specifically, the proposed method can improve the SNR by about 2.7 dB over that of FDICA and by about 6.2 dB over that of TDICA, for an average of 12 speaker-combinations.

7 Acknowledgement

References

[1] T. W. Lee, *Independent component analysis*, Kluwer academic publishers, 1998.

[2] N. Murata and S. Ikeda, "An on-line algorithm for blind source separation on speech signals," *Proc. of 1998 International Symposium on Nonlinear Theory and Its Application (NOLTA98)*, pp.923–926, Sept. 1998.

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

[4] H. Saruwatari, T. Kawamura, and K. Shikano, "Blind source separation for speech based on fast-convergence algorithm with ICA and beamforming," *Proc. Eurospeech2001*, pp. 2603–2606, Sept. 2001.

[5] M. Kawamoto, K. Matsuoka, N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, 22, pp.157–171, 1998.

[6] S. Araki, S. Makino, T. Nishikawa, H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive of speech," *Proc. ICASSP2001*, pp.2737–2740, May 2001.

[7] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," *Proc. ICASSP2000*, pp.3140–3143, June 2000.

[8] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, vol.8, no.3, pp.320–327, May 2000.