A COMBINED APPROACH OF ARRAY PROCESSING AND INDEPENDENT COMPONENT ANALYSIS FOR BLIND SEPARATION OF ACOUSTIC SIGNALS

Futoshi Asano[1], Shiro Ikeda[2], Michiaki Ogawa[3], Hideki Asoh[1], Nobuhiko Kitawaki[3]

1)Electrotechnical Laboratory, 2)PREST,JST, 3)Tsukuba University, Japan
Email:asano@etl.go.jp

ABSTRACT

In this paper, two array signal processing techniques are combined with independent component analysis to enhance the performance of blind separation of acoustic signals in a reflective environment such as rooms. The first technique is the subspace method which reduces the effect of room reflection. The second technique is a method of solving permutation, in which the coherency of the mixing matrix in adjacent frequencies is utilized.

1. INTRODUCTION

When applying blind source separation (BSS) based on independent component analysis (ICA) to an acoustical mixture problem such as a number of people talking in a room, the performance of the BSS system is greatly reduced by the effect of the room reflections/reverberations and ambient noise [1]. In array signal processing, the authors previously proposed a method based on the subspace method for reducing the effect of room reflections and ambient noise [2]. The subspace method works as a self-organizing beamformer and, therefore, can be used in the framework of BSS [3]. In this paper, a combined approach of the subspace method and ICA is proposed.

For combining the subspace method with ICA, the frequency-domain ICA [4] must be employed, since the subspace method works in the frequency domain. The biggest obstacle in the frequency-domain ICA is the permutation problem. In the frequency-domain processing for a convolutive mixture, different permutations at different frequencies lead to re-mixing of signals in the final output. A method for solving the permutation using the correlation between the spectral envelope at different frequencies was proposed [4], but was reported to sometimes fail when the input signals had similar envelopes [1]. In this paper, a new approach for solving the permutation problem termed Inter-Frequency Coherency (IFC) is proposed.

2. MODEL OF SIGNAL

Let us consider the case when there are D sound sources in the environment. By observing this sound field with M microphones and taking the short-term Fourier transform (STFT) of the microphone inputs, we obtain the input vector $\mathbf{x}(\omega,t) = [X_1(\omega,t), \cdots, X_M(\omega,t)]^T$ where $X_m(\omega,t)$ is STFT of the input signal in the tth time frame at the mth microphone. By taking STFT, the convolutive mixture problem is reduced to a complex but instantaneous mixture problem. The symbol $\cdot^T$ denotes the transpose. In this paper, the input signal is assumed to be modeled as

$$\mathbf{x}(\omega,t) = \mathbf{A}(\omega)\mathbf{s}(\omega,t) + \mathbf{n}(\omega,t). \qquad (1)$$

The $M \times D$ matrix $\mathbf{A}(\omega)$ is termed the mixing matrix, its (m,n) element, $A_{m,n}(\omega)$, being the transfer function from the nth source to the mth microphone as

$$A_{m,n}(\omega) = H_{m,n}(\omega)e^{-j\omega\tau_{m,n}}. \qquad (2)$$

The symbol $H_{m,n}(\omega)$ is the magnitude of the transfer function. The symbol $\tau_{m,n}$ denotes the propagation time from the nth source to the mth microphone. Vector $\mathbf{s}(\omega,t)$ consists of the source spectra as $\mathbf{s} = [S_1(\omega,t), \cdots, S_D(\omega,t)]^T$. The first term, $\mathbf{A}(\omega)\mathbf{s}(\omega,t)$, expresses the directional components. On the other hand, the second term, $\mathbf{n}(\omega,t)$, is a mixture of less-directional components, which includes room reflections and ambient noise.

3. BSS SYSTEM

A block diagram of the system is depicted in Fig. 1. First, the subspace method is applied to the input vector $\mathbf{x}(\omega,t)$ to obtain the subspace filter $\mathbf{W}(\omega)$. In this stage, room reflections and ambient noise are reduced in advance of the application of ICA. It should be noted that the node of the filter network is reduced from M to D in this stage as depicted in Fig. 1.

Then, the instantaneous ICA is applied to the output of the subspace stage, $\mathbf{y}(\omega,t) = \mathbf{W}(\omega)\mathbf{x}(\omega,t)$, to

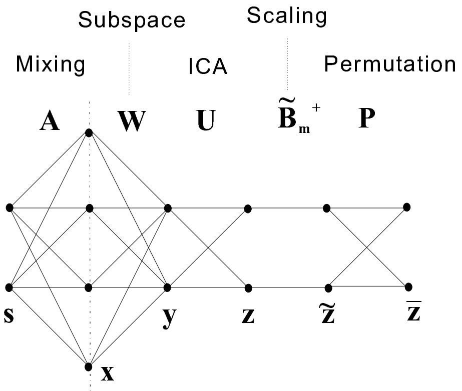


Fig. 1. Proposed BSS filter network.

obtain the filter matrix $\mathbf{U}(\omega)$. In this paper, the Infomax algorithm with feed-forward architecture [5] extended to complex data [4] is used. The learning rule is written as

$$\mathbf{U}(\omega, t+1) = \mathbf{U}(\omega, t) + \eta \left[\mathbf{I} - \varphi(\mathbf{z}(\omega, t))\mathbf{z}^H(\omega, t)\right] \mathbf{U}(\omega, t) \tag{3}$$

where $\mathbf{z}(\omega, t) = \mathbf{U}(\omega)\mathbf{y}(\omega, t)$. The score function for the complex data $\varphi(\mathbf{z})$ is defined as [4]

$$\varphi(\mathbf{z}) = 2\tanh(G \cdot \mathrm{Re}(\mathbf{z})) + 2\jmath \tanh(G \cdot \mathrm{Im}(\mathbf{z})). \tag{4}$$

The matrix $\mathbf{I}$ is an identity matrix. The symbol $\cdot^H$ denotes the Hermitian transpose. The constant η is termed the learning rate. The symbol G is the gain constant for the nonlinear score function, assuming that the magnitude of $\mathbf{y}(\omega, t)$ is normalized.

For the sake of convenience, the product of $\mathbf{W}(\omega)$ and $\mathbf{U}(\omega)$,

$$\mathbf{B}(\omega) = \mathbf{W}(\omega)\mathbf{U}(\omega), \tag{5}$$

is termed the separation filter.

After obtaining this separation filter, the permutation and the scaling problem must be solved. In this stage, the output of the separation filter is processed with the permutation matrix $\mathbf{P}(\omega)$ and the scaling matrix $\tilde{\mathbf{B}}_m^+(\omega)$. The scaling matrix $\tilde{\mathbf{B}}_m^+(\omega)$ is a $D \times D$ diagonal matrix $\tilde{\mathbf{B}}_m^+(\omega) = \mathrm{diag}[B_{m,1}^+, \cdots, B_{m,D}^+]$ where $B_{m,n}^+$ denotes the (m, n)th element of the pseudoinverse of $\mathbf{B}(\omega)$ [4]. The permutation matrix $\mathbf{P}(\omega)$ is described in Section 5.

Using the matrices obtained above, the final filtering matrix in the frequency domain can be written as

$$\mathbf{F}(\omega) = \mathbf{P}(\omega)\tilde{\mathbf{B}}_m^+(\omega)\mathbf{B}(\omega). \tag{6}$$

This filter matrix $\mathbf{F}(\omega)$ is transformed into the time domain, and the input signal is processed with the time-domain filter network [4].

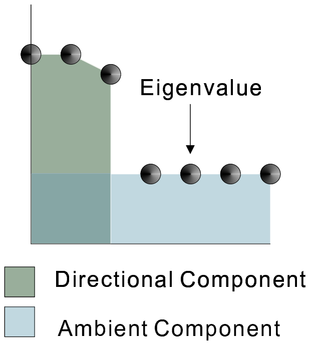


Fig. 2. Typical eigenvalue distribution.

4. SUBSPACE METHOD

4.1. Properties of Spatial Correlation Matrix

The spatial correlation matrix is defined as $\mathbf{R}(\omega) = E[\mathbf{x}(\omega, t)\mathbf{x}^H(\omega, t)]$. The frequency index ω is omitted in this section for the sake of simplicity in notation. Assuming that $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are uncorrelated, $\mathbf{R}$ can be written as

$$\mathbf{R} = \mathbf{A}\mathbf{Q}\mathbf{A}^H + \mathbf{K}, \tag{7}$$

where $\mathbf{Q} = E[\mathbf{s}(t)\mathbf{s}^H(t)]$ and $\mathbf{K} = E[\mathbf{n}(t)\mathbf{n}^H(t)]$. When $\mathbf{n}(t)$ includes room reflections of $\mathbf{s}(t)$, $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are correlated and the above assumption does not hold. However, when the window length of STFT is short and the time interval between the direct sound and the reflection exceeds this window length, this assumption holds to some extent in a practical sense.

By taking the generalized eigenvalue decomposition[6] of $\mathbf{R}$ as $\mathbf{R} = \mathbf{K}\mathbf{E}\Lambda\mathbf{E}^{-1}$, we have the eigenvector matrix $\mathbf{E} = [\mathbf{e}_1, \cdots, \mathbf{e}_M]$ and the eigenvalue matrix $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_M)$, where $\mathbf{e}_m$ and λ_m are the eigenvector and the eigenvalue, respectively. The eigenvalues and eigenvectors have the following properties [2, 6]:

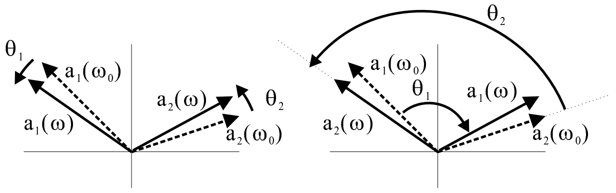
P1: The energy of the D directional signals $\mathbf{s}(t)$ is concentrated on the D dominant eigenvalues.

P2: The energy of $\mathbf{n}(t)$ is equally spread over all eigenvalues.

P3: $\Re(\mathbf{A}) = \Re(\mathbf{E}_s)$.

P4: $\Re(\mathbf{A}) = \Re(\mathbf{E}_n)^\perp$.

The matrices, $\mathbf{E}_s = [\mathbf{e}_1, \cdots, \mathbf{e}_D]$ and $\mathbf{E}_n = [\mathbf{e}_{D+1}, \cdots, \mathbf{e}_M]$, consist of the eigenvectors for the D dominant eigenvalues and those for the other $M - D$ eigenvalues, respectively. The notation $\Re(\mathbf{A})$ denotes the space spanned by the column vectors of $\mathbf{A}$. The notation $\Re(\mathbf{E}_n)^\perp$ denotes the orthogonal complement of $\Re(\mathbf{E}_n)$. The subspaces $\Re(\mathbf{E}_s)$ and $\Re(\mathbf{E}_n)$ are termed signal subspace



(a) Correct Perm. (b) Incorrect Perm.

Fig. 3. Rotation of the location vector. (a) Correct permutation; (b) Incorrect permutation.

and noise subspace, respectively. A typical eigenvalue distribution and the corresponding energy distribution that reflects the properties P1 and P2 are depicted in Fig. 2.

4.2. Subspace Filter

The subspace filter is defined as

$$\mathbf{W} = \Lambda_s^{-1/2}\mathbf{E}_s^H, \tag{8}$$

where $\Lambda_s = \mathrm{diag}(\lambda_1, \cdots, \lambda_D)$. According to the properties P1 and P3, the directional component $\mathbf{As}(t)$ belongs to the signal subspace. On the other hand, using the properties P2-P4, the ambient component $\mathbf{n}(t)$ can be split as $\mathbf{n}(t) = \mathbf{n}_s(t) + \mathbf{n}_n(t)$ where $\mathbf{n}_s(t)$ and $\mathbf{n}_n(t)$ denote the components which belong to the signal subspace and the noise subspace, respectively. Due to the orthogonality in P4, the component $\mathbf{n}_n(t)$ in the noise subspace is canceled by the subspace filter $\mathbf{W}$. In this regard, the subspace filter is equivalent to the delay-and-sum beamformer [2].

5. PERMUTATION

5.1. Structure of mixing matrix

When the mixing matrix $\mathbf{A}(\omega)$ has the form in (2), the nth column vector (location vector of the nth source) in the mixing matrix at the frequency ω and that at the adjacent frequency $\omega_0 = \omega - \Delta\omega$ are written as

$$\begin{aligned} \mathbf{a}_n(\omega) &= [e^{-j\omega\tau_{1n}}, \cdots, e^{-j\omega\tau_{Mn}}]^T \\ \mathbf{a}_n(\omega_0) &= [e^{-j(\omega-\Delta\omega)\tau_{1n}}, \cdots, e^{-j(\omega-\Delta\omega)\tau_{Mn}}]^T \end{aligned} \tag{9}$$

Here, $H_{m,n}(\omega) = 1$ in (2) is assumed for the sake of simplicity. From (9), it can be known that the location vector $\mathbf{a}_n(\omega)$ is $\mathbf{a}_n(\omega_0)$ being rotated by the angle θ_n as depicted in Fig. 3(a). Based on this relation (coherency) of the location vectors at the adjacent frequencies, the relation of the mixing matrix can be written as $\mathbf{A}(\omega) = \mathbf{T}(\omega, \omega_0)\mathbf{A}(\omega_0)$, where the matrix $\mathbf{T}(\omega, \omega_0)$ is the rotation matrix [7]. When the difference in frequency $\Delta\omega$ (frequency resolution of STFT) is sufficiently small, $\mathbf{T}(\omega, \omega_0) \simeq \mathbf{I}$, and the rotation angle θ_n is small.

5.2. Method for solving permutation (IFC)

Based on this, it is expected that θ_n is the smallest for the correct permutation as depicted in Fig. 3. Permutation is solved so that the sum of the angles $\{\theta_1, \cdots, \theta_D\}$ between the location vectors in the adjacent frequencies is minimized. An estimate of the mixing matrix can be obtained as the pseudoinverse of the separation matrix $\mathbf{B}(\omega)$ as $\hat{\mathbf{A}}(\omega) = \mathbf{B}^+(\omega)$. Let us denote the mixing matrix multiplied by the arbitrary permutation matrix $\mathbf{P}$ as $\bar{\mathbf{A}}^T(\omega) = \mathbf{P}\hat{\mathbf{A}}^T(\omega)$. The permutation $\mathbf{P}\hat{\mathbf{A}}^T(\omega)$ exchanges the row vectors of $\hat{\mathbf{A}}^T(\omega)$ (the column vectors of $\hat{\mathbf{A}}(\omega)$). The column vectors of $\bar{\mathbf{A}}(\omega)$ are denoted as $\{\bar{\mathbf{a}}_1(\omega), \cdots, \bar{\mathbf{a}}_D(\omega)\}$. The cosine of the angle θ_n between $\bar{\mathbf{a}}_n(\omega)$ and $\bar{\mathbf{a}}_n(\omega_0)$ is defined as

$$\cos\theta_n = \frac{\bar{\mathbf{a}}_n^H(\omega)\bar{\mathbf{a}}_n(\omega_0)}{||\bar{\mathbf{a}}_n(\omega)|| \cdot ||\bar{\mathbf{a}}_n^H(\omega_0)||}. \tag{10}$$

By using this, the permutation matrix is determined as

$$\hat{\mathbf{P}} = \arg\max_{\mathbf{P}} F(\mathbf{P}), \tag{11}$$

where the cost function $F(\mathbf{P})$ is defined as

$$F(\mathbf{P}) = \frac{1}{D}\sum_{n=1}^{D}\cos\theta_n. \tag{12}$$

5.3. Confidence Measure

Since the permutation at frequency ω is determined based on only the information of the two adjacent frequencies, ω and ω_0, and the permutation is solved iteratively with increasing frequency, once the permutation at the certain frequency fails, the permutation in the succeeding frequencies may also fail. To prevent this, the reference frequency ω_0 is extended to the frequency range $\omega_0 = \omega - k \cdot \Delta\omega$, for $k = 1, \cdots, K$. The cost function (12) is calculated at all K frequencies in this range. Let us denote the value of the cost function at $\omega_0 = \omega - k \cdot \Delta\omega$ as $F(\mathbf{P}, k)$.

Next, a confidence measure for $F(\mathbf{P}, k)$ is considered. When the largest value of the cost function is close to that with other permutations, it is difficult to determine which permutation is correct and the value of $F(\mathbf{P}, k)$ is not reliable. Based on this, the following confidence measure is defined:

$$C(k) = \max_{\mathbf{P}\in\Omega}[F(\mathbf{P}, k)] - \max_{\mathbf{P}\in\Omega'}[F(\mathbf{P}, k)]. \tag{13}$$

Here, Ω denotes the set of all possible $\mathbf{P}$ while Ω' denotes Ω without $\hat{\mathbf{P}} = \arg\max_{\mathbf{P}\in\Omega}[F(\mathbf{P}, k)]$. The appropriate reference frequency ω_0 is determined as $\omega_0 = \omega - \hat{k} \cdot \Delta\omega$ with $\hat{k} = \arg\max C(k)$. The permutation is then solved using the information at this reference frequency as $\hat{\mathbf{P}} = \arg\max_{\mathbf{P}\in\Omega}[F(\mathbf{P}, \hat{k})]$.

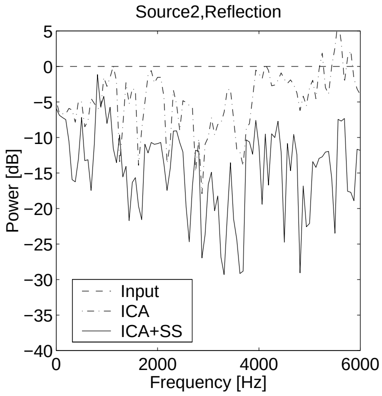


Fig. 4. Spectra of reflection at the input/output.

6. EXPERIMENT

A signal separation experiment was conducted in an ordinary meeting room with a reverberation time of 0.4 s. The sound sources (loudspeakers) were located in the front (0°) and in the right (60°) with a distance of 1 m from the array. The microphone array was circular in shape with a diameter of 0.5 m and $M = 8$.

Figure 4 shows the spectra of the reflection at the input/output of the system (normalized by the input spectrum). From this, it can be seen that the reflections were reduced by 10-15 dB by the subspace method. Figure 5 shows the results of the automatic speech recognition (ASR). For the cases ICA and ICA+SS, the correct permutation was given. As can be seen from this figure, the recognition rate was improved by around 18% by employing the subspace method.

Figure 6 shows the theoretical value of the cost function $F(\mathbf{P}, k)$ for the correct and incorrect permutation. From this, it can be seen that $F(\mathbf{P}, k)$ shows a high value for the correct permutation. In Fig. 5, the ASR rate for the case when the permutation is solved by IFC is also shown (ICA+SS/IFC). From this, reduction of the ASR rate by employing IFC is small (around 4%).

7. CONCLUSION

The performance of BSS using ICA in a reflective environment was improved by combining ICA with array processing techniques. As a pre-processor, the subspace method was employed to reduce the effect of room reflections. This method reduced the reflections by around 10 dB and improved the ASR rate by around 18%. As a post-processor, a method for solving the permutation based on the coherency of the mixing matrix was proposed. The permutation error in ASR was reduced to 4% compared with that for the conventional method reported in [1] (around 18 %).

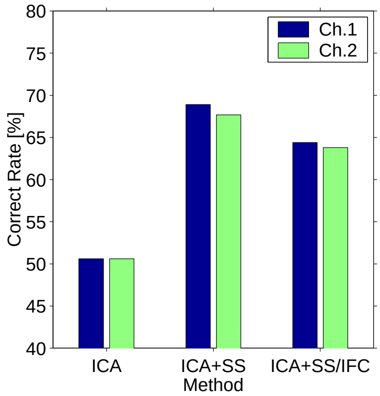


Fig. 5. Comparison of ASR rate for ICA only, ICA + SS (Subspace) and ICA+SS/IFC.

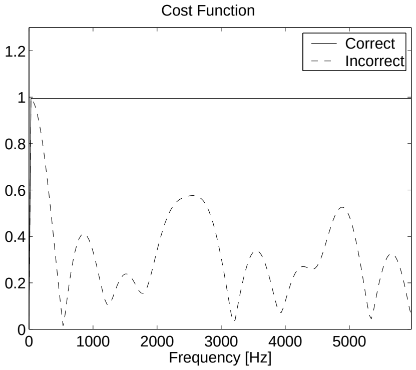


Fig. 6. Theoretical value of the cost function $F(\mathbf{P}, k)$.

8. REFERENCES

[1] F. Asano and S. Ikeda, "Evaluation and real-time implementation of blind source separation system using time-delayed decorrelation," in Proc. ICA2000, Jun. 2000, pp. 411–415.

[2] 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 Sep. 2000.

[3] F. Asano, Y. Motomura, H. Asoh, and T. Matsui, "Effect of pca filter in blind source separation," in Proc. ICA2000, June 2000, pp. 57–62.

[4] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in Proc. ICA'99, 1999, pp. 365–371.

[5] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," Neural Computation, vol. 7, pp. 1129–1159, 1995.

[6] R. Roy and T. Kailath, "Esprit - estimation of signal parameters via rotational invariance techniques," IEEE Trans. ASSP, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[7] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," IEEE Trans. ASSP, vol. 33, no. 4, pp. 823–831, 1985.