BLIND SOURCE SEPARATION IN REFLECTIVE SOUND FIELDS

Futoshi Asano, Shiro Ikeda, Michiaki Ogawa, Hideki Asoh, Nobuhiko Kitawaki

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

ABSTRACT
In this paper, the effect of room reflection on blind source separation is investigated. The higher order re-
verberation (reverberation) can be reduced in advance of blind separation by using the subspace method. On
the other hand, as for the lower order reflection (early re-
lection), it is shown by experiments that the early re-
lection has little effect on the separation performance.

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 [1]. In
this paper, the room reflection is divided into early re-
versation and reverberation, and their effect on BSS is
investigated.

The reverberation consists of higher order reflec-
tions and has lower directivity. The authors previously
proposed a method based on the subspace method for
reducing reverberation in the field of array processing
[2]. The subspace method works as a self-organizing
beamformer and, therefore, can be used in the frame-
work of BSS [3]. A combined approach using both the
subspace method and ICA has been proposed [6].

On the other hand, early reflection consists of lower
order (mostly first order) reflections with high direct-
ity. In this paper, experiments were employed to
investigate how early reflection is treated in the ICA
framework.

2. MODEL OF SIGNAL
Let us consider the case when there are \( N \) sound sources
in the environment. By observing this sound field with
\( M \) microphones and taking the short-term Fourier trans-
form (STFT) of the microphone inputs, we obtain the
input vector \( \mathbf{x}(\omega, t) = [X_1(\omega, t), \ldots, X_M(\omega, t)]^T \) where
\( X_m(\omega, t) \) is STFT of the input signal in the \( t \)th time
frame at the \( m \)th microphone. It is assumed that \( M \gg \)
\( N \). By using STFT, the convolutive mixture problem is
reduced to a complex but instantaneous mixture problem [4]. The symbol \( ^T \) denotes the transpose.

In this paper, the input signal is modeled as a mix-
ture of the direct sound, the early reflection, and the
reverberation as

\[
\mathbf{x}(\omega, t) = \mathbf{A}(\omega)\mathbf{s}(\omega, t) + \mathbf{A}_e(\omega)\mathbf{s}(\omega, t) + \mathbf{n}(\omega, t). \tag{1}
\]

In the first term, the \( M \times N \) matrix \( \mathbf{A}(\omega) \) is termed
the mixing matrix, its \((m,n)\) element, \( A_{m,n}(\omega) \), being
the transfer function of the direct path from the \( n \)th
source to the \( m \)th microphone as

\[
A_{m,n}(\omega) = |A_{m,n}(\omega)| e^{-j\omega \tau_{m,n}}. \tag{2}
\]

The symbol \( |A_{m,n}(\omega)| \) is the magnitude of the trans-
fer function. The symbol \( \tau_{m,n} \) denotes the propaga-
tion time from the \( n \)th source to the \( m \)th microphone.
Vector \( \mathbf{s}(\omega, t) \) consists of the source spectra as
\( \mathbf{s} = [S_1(\omega, t), \ldots, S_N(\omega, t)]^T \). The first term, \( \mathbf{A}(\omega)\mathbf{s}(\omega, t) \),
expresses the direct components.

The second term \( \mathbf{A}_e(\omega)\mathbf{s}(\omega, t) \) denotes the early re-
flexion. In this paper, the early reflection is defined as a
portion of the reflections whose delay relative to the
direct sound is within the window length of STFT.
Generally, the early reflection has multiple paths. Thus,
the element of \( \mathbf{A}_e(\omega) \) will take the form

\[
A_{e_m,n,i}(\omega) = \sum_i |A_{e_m,n,i}(\omega)| e^{-j\omega \tau_{m,n,i}}, \tag{3}
\]

where the subscript \( i \) denotes the path number. Since
\( A_e(\omega)\mathbf{s}(\omega, t) \) is a filtered replica of \( \mathbf{s}(\omega, t) \), it is highly
correlated with the direct sound \( \mathbf{A}(\omega)\mathbf{s}(\omega, t) \).

The reverberation is defined as the rest of the reflec-
tions with a delay greater than the window length.
Based on this definition, the reverberation term, \( \mathbf{n}(\omega, t) \),
can be written as

\[
\mathbf{n}(\omega, t) = \sum_d \mathbf{A}_r(\omega, d)\mathbf{s}(\omega, t - d) \tag{4}
\]

where \( \mathbf{A}_r(\omega, d) \) is the mixing matrix for the reverberation.
As shown in (4), \( \mathbf{n}(\omega, t) \) consists of the filtered
replica of the signal in the previous frames and, thus, has small- or zero-coherence with the direct sound and the early reflection. A typical example of a situation with small coherence is a consonant frame overlapped by the reflection of a previous vowel. Therefore, \( n(\omega, t) \) functions rather as random additive noise. Also, since \( n(\omega, t) \) includes a large number of reflections, its directivity and, hence, the coherency between the microphones is assumed to be low.

3. BSS SYSTEM

A block diagram of the BSS system previously proposed by the authors [6] is depicted in Fig. 1. First, the subspace method is applied to the input vector \( x(\omega, t) \) to obtain the subspace filter \( W(\omega) \). In this stage, reverberation is reduced in advance of the application of ICA. It should be noted that the node of the filter network is reduced from \( M \) to \( N \) in this stage as depicted in Fig. 1.

Then, the instantaneous ICA is applied to the output of the subspace stage, \( y(\omega, t) = W(\omega)x(\omega, t) \) to obtain the filter matrix \( U(\omega) \). In this paper, the Infomax algorithm with feed-forward architecture [5] extended to complex data [4] is used. The learning rule is written as

\[
U(\omega, t+1) = U(\omega, t) + \eta \left[ I - \varphi(z(\omega, t)) z^H(\omega, t) \right] U(\omega, t) \tag{5}
\]

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

\[
\varphi(z) = 2 \tanh(G \cdot \text{Re}(z)) + 2j \tanh(G \cdot \text{Im}(z)) \tag{6}
\]

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

For the sake of convenience, the product of \( W(\omega) \) and \( U(\omega) \),

\[
B(\omega) = W(\omega)U(\omega), \tag{7}
\]

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 \( P(\omega) \) and the scaling matrix \( B_m(\omega) \). The scaling matrix, \( B_m(\omega) \) is an \( N \times N \) diagonal matrix \( B_m(\omega) = \text{diag}(B_{m,1}^+, \cdots, B_{m,N}^+) \) where \( B_{m,n}^+ \) denotes the \((m, n)\)th element of the pseudoinverse of \( B(\omega) \) [4]. A method of obtaining the permutation matrix \( P(\omega) \) has been proposed by the authors [6].

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

\[
F(\omega) = P(\omega)B_m^+(\omega)B(\omega). \tag{8}
\]

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

4. SUBSPACE METHOD

In this section, the subspace method is introduced to reduce the room reverberation. For the sake of simplicity, the early reflection \( A_e(\omega)s(\omega, t) \) is omitted in this section.

4.1. Properties of Spatial Correlation Matrix

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

\[
R = AQA^H + K \tag{9}
\]

where \( Q = E[s(t)s^H(t)] \) and \( K = E[n(t)n^H(t)] \). According to the model described in Section 2, this assumption holds to some extent in a practical sense.

By taking the generalized eigenvalue decomposition of \( R \) as \( R = KEA E^{-1} \), we have the eigenvector matrix \( E = [e_1, \cdots, e_M] \) and the eigenvalue matrix \( \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_M) \), where \( e_m \) and \( \lambda_m \) are the eigenvector and the eigenvalue, respectively. Based on the model in Section 2, the eigenvalues and eigenvectors have the following properties [2, 7]:

P1: The energy of the \( N \) directional signals \( s(t) \) is concentrated on the \( N \) dominant eigenvalues.

P2: The energy of \( n(t) \) is equally spread over all eigenvalues.
P3: $\Re(A) = \Re(E_s)$.

P4: $\Re(A) = \Re(E_n)^\perp$.

The matrices $E_s = [e_1, \ldots, e_N]$ and $E_n = [e_{N+1}, \ldots, e_M]$ consist of the eigenvectors for the $N$ dominant eigenvalues and those for the other $M-N$ eigenvalues, respectively. The notation $\Re(A)$ denotes the space spanned by the column vectors of $A$. The notation $\Re(E_n)^\perp$ denotes the orthogonal complement of $\Re(E_n)$. The subspaces $\Re(E_s)$ and $\Re(E_n)$ are termed signal subspace and noise subspace, respectively.

4.2. Subspace Filter

The subspace filter is defined as

$$ W = \Lambda_s^{-1/2} E_s^H $$

where $\Lambda_s = \text{diag}(\lambda_1, \ldots, \lambda_N)$. According to the properties P1 and P3, the directional component $A_s(t)$ belongs to the signal subspace. On the other hand, using the properties P2-P4, the ambient component $n(t)$ can be split as $n(t) = n_s(t) + n_n(t)$ where $n_s(t)$ and $n_n(t)$ denote the components which belong to the signal subspace and the noise subspace, respectively. Due to the orthogonality in P4, the component $n_n(t)$ in the noise subspace is canceled out by the subspace filter $W$. When there is early reflection, the above discussion holds by replacing $A(\omega)$ with

$$ A'(\omega) = A(\omega) + A_s(\omega) $$

5. BEHAVIOR ON EARLY REFLECTIONS

In this section, reverberation is omitted for the sake of simplicity. In this case, the mixing system becomes $x(\omega, t) = A'(\omega)s(\omega, t)$. In this section, the solution of ICA for this mixing system is compared with the conventional inverse problem and their behaviors are analyzed.

6. EXPERIMENT

6.1. Experimental Condition

A signal separation experiment was conducted in a ordinary meeting room with a reverberation time of 0.5 s. The configuration of the sound sources (loudspeakers) and the microphone array is depicted in Fig. 2. The microphone array was circular in shape with a diameter of 0.5 m and $M = 8$. In Fig. 2, the early reflection is low in Condition A while it is high in Condition B.
6.2. Reverberation

Figure 3 shows the spectra of the reverberation at the input/output of the system (normalized by the input spectrum) for Condition A. From this, it can be seen that the reverberation was reduced by 10-15 dB by the subspace method. Figure 4 shows the results of the automatic speech recognition (ASR). From the results for Condition A, it can be seen that the ASR score was improved by around 15% by the subspace method.

6.3. Early Reflection

Figure 5 shows the impulse response from source #1 to output channels #1 (straight) and #2 (cross) of the coupled system $A'(\omega)B(\omega)$. For ease of viewing, the reverberation is omitted when calculating the impulse responses. From Fig. 5(c), it can be seen that the cross-talk is almost canceled out by ICA. On the other hand, from Fig. 5(b), it can be seen that not only the direct sound but also the early reflection was passed through the straight channel as expected from the discussion in Section 5.

From the results of ASR in Fig. 4, it can be seen that the performance of separation for Condition B is similar to that of Condition A, and the early reflection remaining at the straight channel has little effect on the ASR rate.

7. CONCLUSION

The performance of BSS using ICA in a reflective environment was investigated. As for the reverberation, the performance of BSS was improved by 15-20% by combining ICA with the subspace method. As for the early reflection, the cross-talk for the early reflection is also canceled out by ICA. In the straight channel, early reflection may remain but has little effect on ASR.

8. REFERENCES


