FAST SOURCE SEPARATION BASED ON SELECTION OF EFFECTIVE TEMPORAL FRAMES

Y. Mizuno\textsuperscript{1}, K. Kondo\textsuperscript{1,2}, T. Nishino\textsuperscript{3}, N. Kitaoka\textsuperscript{1}, K. Takeda\textsuperscript{1}
\textsuperscript{1}Nagoya University JAPAN, \textsuperscript{2}Yamaha Corporation JAPAN, \textsuperscript{3}Mie University JAPAN

Abstract
A faster computational method for performing frequency domain independent component analysis (FDICA) using a dodecahedral microphone array is proposed. Source separation with FDICA uses the spectrum of observed signals and estimates separation filters for each frequency. However, this technique is complex and requires high computational resources. In this paper, a method of selecting temporal frames which are effective for training the separation filters is proposed and evaluated. The log power spectrum and the kurtosis of amplitude distribution are employed as selection criteria. Experimental results showed that the proposed method could achieve faster computation with lower computational complexity.

Outline of entire separation processes

Observed signals

1. Subspace method
The dimension of input signals is reduced to the number of desired separation signals.

2. Temporal frame selection
Effective temporal frames are selected.

3. Iterative estimation with ICA
Separation matrices are trained so that the following covariance matrix would be a diagonal matrix.

\[ R_x(f) = E_x[\Phi(Y(f, \tau))Y^H(f, \tau)] \]
\( \Phi \) : non-linear transformation

4. Solving permutation problem
Permutation problem is solved by an arrival direction using the DHMA. [Ogasawara, 2011]

Proposed method

Temporal frames that have many sources and signal power are assumed to be effective.

Source number is evaluated by kurtosis \( K \) of amplitude distribution.

\[ K = E\left(\frac{X - \mu}{\sigma}\right)^4 \]
\( \mu \) : mean \( \sigma \) : standard deviation

- Kurtosis can evaluate the non-Gaussianity of signals. [Hyvarinen, 2001]
- The more sources are mixed the smaller non-gaussianity of the sources become. (cf. Central limit theorem)

Examples of the kurtosis value

(a) One source
(b) Three sources
(c) Twelve sources

Kurtosis becomes smaller for mixtures of larger number of sources.

Selection value \( C \)

\[ C = \beta \overline{P} - (1 - \beta) \overline{K} \]

Normalized power \( \overline{P} = \frac{P - \mu_P}{\sigma_P} \)

Normalized kurtosis \( \overline{K} = \frac{K - \mu_K}{\sigma_K} \)

Temporal frames with high \( P \) and low \( K \) are selected.

Source separation experiment
We compared the following two methods.
1. Separation using beginning frames (Conventional method)
2. Separation using selected frames (Proposed method)

Experimental conditions

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling frequency</td>
<td>40 kHz</td>
</tr>
<tr>
<td># of microphones</td>
<td>60</td>
</tr>
<tr>
<td>Sound source</td>
<td>12 speech sound sources (observed signal is 4 sec)</td>
</tr>
<tr>
<td>SIR improvement</td>
<td>138 ms</td>
</tr>
<tr>
<td>Reverberation time</td>
<td>(Low reverberation room)</td>
</tr>
<tr>
<td>Window size</td>
<td>1024 points (25.6 ms)</td>
</tr>
<tr>
<td>Weighting coefficient</td>
<td>( \beta = 0.4 )</td>
</tr>
</tbody>
</table>

Experimental set-up

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of selected frames</td>
<td>400</td>
</tr>
<tr>
<td>Spectrogram</td>
<td></td>
</tr>
</tbody>
</table>

Experimental results

- Separation quality
- Spectrogram
- Processing time

Future works
- Examining criteria for selection of temporal frames using other features of observed signals.
- Implementation of a real-time multi-source separation system using a dodecahedral microphone array.

Summary
- FDICA source separation using dodecahedral microphone array.
- Reducing computational complexity method was proposed.
- Proposed method was very effective and could reduce computational complexity without degrading separation quality.