Separation matrix optimization using associative memory model for blind source separation

Motoi Omachi, Tetsuji Ogawa and Tetsunori Kobayashi (Waseda University, Japan); Masaru Fujieda and Kazuhiro Katagiri (Oki Electric Industry Co., Ltd., Japan)

Abstract

Objective:
- Linear blind source separation (BSS) yielding high quality and less distorted speech

Approach:
- Optimization of a linear separation matrix using neural network-based associative memory model (AMM)

Result (Simultaneous speech separation):
- Residual distortion caused by independent vector Analysis (IVA) can be reduced.

Conventional linear BSS (e.g., ICA, IVA)

- Estimate linear separation matrix defeating effect of mixing matrix.
- Assume that source signals are statistically independent.
- Does not take account of property of source signals.

Proposed linear BSS

- Separation matrix is estimated by iterating following two steps:
  1. Reference signal estimation using AMM
  2. Separation Matrix optimization

Source signals \( S \)
<table>
<thead>
<tr>
<th>Mixing matrix ( A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed signals ( Z )</td>
</tr>
<tr>
<td>Separated signals ( Y )</td>
</tr>
</tbody>
</table>

Z = AS

\[ Y = WZ \]

Algorithm

1. Reference signal estimation using AMM
   - Convolutional neural network can remove residual distortion included in spectra of separated signal.

2. Separation Matrix optimization
   - Separation matrix is modified using a linear projection matrix.
     \[
     \begin{bmatrix}
     \tilde{Y}_1(\omega, \tau) \\
     \tilde{Y}_2(\omega, \tau)
     \end{bmatrix} = \begin{bmatrix}
     M_{11}(\omega) & M_{12}(\omega) \\
     M_{21}(\omega) & M_{22}(\omega)
     \end{bmatrix} \begin{bmatrix}
     Y_1(\omega, \tau) \\
     Y_2(\omega, \tau)
     \end{bmatrix}
     \]
   - Projection matrix is optimized with a gradient descend method.
     \[
     f(\omega) = \sum_{n=1}^{N_s} \sum_{l=1}^{N_f} \log |S_n(\omega, l)|^2 - \log |M_{nj}Y_j(\omega, l)|^2
     \]

Source separation experiment

Acoustic Environment

Dataset: Speech materials
- Training: Japanese phoneme balanced sentence database [4 females]
- Dev.: 9 speaker pairs, 50 utterances for each pair
- Test: 1 speaker pair, 52 utterances

Speech-to-distortion ratio (SDR)
- Our method can reduce residual distortions caused by IVA

Phoneme error rate (DNN/ triphones + bigram)
- Our method can improve speech recognition performance