Multi-Shift Principal Component Analysis based Primary Component Extraction for Spatial Audio Reproduction

Jianjun He, and Woon-Seng Gan \{jhe007@e.ntu.edu.sg, ewsgan@ntu.edu.sg\}

Digital Signal Processing Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

### Motivation

**Objective:** to obtain a new representation of sound scenes in digital media, which is both flexible and efficient in spatial audio reproduction for any playback systems.

- Existing sound scene representations:
  - Channel-based
    - Conventional, for a specific playback system;
    - Lacks the flexibility to support different playback configurations.
  - Object-based
    - Emerging, for any playback system;
    - Lacks the efficiency: large storage and high transmission bandwidth.

- Primary-ambient based representation
  - Inspired by human auditory system;
  - Facilitates flexible and efficient rendering of immersive spatial audio.

- Primary-ambient extraction (PAE) from the channel-based audio (e.g., stereo).
  - Existing approaches: mainly for one dominant source in primary components;
  - Subband techniques: problematic for overlapping spectra;
  - PAE with multiple sources (different directions) not well studied.

### Stereo Signal Model

<table>
<thead>
<tr>
<th>Signal = Primary + Ambient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary correlated $p_i = i p_i$</td>
</tr>
<tr>
<td>Ambient uncorrelated $a_i \perp a_i$</td>
</tr>
<tr>
<td>Primary ambient uncorrelated $a_i \perp p_i$</td>
</tr>
<tr>
<td>Ambient power balanced $p_i = p_i$</td>
</tr>
</tbody>
</table>

### PCA based PAE

- **To account for the partial primary correlation (0-lag) caused by the inter-channel time difference (ICTD) $\tau$.**

### Shifted PCA based PAE

- **To account for the partial primary correlation (0-lag) caused by the inter-channel time difference (ICTD) $\tau$.**

### Experimental Settings

**Primary components:** speech and music;
**Ambience:** white Gaussian noise; Equal power among speech, music, ambience; ICTD range: ±500 lags, (-2ms for fs=44.1 kHz);

- Approaches:
  - PCA, SPCA;
  - MSPCA-T, MSPCA ($a = 2, 10$)

### CONCLUSIONS

1. Proposed multi-shift PCA to handle multiple sources in primary component extraction;
2. MSPCA with typical structure (selected shifts), but its performance is degraded when ICTD estimation is inaccurate;
3. MSPCA with consecutive structure is more robust, by applying weights on every shifted versions.
4. The weighting method for different shifts is critical; in general, applying a proper exponent of the ICC yields good (objective and subjective) performance.