Exploiting non-negative matrix factorization for binaural sound source localization in the presence of directional interference

Ingvi Örnolfsson*, Torsten Dau*, Ning Ma† and Tobias May*

*Department of Health Technology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
†Department of Computer Science, University of Sheffield, Sheffield S1 4DP, United Kingdom
Why study sound localization?

- Beamforming for hearing aids
- Social robots
- Auditory scene analysis
Task to solve:
- Predict azimuth angle $\phi$ of a target speech source
- An additional masking source will be present at a different angle

Problem:
- The directional masking source provides competing ITD and ILD cues

Solution:
- Introduce a separation stage that indicates where speech is dominating the mixture
Research questions

1. Which performance metric should be used for NMF validation?

2. What material should be used for NMF training?

3. Can NMF beat a learning-free approach?

- NMF-based separation will be compared the APAB algorithm

Summary of algorithms

<table>
<thead>
<tr>
<th>NMF variants</th>
<th>Validation metric</th>
<th>Training material</th>
<th>Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMF</td>
<td>PC</td>
<td>Anechoic</td>
<td></td>
</tr>
<tr>
<td>NMF-SDR</td>
<td>SDR</td>
<td>Anechoic</td>
<td></td>
</tr>
<tr>
<td>NMF-REV</td>
<td>PC</td>
<td>Reverberant</td>
<td></td>
</tr>
</tbody>
</table>

Baselines

- **Loc. Only**: Separation stage is excluded.
- **Oracle**: Separation is based on oracle information about source activity.
- **APAB**: Separation is obtained via a variance-based noise reduction technique.
## Evaluation methodology

<table>
<thead>
<tr>
<th></th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targets</td>
<td>TIMIT-TRAIN</td>
<td>TIMIT-TEST</td>
</tr>
<tr>
<td>Maskers</td>
<td>NatNoises-VAL</td>
<td>NatNoises-TEST &amp; ICRA</td>
</tr>
<tr>
<td>Rooms</td>
<td>Surrey anechoic</td>
<td>All five Surrey rooms</td>
</tr>
<tr>
<td>Azimuths</td>
<td>-90° to 90°, 5° steps</td>
<td></td>
</tr>
<tr>
<td>Long-term TMRs</td>
<td>-15dBA to 15dBA, 5dBA steps</td>
<td></td>
</tr>
<tr>
<td>Evaluation metric</td>
<td>PC or SDR, depending on algorithm</td>
<td>PC</td>
</tr>
</tbody>
</table>

SDR: See reference [11] in the paper*  
PC: 1 if correct azimuth, 0 otherwise
Results

Stationary noise (ICRA01)

Nonstationary noise (ICRA05)
Results

Stationary noise (ICRA01)

Nonstationary noise (ICRA05)
Conclusion - research questions revisited

1. Which performance metric should be used for NMF validation?

   ![](diagram1.png)

2. What material should be used for NMF training?

   ![](diagram2.png)

3. Can NMF beat a learning-free approach?
   - In nonstationary noise, yes!
   - In stationary noise, NMF is slightly worse than the APAB algorithm.
Study further variations on the NMF
Different source separation strategies: DNN, ICA, etc.
What about a different combination stage?
Use human performance as a baseline
Thank you for your time!

Please join us for our poster session:

**AUD-8: Audio and Speech Source Separation 4: Multi-Channel Source Separation**

Wednesday, 9 June from 13:00 to 13:45 in Eastern Daylight Time