SOURCE SEPARATION FOR DIALOGUE ENHANCEMENT IN BROADCAST APPLICATIONS

Jouni Paulus / 23.5.2018
Why Dialogue Enhancement in broadcast?

- Low dialogue intelligibility is major source for complaints
  - SS-GB: ca. 100 complaints / 24 hours [1]
  - Wonders of the Universe [2] similar numbers
    - Re-mixed and re-broadcasted
- Still intelligible, but “too loud background”
- DE: Allow end-user to adjust dialogue / background mixing ratio [3]

Why to adjust the mixing level?

- Hearing impairments, >50% of 65-years [1]
- Listening environment
- Non-native language [2], dialect, pronunciation
- Personal preferences

Personal preferences

Providing Dialogue Enhancement functionality

- Object-based audio
  - Transport dialogue and background as a separate tracks
  - Mixing at the receiver
  - E.g., MPEG-H
- Legacy content with only mixes stored
- Source separation?
  - Armstrong: "From the above examples it can be seen that current audio processing techniques cannot significantly improve the intelligibility of speech in noise, if at all." [1]

Source separation for broadcast content

- Split mixture into estimates of dialogue and background
- Work on speech enhancement
- Various principles
  - Spatial location, e.g., “centered dialogue”
  - Dictionaries, e.g., semi-supervised NMF
- One tool for all?
- Deep neural networks
  - Fusion [1] [2]
  - Separation [2]

Challenging content

- Number of channels
- Background sounds
- Mixing process
- Dialogue itself

Difficult to obtain high separation quality for everything
Required amount of separation

- Moderate mixing ratio adjustments are enough
  - Cohen (BBC): 1.4 dB change [1]
  - Brand: +20% intelligibility / dB [2]
- Dialogue enhancement: *adjust* mixing ratio
- Benefit: mixing hides artifacts
- Trade-off: audio quality vs. adjustment

Quality vs. adjustment

### Example

- Beginning sequence of “Wonders of the Universe”, BBC, 2011

<table>
<thead>
<tr>
<th></th>
<th><strong>Stereo</strong></th>
<th><strong>Mono</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Original mix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft adjustment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard adjustment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016 Re-mix</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- Need for dialogue enhancement
  - Intelligibility
  - Aesthetic preferences
- Object-based audio in broadcast
- Source separation for legacy content for obtaining objects
- Imperfect separation not necessarily a problem

Source separation enabling DE application for legacy content