

---

# 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]

[1] <http://www.bbc.co.uk/news/entertainment-arts-39038406>

[2] <https://www.theguardian.com/tv-and-radio/tvandradioblog/2011/mar/15/tv-background-music-wonders-of-the-universe>

[3] Fuchs, H. and Oetting, D., “Advanced Clean Audio Solution: Dialogue Enhancement”, SMPTE Motion Imaging Journal, 2014.



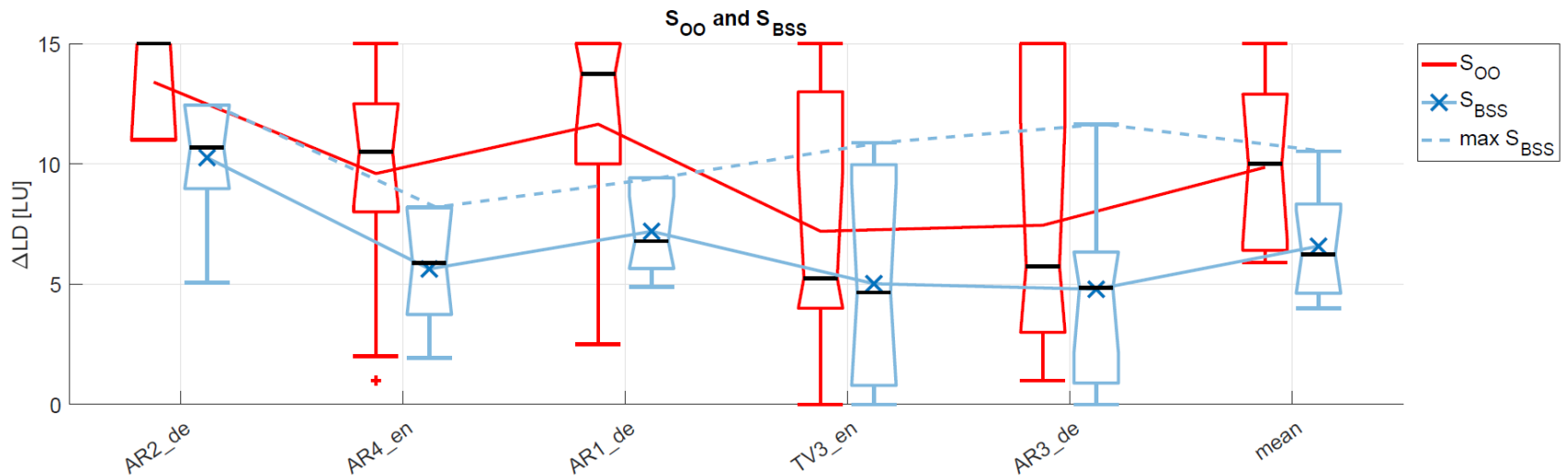
# Why to adjust the mixing level?

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

[1] Shirley, B., and Oldfield R., "Clean Audio for TV broadcast: An Object-Based Approach for Hearing-Impaired Viewers", JAES, vol. 63, no 4, 2015.

[2] Warzybok A, et al., "Influence of the linguistic complexity in relation to speech material on non-native speech perception in noise", DAGA, 2010.

# Personal preferences



Torcoli M., et al., "The Adjustment / Satisfaction Test (A/ST) for the Evaluation of Personalization in Broadcast Services and its Application to Dialogue Enhancement", IEEE Transactions on Broadcasting. 2018.



# 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]

[1] Armstrong M., "Audio Processing and Speech Intelligibility: a literature review", BBC Research White Paper WHP190, April 2011

---



# 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]

[1] Manilow E., et al. “Predicting algorithm efficacy for adaptive multi-cue source separation”, WASPAA 2017.  
[2] Grais, E. M., et al., “Single Channel Audio Source Separation using Deep Neural Network Ensembles”, 140th AES Convention, Paris, 2016.



# Challenging content

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

➔ Difficult to obtain high separation quality for everything



# Required amount of separation

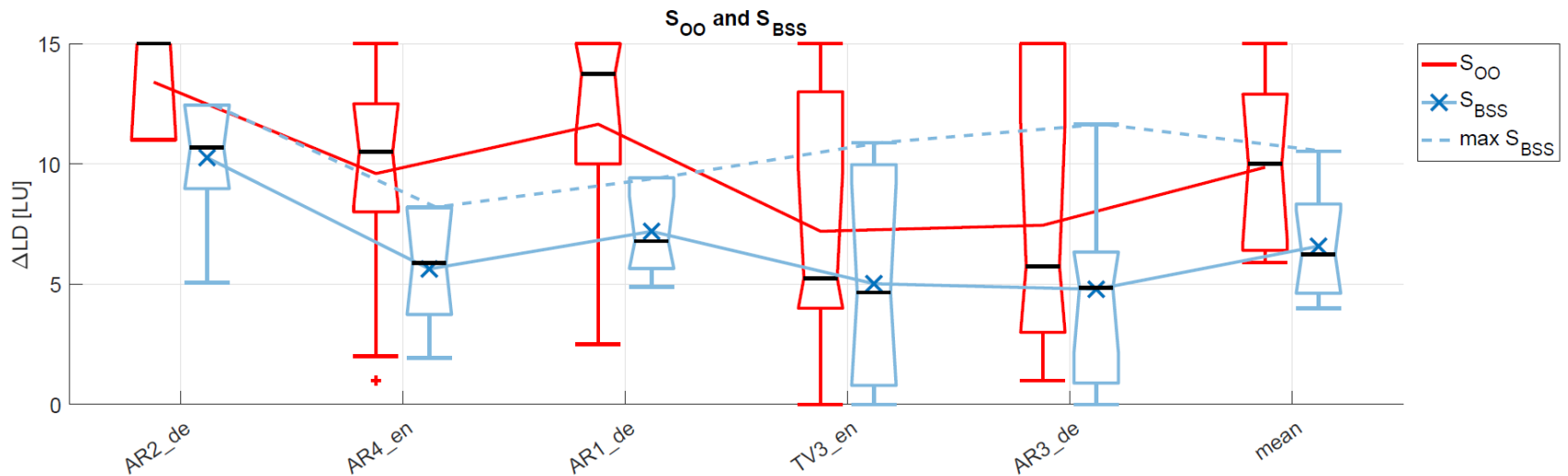
- Moderate mixing ratio adjustments are enough
  - Cohen (BBC): 1.4 dB change [1]
  - Brand: +20% intelligibility / dB [2]
- Dialogue enhancement: *adjust* mixing ratio
- Benefit: mixing hides artifacts
- Trade-off: audio quality vs. adjustment

[1] <http://www.bbc.co.uk/blogs/tv/2011/03/is-the-background-music-too-loud.shtml>

[2] Brand, T., "Analysis and optimization of psychophysical procedures in audiology", PhD thesis, University of Oldenburg, 2000.










# Quality vs. adjustment



Torcoli M., et al., "The Adjustment / Satisfaction Test (A/ST) for the Evaluation of Personalization in Broadcast Services and its Application to Dialogue Enhancement", IEEE Transactions on Broadcasting. 2018.

# Example

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

	Stereo	Mono
Original mix		
Soft adjustment		
Hard adjustment		
2016 Re-mix		



# 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