Perceptual Evaluation of Source Separation for Remixing Music

H. Wierstorf [1] D. Ward [1] E. M. Grais [1] M. D. Plumbley [1]
R. Mason [2] C. Hummersone [2]

[1] Centre for Vision, Speech and Signal Processing, University of Surrey
[2] Institute of Sound Recording, University of Surrey

143rd AES Convention

UNIVERSITY OF SURREY

Source separation for music

Reference: vocals others mixture

Source separation: vocals others

How to talk about source separation?

- **Sound quality**: artifacts and distortion added
- **Interference**: not perfect separation achieved

Source separation for music

How to evaluate source separation?

- **BSS eval**: signal decomposition and energy ratios[1]
- PEASS: signal decomposition and auditory model[2]

Open questions

- Correlation with perception has been questioned[3]

[1] Vincent, et al. (2006), *IEEE TASLP*, doi: 10.1109/TSA.2005.858005
[2] Emiya, et al. (2011), *IEEE TASLP*, doi: 10.1109/TASL.2011.2109381
[3] e.g. Gupta, et al. (2015), *WASPAA*, doi: 10.1109/WASPAA.2015.7336923

BSS eval

Decompose signal into different components

$$s_{\text{estimated}} = s_{\text{original}} + e_{\text{interferer}} + e_{\text{artifacts}}$$

$$\text{SAR} = 10 \log_{10} \frac{||s_{\text{original}} + e_{\text{interferer}}||^2}{||e_{\text{artifacts}}||^2}$$

$$\text{SIR} = 10 \log_{10} \frac{||s_{\text{original}}||^2}{||e_{\text{interferer}}||^2}$$

Source separation for music

Reference: vocals others mixture

Source separation: vocals others

How to talk about source separation?

- **Sound quality**: artifacts and distortion added
- **Interference**: not perfect separation achieved

Source separation for music

Reference: vocals others mixture

Source separation: vocals others mixture

How to talk about source separation?

- **Sound quality**: artifacts and distortion added
- **Interference**: not perfect separation achieved

Remixing using source separation

- Modify component levels[4]
- Change positions (upmix)[5]
- Change frequency content[6]
- Add effects[7]
- Mashups

[4] Itoyama, et al. (2009), *ISMIR*, pp. 133–138
[5] Cobos, et al. (2008), *ISCCSP*, doi: 10.1109/ISCCSP.2008.4537423
[6] Yoshii, et al. (2005), *WASPAA*, doi: 10.1049/ic.2005.0733
[7] Woodruff, et al. (2006), *ISMIR*, pp. 314–319

Evaluation of remixes

- Evaluate the actual remix
- Problem if only asked for preference or naturalness[8]
- Enable for adjustment by listeners[9]
- Trade-off between artifacts and level increase[10]
- Predictions with BSS eval?

[8] Gillet and Richard (2005), *WASPAA*, doi: 10.1109/ASPAA.2005.1540232
[9] Yoshii, et al. (2005), *WASPAA*, doi: 10.1049/ic.2005.0733
[10] Pons, et al. (2016), *JASA*, doi: 10.1121/1.4971424

Experiment

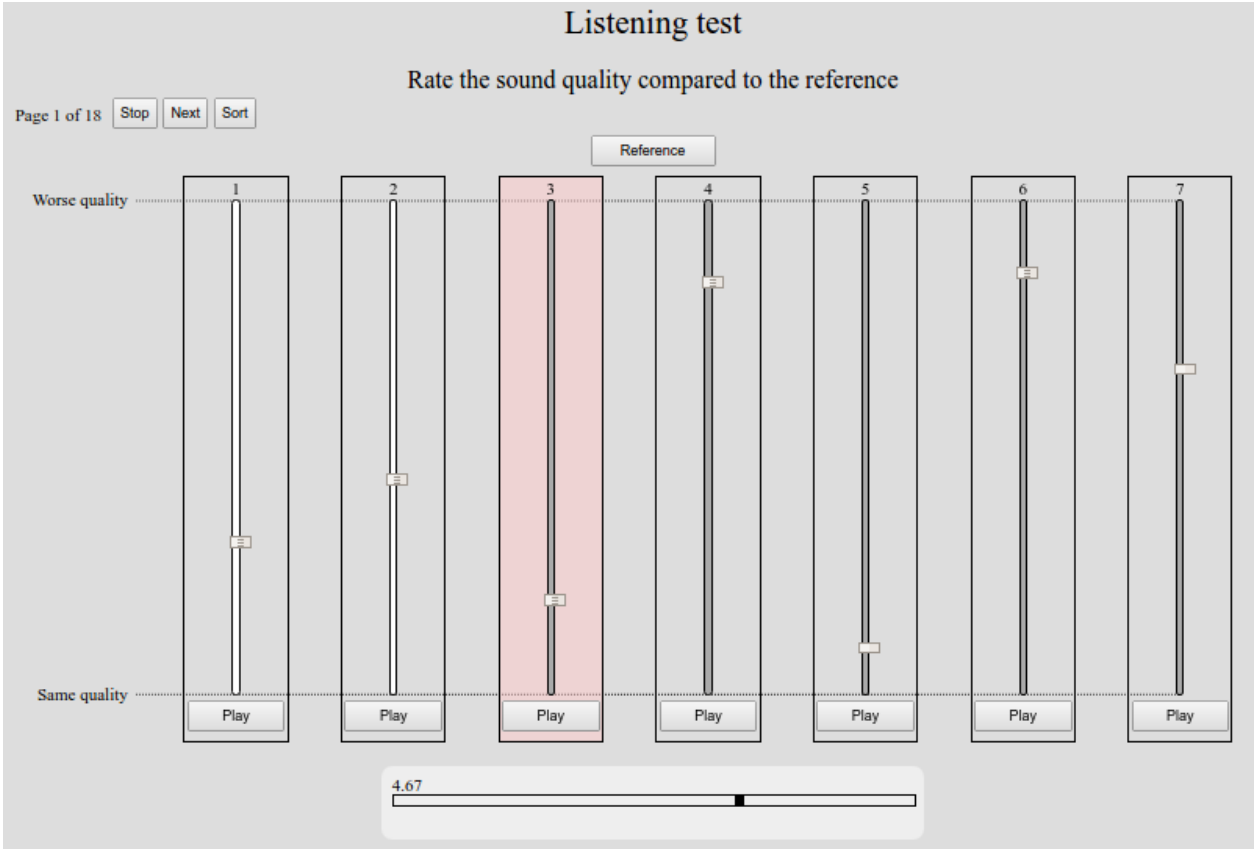
- Start with reference mix
- Introduce changes in level of vocals
- Rate **sound quality** and **loudness balance**
- Look for correlations with SAR and SIR

Experiment

Loudness balance describes the relation of the overall loudness of the vocals to the overall loudness of the remaining instruments. It does not include short and abrupt changes in loudness that you might experience for some test sounds. It is more considered with the general balance of the vocals and the accompanying instruments.

Experiment

MUSHRA inspired experiment using Web Audio Evaluation Tool[11]



[11] Jillings, et al. (2015), *SMC*, github: BrechtDeMan/WebAudioEvaluationTool

Experiment

- 2 tasks: **sound quality** and **loudness balance**
- 5 source separation algorithms
- 6 songs (converted to mono)
- 3 remixes, level of vocal (0 dB, 6 dB, 12 dB)
- 3 anchor and references for every task
- loudness anchor: vocals −14 dB
- quality anchor: artifacts, distortions, 3.5 kHz low pass
- 15 participants

Stimuli

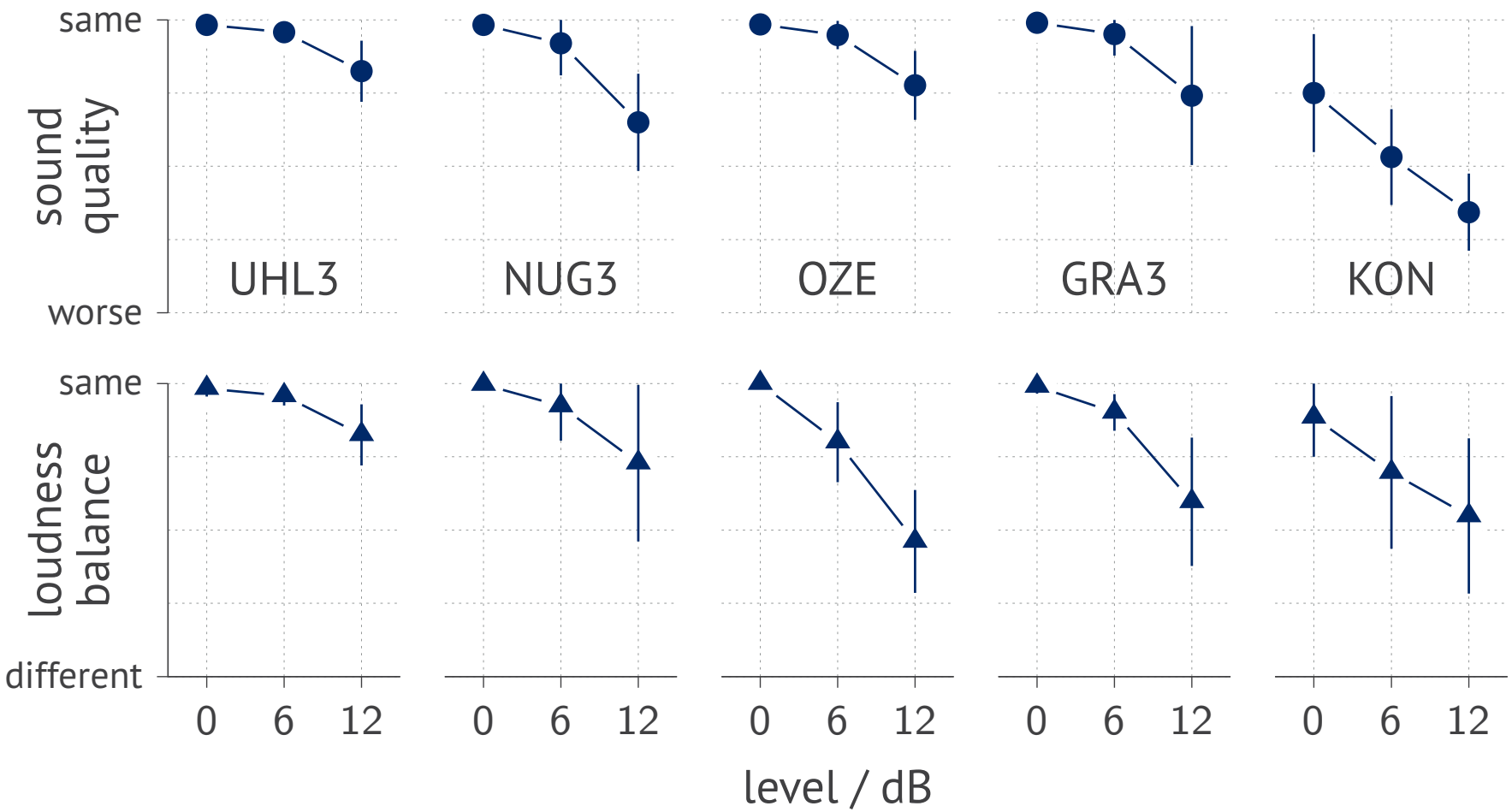
- Signal separation evaluation campaign (SiSEC)[12]
- The MUS task includes 23 algorithms and 100 mixed songs[13]

SAR:	7.7	6.1	2.8	6.3	−3.4
SIR:	10.2	11.1	8.8	6.2	7.0
Vocal:	UHL3	NUG3	OZE	GRA3	KON

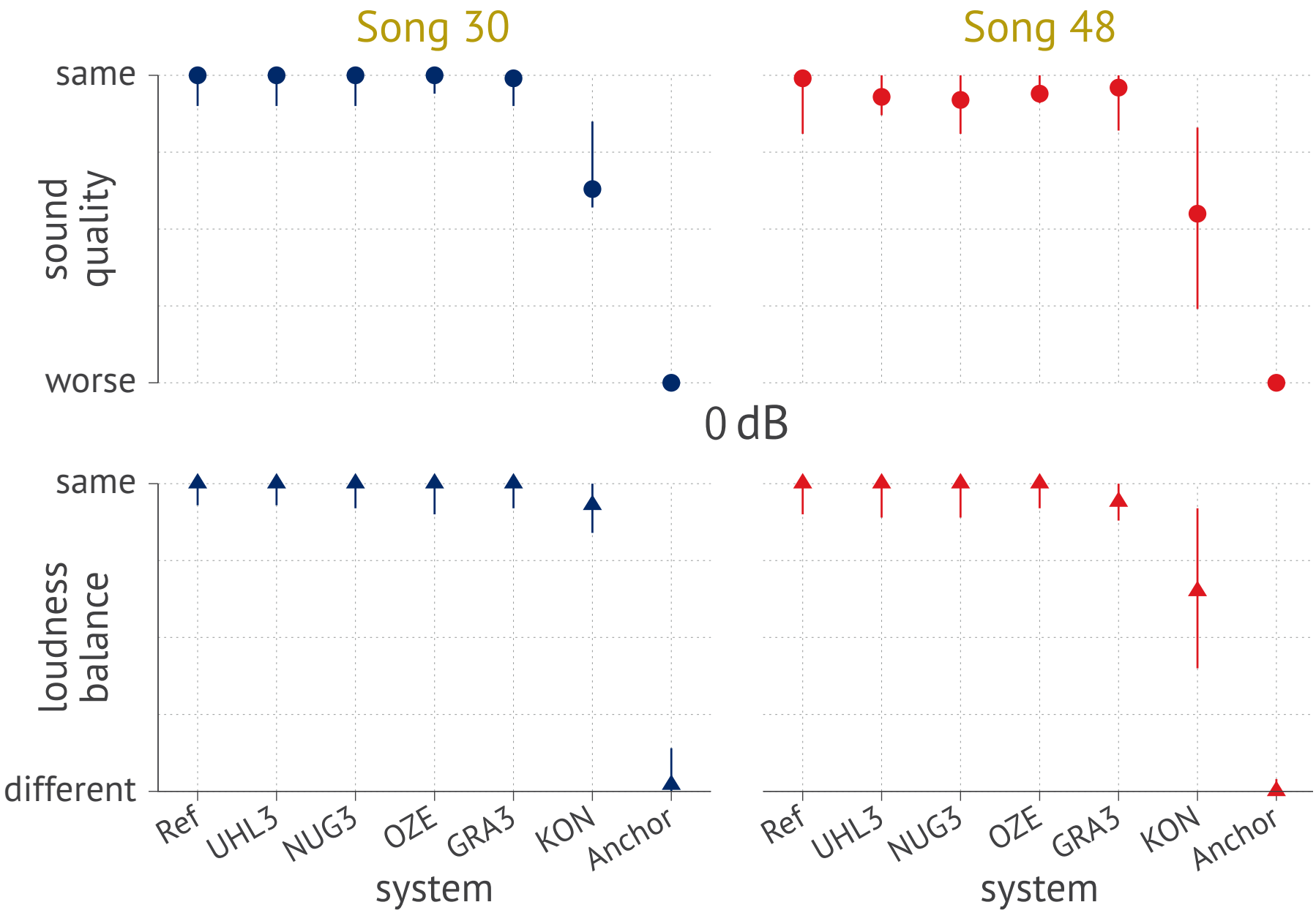
[12] Liutkus, et al. (2017), *LVA/ICA*, doi: 10.1007/978-3-319-53547-0_31
[13] https://www.sisec17.audiolabs-erlangen.de

Results

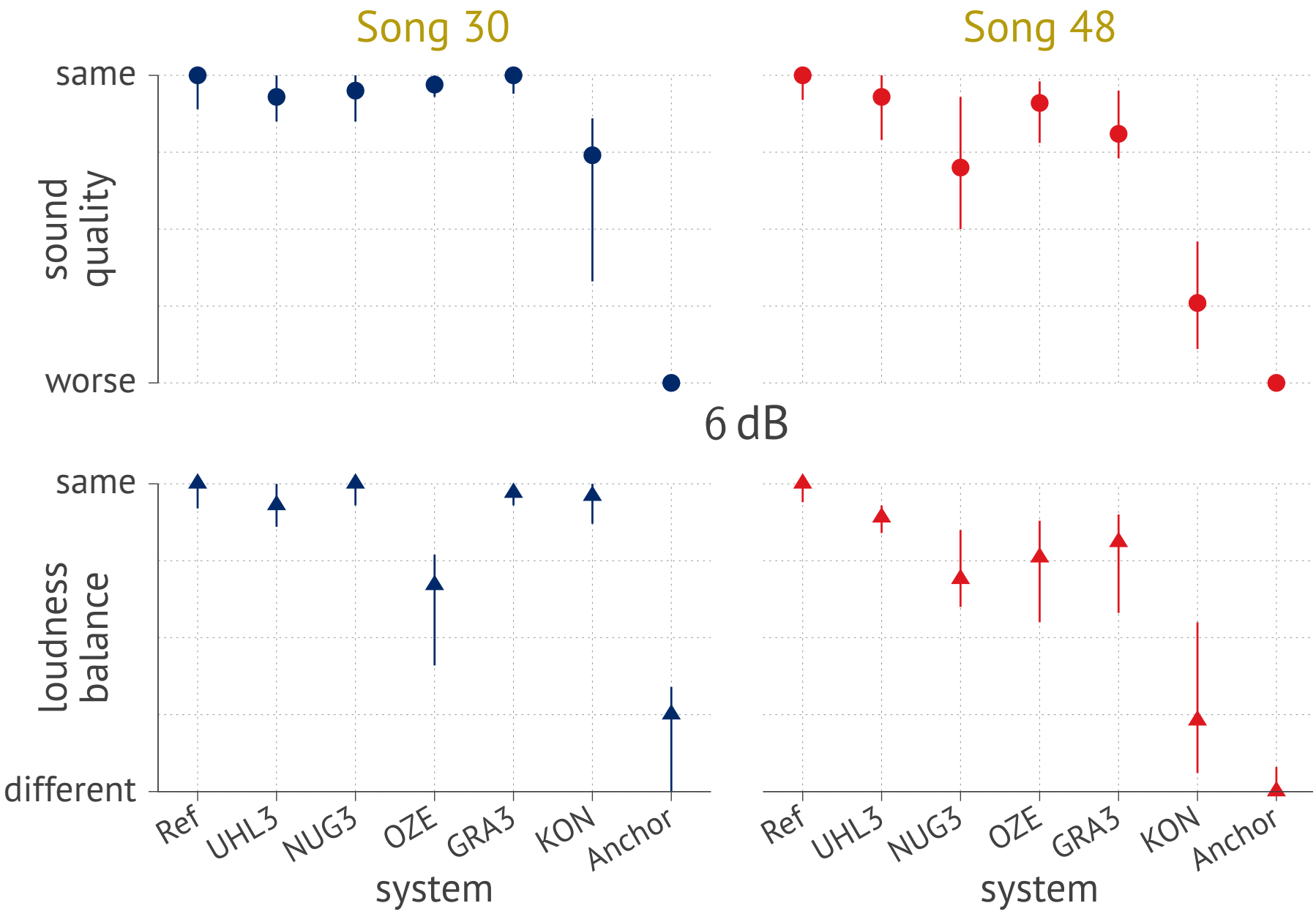
Average across medians of every song



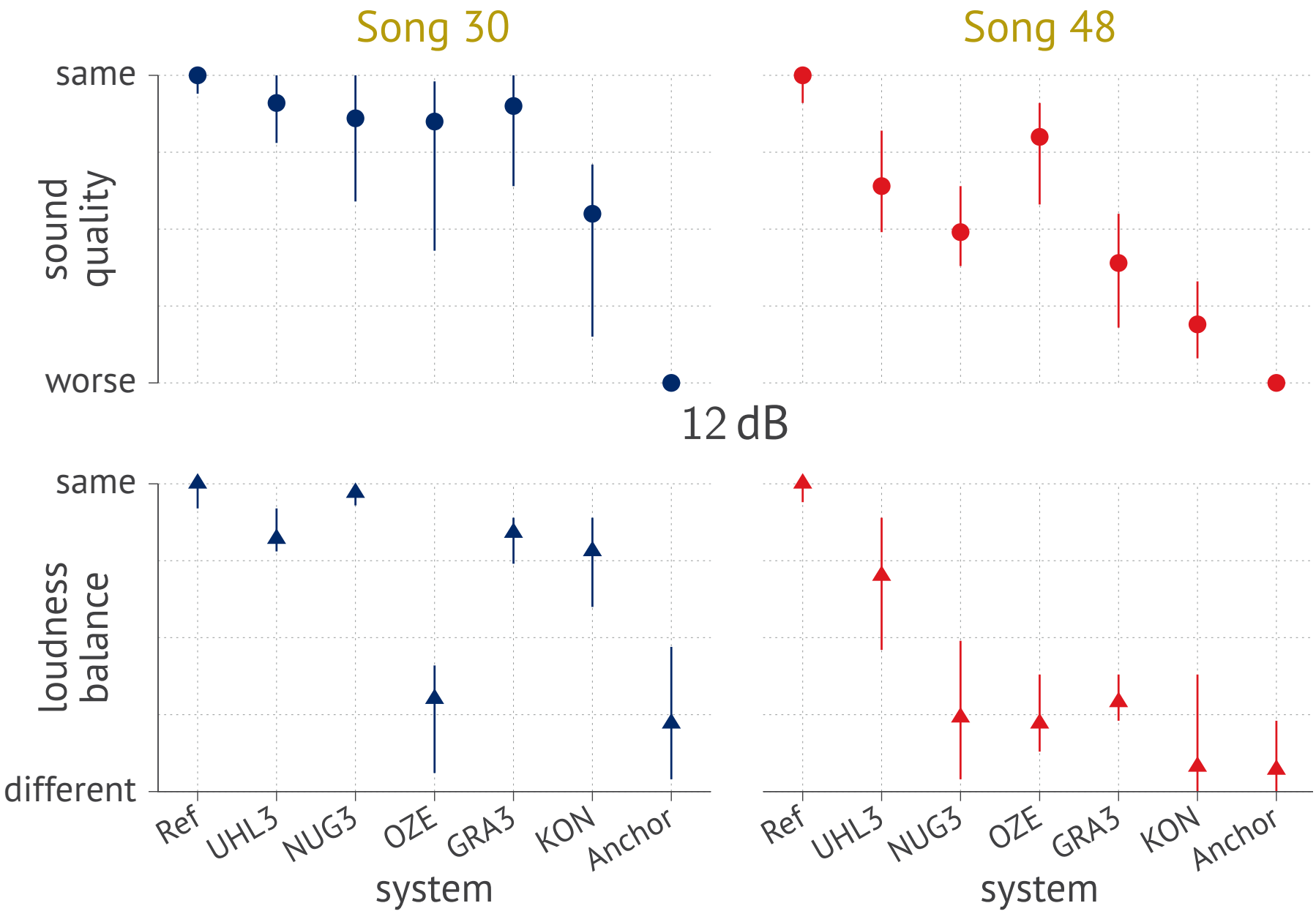
Influence of song



Influence of song



Influence of song

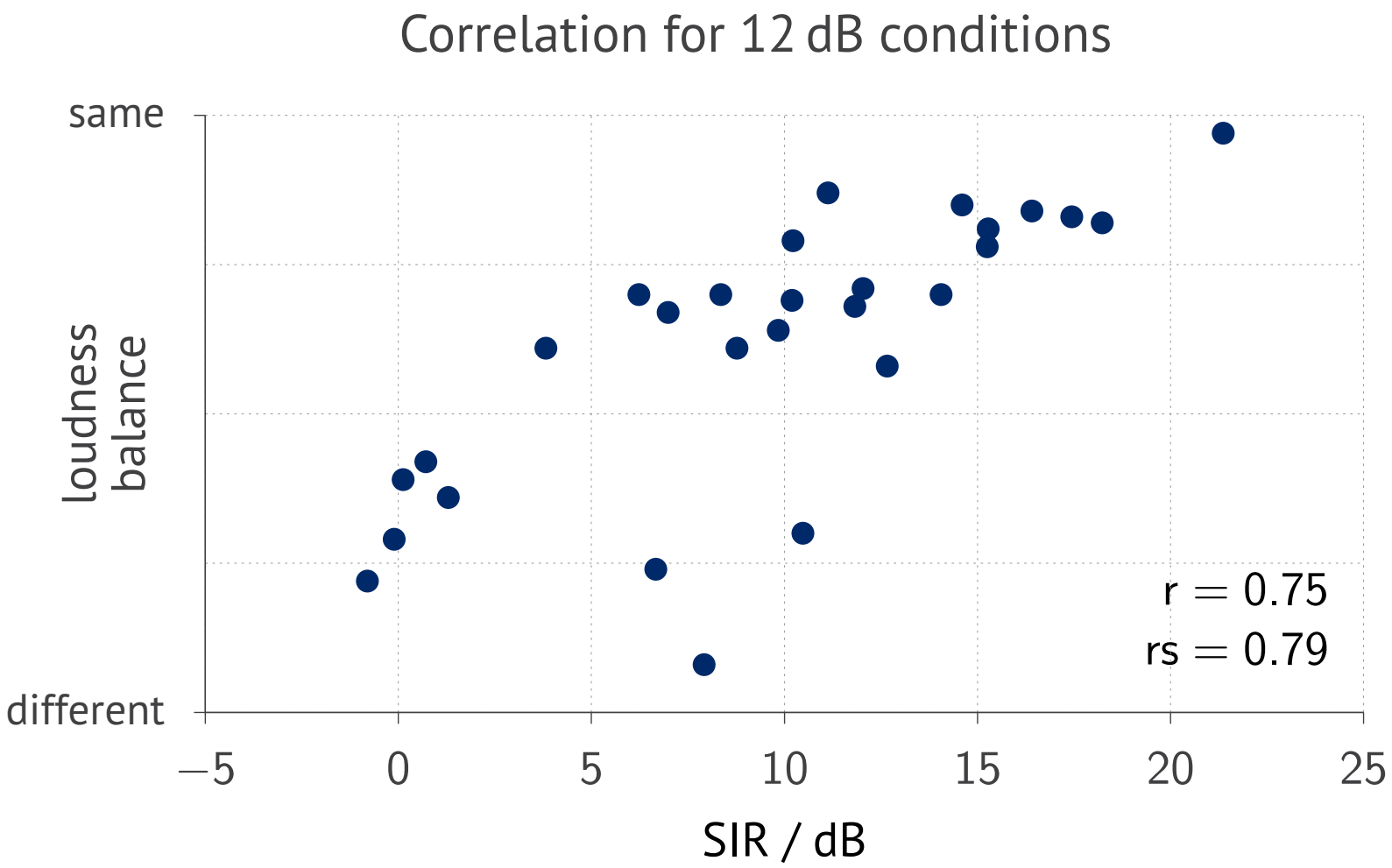


Influence of song

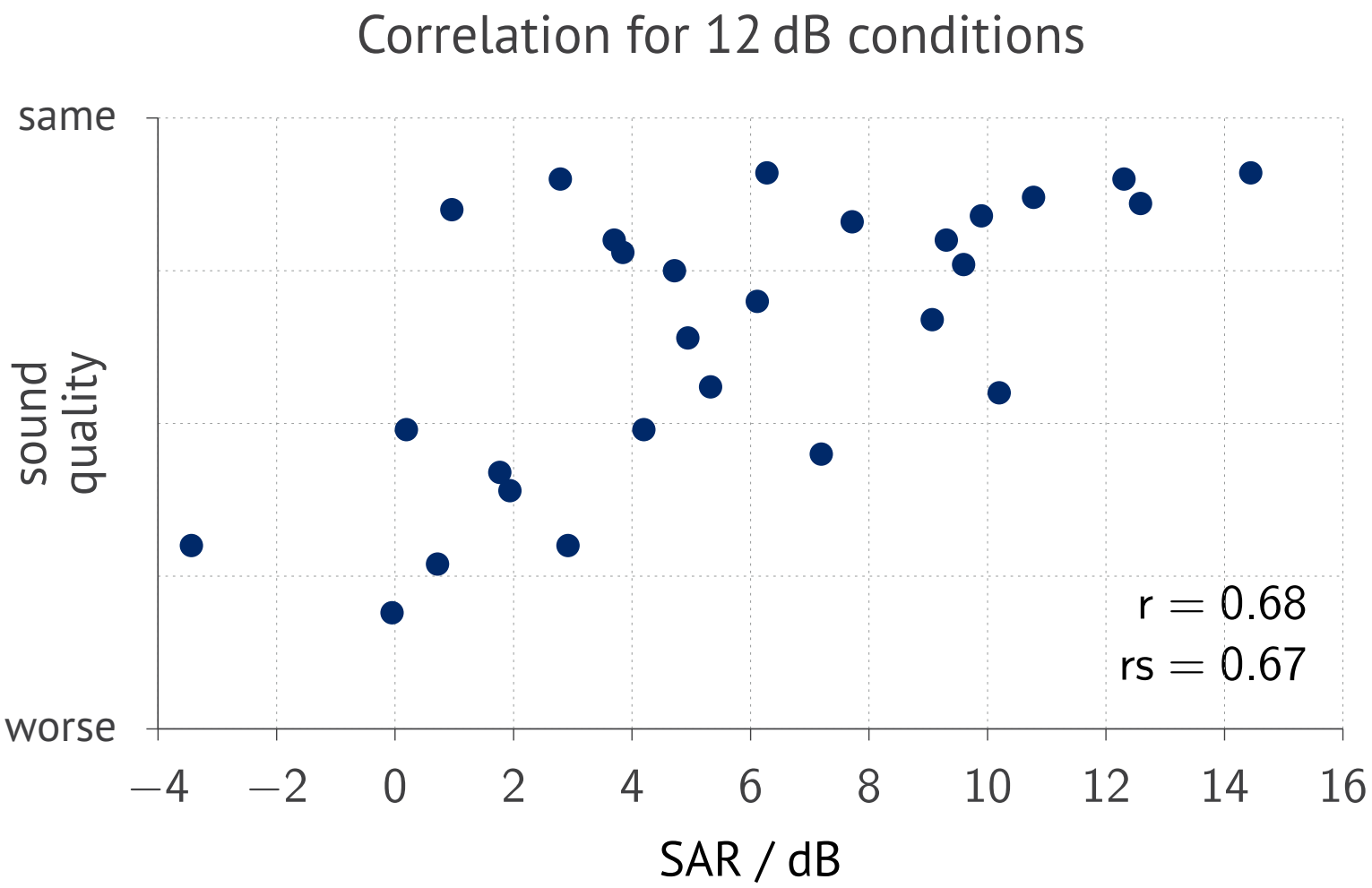
Connected to level balance of original mix?

- Song 30, level balance: 1.7 dB
- Song 48, level balance: -5.7 dB
- Weak correlation with both results for 12 dB
- Two songs were worse in level balance than song 48

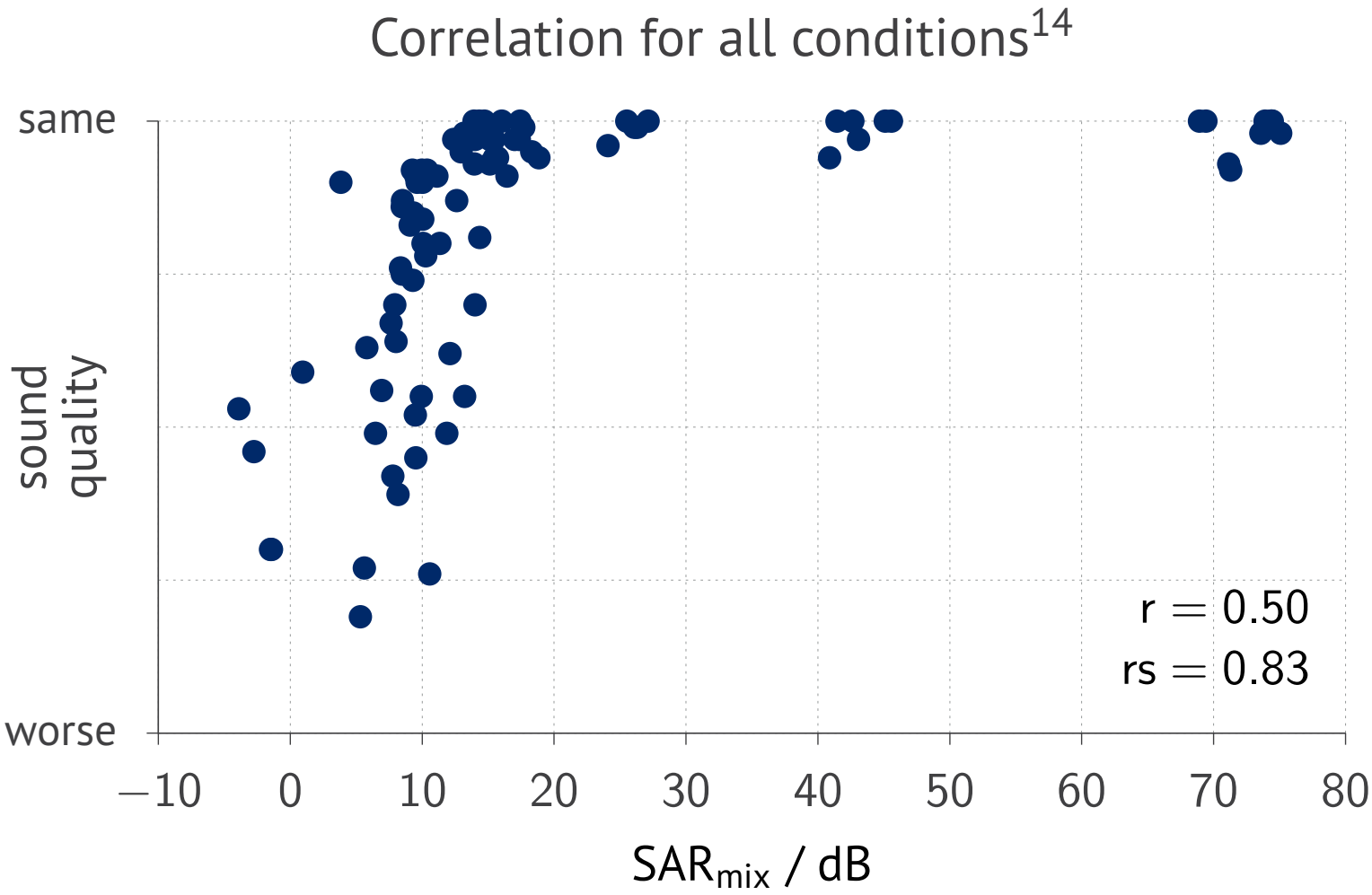
BSS eval and remixes



BSS eval and remixes



BSS eval and remixes



[14] Liu et al. (2015), *EUSIPCO*, doi: 10.1109/EUSIPCO.2015.7362551

Conclusions

- Source separation methods suitable for level remixing
- Trade off between achievable level and sound quality
- Maximum reachable level
- BSS eval can be used to pick algorithm
- Connection to adjustment experiments?

https://hagenw.github.io

http://cvssp.org/events/lva-ica-2018