

HRTF Individualised Mag-LS and COMPASS Ambisonics-To-Binaural Rendering: Overall Perceived Quality for Pre-Conditioned Listeners

Frank Schultz¹, Shaimaa Doma², Sascha Spors¹, Janina Fels²

¹Institute of Communications Engineering, University of Rostock, Germany

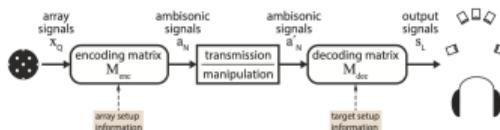
²Institute for Hearing Technology and Acoustics, RWTH Aachen, Germany

154th AES Convention - AES Europe 2023
Spatial Audio Session, May 15th 11am EEST

Higher-Order-Ambisonics-To-Binaural Rendering (HOA2Bin) for 3DOF



→ non-parametric decoder



→ parametric decoder

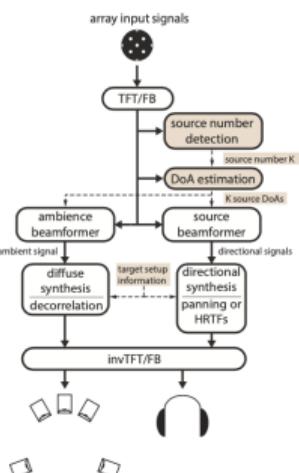


Fig. 4.2, 5.2 from Politis, 2016

→ separate acquisition of sound fields and of HRTFs

→ dynamic rendering via HOA scene rotation

→ convenient selection of different HRTF datasets

→ use of individual or individualised HRTFs aka **personalised HOA2Bin**

→ do individual HRTFs outperform at lower order HOA2Bin ?

→ in this study we ask for **overall perceived quality**, cf. **HRTF selection scenario**

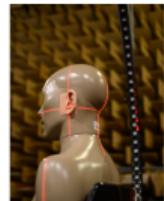
Experiment Preparation I

- measurements
- post-processing
- @ Aachen Labs
- with ITA Toolbox

→ 2 HATS

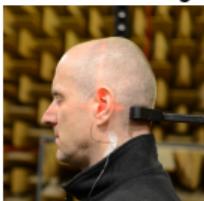


Achen



KEMAR

→ 13 subjects



HRTFs



Schultz, Doma et al.

Four screenshots of the ITA Toolbox software interface showing different decoding and processing plugins:

- SPARTA| Rotator
- SPARTA| AmbiBIN
- COMPASS| Binaural
- SPARTA| MultiConv

Each plugin has various controls for parameters like Order, Format, Norm, and various rotation and convolution settings.

leomccormack.github.io/sparta-site/, v1.6.2

A screenshot of the ABComparison VST plugin interface. It shows a list of choices (6) and channels (2 ch). The "Switch mode" is set to "Exclusive Solo". The "FadeTime" is 100 ms. There are six buttons labeled:

- My HRTF max HOA (highlighted in green)
- My HRTF current HOA
- KEMAR HATS
- Achen HATS
- Random Human 1
- Random Human 2

github.com/DanielRudrich/ABComparisonPlugin, v1.4.0

- VST host Reaper
- controlled via OSC / reapy
- from Jupyter Notebook GUI
- Supperware HeadTracker 1
- Sennheiser HD650 / RME FF UC
- ITU Rec. BS.1770 SPL calibrated

Experiment Preparation II

spatial-aliasing free HOA scenes via shoebox room modelling with RAVEN

female speech (German = native)

10 s = 5 short sentences

empty **seminar room**

($10 \times 7 \times 4$) m³, T₆₀ ≈ 0.6 s

speaker lowered, slightly off-centred

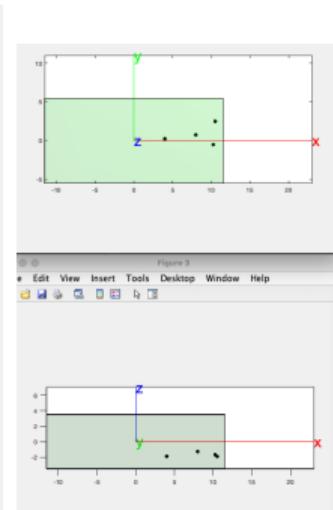
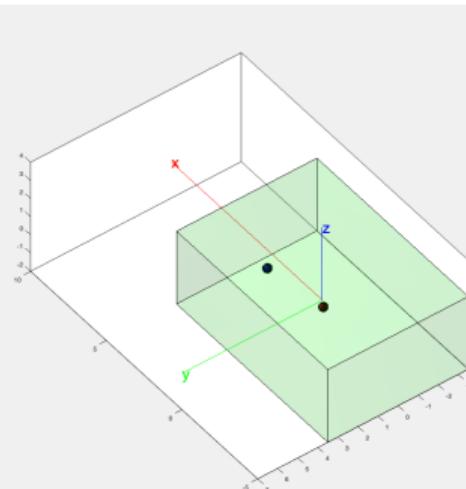
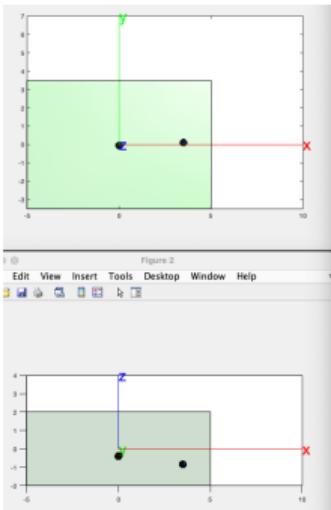
music trio (double bass, clarinet, guitar)

11 s jazz standard (Therry and Katz, 2019)

empty chamber music **concert hall**

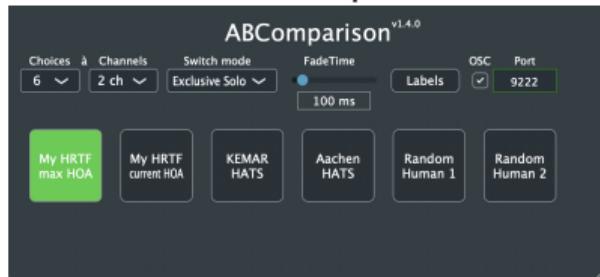
($23 \times 11 \times 7$) m³, T₆₀ ≈ 1.3 s

solo clarinet elevated, all slightly off-centred

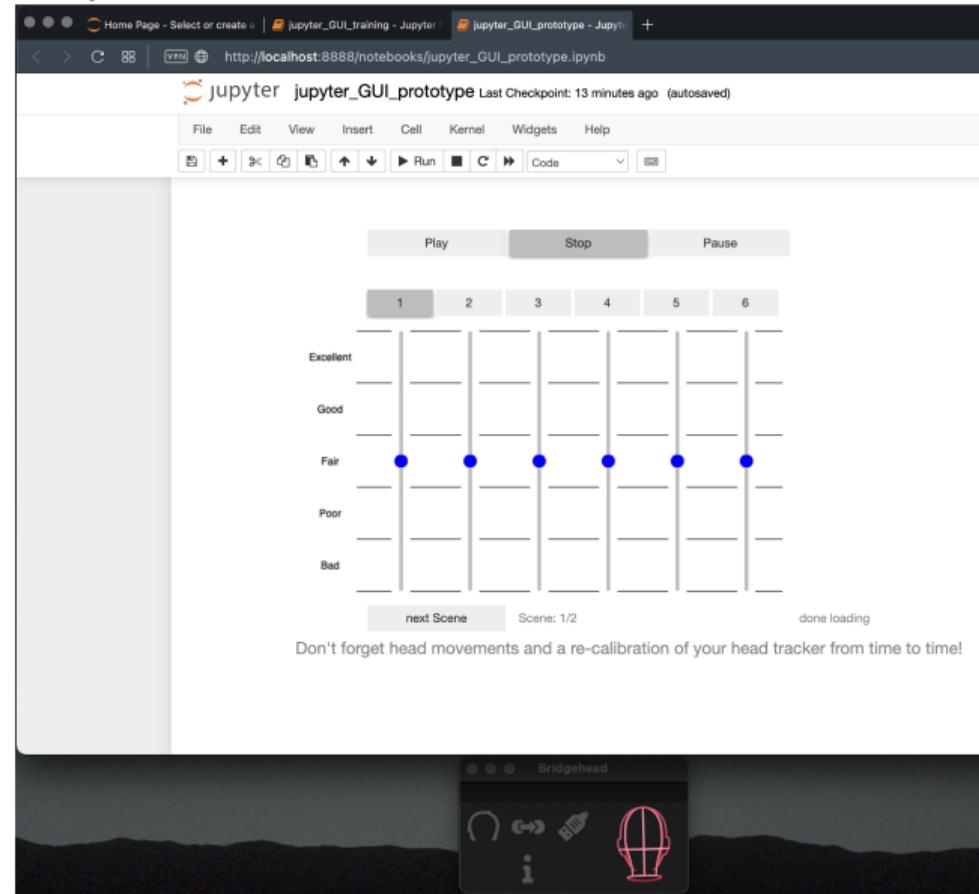


Listening Test Design - General Aspects, cf. Schultz et al., 2023b

- rate overall perceived quality
- 6 HRTFs for each panel



- like a **HRTF selection paradigm**
- **no anchor, no reference**
- **panel conditions:**
 - HOA music vs. HOA speech
 - Mag-LS vs. COMPASS
 - HOA Order: 7, 3,2,1
- thus $(2 \times 4 + 2 \times 3) \times 6$
= (14 panels) \times 6 ratings per test



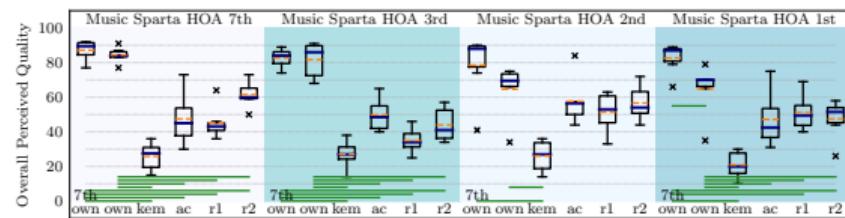
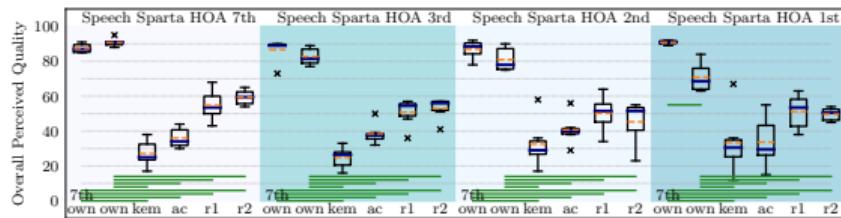
Listening Test Design: Pre-Conditioning / Bias for One Test Subject

in this paper Schultz et al., 2023a:

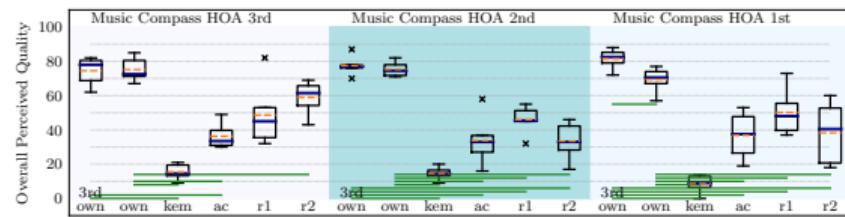
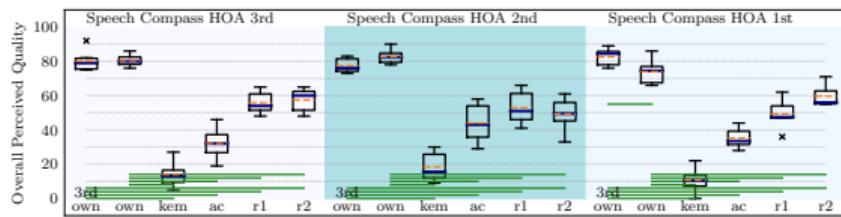
- 1 pre-conditioned, biased, trained participant performed 6 complete tests, thus: $14 \times 6 \times 6$ quality ratings
- participant had profound expectation for highest externalisation and localisation quality, producing highest load to overall perceived quality, followed by timbre, room impression and their dynamic aspects
- participant had – although randomized presentation order – knowledge about HOA-to-binaural rendering variations (Mag-LS vs. Compass, different HOA orders)
- participant had – although randomized presentation order – knowledge about HRTF variations (2x own, 2x generic, 2x random human)
- this pre-condition might shift the HRTF selection to a HRTF detection paradigm

Results Overview: Ratings for Overall Perceived Quality

→ non-parametric rendering: Sparta MagLS (Schörkhuber, Zaunschirm, and Höldrich, 2018)



→ parametric rendering: COMPASS (Politis, Tervo, and Pulkki, 2018)



→ robust statistics for dependent group comparisons with rmmcppb() (Wilcox, 2022)

→ overall $\alpha = 0.05$ level per panel (for 15 comparisons), no comparisons between panels

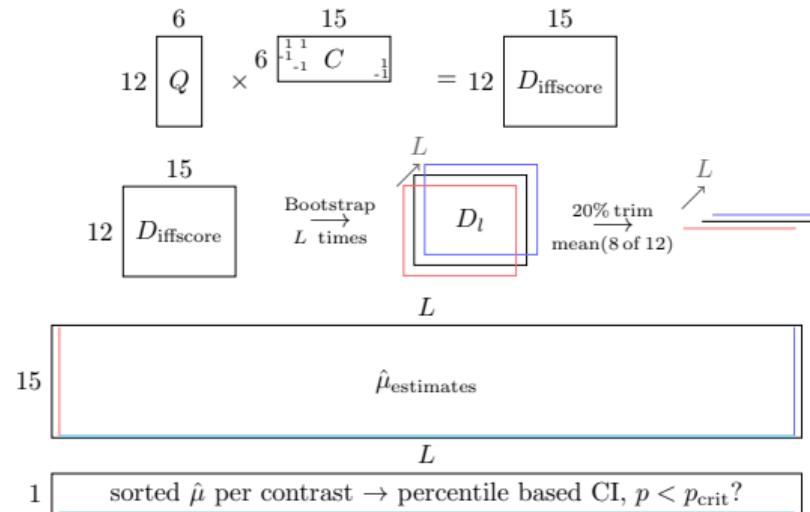
→ green bars indicate only the significant results with respect to individual HRTFs

Robust statistics for dependent group comparisons rmmcppb()

see Wilcox, 2022; Campopiano and Wilcox, 2020

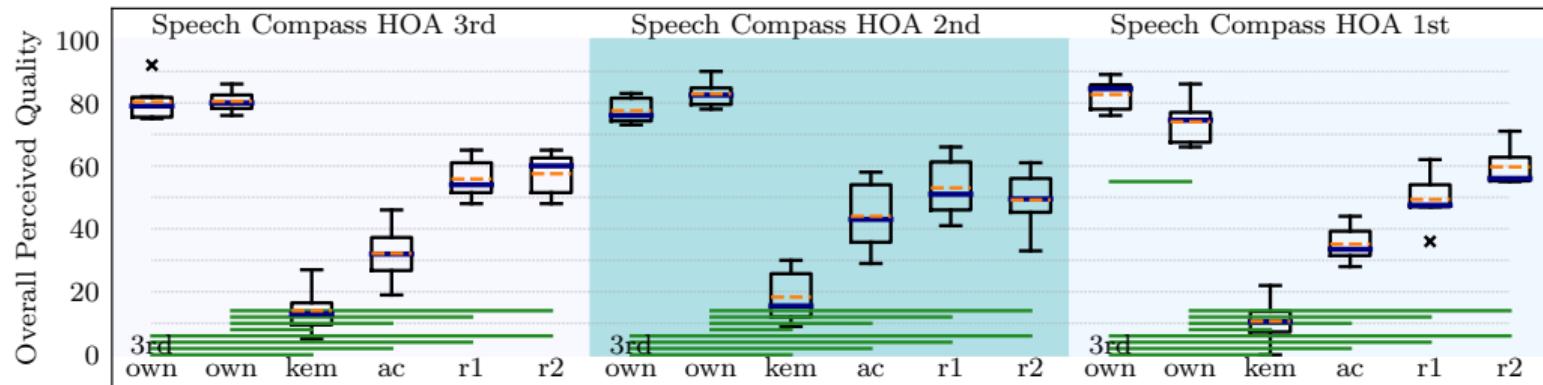
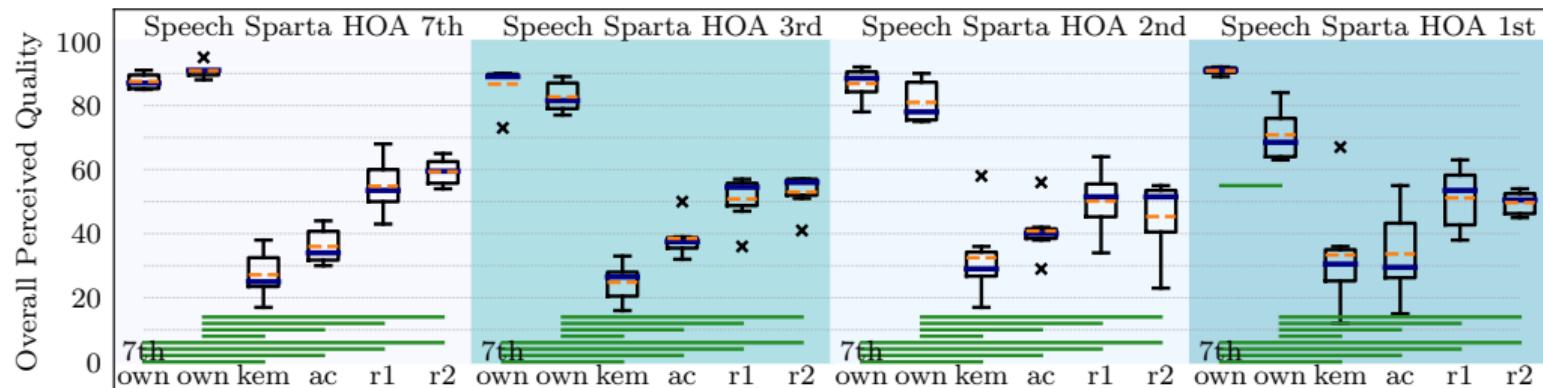
percentile bootstrap with 20% trimmed mean and $\alpha_{\text{total}} = 0.05$ per panel (note the typo in paper Schultz et al., 2023a: *for each group comparison*, which is obviously wrong)

in the example below: 12 subjects on 6 HRTFs, i.e. 15 possible group comparisons per panel

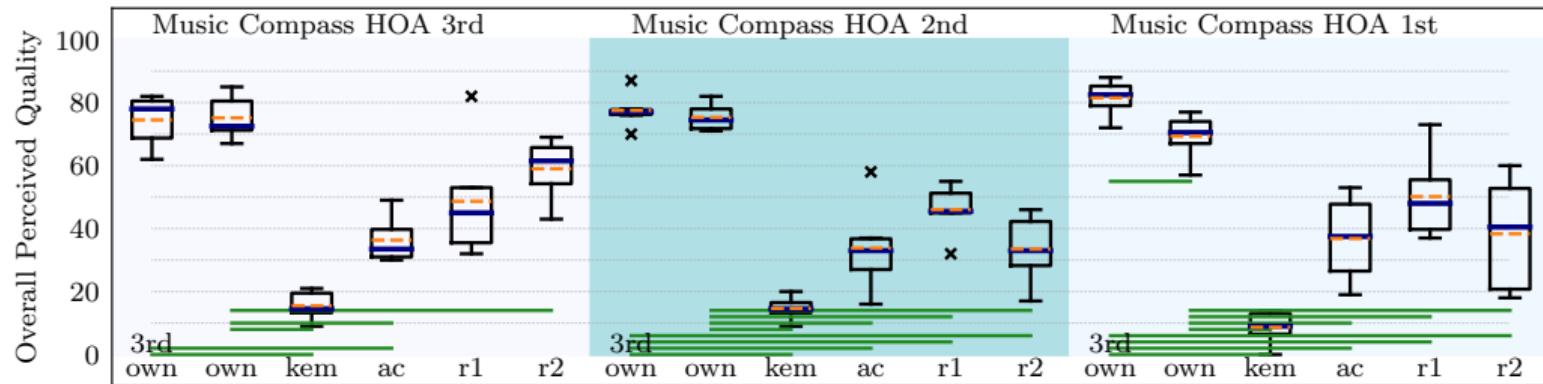
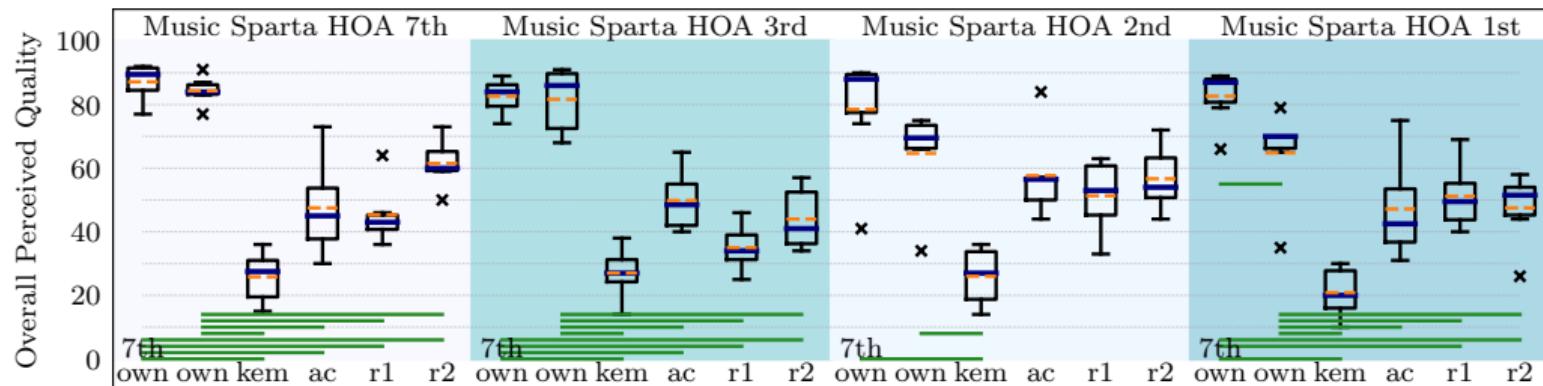


we do this for all 14 panels in our study

Results Speech: Ratings for Overall Perceived Quality

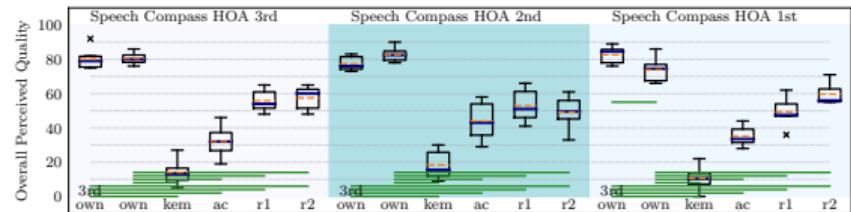


Results Music: Ratings for Overall Perceived Quality



Conclusion

- overall perceived quality for personalised HOA2Bin
- no anchor & no reference design, cf. HRTF selection scenario
- today: 6 test repetitions for one biased, trained participant, cf. detection scenario?!
- significant higher quality for individual HRTFs than for generic / random human HRTFs
- need of individual HRTFs?! yes / no... depends on application / audio scene / multi-modality / expectation / training / listening test design / HpTF EQing ...



[project repo](#) (raw data and statistical evaluation, paper and slides):

<https://github.com/spatialaudio/paper-aes154-individual-hrtf-hoa2binaural>

[GUI minimal example](#)

<https://doi.org/10.5281/zenodo.7912849>

References

-  Campopiano, Allan and Rand R. Wilcox (2020). "Hypothesize: Robust Statistics for Python". In: *J. Open Source Softw.* 5.50, p. 2241.
-  Politis, Archontis (2016). "Microphone array processing for parametric spatial audio techniques". PhD thesis. Aalto University.
-  Politis, Archontis, Sakari Tervo, and Ville Pulkki (2018). "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes". In: *Proc. IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP), Calgary*, pp. 6802–6806.
-  Schörkhuber, Christian, Markus Zaunschirm, and Robert Höldrich (2018). "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares". In: *Proc. 44th German Annu. Conf. Acoust. (DAGA), Munich*, pp. 339–342.
-  Schultz, Frank et al. (2023a). "HRTF Individualised Mag-LS and COMPASS Ambisonics-To-Binaural Rendering: Overall Perceived Quality for Pre-Conditioned Listeners". In: *Proc. 154th Audio Eng. Soc. Conv.*
-  Schultz, Frank et al. (2023b). "Overall Perceived Quality for Mag-LS And Compass Ambisonics-To-Binaural Rendering Using Individualised vs. Non-Individualised HRTFs". In: *TBA as journal article*.
-  Therry, David and Brian F.G. Katz (2019). "Anechoic audio and 3D-video content database of small ensemble performances for virtual concerts". In: *Proc. 23rd Intl. Congress Acoustics, Aachen*.
-  Wilcox, Rand R. (2022). *Introduction to Robust Estimation and Hypothesis Testing*. 5th. Academic Press.