I3DA 2021

University of HUDDERSFIELD
Inspiring global professionals

Subjective evaluation of spatial analysis and synthesis methods using different microphone arrays

Alan Pawlak [(1)]

Hyunkook Lee [(1)]

Aki Mäkivirta [(2)]

Thomas Lund [(2)]

[(1)] University of Huddersfield, Applied Psychoacoustics Lab, alan.pawlak@hud.ac.uk

[(2)] Genelec Oy, www.genelec.com

I3DA Conference 8-10 September, 2021

Project goal

To determine the most accurate spatial analysis and synthesis method

Spatial Analysis

Binaural Room Impulse Response

Synthetic Binaural Room Impulse Response

Perceptual evaluation

Background

Methods

Spatial Decomposition Method (SDM)

B-Format Spatial Decomposition Method (B-format SDM)

Higher-Order Spatial Impulse Response Rendering (HO-SIRR)

Previous research and limitations

Inconsistent results from the previous subjective tests

It is unclear which method works the best in the real-life scenario

Improved B-format SDM – Ambisonic SDM (ASDM)

Use of a dedicated omnidirectional pressure signal located at the centre of the array

Aims

To put the discussion on the scientific ground

To extend the previous research by employing additional conditions

To provide new knowledge and insights about the systems of interest

Listening test

Methods

- Spatial Decomposition Method (SDM)
- Ambisonic Spatial Decomposition Method (ASDM)
- Higher-Order Spatial Impulse Response Rendering (HO-SIRR)

Microphone array

- Spaced array (6OM1)
- Spherical array (em32)

Pressure signal for SDM

- Microphone from the array
- Microphone at the centre of the array

Source positions

- +30°
- +90°
- +135°

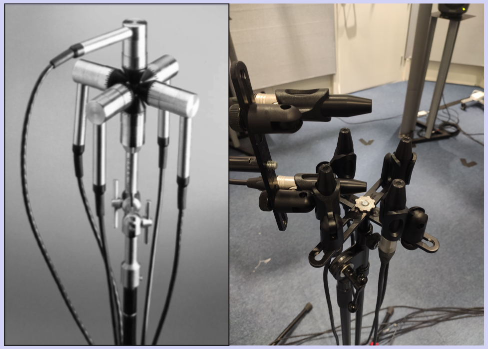
Source type

- Bongo
- Orchestra
- Speech

Room measurements

6OM1 array

- 100mm spacing
- Based on GRAS 50VI
- Six Line Audio OM1 microphones
- Smallest DOA error for SDM*
- Smallest perceptual difference to the ref**



* Amengual Garí et al., 2021, ** Ahrens, 2019

Line Audio OM1

- Reference omnidirectional measurement
- Used as a centre omnidirectional pressure signal for the SDM conditions.
- (20 Hz to 20kHz, ± 1dB)

Eigenmike em32

- 32-element spherical microphone array
- Used to obtain spherical harmonic components for ASDM and HO-SIRR
- Raw SRIR from em32 were used with original SDM

Neumann KU100

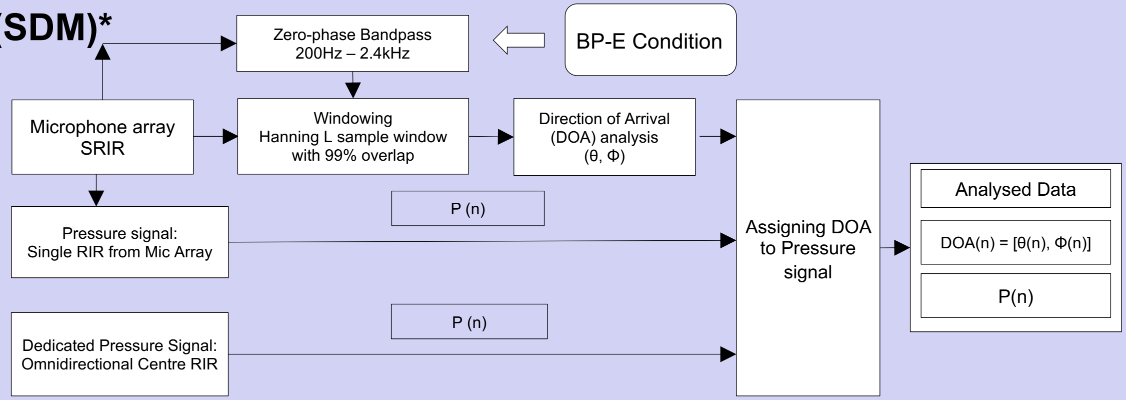
- Reference BRIR
- Served as a reference in the listening test
- All measurements captured at APL's ITU-R BS.1116-compliant critical listening room (6.2m x 5.6m x 3.8m)

Test conditions

Spatial Decomposition Method (SDM)*

- Conditions:
  - 6OM1
  - em32
  - 6OM1+Omni
  - em32+Omni
  - 6OM1+Omni+BP-E
  - em32+Omni+BP-E

*** Tervo et al., 2013**



Spatial Decomposition Method: Omni+BP-E Condition

- BP-E stands for **Band Pass – Enforcement.** It is original SDM with two optimisations *
  - DOA enforcement for the Direct Sound (2.6ms)
  - Band-limited TDOA-based DOA estimation
- Spatial aliasing can cause errors in TDOA-based DOA estimations. It is recommended to apply a low-pass filter to the impulse responses prior to the estimation. Cut-off should be set to fc = c/2d **
- **For band-limited DOA estimation we suggest to use an upper bound proposed by Benesty et.al****
- Spacing between two closest sensors in 6OM1 is 7.07cm, hence **fc = 2433Hz**

*** Amengual Garí et al., 2021, ** Benesty et al., 2008, p. 189-190**

Test conditions

Higher Order Spatial Impulse Response Rendering (HO-SIRR)*

- Conditions:
  - HO-SIRR
  - HO-SIRR+diffuse

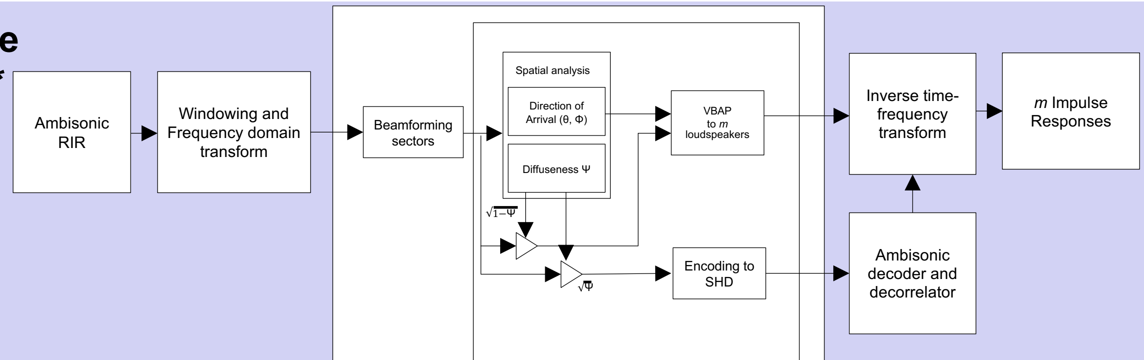
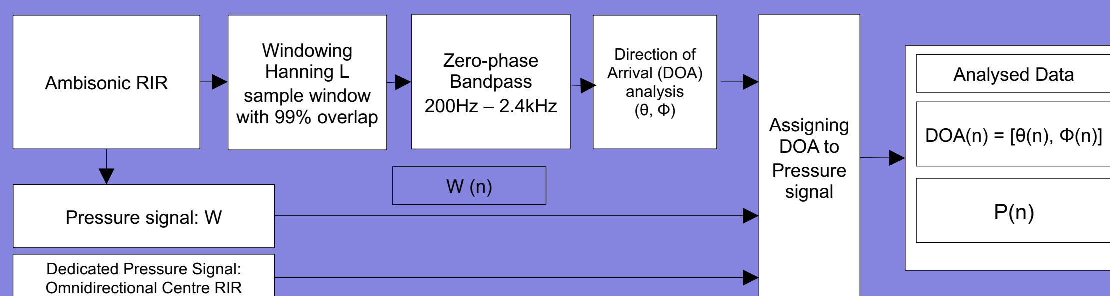


Diagram adopted from *

* **McCormack, Pulkki, et al., 2020**

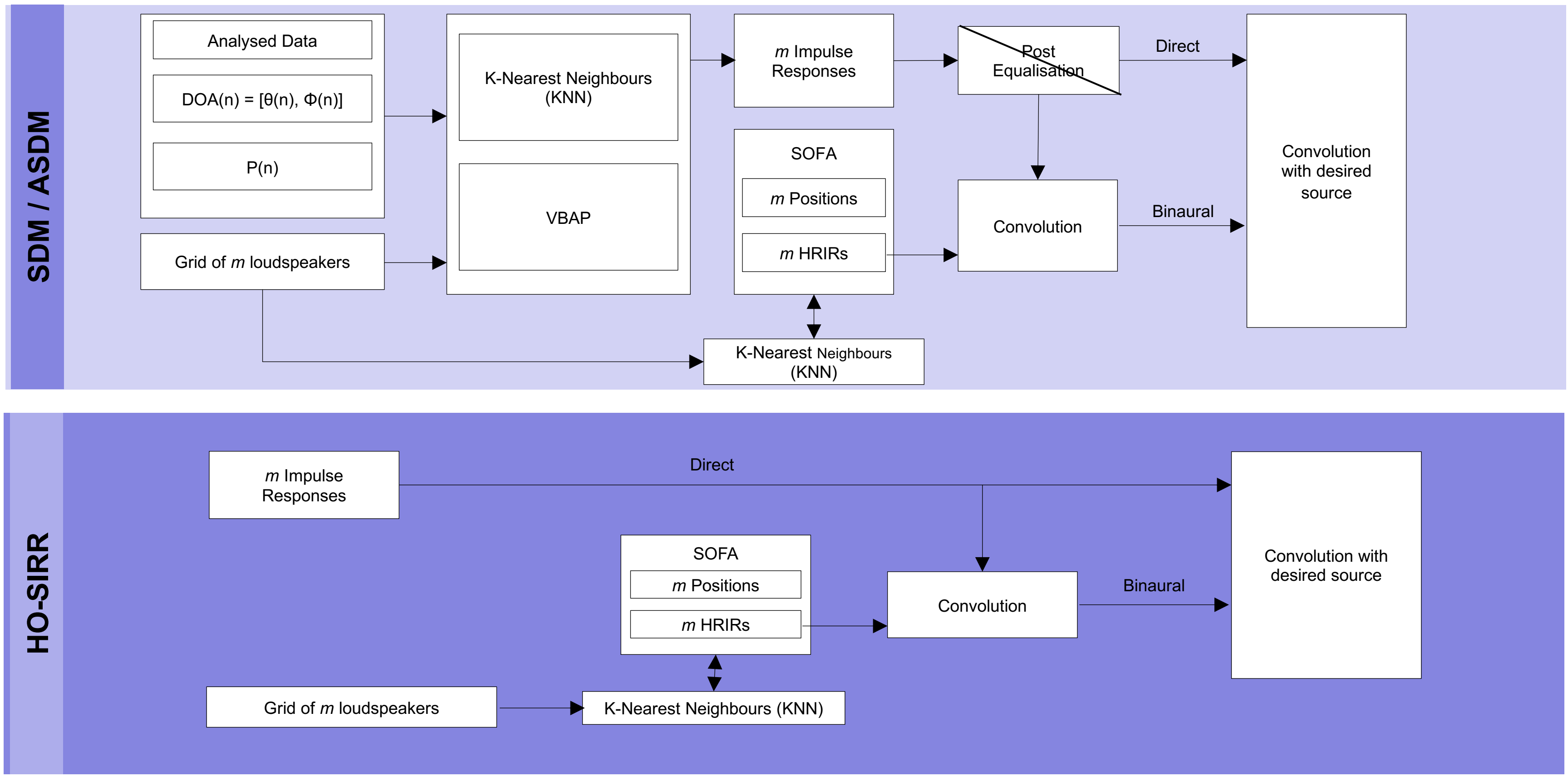
Ambisonic Spatial Decomposition Method (ASDM)*

- Conditions:
  - ASDM
  - ASDM+Omni
- The bandwidth for a PIV-based DOA consistent with BP-E condition



* **Zaunschirm et al., 2020**

Rendering

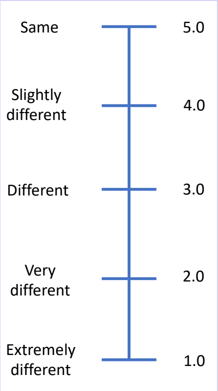

SDM / ASDM
Analysed Data
DOA(n) = [θ(n), Φ(n)]
P(n)
Grid of m loudspeakers
K-Nearest Neighbours (KNN)
VBAP
m Impulse Responses
SOFA
m Positions
m HRIRs
K-Nearest Neighbours (KNN)
Post Equalisation
Convolution
Direct
Binaural
Convolution with desired source
HO-SIRR
m Impulse Responses
Direct
SOFA
m Positions
m HRIRs
Convolution
Binaural
Convolution with desired source
Grid of m loudspeakers
K-Nearest Neighbours (KNN)

Experimental design

Test protocol

- Listening test performed remotely using HULTIGEN v2*
- MUSHRA-like methodology, with no anchor.
- Five-grade similarity scale with five semantic labels.

*** Johnson & Lee, 2020**



Test protocol

- Tested attributes: (i) **Spatial Fidelity** , (ii) **Timbral Fidelity.**
- Fidelity defined as “trueness of reproduction quality to that of the original”.
- Six sessions per attribute, 20 minutes per session.

*** Zielinski et al., 2005**

Test scenarios

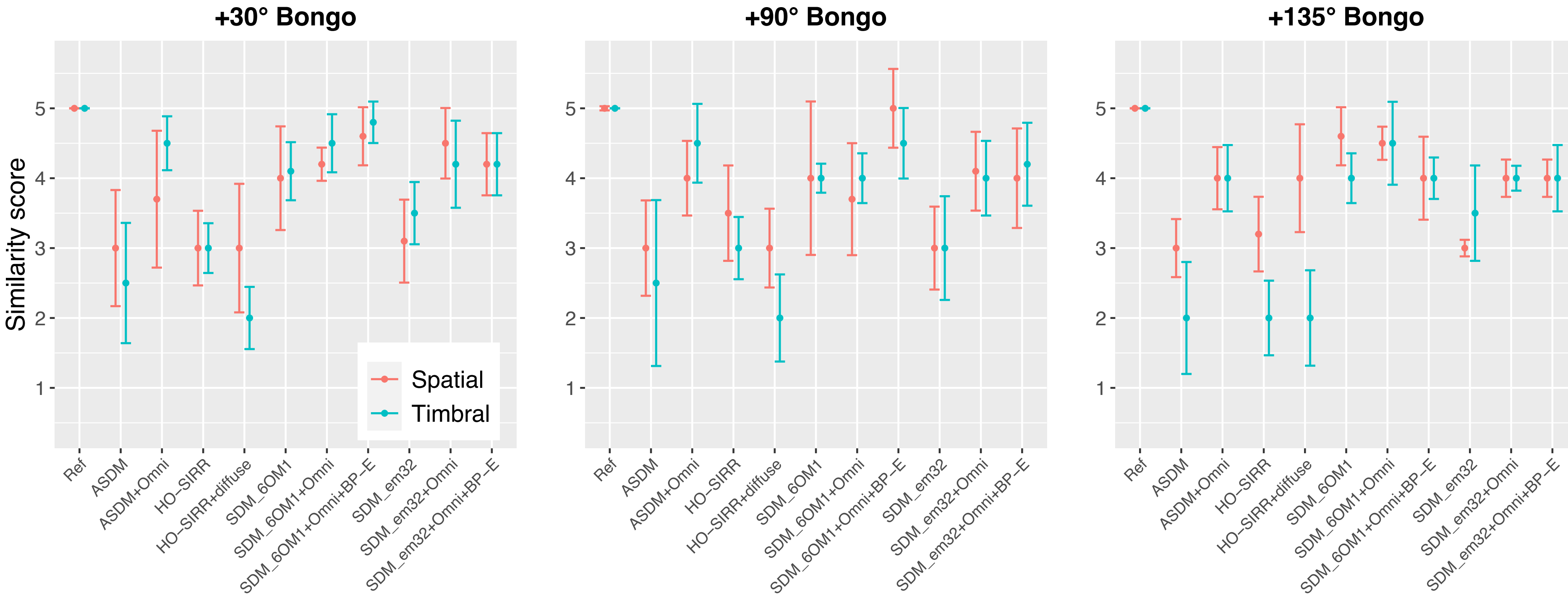
- Three static source positions: **+30°**, **+90°**, **+135°.**
- Source types:
  - **Bongo** from B&O ”Music for Archimedes” CD.
  - **Speech** from B&O ”Music for Archimedes” CD.
  - **Orchestra** from “Denon anechoic orchestral music recording” CD.

Reproduction system

- Binaural rendering was performed using 2702 KU100 HRIRs sampled on the Lebedev grid*.
- AKG K702 headphones.
- Inverse filter measured using obtained using KU100 binaural head.
- Loudness calibration procedure using a hand-rubbing file calibrated to 67dB LAeq**.

*** Bernschütz, 2013 , ** Lee et al., 2021**

Test results – Bongo – Medians and 95% CI



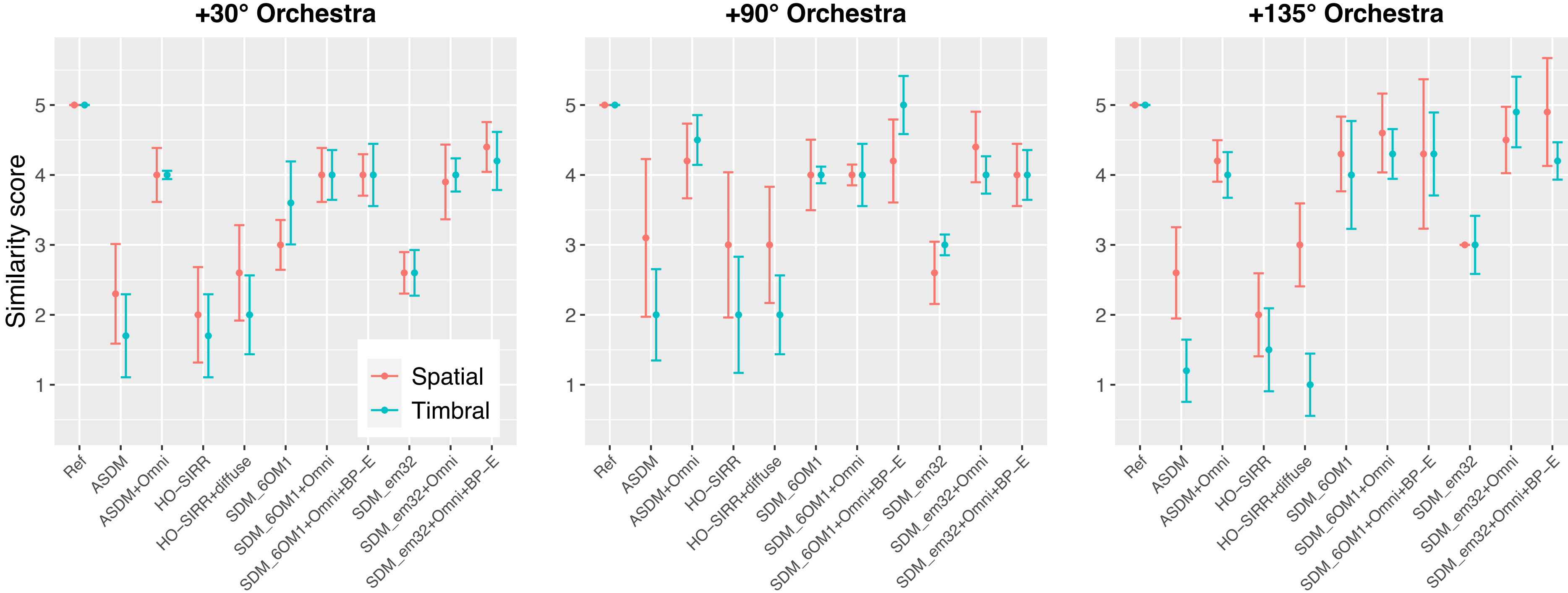
ASDM, HO-SIRR and HO-SIRR+diffuse for most of the time performed between 3.0 (Different) and 2.0 (Very different) for both spatial and timbral fidleities.

SDM em32 condition performs around the centre point of the scale for both timbral and spatial fidelities.

SDM conditions using omnidirectional microphone located at the centre tend to perform at higher end of the scale ("Slightly different" or "Same").

No sig. diff. w.r.t. reference:
Spatial +30°: 6OM1+Omni+BP-E, em32+Omni; **Spatial +90**: 6OM1+Omni+BP-E
Timbral +30: 6OM1+Omni+BP-E;
Timbral +135: 6OM1+Omni.

Test results – Orchestra – Medians and 95% CI



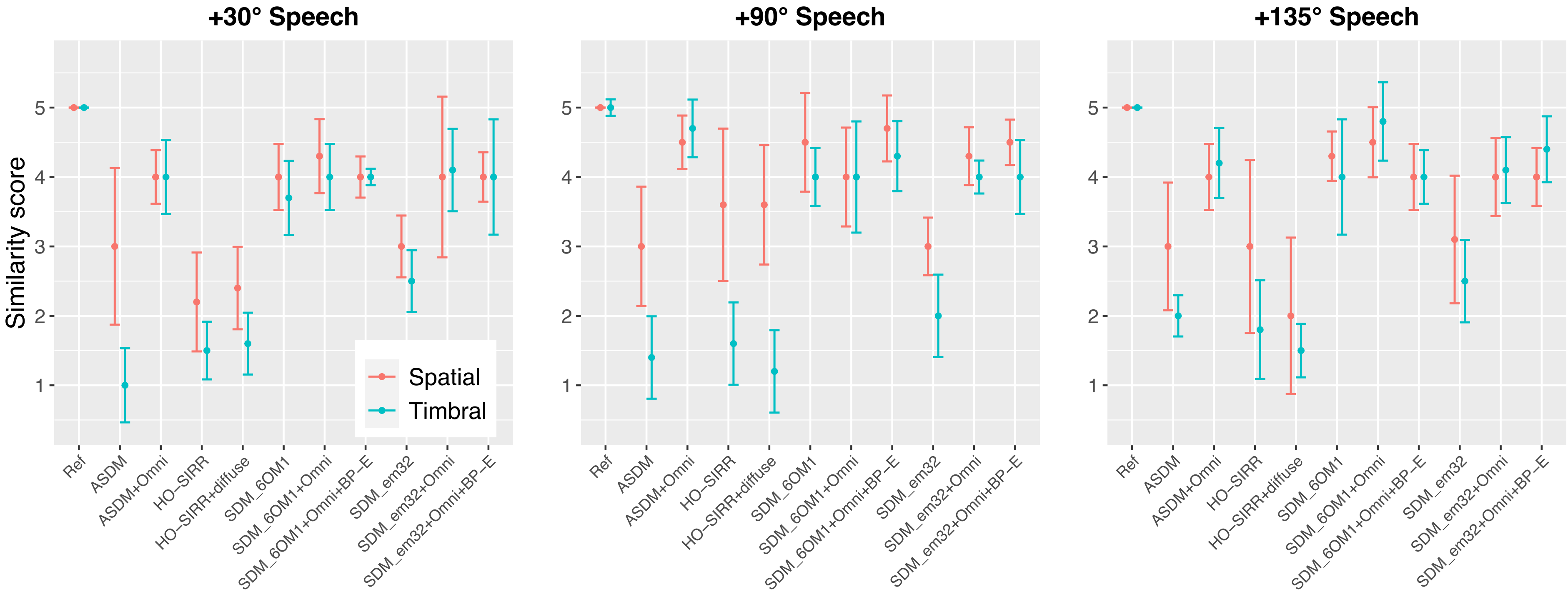
ASDM, HO-SIRR and HO-SIRR+diffuse for most of the time scored between the centre of the lower end of the scale.

SDM em32 condition performs between 3.0 (Different) and 2.0 (Very different) timbral and spatial fidelities.

SDM conditions using omnidirectional microphone located at the centre tend to perform at higher end of the scale ("Slightly different" or "Same").

No sig. diff. w.r.t. reference: **Spatial +90**: ASDM+Omni, 6OM1+Omni+BP-E; **Spatial +135**: SDM conditions using omni mic; **Timbral +90**: ASDM+Omni, 6OM1+Omni+BP-E; **Timbral +135:** em32+Omni.

Test results – Speech – Medians and 95% CI



For timbral fidelity ASDM and HO-SIRR score at the low end of the scale 1.0 to 2.0. (Extremely/Very Different). For spatial fidelity these methods performed around 3.0 (“Different”)

SDM em32 condition performs between 3.0 (Different) for spatial fidelity and 2.0 (Very different) timbral fidelity.

SDM conditions using omnidirectional microphone located at the centre tend to perform at higher end of the scale (”Slightly different” or “Same”).

No sig. diff. w.r.t. reference:
Spatial +90: ASDM+Omni, 6OM1, 6OM1+Omni+BP-E; **Spatial +135**: 6OM1+Omni; **Timbral +90**: ASDM+Omni; **Timbral +135:** 6OM1, 6OM1+Omni, em32+Omni+BP-E

Conclusions

- **The conditions employing SDM and high-quality omni microphone as a pressure signal performed better than any other tested methods.**
- The use of Eigenmike em32 with the original SDM did not significantly improve perceived timbral and spatial fidelities.
- The use of a dedicated pressure signal from a high-quality omni microphone was beneficial especially for the ASDM and SDM em32 conditions.
- DOA Enforcement for the direct sound and estimating the DOA in a specific bandwidth was found to be beneficial only in two conditions
- Some systems were not significantly different to the reference. This appeared to be dependent on the source position and source type.

Future work

- The next step is to evaluate more subjects to increase the sample size, thus increase the statistical confidence of the experiment.
- An objective analysis of the synthesised stimuli will be performed to support subjective results and to get a better understanding of the evaluated methods.
- Further study could be conducted employing the same setup in more reverberant rooms e.g., lecture hall and concert hall.

References

- Ahrens, J. (2019). Perceptual Evaluation of Binaural Auralization of Data Obtained from the Spatial Decomposition Method. Proc. IEEE International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2019-Octob, 65–69.
- Amengual Garí, S. V., Arend, J. M., Calamia, P. T., & Robinson, P. W. (2021). Optimizations of the Spatial Decomposition Method for Binaural Reproduction. Journal of the Audio Engineering Society, 68(12), 959–976.
- Benesty, J., Chen, J., & Huang, Y. (2008). Springer Topics in Signal Processing: Microphone Array Signal Processing. Springer.
- Bernschütz, B. (2013). A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100. Fortschritte der Akustik – AIA-DAGA 2013, 592–595.
- Johnson, D., & Lee, H. (2020). Huddersfield universal listening test interface generator (HULTI-GEN) version 2. 149th Audio Engineering Society Convention 2020, AES 2020, 2–5.
- Lee, H., Johnson, D., & Zieliński, S. The effect of direct-to-reverberant energy ratio on frontback confusion in binaural reproduction. In: International conference on immersive and 3D audio. 2021, September. https://www.i3da.eu/
- McCormack, L., Pulkki, V., Politis, A., Scheuregger, O., & Marschall, M. (2020). Higher- Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution. Journal of the Audio Engineering Society, 68(5), 338–354.
- Tervo, S., Pätynen, J., Kuusinen, A., & Lokki, T. (2013). Spatial decomposition method for room impulse responses. Journal of the Audio Engineering Society, 61(1-2), 17–28.
- Zaunschirm, M., Frank, M., & Zotter, F. (2020). Binaural rendering with measured room responses: First-order ambisonic microphone vs. dummy head. Applied Sciences (Switzerland), 10(5).
- Zielinski, S. K., Rumsey, F., Kassier, R., & Bech, S. (2005). Comparison of basic audio quality and timbral and spatial fidelity changes caused by limitation of bandwidth and by down-mix algorithms in 5.1 surround audio systems. Journal of the Audio Engineering Society, 53(3), 174-192.

Thanks!

- Presented research presented was funded by Genelec Oy and the Univeristy of Huddersfield.
- Thanks to everyone who took part in the listening test.
- Thanks to Prof. Tapio Lokki, Nils Meyer-Kahlen and Neo Kaplanis for the discussion on SDM, and Leo McCormack for his comments and insights into the HO-SIRR method.

https://www.hud.ac.uk/apl

https://www.facebook.com/applied.psychoacoustics.lab/

https://www.linkedin.com/in/alanpawlak/