Audio Engineering Society

Conference Paper 31

Presented at the AES 5th International Conference on Audio for Virtual and Augmented Reality
2024 August 19–21, Redmond, WA, USA

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Sound Sphere 2: A High-resolution HRTF Database

Michaela Warnecke, Samuel Clapp, Zamir Ben-Hur, David Lou Alon, Sebastià V. Amengual Garí, and Paul Calamia

Meta Reality Labs Research, Redmond, WA, 98052, USA

Correspondence should be addressed to Paul Calamia (pcalamia@meta.com)

ABSTRACT

This publication describes a high-resolution database of 78 head-related transfer functions (HRTFs) that were collected during the 2022 Audio Engineering Society AVAR conference; the database is freely available for download (https://facebookresearch.github.io/SS2_HRTF). HRTFs were collected in a purpose-built anechoic chamber housing a vertically oriented, motorized arc with a 2-meter-radius, which contains 54 loudspeakers spaced every 3° in elevation. Participants were aligned using a height-adjustable platform and wall-mounted cross lasers; their head position and orientation were tracked in real-time during the measurement. Inter- and intra-participant error analyses across more than 1600 incident angles per participant indicate high precision in measurements. In addition to being a high-resolution and a high-precision HRTF database, this database includes the corresponding equalization filters for both commercial headphones and a VR headset, as well as some demographic information for each participant. A relatively large number of the participants in this database are researchers from the spatial audio research community (who participated in the AVAR 2022 conference). The authors hope that publishing this database may facilitate cross-lab research in the future.

1 Introduction

Head-related impulse responses (HRIRs) are acoustic descriptors of the sound scattering introduced by the torso, head and ears for sounds emanating from a specific location in space. They can be converted into the frequency domain, yielding direction-dependent transfer functions called head-related transfer functions (HRTFs). HRTFs play a central role in creating spatial audio and are vital for achieving realism in virtual sound spaces.

When rendering sounds over headphones, listeners can perceive sound direction via binaural cues, such as interaural time and level differences (ITDs and ILDs, respectively) [1], but HRTFs are needed to create the perception of sound in space. Spatial rendering using generic (non-individualized) HRTFs often creates perceptual artifacts, such as front-back confusions, angular distortion in elevation perception, and weak externalization [2, 3, 4]. For convincing spatial rendering of an acoustic environment, individual HRTFs are necessary.

To date, several publicly available HRTF databases exist (for example, see various databases available on-

line[1]), with data on as few as 18 to as many as 220 participants. Depending on the database, HRTF measurements, anthropometric measures, 3 dimensional (3D) scans or HRTF simulations are available for at least a subset of the participants included in the respective database. To date, only one database includes information on some participants' gender (RIEC[2] published the gender for 50% of their database), and no database has published information on participants' ethnicity or age, despite the fact that ethnicity, age and gender all contribute to differences in facial shapes [5, 6]. Differences in facial and torso anthropometry can influence differences in HRTFs across these variables. Note, however, that at least one research approach has used these markers in HRTF modeling via an internal database [7].

In this paper we present a new, high-resolution HRTF database collected during the Audio Engineering Society's 4th International Conference on Audio for Virtual and Augmented Reality (AVAR) in 2022. A total of 78 participants were measured during the three-day conference, ranging in age, ethnicity, and sex. The median age of all participants was 34 years of age, 21% of participants were female, and 68% identified as Caucasian/White, 17% identified as Asian, 9% identified as Multi-Racial, 3% identified as Black and 1% identified as Latin/Hispanic. Acoustic re-measurements for eight participants are included for repeatability analysis, along with data for the KEMAR, B&K HATS and KU100 mannequins. In addition to acoustically measured HRIRs and their frequency domain representation as HRTFs, the database contains headphone impulse responses (HpIRs) for Sennheiser HD650s and Meta Quest2 impulse responses (QIRs) for all participants and the HATS mannequin, as well as limited demographic information.

2 Method

2.1 Setup and Equipment

HRIRs were measured in an anechoic chamber (interior dimensions: 4.5 x 7.5 x 4.5 m), purpose-built to conduct high-quality, high-resolution far-field acoustic measurements. 80 cm fiberglass wedges render the chamber anechoic down to 100 Hz. The noise floor is 10 dB below the threshold of hearing. A motorized, vertically oriented semicircular arc (Sigma Design, WA, USA), 2 m in radius, can rotate freely in either direction with a top speed of 12 RPM. For the purpose of collecting acoustic measurements from a spherical surface, the arc is outfitted with 54 loudspeakers (Meyer Sound MM-4XP, with custom firmware designed to reduce self noise, Meyer Sound Laboratories, CA, USA) spaced every 3° in elevation from $+90°$ to $-69°$; the speakers can be aimed with an accuracy of $0.1°$ in elevation. Due to the continuous positioning capability of the arc, the azimuth resolution can be chosen arbitrarily with an angular accuracy of $0.1°$. At the base of the arc, a motorized, height-adjustable, custom-built platform with knee- and backrests allows researchers to utilize wall-mounted cross lasers for precise centering of participants (see Section 2.2 and Fig. 1). Throughout the measurement process, 4 OptiTrack cameras (NaturalPoint, OR, USA), mounted outside the sound-incident area, monitor reflective infrared (IR) beads attached to a wig cap worn by the participant, as shown in Fig. 1C. Information about the participant's head position can be streamed in real-time both to the researchers outside the anechoic chamber and to a display inside the chamber, allowing participants to "self-correct" their head position during the measurement session.

2.2 Participant Preparation

Prior to the acoustic measurements, participants signed IRB-approved consent forms and received a detailed explanation of the study procedure. For all measurements, participants were asked to remove any jewelry, badges, or similar items they wore around their upper torso. Subsequently, in order to standardize measurements and control head position, participants were asked to put on a single-use wig cap, over which they then placed a second cap with reflective IR markers attached to it. In cases where participants had longer hair that did not fit underneath the wig cap, they were asked to tie their hair at the base of their skull, leaving only a ponytail visible outside the cap. Lastly, research assistants (RAs) prepared a set of single-use commercially-available ear plugs (Comply Foam Isolation+, Hearing Components Inc., MN, USA) with in-ear microphones (Knowles FG-23329-D65 omnidirectional electret, Knowles Electronics, NY, USA) housed inside custom-designed microphone casings. Participants then inserted the microphone-outfitted ear

[1] https://www.sofaconventions.org/mediawiki/index.php/Files

[2] https://www.riec.tohoku.ac.jp/pub/hrtf/hrtf_data.html



Fig. 1: Examples of the HRIR measurement system: (A) Participant aligned at the radial center of the loudspeakers' arc. The laser-protection glasses are removed for the data collection; (B) Blocked in-ear microphone placement with wig cap to prevent hair occlusion; (C) Participant alignment at the radial center of the arc using cross lasers.

plugs into their ear canals, and RAs checked for correct placement of the microphones, including a flush lining with the ear-canal entrance.

Following the above-described preparation for the measurement session, the participant entered the anechoic chamber and sat in the custom-built chair that is fastened to the motorized stage (see Fig. 1A). RAs then helped the participant to settle, enabled any security/safety measures, and raised the stage until the participant reached the correct height at the radial center of the loudspeaker arc. Finally, with the help of the cross-lasers, the RAs confirmed the participant's position and alignment with respect to the coordinate axes of the measurement system. Next, the participant received information about how their head position and orientation would be tracked via the real-time tracking system installed in the anechoic chamber, and instructions on how to identify and correct for movement as necessary.

2.3 Measurements

2.3.1 HRIRs

After participant preparation and alignment, the chamber doors were closed from the outside and the measurements began. All data were played and captured at a sampling rate of 48 kHz using the multiple exponential sweep method [8] that interleaved logarithmic swept-sine signals of 250 ms duration from 200 Hz to 20 kHz. The level of the sweeps at the participant's head position was 84 dBSPL at 1 kHz. Raw HRIRs were deconvolved from the captured sine sweeps using the original sweep stimulus. Free-field equalization was done in post-processing (see Sections 2.3.2 and 2.4). Each measurement procedure took approximately 8-12 minutes, depending on the head motion of the participant (see Section 3.1). HRIRs were collected for source positions on a modified Lebedev grid [9], quantized to every 6° in azimuth and every 3° in elevation (from $+90°$ to $-69°$ per the arc design). Including the 0°-elevation point at every azimuth, a total of 1625 directions were collected.

All equipment was cleaned and disinfected between participants. Single-use items, such as ear plugs, were discarded.

2.3.2 Free-Field Impulse Response

The free-field response of each of the 54 loudspeakers in the arc was captured by both the left and the right in-ear microphones, individually mounted to a foam-wrapped pole to limit unwanted scattering, and positioned at the radial center of the arc. The same multiple exponential sweep procedure and stimulus as described above in Section 2.3.1 was used for the free-field measurements, with the arc in the 0°-azimuth position.

Raw impulse responses were deconvolved from the captured sine sweeps using the original stimulus, and windowed to obtain the final free-field impulse responses between each loudspeaker and each microphone (see Section 2.4 for details). This procedure was performed at the beginning and at the end of each recording day for HRTF equalization and to monitor any changes in the measurement system's performance.

2.3.3 Headphone Impulse-Response Measurements

After HRIR measurements had been collected, the motorized platform was lowered and participants were allowed to move. With the binaural microphones still inserted, we then recorded HpIRs (Sennheiser HD650s, Sennheiser Electronic Corporation, CT, USA and Quest2, Meta, CA, USA) for each participant, using the same sweep signal as described above (2.3.1). This measurement was repeated 3 times, where the participant doffed and donned the headphones/Quest each time to capture fit-to-fit variations.

The measured HpIRs were then windowed to a length of 2,048 samples with fade-in and fade-out applied via Hann windows of 410 samples. To calculate the Headphone Equalization (HpEq) filters, the windowed HpIRs were transformed to the frequency domain via the discrete Fourier transform. Then, a minimum-phase HpEq filter was constructed by inverting the magnitude response and using the Hilbert transform to generate the phase response [10]. Regularization was applied to prevent ringing artefacts, which could be caused by high amplification of low magnitude values after inversion [11]. HpIRs and their corresponding HpEq filters are available in the database for both measured devices.

2.4 Post-processing

The raw HRIRs were equalized in post-processing with the free-field measurements described in Section 2.3.2, using the appropriate data from the left or right microphone. Each HRIR was equalized with the free-field IR from the corresponding loudspeaker. The HRIRs and free-field IRs were trimmed to 384 samples (8 ms at a 48 kHz sampling rate), windowed with a right-hand Tukey window, and converted to the frequency domain with a 1536-point DFT. Division in the frequency domain was followed by an inverse DFT and a small circular shift to avoid anticausal HRIRs. The resulting HRIRs were low-pass filtered with a cutoff frequency of 16 kHz, using a zero-phase application of a Butterworth filter, trimmed back to 384 samples, and windowed with a right-hand Tukey window.

HRIR data files were saved in SOFA (spatially-oriented format for acoustics).

3 Database Verification and Analysis

To assess the quality of the HRIR measurement system, as well as the quality of the measured HRIRs, 3 stages of evaluation were performed and are detailed below: on-line inspection, repeated measurement evaluation, outlier HRIR detection.

3.1 Online Inspection

In order to verify that the measured data is free of artifacts due to participant movements or participant related noises (e.g. heavy breathing, throat clearing) during the HRIR and HpIR measurements, the participants' position was continuously tracked and measurements were visually inspected. If a participant's head position changed by more than 1° in yaw, pitch, or roll during a measurement in a single position of the loudspeakers' arc, the measurement was paused and restarted at the current arc position once the participant had re-aligned him/herself (instructions were provided verbally and displayed on a monitor). Similarly, if the head position was found to have deviated from the initial, aligned position by more than 2° in yaw, pitch, or roll, the participant was asked to re-align her/himself before continuing. In addition, auto-generated plots of the recorded left and right microphone signals, as well as ITD and ILD statistics of the HRIRs and spectrograms of the HpIRs, were presented in real-time to the RA. The RA visually inspected theses plots to verify that the collected data are free of significant artifacts, noise, or other errors.

3.2 Repeated Measurement Evaluation

To characterize the precision of the HRIR measurement system and measurement procedure, 8 participants were repeatedly measured 4 times. The first measurement was considered the reference measurement and the following measurements were conducted as a separate measurement session (with the participant departing from the system and then returning to restart the whole procedure). The spectral difference between the HRTF and the reference measurement and each of the 3 repeated measurements was calculated for each frequency and sound source direction. The mean difference between the reference measurement and the 3 repetitions across all 8 participants and all directions is presented in Figure. 2 as a function of frequency, as well as the 95th percentile for both right and left ears.

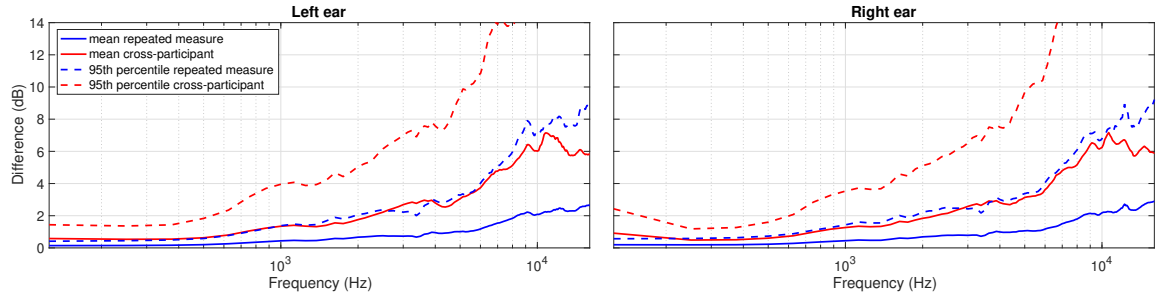


Fig. 2: Spectral difference between HRTF measurements, with 4 repeated measurements of the same participant in blue and cross-participant difference in red. Solid line represents the mean across all evaluated participants and directions, while the 95th percentile is represented with a dashed line. Left Panel for left ear and right panel for right ear.

To assess the repeatability error in the context of HRTF variability across participants, the same errors were computed by randomly comparing each of the 8 reference measurements with randomly selected repetitions of other participants. The mean repeated measurement error level increases with frequency with 1 dB and 2 dB error at frequencies of 5 kHz and 10 kHz, respectively. The mean repeated measurement error is lower than the mean cross-participant error with 3 dB and 6 dB error at frequencies of 5 kHz and 10 kHz, respectively. Interestingly, the 95th percentile repeated measured error is similar to the levels of mean cross-subject error. These results indicate that the variability due to the limitation of the measurement system precision is lower than the variability due to cross-participants HRTF differences.

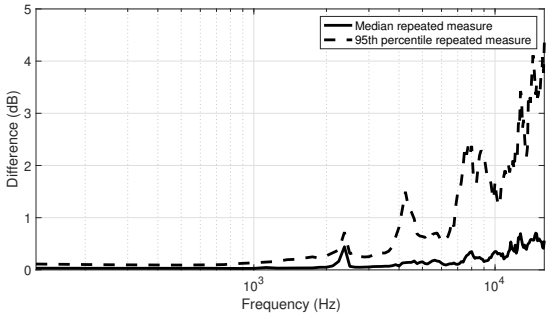


Fig. 3: Spectral difference between HRTF measurements of the B&K HATS, with 4 repeated measurements. Solid line represents the median across horizontal plane, while the 95th percentile is represented with a dashed line.

Another insight regarding the precision of the HRIR measurement system can be provided by comparing the current results in Figure 2 with the results in Figure 8 of the Sonicom HRTF dataset [12]. The Sonicom database presents a similar repeatability error that is based on 5 repeated measurements, with a comparable mean repeated measurement error, but a higher 95th percentile repeated measurement error, showing more than 5 dB at 5 kHz. A futher comparison with the HUTUBS HRTF dataset [13], where the spectral difference between repeated measurements of B&K HATS is evaluated (Figure 3), reveals that a higher precision is achievable when utilizing a median repeated measurement error; achieving errors below 1 dB up to 16 kHz in the horizontal plane, comparable with the precision reported for HUTUBS.

3.3 Outlier HRIR Detection

To identify suspicious outliers in the measured HRIR we adopt the Tukey Outlier approach [14] to examine HRTF magnitude, ITD, ILD and loudness. First, the full database statistics were calculated for each one of the metrics. The ITDs were evaluated for each HRIR direction for elevations from -30° to 30° using the IACCE method, as described in [15]. The ILDs were computed for the same directions in auditory filters of equivalent rectangular bandwidth (ERB) using the `AKerbILD` function in AKtools [16], and were averaged over frequencies between 3-10 kHz. The loudness was computed for all directions using the integrated loudness algorithms from ITU-R BS.1770-4 [17], for an input signal of white noise. Then, a score s was computed for each metric by comparing each individual HRIR

metrics to the database statistics:

$$s = \begin{cases} 0, & v \in [Q1, Q3] \\ \frac{v-Q3}{Q3-Q1}, & v > Q3 \\ \frac{Q1-v}{Q3-Q1}, & v < Q1 \end{cases}, \quad (1)$$

where v is the metric value, and $Q1$ and $Q3$ are the 1st and 3rd quartiles. For the HRTF magnitude, the 99th percentile of s over frequencies (up to 15 kHz) and directions was used as the *Outlier score*. For ITD, ILD and loudness, the maximum value of s across directions was used.

HRIRs with a value of the final *Outlier scores* that exceeded 3 for the magnitude, ILD and loudness, or 6 for the ITD, were considered as suspicious outliers, and were manually observed and listened to (using SPARTA Binauraliser [18]). Figure 4 presents the *Outlier score* for all subjects and metrics. These Tukey-Outlier-based evaluations together with the manual examination is proposed as a final validation stage of the dataset quality.

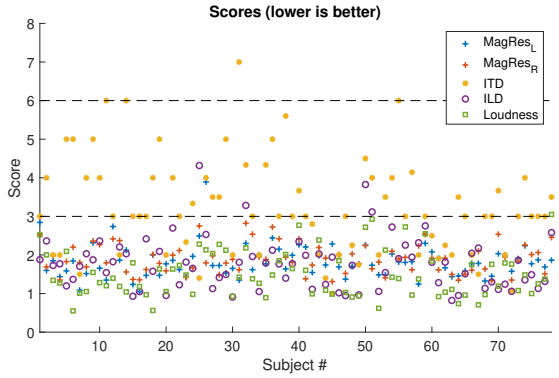


Fig. 4: The *Outlier scores*: scores to detect HRIR outliers. 4 metrics were computed for each subject: Magnitude (left and right ears, ITD, ILD and loudness. The score represents the maximum (or 99% percentile for the magnitude metric) distance from the intequartile range compared to all the database statistics.

4 Database structure

The database is available to download from `https://facebookresearch.github.io/SS2_HRTF`. The data include:

- `HRIRs.zip`: 78 unique human HRIRs.
- `HRIRs_Repeated_Measurements.zip`: 32 HRIRs, 8 humans × 4 repetitions.
- `HRIRs_mannequins.zip`: 12 HRIRs, 3 mannequins (KEMAR, HATS, KU100) × 4 repetitions.
- `Hp_Filters.zip`: 79 folders (78 humans + HATS) with 6 subfolders (2 headsets, Quest2 and HD650, × 3 repetitions[3]) that contains:
  - `HpIR.wav`: the headset impulse responses (2ch, left/right, 2048 taps @ 48 kHz).
  - `HpEq.wav`: the minimum phase equalization filters (2ch, left/right, 400 taps @ 48 kHz).
  - `HpIR_FR.png`: figure that shows the measured impulse responses before and after windowing, and the transfer functions.
  - `Recorded_Signal.png`: figure that shows the raw recorded sweep in time and spectrogram, including an estimation of the SNR.
- `Demographics.xls`: an Excel sheet with demographics information - age, sex and statistics on ethnicity.

All HRIRs are provided in standard SOFA format.

The `Hp_Filters.zip` also contains a `HpIR_Validation.csv` table with computed SNR values and energy differences between the left and right ear recordings. This data was used to validate the measurements.

5 Conclusion

The Sound Sphere 2, a high-resolution HRTF database, with HRTF measurements from 78 participants measured from 1625 directions is publicly available (see details in section 4). In addition to HRTFs, the data set contains HpIR from a commercial VR headset and headphones, as well as demographics such as age and sex for each participant. The database was collected during the 2022 Audio Engineering Society AVAR conference, and a relatively large number of participants in this database are researchers from the spatial audio research community. The authors are hopeful that this publication will contribute to collaborative research in the future.

[3]Not all the data from the 78 participants include the full set of 2 headsets and 3 repetitions.

6 Acknowledgment

The authors would like to acknowledge and thank many colleagues who contributed to designing and building the HRIR measurement system and to conducting and processing the HRIR measurements: Ravish Mehra, Terry Cho, Rob Perin, Kevin Scheumann, Zachery Schramm, Clarissa Munoz, Ho Yi Shek, Laura Alonso Gonzalez, Henry Von Dollen, Justin Lam, Katrina Pirie , Matt Cangie, Mikhail Lev, David Kalamen, Saveliy Baranov, Pete Stirling and Isaac Engel.

References

[1] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[2] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, 94(1), pp. 111–123, 1993.

[3] Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D., "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society*, 44(6), pp. 451–469, 1996.

[4] Simon, L. S., Zacharov, N., and Katz, B. F., "Perceptual attributes for the comparison of head-related transfer functions," *The Journal of the Acoustical Society of America*, 140(5), pp. 3623–3632, 2016.

[5] Farkas, L. G., Katic, M. J., and Forrest, C. R., "International anthropometric study of facial morphology in various ethnic groups/races," *Journal of Craniofacial Surgery*, 16(4), pp. 615–646, 2005.

[6] Zhuang, Z., Landsittel, D., Benson, S., Roberge, R., and Shaffer, R., "Facial anthropometric differences among gender, ethnicity, and age groups," *Annals of occupational hygiene*, 54(4), pp. 391–402, 2010.

[7] Bilinski, P., Ahrens, J., Thomas, M. R., Tashev, I. J., and Platt, J. C., "HRTF magnitude synthesis via sparse representation of anthropometric features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4468–4472, IEEE, 2014.

[8] Majdak, P., Balazs, P., and Laback, B., "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, 55(7/8), pp. 623–637, 2007.

[9] Lebedev, V. I. and Laikov, D. N., "A quadrature formula for the sphere of the 131st algebraic order of accuracy," in *Doklady Mathematics*, volume 59, pp. 477–481, 1999.

[10] Smith III, J. O., "Spectral audio signal processing," *(No Title)*, 2011.

[11] Engel, I., Alon, D. L., Scheumann, K., Crukley, J., and Mehra, R., "On the differences in preferred headphone response for spatial and stereo content," *Journal of the Audio Engineering Society*, 70(4), pp. 271–283, 2022.

[12] Engel, I., Daugintis, R., Vicente, T., Hogg, A. O., Pauwels, J., Tournier, A. J., and Picinali, L., "The sonicom hrtf dataset," *Journal of the Audio Engineering Society*, 71(5), pp. 241–253, 2023.

[13] Brinkmann, F., Dinakaran, M., Pelzer, R., Grosche, P., Voss, D., and Weinzierl, S., "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, 67(9), pp. 705–718, 2019, doi:10/gf8ww6.

[14] Tukey, J. W. et al., *Exploratory data analysis*, volume 2, Springer, 1977.

[15] Brian, K., Rozenn, N., and Sylvain, B., "Subjective investigations of the interaural time difference in the horizontal plane," in *Audio Engineering Society Convention 118*, Audio Engineering Society, 2005.

[16] Brinkmann, F. and Weinzierl, S., "Aktools—an open software toolbox for signal acquisition, processing, and inspection in acoustics," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[17] Series, B., "Algorithms to measure audio programme loudness and true-peak audio level," in *International Telecommunication Union Radiocommunication Assembly*, 2011.

[18] McCormack, L. and Politis, A., "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.