CrossMark

# A multiple model high-resolution head-related impulse response database for aided and unaided ears

Joachim Thiemann* and Steven van de Par

## Abstract

Head-related impulse responses (HRIRs) allow for the creation of virtual acoustic scenes. Since in ideal conditions the human auditory system can localize sounds with a very high degree of accuracy, it is useful to have an HRIR database with high spatial resolution, such that realistic-sounding scenes can be created. In this article, we present an HRIR database with 12722 directions, giving a spatial resolution of 2 arc degree or better, for a sphere covering − 64° elevation to the zenith. Four sets of HRIRs were recorded with different head-and-torso simulators (HATSs), including one with a six-channel bilateral behind-the-ear hearing aid model. The resulting database is available at https://www.uni-oldenburg.de/akustik/mmhr-hrtf and is distributed under a Creative Commons license.

**Keywords:** Head-related impulse responses, Head-related transfer functions, Hearing aid, Auditory virtualization

## 1 Introduction

Humans are very good at localizing sounds relative to their heads. The human brain can localize sounds by taking advantage of the way sound is modified on its path from the source to the ears, especially how sound is modified differently between one ear and the other.

These differences in sound between the ears give rise to binaural cues such as the interaural time difference (ITD) and interaural level difference (ILD). These cues depend on a variety of factors, notably the shape of the head, pinnae, and upper torso. All of these can be affected by the distance the sound travels to each ear, attenuation due to occlusion (the head shadow effect), and reflections (from the upper torso and the pinnae). In addition, the pinna creates direction-dependent spectral cues that can be used by the brain to infer source direction.

As a result, there is no simple (or single) relationship between direction and the binaural and spectral cues; nevertheless, the human brain can use these cues to estimate the location of sound sources with astonishing accuracy. If we then want to simulate acoustic scenes with sources in different directions to a human listener with a high degree

of naturalness, the direction-dependent modification of the sound (to each ear) needs to be recreated.

In this article, we present a set of recorded transfer functions that include all binaural and spectral cues resulting from the interaction of the head with the impeding sound of the source. These are referred to as head-related impulse responses (HRIRs) in time domain or head-related transfer functions (HRTF) in frequency domain. HRIRs are time domain impulse responses that can be used to filter audio signals which results in a binaural signal that is equivalent to the audio signal being received from the direction from which the HRIR was recorded. Given a set of HRIRs with sufficient resolution, a complete acoustic scene can be created in principle using room acoustical simulation methods such as those described in [1–3]. Such simulations can be helpful for the investigation of room acoustical perception and speech perception in various simulated reverberant environments. More specifically, it would also be very helpful to be able to evaluate signal processing algorithms that are developed for hearing aids while operating in (simulated) reverberant multi-talker settings. For this purpose, besides measuring HRIR at the entrance of the ear canal, sets of impulse responses were also measured for microphones of a hearing aid attached to a manikin.

*Correspondence: joachim.thiemann@uni-oldenburg.de
Cluster of Excellence "Hearing4All", University of Oldenburg, Ammerlander Heerstr. 116, 26129 Oldenburg, Germany

Spatial resolution, the density of directions with which the HRIRs are recorded, is a critical aspect of an HRIR database, because it affects how accurately the incident direction can be chosen. The set of impulse responses presented here has been measured with high spatial resolution (2° steps in the horizontal and vertical directions), which is close to the threshold of human perception [4–6] and is sufficient for simulating reverberant environments [7] and smooth movements of sources.

For this study, we obtain HRIRs by fitting a dummy head-and-torso simulator (HATS) with microphones and placing the HATS into an anechoic chamber. Using a test signal emitted from a loudspeaker, the acoustic transfer function (ATF) from the direction of the loudspeaker to the microphones is obtained by comparing the test signal to the signal captured by the microphones. This procedure is repeated while changing the relative position of the loudspeaker and HATS, by using additional loudspeakers, moving the loudspeaker, or changing the orientation of the HATS. In many cases, a combination of these scene modifications is used.

In addition to the signal as it reaches the opening of the ear canal, for one HATS our database also provides the microphone impulse responses of a behind-the-ear (BTE) hearing aid. While these recordings are specific to the hearing aid being measured, the insights gained from such recordings can often be generalized to other similar devices. The BTE hearing aid has three microphones per side, thus the resulting HRTFs consist of 8 channels (hearing aid plus in-ear microphones).

To date, several HRIR databases have been published, differentiating themselves in the details of the HATS model and available directions relative to the head. The work presented here is based on methods developed for the creation and evaluation of the databases presented in [8–10]. Our database differs by having several HATSs measured using the same setup at high spatial resolution and coverage and includes one HATS fitted with a hearing aid, allowing for comparative studies. The preliminary database was previously presented in [11], and we present here an evaluation of an updated version of the full recordings, with comparisons to the database in [8].

## 2 Method

In this section, we describe the physical setup of the HATS, the stimulus, the recording setup, and the post-processing of the recordings to obtain the HRTFs. We also describe the method with which we evaluate the database and compare it to a similar database.

### 2.1 Setup

For the database presented here, we measured HRIRs using a variety of HATSs inside an anechoic chamber, where the probe signals were emitted from loudspeakers mounted on a movable platform. This platform, the two-arc source position (TASP) system, allowed placing of transducers onto any point of an approximately spherical surface, at the center of which the HATS was located.

A total of four HATSs were used, each with their own specific in-ear microphones coupled with appropriate amplifiers. In addition, one HATS (Brüel & Kjær Type 4128C) was fitted with a behind-the-ear (BTE) hearing aid. The details of the four HATSs and their amplifiers are given in Table 1 and Fig. 1.

#### 2.1.1 Brüel & Kjær Type 4128C

The Brüel & Kjær HATS was fitted with Brüel & Kjær type 4158C and 4159C artificial ears and uses the standard built-in in-ear microphones. The analog signal was amplified using a *G.R.A.S. Power Module Type 12AA.*

Two sets of recordings were performed with this HATS, the first without any additional hardware and the second with a set of BTE hearing aid models fitted to the HATS. The hearing aid was the same as in [8], dummies of type *Acuris* provided by Siemens Audiologische Technologie GmbH, with an in-house developed preamplifier. The MMHR database contains the dataset of recordings with the BTE hearing aid models, and is referred to below using the label BKwHA.

#### 2.1.2 G.R.A.S. KEMAR Type 45BB

The G.R.A.S. KEMAR HATS was fitted with artificial ears KB0090 and KB0091, and microphones of type 26AS. Power to the microphones and amplification was also provided by the *G.R.A.S. Power Module Type 12AA.* This dataset is referred to using the label KEMAR.
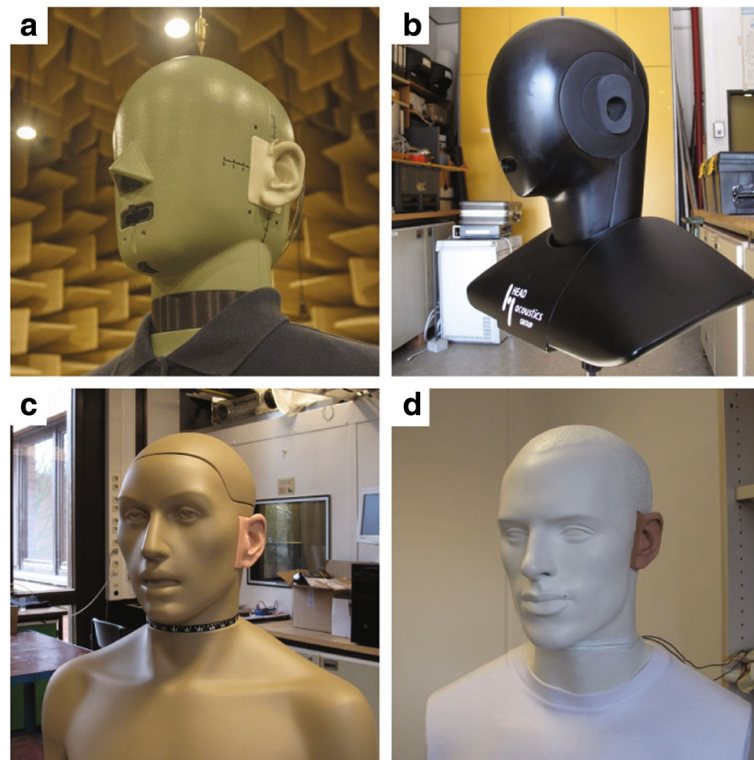
#### 2.1.3 HEAD acoustics HMSII.2

The HEAD Acoustics HMSII.2 HATS was fitted with in-ear microphones supplied by Brüel & Kjær, Type 4165, and a custom power supply and preamp provided by HEAD Acoustics. This dataset is referred to using the label HEAD.

#### 2.1.4 DADEC

The final HATS measured was the custom construction "Dummy Head with Adjustable Ear Canals" (DADEC) [12] with in-ear microphones and preamps custom developed with the intention of simulating the correct position and

**Table 1** HATS and the associated recording equipment used in the recordings

| HATS model | Microphones | Amplifier |
|---|---|---|
| Brüel & Kjær 4128C | Built-in | G.R.A.S. 12AA |
| HEAD Acoustics HMSII.2 | Brüel & Kjær 4165 | Custom |
| G.R.A.S. KEMAR | G.R.A.S. 26AS | G.R.A.S. 12AA |
| DADEC | Custom | Custom |

**Fig. 1** HATS measured for the database. **a** Brüel & Kjær Type 4128C. **b** HEAD acoustics HMSII.2. **c** G.R.A.S. KEMAR Type 45BB. **d** DADEC [12] (University of Oldenburg)

orientation of a human eardrum. In the discussion and figures below, the label DADEC is used for this dataset.

## 2.2 Recording room

The HRIRs were measured in the anechoic chamber of the University of Oldenburg [13]. The room has a volume of 238 m$^3$ and a measured background noise level of 3 dB SPL. Temperature and humidity of the atmosphere in the room were measured continuously during recording in case equipment behaved abnormally due to environmental conditions.

### 2.2.1 TASP

The two-arc-source-positioning (TASP) system shown in Fig. 2 consisted of two metal arcs joined by circular supports at the top and bottom, where the top support was connected to a drive system allowing the structure to be freely rotated around its vertical axis. The extent of rotation was limited only by the length of the cables carrying the transducer signals. The bottom support ensured the arcs were centered. On the arcs, two sleds allowed a pair of transducers to be moved along the arcs, arranged to always be directed at the center of the TASP (and thus the HATS). Extenders allowed the transducers to travel beyond the extent of the arcs; one transducer could move
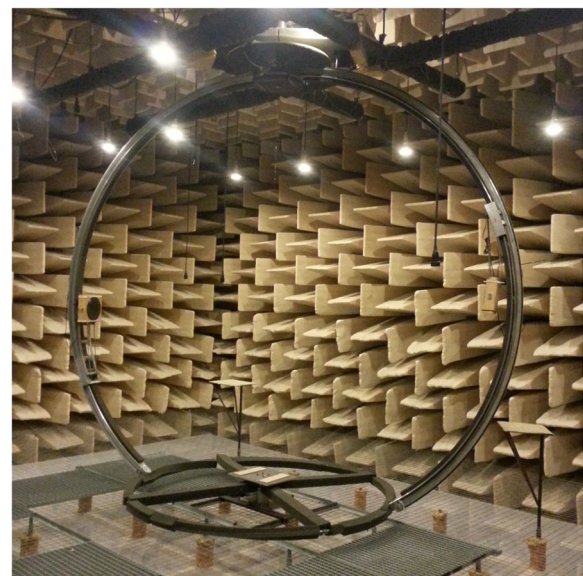


**Fig. 2** The TASP system in the anechoic chamber of the University of Oldenburg. Note the offset sled mounts, allowing for greater coverage for the vertical positioning of the transducers
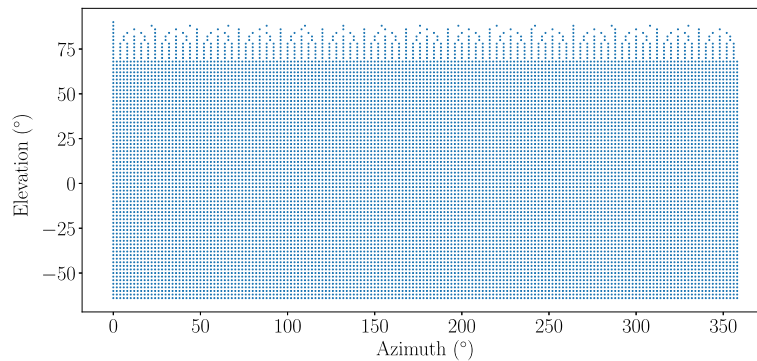
**Fig. 3** HRTF azimuth/elevation grid. From − 64° to 68° , each elevation has 180 points; from 70° to 78° , each elevation has 90 points. Elevations 80° and 82° have 60 points, 84° has 45 points, 86° has 30 points, and 88° elevation has 16 points. The pole is a single point

up to the zenith of the sphere, the other could move to 64° below the horizon of the sphere.

The TASP system was used to position the speakers on a grid of 12,722 points on the partial sphere reachable by the transducers (Fig. 3). The points form a grid of 2° resolution around the horizontal axis (azimuth) and in the vertical direction (elevation). Near the top (elevation > 70°), the azimuthal resolution was reduced since the great-arc distance between points on the grid becomes very small.

The probe signals for measuring the HRIR were emitted using two Manger W05/1 sound transducers, mounted in custom enclosures. Due to the construction of the TASP, one of the transducers had a range of ca. − 35° to 90° elevation, while the other had a range of ca. − 65° to 60°.

### 2.2.2 Recording equipment
Sound playback and capture was performed using a RME ADI-8 QS interface connected via MADI to the host computer, enabling sample-synchronous recording of 8

channels. A MATLAB script controlled audio recording, playback, and the positioning of the TASP system.

### 2.2.3 Probe signal
The stimulus to measure the ATF from the tranducers to the microphones was a sine sweep band-limited between 100 Hz and 21 kHz, as used by Brinkmann et al. [9], the lower limit being dictated by the limits of the transducer. The sweep had a spectral coloration in the form of a low-frequency boost (4 dB/octave, 100–5000 Hz) to compensate for the background noise of the anechoic chamber. The measurements used a sampling rate of 44100 Hz.

### 2.3 Impulse response calculation
The design goal of the database was to provide anechoic HRTFs with very little post-processing, to allow the end-user to do specific post-processing needed for a particular purpose. Specifically, we avoided mirroring and smoothing between adjacent measurements. Post-processing consisted only of windowing and time shifting,
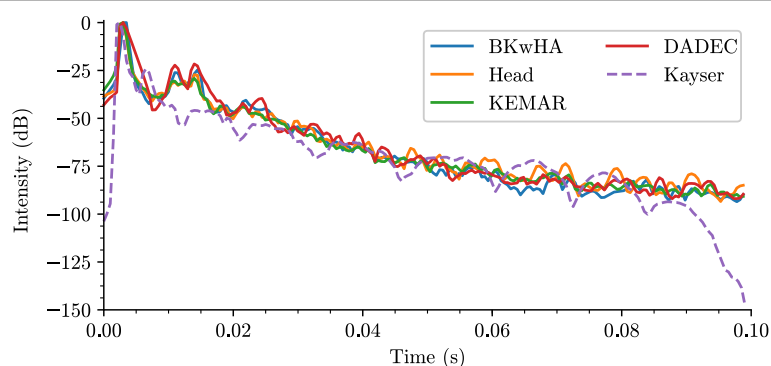


**Fig. 4** Average energy in 1 ms sections of the left ear channel, for an azimuth of − 10°, normalized to the section with highest energy. BKwHA refers to the dataset recorded with the Brüel & Kjær HATS with hearing aid, KEMAR refers to the dataset recorded with the G.R.A.S. KEMAR HATS, Head refers to the dataset recorded with the HEAD acoustic HATS, and DADEC refers to the dataset recorded with the DADEC HATS. Kayser refers to the HRIRs presented in [8]
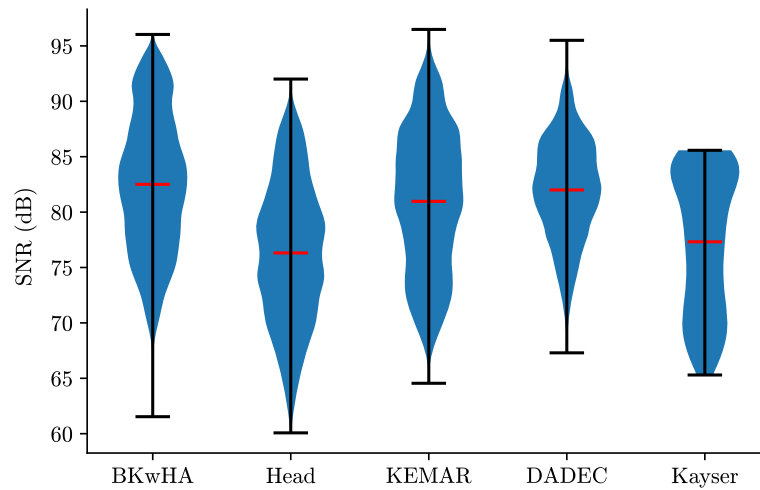
**Fig. 5** Distribution of SNRs for HRIRs in the in-ear channels for the MMHR-HRTF database and the database in [8] ("Kayser"). The MMHR response SNRs show a slightly larger spread than the Kayser database, but it should be noted that the MMHR-HRTF datasets contain more measurements (directions) than the Kayser database
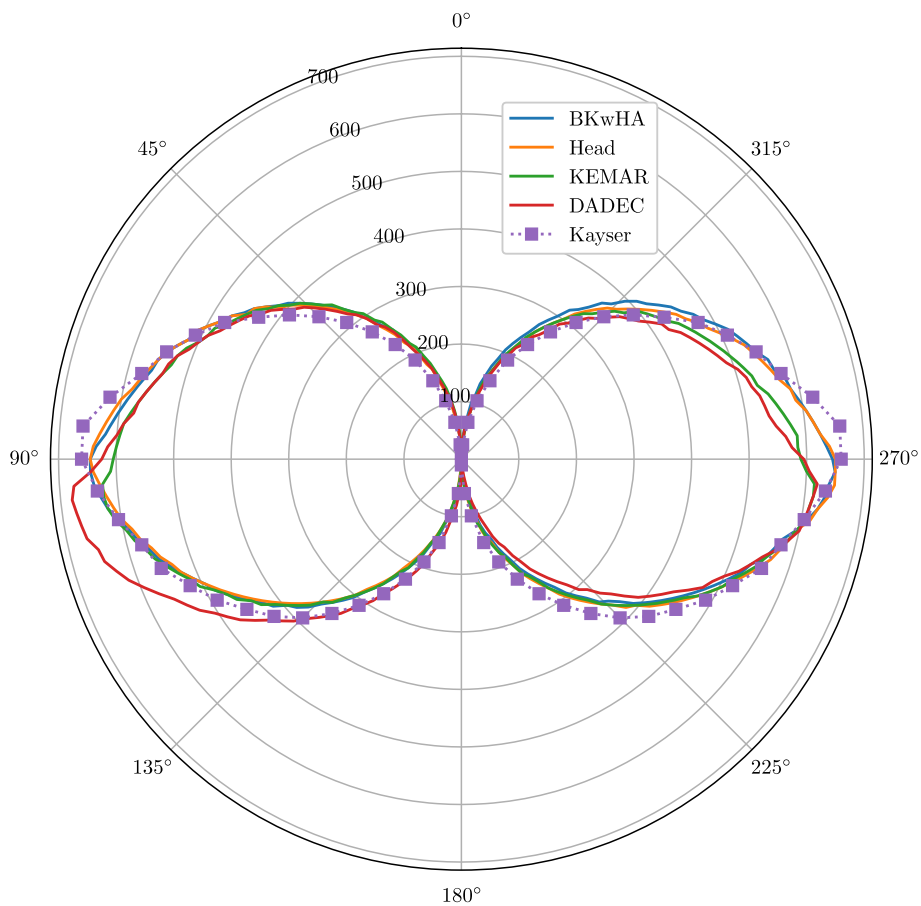


**Fig. 6** Broadband ITD for all HATS at 0° elevation, including reference database in [8], measured in ms and plotted as a function of azimuth. Note that the reference database ("Kayser") is symmetric by construction

with the time shift of each measurement recorded in the database.

For each spatial direction, three sine sweeps were recorded and averaged. Using a probe signal recorded by a reference microphone (located at the TASP center) with the transducer at the same elevation, the time-domain impulse response was computed using spectral division with Tikhonov regularization. For the IR with azimuthal direction $\phi$ and elevation $\theta$, this is described by the function

$$M_{(\phi,\theta)}(\omega) = \frac{G_\theta^*(\omega)D_{(\phi,\theta)}(\omega)}{G_\theta^*(\omega)G_\theta(\omega) + \lambda}, \tag{1}$$

where $\cdot^*$ indicates complex conjugation, $G_\theta(\omega)$ is the probe signal in frequency domain (with index $\omega$), $D_{(\phi,\theta)}(\omega)$ is the recorded signal for position $(\phi,\theta)$, and $\lambda$ is a regularization term to avoid computational noise when $G_\theta(\omega)$ is small. The choice of $\lambda$ is not critical, but should be several orders of magnitude smaller than the average observed $G_\theta(\omega)$. This procedure compensates for linear effects from the transducer.

The resulting transfer function was converted into time domain and truncated to the portion that characterizes the acoustic effects of the ear, the head, and the torso. For typical applications, a set of responses of length 6.66 ms (294 samples) is provided, with the first peak occurring at

0.5 ms. The length of this HRIR is sufficient for perceptually plausible spatialization [14]. In addition, responses of length 100 ms are provided, with the first peak at 3.33 ms. The length of this set of responses was chosen such that the impulse response can drop to the noise floor. In both cases, the window is a hybrid rectangular window with a 10-sample Hann onset followed by a flat section and 10-sample Hann offset.

### 2.4 Database format
The data of the MMHR-HRTF database is stored in the spatially oriented format for acoustics (SOFA) [15] format. This format is specifically designed to store acoustic data (e.g., HRTFs or room impulse responses) with well-defined fields for metadata such as the spatial position of microphones and loudspeakers. The SOFA format was created to rectify the problem of HRTF databases being in custom formats (see, e.g., [16]), which made it difficult to replace one database with another. For our database, the responses of each HATS are stored in a separate SOFA file.

## 3 Results and discussion
The quality of HRIR measurements can be assessed using many different metrics. Here, we compare the impulse responses to those of the database in [8]. Primarily, the
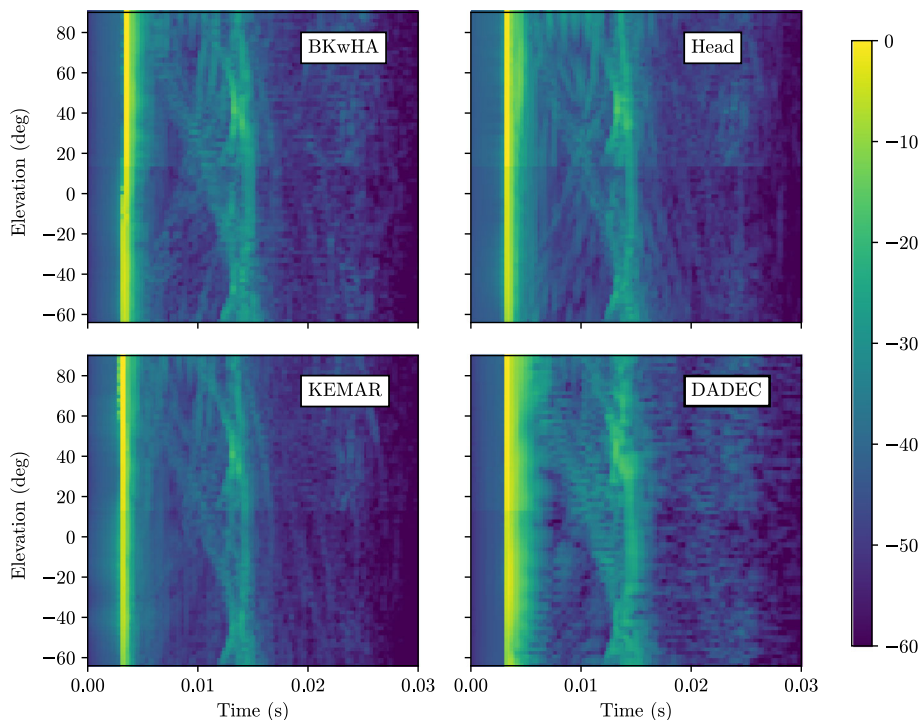


**Fig. 7** Short-term energy in impulse responses over time and elevation for azimuth 0°. Clearly visible is the reflection of the TASP structure at around 10 ms after the peak and weaker reflections caused by the ground plate and upper support structure

evaluation is intended to ensure the impulse responses have similar SNR, and that the interaural properties are as expected.

### 3.1 Background noise level

One method to assess the quality of a HRIR recording is to examine the signal-to-noise ratio. However, the noise level cannot typically be observed directly from the impulse response, but must be estimated from a portion of the response where the desired signal has decayed into the noise floor.

In Fig. 4, we show the average energy of 1 ms sections of the impulse responses over time for the left ear channel, for an azimuth of −10°. The responses are normalized to the peak 1 ms section. We note that the responses all have a similar overall decay characteristic, converging to about 75 dB below the peak, but the responses recorded by Kayser [8] have lower energy at the endpoints both before the leading peak and at the end. This may be attributed to differing processing, and thus to estimate the noise floor, we use a 10 ms window around the 80 ms mark. We also note that the MMHR responses have significant secondary peaks between 10 and 20 ms, at about −25

dB. These are caused by the TASP setup and discussed in detail below.

The distribution of measured SNRs for the in-ear channels is shown in Fig. 5. The MMHR-HRTF database results are generally comparable to the Kayser database, having a higher average SNR but a higher spread. However, for the MMHR-HRTF database, the vertical range of measurements is much higher, so greater variability can be expected (Fig. 6).

### 3.2 Reflections from the recording setup

In Fig. 4, the IRs in the MMHR database show higher energy relative to the Kayser database IR at about 10 ms after the initial peak. The exact location of this additional energy is dependent on the elevation, shown in Fig. 7. This is consistent with the reflections observed by [9], and can be explained as being caused by refections off the TASP structure, which is roughly symmetric around the HATS setup, but not perfectly circular.

In most situations, only the direct part of the HRIR is desired, and for this reason, a shortened version of the MMHR database is available in which impulse responses



**Fig. 8** Comparison of ILDs at 0°, − 30°, − 60°, − 90°, − 120°, and − 150° azimuth
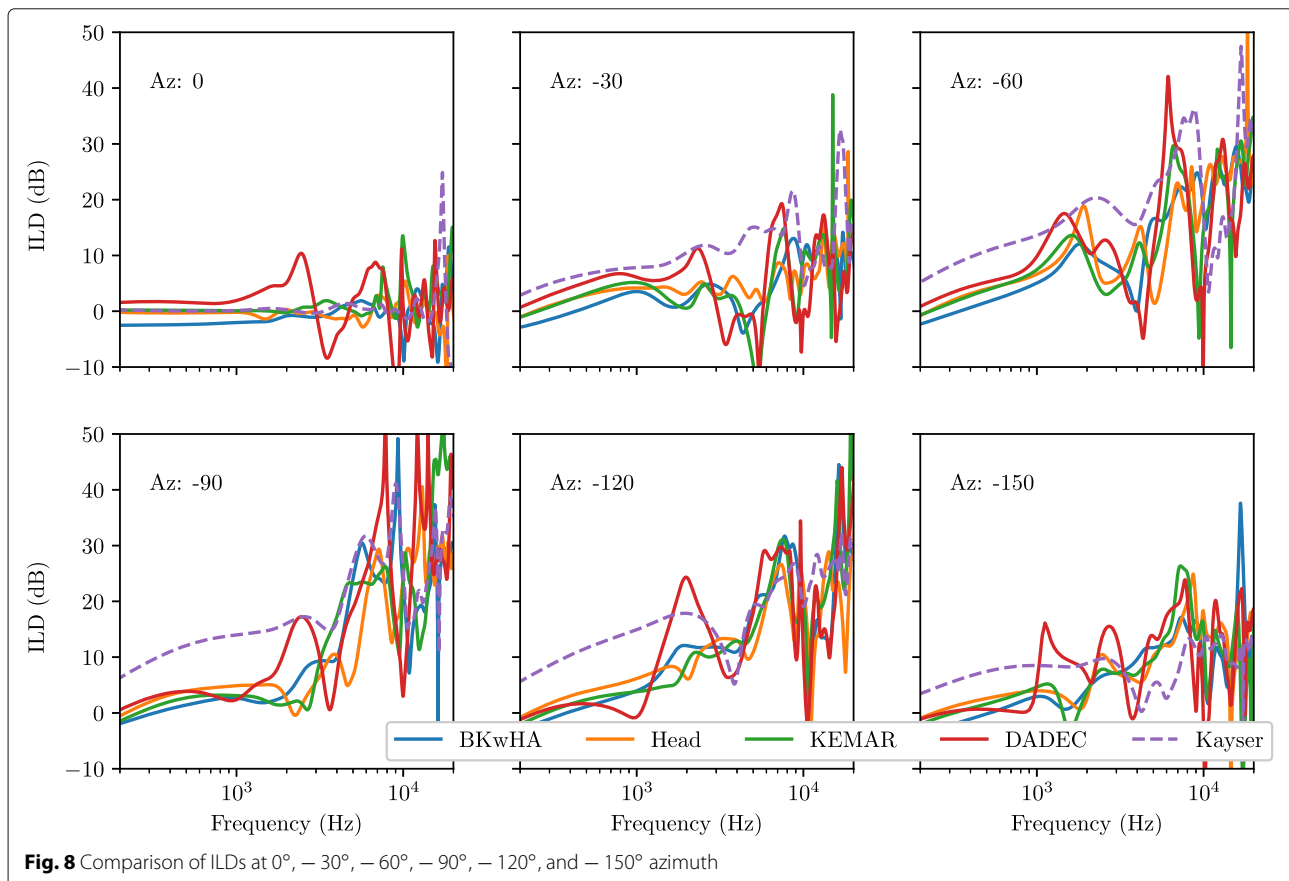
are truncated such that the initial peak is at 0.5 ms and the total HRIR is 6.66 ms (294 samples) long.

### 3.3 Interaural properties

When HRIRs are used to render an acoustic scene, the interaural properties, that is, the difference in the transfer functions between the signals reaching the ears, are critical. While it can be expected that interaural properties differ between different HATSs (due to differences in the geometry), it has also been found that, using identical HATSs, HRTF measurements can vary noticeably between measurements due to the recording setup and methodology [16]. For the MMHR database, we examine some of the key properties to ensure that they are plausible. In particular, we examine the interaural time differences (ITDs) and interaural level differences (ILDs).

Figure 6 shows the ITD obtained by measuring the difference in the onset between the channels, where the onset is defined as the first sample whose magnitude exceeds 10 dB below the overall peak value [17]. To compute ITDs with sub-sample accuracy, the HRIRs were upsampled to 200 kHz.

The plot shows that ITDs are close to each other even for different head models. Only at the extreme lateral positions (near 90° and 270°) do the curves deviate significantly; at the rear, even where the difference in ITD is above 50 $\mu$s, it is equivalent to a small directional shift. We also note that the MMHR ITDs are in general asymmetric, especially the SP HATS. Note that the HRIRs of the Kayser database are forced to be symmetric by mirroring.

In Fig. 8, we compare the ILDs for the various HATSs, with the Kayser HRIR again as comparison. ILDs were calculated using a Fourier transform of 128 samples windowed with a Hann window centered on the initial peak of the IR. As expected, there are large variations in the ILDs between datasets. Noteworthy is that, while the MMHR measurements (BKwHA, Head, KEMAR, and DADEC) are similar to each other, they are up to 8 dB lower than the Kayser HRIRs at low frequencies. This difference is likely due to the recording setup. However, the deviation is within expected bounds.

### 4 Conclusion

The MMHR-HRIR database introduced here is a set of binaural and bilateral hearing aid impulse responses recorded using a set of four different HATSs. The HRIRs were recorded at high spatial resolution to enable the simulation of movement in an acoustic scene. The impulse responses are made available both at a length of 100 ms and 6.66 ms. The longer responses allow for better simulation of environments with long reverberant tails. In contrast, the shorter responses make spatialization more computationally efficient and contain fewer recording artifacts.

The database was evaluated by comparing the responses with an earlier database [8] recorded using one of the HATSs used in the MMHR-HRIR database. We find that the MMHR-HRIR database has similar SNR and ITD behavior, only the ILDs show a noticeable difference. However, it appears that the differences do not affect the utility of the database in any way.

The MMHR-HRIR database is distributed under a Creative Commons (CC-BY 3.0) license and can be downloaded from https://www.uni-oldenburg.de/akustik/mmhr-hrtf. The data is provided in the SOFA format, allowing the MMHR-HRIR database to be used in compliant software with minimal modifications. A MATLAB API for the SOFA format can be obtained from [15].

#### Authors' contributions
The conception of the MMHR-HRTF database originated in discussions by the authors. The data was recorded by JT with guidance by SP. The manuscript was written by JT in consultation with SP. Both authors read and approved the final manuscript.

#### Competing interests
The authors declare that they have no competing interests.

### Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### References
1. D. Schröder, Physically based real-time auralization of interactive virtual environments. PhD thesis, RWTH Aachen (2011). http://publications.rwth-aachen.de/record/50580
2. M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. (Springer, Berlin, 2008)
3. T. Wendt, S. van de Par, S. D. Ewert, Perceptually plausible acoustics simulation of single and coupled rooms. J. Audio Eng. Soc. **62**(11), 748–766 (2016)

4. J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. (MIT Press, Cambridge, 1996)
5. B. C. J. Moore, *An introduction to the psychology of hearing 5th edn.* (Academic Press, Cambridge, 2003)
6. A. W. Mills, On the minimum audible angle. J. Acoust. Soc. Amer. **30**(4) (1958). https://doi.org/10.1121/1.1909553
7. S. Klockgether, S. van de Par, A model for the prediction of room acoustical perception based on the just noticeable differences of spatial perception. Acta Acustica U. Acustica. **100**, 964–971 (2014)
8. H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, B. Kollmeier, Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. EURASIP J. Appl. Sig. Proc. (2009). https://doi.org/10.1155/2009/298605
9. F. Brinkmann, A. Lindau, S. Weinzierl, S. van de Par, M. Mueller-Trapet, R. Opdam, M. Vorlander, A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations. J. Audio Eng. Soc. **65**(10), 841–848 (2017)
10. F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, S. van de Par, in *Proceedings AIG-DAGA, (Merano, Italy)*. A high resolution head-related transfer function database including different orientations of head above the torso (Deutsche Gesellschaft für Akustik e.V., Merano, 2013)
11. J. Thiemann, A. Escher, S. van de Par, in *Proceedings DAGA, (Nürnberg, Germany)*. Multiple model high-spatial resolution HRTF measurements (Deutsche Gesellschaft für Akustik e.V., Nürnberg, 2015)
12. M. Hiipakka, M. Tikander, M. Karjalainen, Modeling the external ear acoustics for insert headphone usage. J. Audio Eng. Soc. **58**(4), 269–281 (2010)
13. J. Otten, Factors influencing acoustical localization. PhD thesis, University of Oldenburg (2001). http://oops.uni-oldenburg.de/335
14. E. Rasumow, Synthetic reproduction of head-related transfer functions by using microphone arrays. PhD thesis, University of Oldenburg (2015). http://oops.uni-oldenburg.de/2404
15. P. Majdak, M. Noisternig, *AES69-2015: Aes standard for file exchange - spatial acoustic data file format*. (Audio Engineering Society, 2015)
16. A. Andreopoulou, D. R. Begault, B. F. G. Katz, in *IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 5*. Inter-Laboratory Round Robin HRTF Measurement Comparison, (2014), pp. 895–906. httpd://doi.org/10.1109/JSTSP.2015.2400417
17. B. F. G. Katz, M. Noisternig, A comparative study of interaural time delay estimation methods. J. Acoust. Soc. Amer. **135**(6), 3530–3540 (2014)