KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Arnout Roebben, Giuliano Bernardi, Jan Wouters, Toon van Waterschoot, and Marc Moonen Title: HIDVAS: A Hearing Instrument Dataset in Various Acoustical Scenarios for Algorithm Evaluation and Training Journal: Journal on Audio, Speech, and Music Processing Pages: 1-30
Archived version	☐ Submitted version ■ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ☐ EU H22020 ☐ EU Horizon Europe ■ FWO ■ KULeuven Internal Funds ■ VLAIO ☐ AI ■ Other
Author contact	arnout.roebben@esat.kuleuven.be



HIDVAS: A Hearing Instrument Dataset in Various Acoustical Scenarios for Algorithm Evaluation and Training

Arnout Roebben[1*†], Giuliano Bernardi[1†], Jan Wouters[2], Toon van Waterschoot[1], Marc Moonen[1]

[1]Department of Electrical Engineering, ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg 10, Leuven, 3001, Belgium.
[2]Department of Neurosciences, ExpORL, KU Leuven, Onderwijs en Navorsing 2 (O&N2), Herestraat 49, Leuven, 3000, Belgium.

*Corresponding author(s). E-mail(s): arnout.roebben@kuleuven.be;
Contributing authors: giuliano.bernardi@kuleuven.be;
jan.wouters@kuleuven.be; toon.vanwaterschoot@kuleuven.be;
marc.moonen@kuleuven.be;
†These authors contributed equally to this work.

Abstract

To evaluate the performance of audio signal processing algorithms and to train data-driven algorithms, e.g., as applied in hearing instruments, either simulated or recorded data can be used. While large batches of simulated data can be generated using mathematical models, recorded data provide a more adequate representation of real-life scenarios. Therefore, in this paper, the Hearing Instrument Dataset in Various Acoustical Scenarios (HIDVAS) is introduced. This dataset consists of both impulse responses and audio recordings using eight external loudspeakers, two external microphones, and a dummy head. On this dummy head behind-the-ear (BTE) hearing instrument shells with two microphones per shell are mounted, and in the dummy head's ears receiver-in-canal (RIC) hearing instrument loudspeakers are inserted. The dummy head also contains microphones located at its eardrum. The impulse responses have been computed from a swept-sine recording for each microphone-loudspeaker pair, and the audio recordings have been obtained by playing back audio (male and female speech, speech shaped noise, singing voice, stringed instrument, wind instrument, and percussion instrument) through each individual loudspeaker and recording simultaneously using all microphones. These recordings have been repeated for four

hearing instrument domes (open, semi-open, closed, and no-RIC) in three reverberation conditions in one room ($T_{30} = 0.09\,s$, $T_{30} = 0.47\,s$, and $T_{30} = 0.73\,s$), and in one reverberation condition in a different room ($T_{30} = 1.48\,s$). The usage of the dataset as a 'hearing instrument in a box' is exemplified with three example use cases. To this end, the ipsilateral feedback paths, the ratio of direct-to-reverberant sound energy in an external microphone close to the sound source versus in the BTE hearing instrument microphones, and the leakage from external sound sources to the eardrum are studied. The feedback and leakage for the open and semi-open domes are similar and substantially larger than for the closed dome. The ratio of the direct-to-reverberant sound energy in the BTE hearing instrument microphones is lower than in the external microphone, and this difference increases with the reverberation time. The dataset is available at [1, 2].

Keywords: Impulse responses dataset, Audio recordings dataset, Hearing instrument, Behind-the-ear hearing instrument, Receiver-in-canal hearing instrument, Multi-microphone, Multi-loudspeaker, Audio signal processing

1 Introduction

To evaluate the performance of audio signal processing algorithms and to train data-driven algorithms, e.g., as applied in hearing instruments [3], either simulated or recorded data can be used [4, 5]. As the generation of simulated data does not require any recording equipment, large batches of simulated data can be generated. Nevertheless when using these simulated data, the effects of room geometry and absorption are modelled, often requiring simplifying assumptions to make the generation of the simulated data tractable, e.g., assuming the room to be a perfect rectangular 'shoebox' [4, 5]. While the process of obtaining recorded data can be time-consuming, it allows to obtain adequate representations of real-life scenarios with non-regular room geometries and varying absorption properties. Consequently, recorded data allow to obtain a more adequate representation of real-life scenarios to characterise an algorithm's performance or to train a data-driven algorithm.

The generation of such recorded data tailored to acoustic hearing instruments commonly employs human listeners or a dummy head, i.e., a head and torso with ears on which hearing instruments can be mounted. Dummy heads offer advantages such as repeatability and the usage of standardised components, while human listeners offer added realism and allow to quantify intersubject variability [6]. Datasets that utilise such a dummy head or human listener for hearing instrument applications differ in 1) the loudspeaker configuration, 2) the microphone configuration, 3) the nature of the recorded material, and 4) the recording room conditions. First, regarding the loudspeaker configuration, loudspeakers can be mounted on an earpiece and inserted into the dummy head's or human listener's ears replicating hearing instrument loudspeakers (commonly referred to as receivers) [7, 8]. Alternatively, external loudspeakers can be used replicating external sound sources [9–22]. Both loudspeaker

configurations can also be used simultaneously [6, 23, 24]. Second, regarding the microphone configuration, microphones can be mounted on the earpiece replicating hearing instrument microphones [11, 13, 24], microphones can be inserted into the dummy head's or human listener's ear canal to record the signal close to the eardrum [16–18], or external microphones can be used. Different microphone configurations can also be used simultaneously, such as considering both microphones on the earpiece and close to the eardrum [6–8, 12, 14, 19–23], or considering all microphone configurations simultaneously [9, 10, 15]. Third, regarding the nature of the recorded material, impulse responses can be captured, representing a linear model between each considered loudspeaker-microphone pair [6–8, 13–24]. While this linear model is commonly assumed in algorithm design, e.g., [3, 24], and provides an adequate description of the system, it nevertheless fails to fully capture the non-linearities in the playback and recording equipment, e.g., in the microphones and loudspeakers. Consequently, audio can also be directly recorded, or both impulse responses and audio can be recorded [9–12]. Finally, regarding the recording room conditions, the recording procedure can either be performed in one [6–8, 11, 12, 15–24] or multiple conditions [9, 10, 13, 14] thereby, e.g., varying the reverberation time. To this end, the reverberation time can be varied by considering different rooms [10, 13, 14], or by varying the absorption properties within a room [9]. To remove the room-dependent effects, anechoic rooms can be considered [6, 14, 15, 17–20, 22, 23].

In this paper, a dummy head is equipped with behind-the-ear (BTE) hearing instrument shells, devices mounted behind the ear of the dummy head that contain the hearing instrument microphones, as well as receiver-in-canal (RIC) hearing instrument loudspeakers, devices where the loudspeaker signal is transmitted electrically to a loudspeaker in the ear. This configuration allows all aforementioned degrees of freedom to be considered simultaneously

1. by utilising both external loudspeakers replicating external sound sources, and RIC hearing instrument loudspeakers mounted on the dummy head's earpiece, replicating the hearing instrument loudspeakers, with four different domes (open, semi-open, closed, and no-RIC);
2. by utilising microphones on the dummy head's earpiece replicating BTE hearing instrument microphones (using which software directional microphones could be configured), as well as utilising microphones at the eardrum, and external microphones;
3. by both computing impulse responses from swept-sine recordings between each microphone-loudspeaker pair and recording audio (male and female speech, speech shaped noise, singing voice, stringed instrument, wind instrument, and percussion instrument);
4. by replicating the recording procedure in four different reverberation times ($T_{30} = 0.09\,\mathrm{s}$, $T_{30} = 0.47\,\mathrm{s}$, $T_{30} = 0.73\,\mathrm{s}$, and $T_{30} = 1.48\,\mathrm{s}$).

The resulting dataset will be referred to as the Hearing Instrument Dataset in Various Acoustical Scenarios (HIDVAS).

As all degrees of freedom are explored simultaneously and all loudspeaker-microphone pairs are considered, e. g., also considering the contralateral feedback paths

between hearing instrument loudspeakers and microphones, the dataset facilitates its usage as a 'hearing instrument in a box'. To further demonstrate this, in this paper, three example use cases are explored, thereby also motivating the different components (e.g., the multitude of domes, the use of external microphones) included in the dataset. A feedback use case explores the effect of the dome type and reverberation time on the feedback paths between RIC hearing instrument loudspeakers and BTE hearing instrument microphones, an assisted listening device use case explores the direct-to-reverberant ratio (DRR) both in the external microphone close to the external sound source and in the BTE hearing instrument microphones, and a leakage use case explores the propagation of sound from an external source to the eardrum for different domes.

In [6], a dataset of impulse responses with a dummy head and different human listeners has been recorded in an anechoic chamber, thereby also considering microphones at the earpiece and eardrum, loudspeakers at the earpiece, and external loudspeakers. This setup also facilitates the 'hearing instrument in a box' application. Nevertheless, the dataset as presented here is complementary as it also includes external microphones, provides direct recordings of audio signals next to impulse responses, and considers multiple reverberation times. Additionally, while in [6] an in-the-ear (ITE) hearing instrument was used, thereby studying devices where both the microphones and loudspeakers are part of the same rigid structure to be inserted into the ear canal, in the present dataset BTE hearing instrument microphones are used together with RIC hearing instrument loudspeakers.

Finally, while in [6] the effect of the hardware is removed, as the primary focus in [6] is to study the properties of the acoustic paths, in the present dataset the hardware effects are preserved in the recordings. Indeed, the intended goal of the presented dataset is to provide realistic data for algorithm evaluation and training, not to provide and study acoustic impulse responses. The inclusion of the hardware adds realism for evaluation and training as the hardware-in-the-loop affects the recordings in a real-life hearing instrument, and as the effect of hardware and acoustics cannot be disentangled in practice. When recording data for this algorithm evaluation and training, the relative differences between the microphones should be preserved, such that a recording procedure is adhered to to achieve a comparable recording level across microphones. Other recently recorded and documented datasets also incorporate such hardware effects and follow a similar procedure, e.g., [9, 10, 14, 25], and have been used for algorithm evaluation and training, e.g., [26].

The paper is structured as follows. In Section 2, the recording rooms, and microphone and loudspeaker configuration are described. In Section 3, the audio equipment is discussed, and in Section 4, the recorded impulse responses and audio signals are described. In Section 5, the dataset is analysed in terms of reverberation time, and the error between the linear impulse response model and the recorded audio. Finally, in Section 6 the dataset is further motivated in terms of the three example use cases.

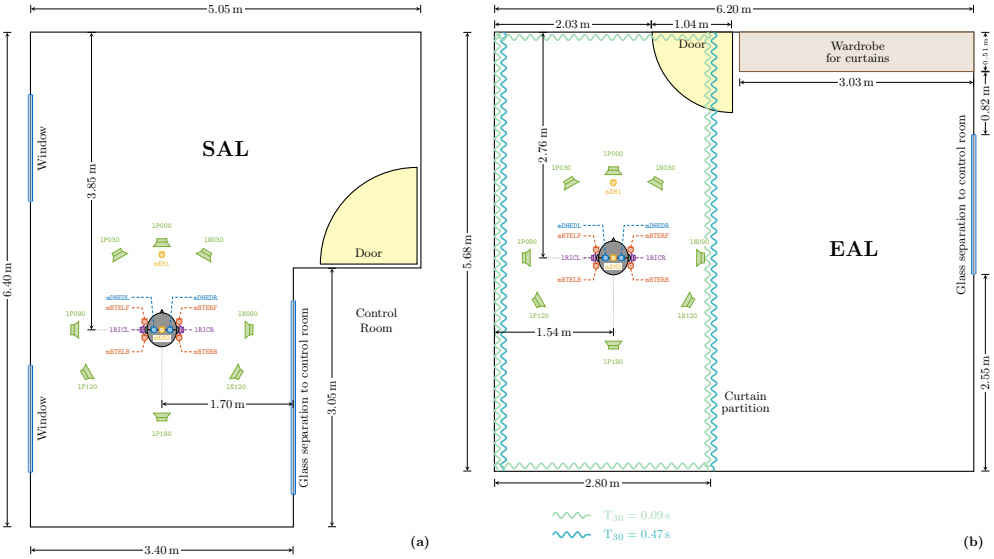


Fig. 1: Floor plans of the SONORA Audio Laboratory (SAL) (a) room and ExpORL Audio Laboratory (EAL) (b) room, on which the positions of the microphones and loudspeakers are indicated. For the floor plan of the EAL, the curtains used for the two lowest reverberation times, i. e., $T_{30} = 0.09\,s$ and $T_{30} = 0.47\,s$, are schematically indicated. The curtain covering the ceiling for $T_{30} = 0.09\,s$ has been omitted from the floor plan.

2 Setup

2.1 Recording Rooms

2.1.1 SONORA Audio Laboratory (SAL)

The floor plan of the SONORA Audio Laboratory (SAL) [27], located at the Department of Electrical Engineering (ESAT) of KU Leuven (Heverlee, Belgium), is shown in Fig. 1a. The floor plan of this 3.75 m-tall room is consistent with the description, schematics, and figures provided in [10], where further characteristics of the room have been detailed. Due to the storage of additional material and equipment within the room, the reverberation time is $T_{30} = 1.48\,s$, estimated as described in section 5, which differs from $T_{20} = 2.1\,s$ as reported in [10].

2.1.2 ExpORL Audio Laboratory (EAL)

The floor plan of the ExpORL Audio Laboratory (EAL), located at the Department of Neuroscience (ExpORL) of KU Leuven (Leuven, Belgium), is shown in Fig. 1b. The reverberation time of this 2.59 m-tall room can be varied by covering the walls and ceiling using curtains. Additionally, the room can be split in two using a curtain partition. As schematically indicated in Fig. 1b, the setup was positioned in the left side of the room when facing the door; as also apparent from a picture of the setup



Fig. 2: Picture of the setup in the ExpORL Audio Laboratory (EAL) leading to T_{30}=0.47 s.

in Fig. 2. Three different curtain configurations were used, resulting in different T_{30}-values, which were estimated as described in Section 5:

- T_{30} = 0.09 s: The curtain partition was used to split the room in two, and all walls in the left side of the room were covered using the curtains as shown in Fig. 1b. Additionally, the ceiling was covered using a curtain.
- T_{30} = 0.47 s: The curtain partition was used to split the room in two. The 5.68 m-long sides of the rectangular left side of the room were covered using the curtains, while the 2.80 m-long sides were left uncovered as shown in Fig. 1b. The ceiling was left uncovered as well.
- T_{30} = 0.73 s: The curtain partition was left open, and no curtains covered any of the walls or ceiling.

2.2 Microphone and Loudspeaker Configuration

Fig. 3 shows the microphone and loudspeaker configuration as positioned both in the SAL and EAL, and Table 1 contains the accompanying descriptions of the microphone and loudspeaker naming convention. The centre of the Cortex MK II artificial head and torso is considered as the origin in this configuration. This Cortex MK II artificial head and torso is a dummy head with cavities to model the ear canal and on which different models of the outer ear can be mounted. At the end of the cavity on each

side of the head, the dummy head contains a microphone (`mDHEDL` and `mDHEDR`) to record signals at the eardrum. These microphones are MTG condenser microphones, and were located at a height of 1.2 m in accordance with the height of the dummy head and correspondingly in accordance with the height of a person sitting on a chair.

On top of the dummy head ears, type SP15 BTE hearing instruments provided by Cochlear Ltd. were mounted. Each of these BTE hearing instruments contains two microphones (`mBTELF`, `mBTELB`, `mBTERF`, and `mBTERB`). Two Sonion Steel Series XP RIC hearing instrument loudspeakers, as shown in Fig. 4b, were inserted in the ear canal to serve as hearing instrument loudspeakers (`lRICL` and `lRICR`). The insertion of the RIC hearing instrument loudspeakers was done in a way that the posterior tip of the RIC hearing instrument loudspeaker was aligned with the entrance of the ear canal of the dummy head. To keep the RIC hearing instrument loudspeakers in place, three different domes with variable occlusion were used as shown in Fig. 4c. Ranging from least to most occluding, the Phonak open dome large 4.0, Phonak open dome large 3.0, and Phonak power dome large 3.0 were used. These domes will be further referred to in terms of occlusion as open, semi-open, and closed respectively. A condition without insertion of the RIC hearing instrument loudspeakers was also considered, and will be referred to as the no-RIC condition. A closeup of the left BTE hearing instrument microphones and RIC hearing instrument loudspeaker mounted on the dummy head is shown in Fig. 4a.

Eight Genelec 8030CP loudspeakers, used to model external sound sources, were positioned on loudspeaker stands on a circular grid with a 1 m radius at the angular positions 0°, ±30°, ±90°, ±120°, and 180°. Herein, 0° is defined as the frontal direction of the dummy head, i. e., the direction the dummy head is facing. The corresponding loudspeaker positions are labeled as `lP000`, `lP030`, `lN030`, `lP090`, `lN090`, `lP120`, `lN120`, and `lP180`. Herein, the `P` and `N` naming convention refers to positive and negative angles respectively, and the three succeeding digits refer to the angle at which the loudspeaker is positioned. As these loudspeakers are external to the dummy head configuration, they will be referred to as external loudspeakers. The external loudspeakers were located at a height of 1.1 m.

Finally, two AKG CK32 microphones, external to the dummy head configuration and hence referred to as external microphones, were considered. These microphones are consequently referred to as `mXM1` and `mXM2`. To this end, `mXM1` was positioned close to the frontal loudspeaker at the same height as this loudspeaker at a distance of 0.17 m to model the use of assisted listening devices. This `mXM1` microphone was oriented horizontally towards the `lP000` loudspeaker. The microphone `mXM2` was positioned on top of the dummy head above its centre position to allow for microphone recordings less influenced by the head shadow effect of the dummy head. This `mXM2` microphone was oriented vertically towards the ceiling.

For further details about loudspeaker and microphone positions, all the 3D coordinates with respect to the centre of the dummy head are provided in a `coordinates.csv` file in the dataset.

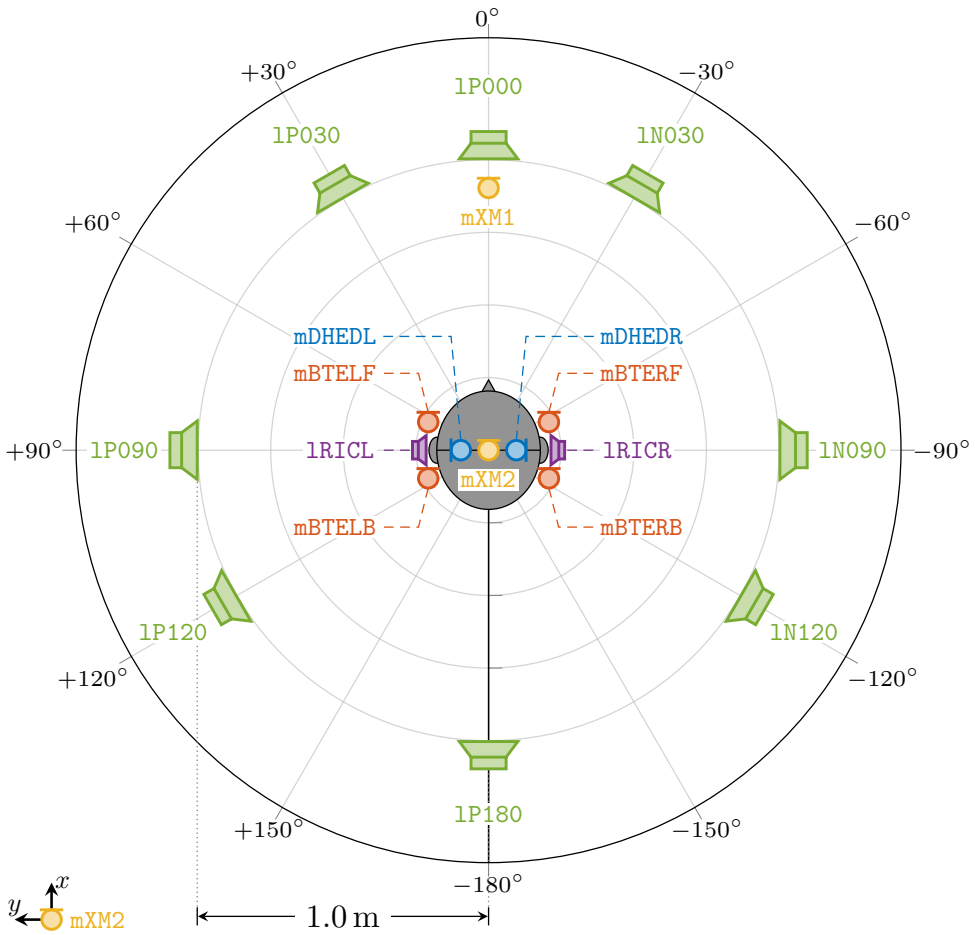


Fig. 3: Plan view of the microphone and the loudspeaker configuration. The coordinate system on the bottom left indicates the convention used for the coordinates given in the `coordinates.csv` file that comes with the dataset. The naming convention of the loudspeakers, indicated with `l` suffix, and of the microphones, indicated with `m` suffix, is detailed in Table 1.

3 Recording Chain

3.1 Hardware

The recording chain is shown in Fig. 5, and was identical for the SAL and the EAL. An iMac running Python 3.10 was used to control the audio equipment for playback and recording as a digital audio workstation. The signals were fed to the playback chain, and fetched from the acquisition chain via an RME Digiface audio interface connected to the iMac via USB. The recorded microphone signals were stored on the iMac at a sampling frequency $f_s = 48\,\text{kHz}$ using a bit depth of 16 bit.

Abbreviation	Description
mBTELF	Microphone behind-the-ear left front
mBTELB	Microphone behind-the-ear left back
mBTERF	Microphone behind-the-ear right front
mBTERB	Microphone behind-the-ear right back
mDHEDL	Microphone dummy head ear drum left
mDHEDR	Microphone dummy head ear drum right
mXMx	Microphone external microphone x
lRICL	Loudspeaker receiver-in-canal left
lRICR	Loudspeaker receiver-in-canal right
lPxxx	Loudspeaker positioned at xxx°
lNxxx	Loudspeaker positioned at $-xxx^\circ$

Table 1: Abbreviations and accompanying descriptions of microphones and loudspeakers as shown in Fig. 3.

For the playback chain, the RME Digiface audio interface was connected to an RME M-32 DA multi-channel digital-to-analog converter. The transfer of the playback signals from the RME Digiface to the RME M-32 DA was done using the ADAT protocol. The output of the RME M-32 DA was then connected to the eight Genelec 8030CP external loudspeakers and the two Sonion Steel Series XP RIC hearing instrument loudspeakers.

For the acquisition chain, the RME Digiface audio interface was connected to an RME Micstasy preamplifier/converter. The transfer of the recorded signals from the RME Micstasy to the RME Digiface was done using the ADAT protocol. The AKG CK32 external microphones were connected directly to the RME Micstasy, while the BTE hearing instrument microphones and dummy head ear drum microphones first passed through dedicated preamplifiers provided by Cochlear Ltd. and Cortex Instruments respectively. The Cortex Instruments preamplifier was the original preamplifier accompanying the Cortex MK II artificial head and torso.

As for the clock signals, the RME Digiface was used as the primary clock, to which the RME M-32 DA was connected as the secondary clock via an ADAT cable. In turn, the RME Micstasy was connected to the RME M-32 DA to receive the clock signal via a BNC cable.

3.2 Gain Adjustment

Prior to the measurements in the SAL and the EAL rooms, the software and hardware gains of the loudspeakers and microphones were determined and the hardware delay was compensated for. This procedure aims at achieving comparable recording levels across the microphones in order to maintain comparable relative differences between microphones, which is important for the intended goal of algorithm evaluation and training [10, 14]. Other hardware effects were not compensated for as the dataset functions as a 'hearing instrument in a box', where the hardware-in-the-loop does affect the recordings, and to allow for a plug-and-play usage of the dataset. Indeed, the goal of the dataset is not to provide acoustic impulse responses, which would

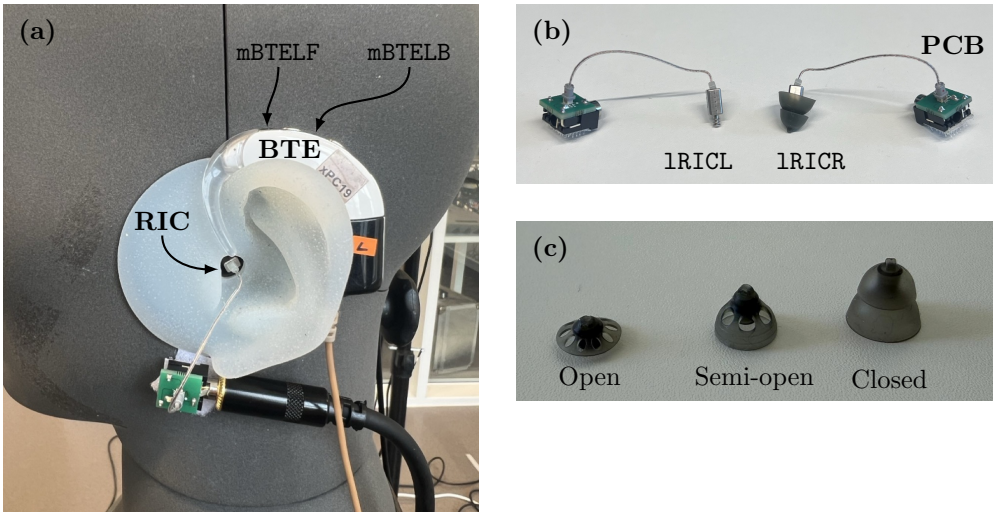


Fig. 4: (a) Closeup of the receiver-in-canal (RIC) hearing instrument loudspeaker and behind-the-ear (BTE) hearing instrument microphones placed on the left ear of the dummy head. (b) Closeup of the left RIC hearing instrument loudspeaker and right RIC hearing instrument loudspeaker (without and with dome) mounted on their respective printed circuit boards (PCBs). (c) Closeup of the RIC hearing instrument loudspeaker domes (open, semi-open, closed).

require hardware to be switched in for algorithm evaluation and training. The following procedure was used:

1. **External Loudspeaker Gain Adjustment:** The software gains of the external loudspeakers, positioned on a circle at 1 m distance from the centre of the circular grid (Fig. 3), were adjusted such that each loudspeaker achieved an equal pressure at the centre of this grid. To this end, a B&K Type 2250 sound pressure level (SPL) meter was placed at the centre of this circular grid at a 1.2 m height, i. e., at the same height as the dummy head ear drum microphones and at the same distance from each of the external loudspeakers, by temporarily removing the dummy head. The SPL meter was tilted upwards. The software gain values in the control software were adjusted to ensure an equal output level at the SPL meter of 85 dB SPL for each of the individual external loudspeakers playing back speech shaped noise (SSN), resulting in the software gain values used throughout the recordings. These gain values are reported in [1, 2]. To verify these gain values, as per [25], the power in the reverberant tail $P_{\text{rev.}}$ for each of these loudspeakers to the `mXM2` microphone was calculated. For a fixed room, this reverberant power only depends on the power of the loudspeaker signal, and on hence the associated software gain [28]. As shown in Fig. 6, the reverberant power for each of these impulse responses is in agreement, with a maximum deviation between any two loudspeakers of 1.6 dB.
2. **Microphone Gain Adjustment:** The microphone gains were subsequently adjusted such that a diffuse sound field resulted in the same overall recording level

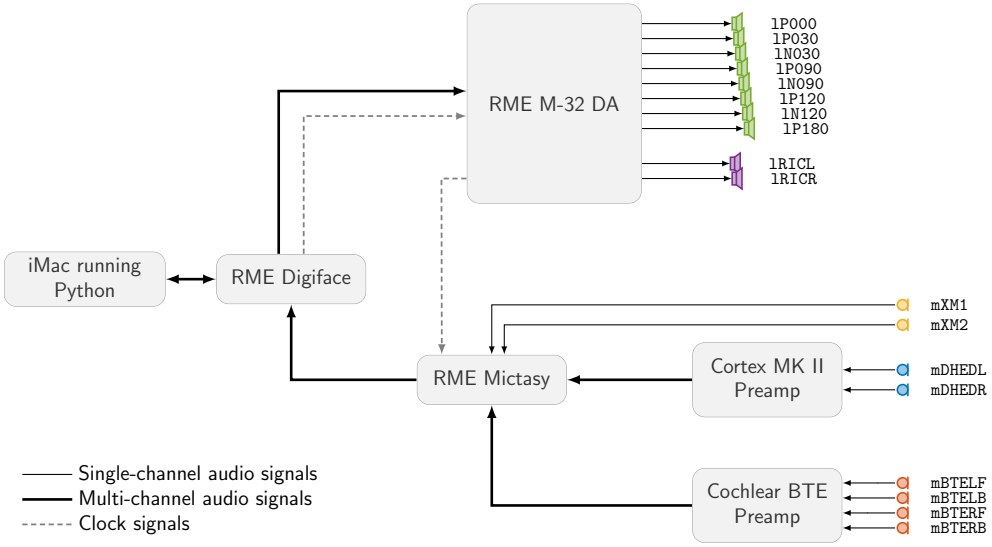


Fig. 5: Recording chain indicating the different components used for the acquisition and playback of audio signals.

in each of the microphones. To this end, after loudspeaker gain adjustment, the dummy head itself, containing the dummy head ear drum microphones and BTE hearing instrument, and the external microphones were put in place, while the RIC hearing instrument loudspeakers were still excluded. A diffuse sound field was created by positioning a Genelec 8030CP loudspeaker and a Genelec 8020DPM loudspeaker in opposite corners of the room and tilting those loudspeakers outwards to face the corners, by positioning the SPL meter at 1 m height between the dummy head and the `mXM1` microphone, by adjusting the software gain values in the control software such that each of the additional Genelec 8030CP and 8020DPM loudspeakers separately playing back SSN produced 70 dB SPL at the SPL meter, and finally by playing back the same SSN through both loudspeakers simultaneously. The sound field was verified to be sufficiently diffuse in the measurement area by placing the SPL meter at different positions in the room where the measurement microphones would be placed and checking that the root-mean-square (RMS) power of the recorded signals were within a 1.5 dB SPL tolerance. Using this diffuse sound field, the gains on the RME Micstasy associated to the BTE hearing instrument and dummy head microphones, and to the external microphones were adjusted to ensure that the recordings were at the same level (within a 1.5 dB tolerance) for this diffuse sound field. Additionally, the operating gain was chosen roughly midway the dynamic range of the RME Micstasy for the software gain values of the external loudspeakers as determined in the previous step in order to maximise the dynamic range of the recordings. These gain values are reported in [1, 2].

3. **RIC hearing instrument Loudspeaker Gain Adjustment:** The software gains of the `lRICR` (`lRICL`) loudspeaker were adjusted to achieve an equal pressure at the `mDHEDR` (`mDHEDL`) microphone as with the `lN090` (`lP090`) loudspeaker. To this end, SSN was played back from the `lN090` (`lP090`) loudspeaker and the microphone signal at the corresponding `mDHEDR` (`mDHEDL`) microphone was recorded without the `lRICR` (`lRICL`) loudspeaker inserted. After inserting the `lRICR` (`lRICL`) loudspeaker with the `CL` dome, the microphone signal was recorded again with playback from this `lRICR` (`lRICL`) loudspeaker. The gain values in the control software for the `lRICR` (`lRICL`) loudspeaker were then adjusted so that the RMS power of the microphone signals were within 0.5 dB tolerance to achieve an equivalent sound level at the eardrum.
4. **Hardware Latency Compensation:** The latency of the recording and playback equipment was measured using a calibration pulse signal. In the impulse responses and recorded audio this latency is compensated for, such that the direct component in the recordings matches with the travel time of a sound wave from the source to the microphone.

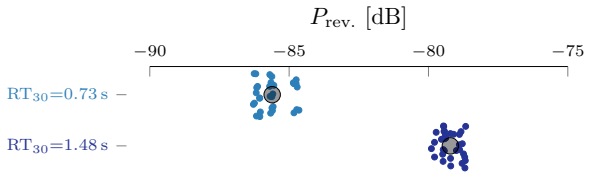


Fig. 6: The power in the reverberant tail for the impulse responses for each of the external loudspeakers to `mXM2` are in agreement, confirming the validity of the loudspeaker gain adjustment. The black dots represent the average across loudspeakers.

4 Recorded Signals

The dataset contains the impulse responses for all microphone-loudspeaker pairs, and audio (male and female speech, speech shaped noise, singing voice, stringed instrument, wind instrument, and percussion instrument) played back by each individual loudspeaker and recorded by all microphones simultaneously. The recording sweptsines and computation of the resulting impulse responses is described in Section 4.1 and the recording of the audio in Section 4.2.

In total, the dataset contains 14 592 files of raw audio data, consisting of 1216 files of recorded sweeps for impulse response measurements and 13 376 files of other recorded audio, 1216 files of computed impulse responses, and 11 files of audio sources, for a total of 15 819 files. The raw audio data amounts to 156.39 h of recordings, with an additional 1.01 h of computed IR data and 0.12 h of source audio data, for a total of 157.53 h of audio data. The dataset occupies 50.34 GB of storage for the raw files, 0.65 GB for the computed impulse response files, and 0.04 GB for the source files, totaling 51.03 GB.

4.1 Impulse Responses

To determine the impulse response for a microphone-loudspeaker pair, the exponential sine sweep method was used, wherein a sine with exponentially increasing frequency excites the system and then the recorded signal is deconvolved with the original swept-sine to obtain the impulse response [29, 30].

For each microphone-loudspeaker pair, two repetitions of an 8 s-long swept-sine followed by 3 s of silence were played as shown schematically in Fig. 7. This 3 s of silence was chosen to be larger than the largest reverberation time of the room (T_{30} = 1.48 s) to avoid intersweep interference Additionally, 1 s of silence was added before the first sweep and 0.5 s of silence was added after the second sweep to allow for latency compensation in case the second sweep was used for deconvolution. Finally, an additional silence of 2 s was inserted between the measurements of any two consecutive microphone-loudspeaker pairs; this was not strictly necessary for the impulse response measurements, but it was retained for consistency with the audio recordings. This presentation structure of the sweeps is schematically illustrated in Fig. 7. The swept-sines were designed to excite a frequency region ranging from 1 Hz to 24 kHz according to [29, 30], although other swept-sines exist, e.g., [31]. Before playback, the gain of the swept-sines was tuned to achieve the same RMS level as the SSN from the Voice Cloning Toolkit (VCTK) corpus (cfr. Section 4.2).

The deconvolution was performed using the first of the recorded swept-sines unless the first recorded swept-sine was noisy (e. g., due to impulsive noise) as identified based on visual inspection. In that case, the second swept-sine was used for the deconvolution[1]. Averaging the impulse responses from both recordings was explicitly avoided as per [30]. The hardware delay in the resulting impulse responses was compensated for as detailed in Section 3.2 such that the direct component of the impulse responses matches with the travel time of sound waves from the loudspeakers to the microphones. The resulting impulse responses were saved at a bit depth of 32 bit. Other hardware effects were not compensated for as the dataset functions as a 'hearing instrument in a box', where the hardware-in-the-loop indeed does affect the recordings.

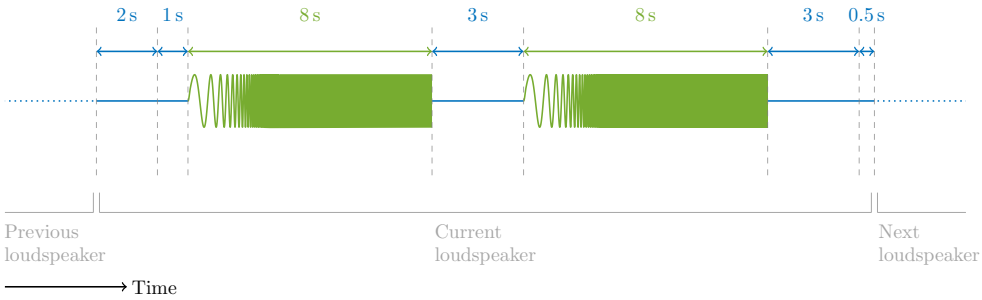


Fig. 7: Presentation structure of the exponential sine sweeps. The silence interleaving the sweeps avoids intersweep interference and allows for latency compensation.

[1]The second swept-sine was only used for 2 of the recordings.

4.2 Audio

Although the impulse responses adequately model the input-output relationship between the loudspeaker and microphone signals, as will be shown in Section 5, there are still non-linear effects in the loudspeakers and microphones which cannot be fully captured by the impulse responses. Consequently, audio is recorded as well, across a wide range of speech and music classes (male and female speech, speech shaped noise, singing voice, stringed instrument, wind instrument, and percussion instrument) by playing back the audio through each individual loudspeaker, one at a time, and recording through all microphones simultaneously.

Four types of speech materials were included in the dataset, i.e., Flemish and English speech as spoken by both female and male speakers. The Flemish female speech materials were sampled from the corpus of the Leuven intelligibility sentences test spoken by a female speaker (LISTf) [32], and the Flemish male speech materials were sampled from the corpus of the Leuven intelligibility sentences test spoken by a male speaker (LISTm) [33]. Both corpora are typically used for audiological testing [32, 33]. The English female and male materials were sampled from the corpus of the Centre for Speech Technology Research (CSTR) VCTK [34], commonly employed in the evaluation of speech enhancement algorithms [35].

More specifically, for the LISTf material, one speech file of duration 54.9 s was created by concatenating the ten sentences from `list01` (`001.wav - 010.wav`) and the first five sentences from `list02` (`001.wav - 005.wav`). Each of the sentences was separated by 2.5 s of silence to avoid intersentence interference, as such interference would disturb an audiological test administered using the recorded material.[2] Similarly, for the LISTm material, one speech file of duration 65.0 s was created by concatenating the ten sentences from `list01` (`001.wav - 010.wav`) and the first five sentences from `list02` (`001.wav - 005.wav`). Again, 2.5 s of silence was inserted between each of the sentences to avoid intersentence interference. Finally, for the VCTK female and male materials, one female and one male speech file of duration 30.6 s and 30.7 s respectively, was created by concatenating randomly sampled sentences from female and male speakers. Contrary to the LISTf and LISTm material, no intersentence silence was added between the sentences of the VCTK materials, as these VCTK materials were included to mimic continuous natural speech.

Three types of SSN, of 20.0 s each, were included in the dataset, i.e., SSN calculated using the LISTf corpus, the LISTm corpus, and the VCTK corpus. These SSN materials were created by modifying the spectrum of white noise to match the average spectrum of the respective corpora, and are commonly employed to mask speech signals [36].

Four types of music materials (singing voice, stringed instrument, wind instrument, and percussion instrument) were included in the dataset in order to cover a range of timbres and instruments classes, e.g., to be used for music processing [37]. The singing voice was selected from the 'Anechoic recordings of Italian opera' corpus [38], and the other music materials from the 'Music for Archimedes' album by Bang & Olufsen [39] (based on [40]).

[2] This value was chosen to be between 1.5 times and 2 times the largest reverberation time of the room (T_{30} = 1.48 s).

More specifically, for the singing voice, one music file (`mu_Soprano.wav`) of duration 30.8 s was created by selecting a part of the original audio file wherein a lyric soprano performs a portion of the aria 'Di tale amor, che dirsi' from *Il Trovatore* by Giuseppe Verdi. For the stringed instrument, wind instrument, and percussion instrument 30.8 s, 31.1 s, and 30.4 s files were created by selecting parts of the original cello (`Track22.wav`), trumpet (`Track37.wav`), and bongo (`Track26.wav`) files respectively, on the 'Music for Archimedes' album. Fade-out was applied to the selected portion of the source audio when necessary.

All speech and music materials were recorded in (hemi-)anechoic rooms [32–34, 38, 39]. Before playback, all the speech and music `.wav` files were additionally normalised to achieve a target of −23 dB on the loudness units full scale (LUFS) according to the ITU-R BS.1770-4 standard [41] using the pyloudnorm toolbox [42]. During playback, a silence of 2 s was inserted between each of the audio files to avoid interfile interference. As for the impulse responses, the hardware delay was compensated for, such that the delay of the audio recordings matches the delay of the impulse responses convolved with the sound sources.

5 Analysis

In Section 5.1, the T_{30} estimates of the reverberation times are discussed to characterise the acoustic properties of the rooms, and in Section 5.2 the recorded audio signals are compared to the signals obtained by convolving the source audio with the impulse responses to validate the applicability of the impulse response model.

5.1 T_{30}

Fig. 8 shows the averaged (i. e., mid) and per-octave-band-frequency estimates (i. e., at 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz) of the T_{30} estimates obtained from the impulse responses between each external loudspeaker and the `mXM2` external microphone according to the methodology specified in the ISO 3382-1 standard [43] and implemented using the IoSR MATLAB Toolbox [44]. This procedure was repeated four times, as there are four dome types and these domes do not affect the reverberation time. The mid T_{30} estimate has been obtained as the average T_{30} across the 500 Hz and 1 kHz octave bands conform the ISO 3382-1 standard [43]. As shown by the individual data points in Fig. 8, little deviation from the mean T_{30} is found across different loudspeaker positions (and domes).

5.2 Impulse Response Versus Recorded Audio

Fig. 9 shows the spectra of source audio, recorded audio, and the source audio convolved with the impulse response. The spectra of recorded and convolved audio match well, and illustrate the effect of smearing due to the reverberation. In order to further confirm the validity of this linear model of the impulse responses, the signal-to-error ratio (SER) is analysed between the recorded audio signal and the source audio signal convolved with the corresponding impulse response [10, 45]. Denoting the recorded audio signal at time k as $a[k]$, the source audio signal as

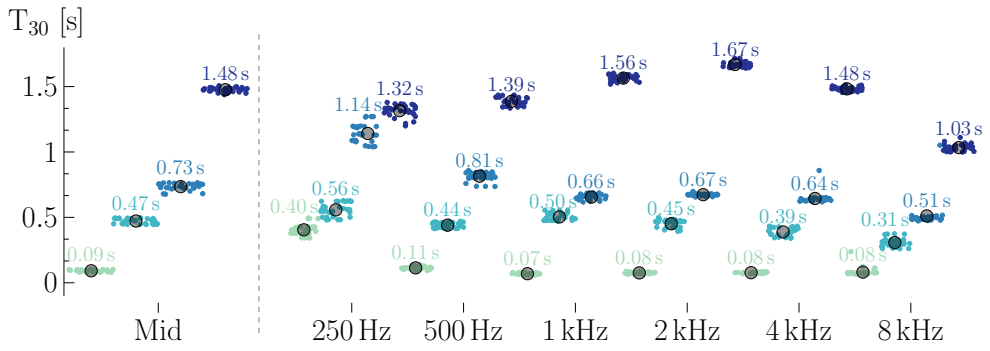


Fig. 8: Estimated T_{30} averaged across the 500 Hz and 1 kHz octave bands (mid) and in six individual octave bands, obtained from the impulse responses between the external loudspeakers and the `mXM2` external microphone. Each impulse response was remeasured four times, as the measurement procedure was repeated for the four dome types and the dome does not affect the reverberation time. Each data point represents the estimate obtained from one impulse response, and the black dots represent the average across loudspeaker positions (and domes). The numeric value of this average is shown as well. The order of the estimates per frequency band and the associated color scheme correspond, from left to right, to $T_{30} = 0.09\,\text{s}$, $T_{30} = 0.47\,\text{s}$, $T_{30} = 0.73\,\text{s}$, and $T_{30} = 1.48\,\text{s}$.

$s[k]$ with $\mathbf{s}[k] = \begin{bmatrix} s[k] & \cdots & s[k-L+1] \end{bmatrix}^\top \in \mathbb{R}^{L\times 1}$, and the impulse response as $\mathbf{h} = \begin{bmatrix} h[0] & \cdots & h[L-1] \end{bmatrix}^\top \in \mathbb{R}^{L\times 1}$, the SER is defined as

$$\text{SER} = 20\log_{10}\left(\frac{\sqrt{\frac{1}{K}\sum_{k=0}^{K-1} a[k]^2}}{\sqrt{\frac{1}{K}\sum_{k=0}^{K-1}\left(a[k]-\mathbf{h}^\top\mathbf{s}[k]\right)^2}}\right), \tag{1}$$

with K the total length of the audio recording. For additional validation, next to the linear model $\mathbf{h}$, the exponential sine sweep also comes with a model $\mathbf{h}_n \in \mathbb{R}^{L_n\times 1}$ for the non-linear distortion in the system [29, 30]. Correspondingly, a linear-to-distortion ratio (LDR) metric is defined as follows

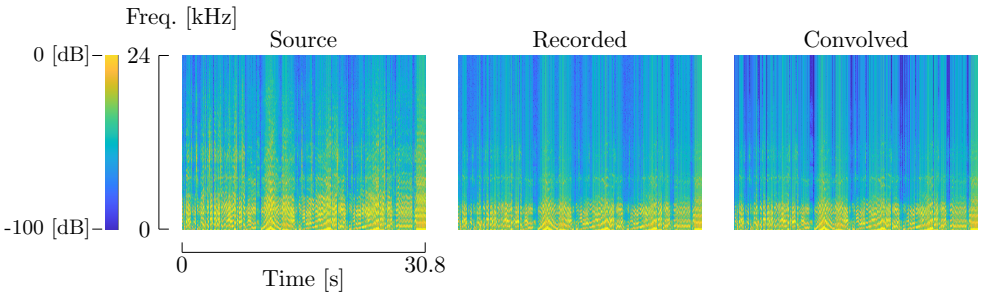
$$\text{LDR} = 20\log_{10}\left(\frac{\sqrt{\frac{1}{L}\sum_{k=0}^{L-1} h[k]^2}}{\sqrt{\frac{1}{L_n}\sum_{k=0}^{L_n-1} h_n[k]^2}}\right). \tag{2}$$

Fig. 10a shows the SER with the VCTK male speech for different loudspeaker-microphone pairs that are recorded at a sufficiently high signal-to-noise ratio (SNR), i. e., the SER between each of the external loudspeakers and both the external microphones, between the `lRICR` (`lRICL`) loudspeaker and the dummy head ear drum

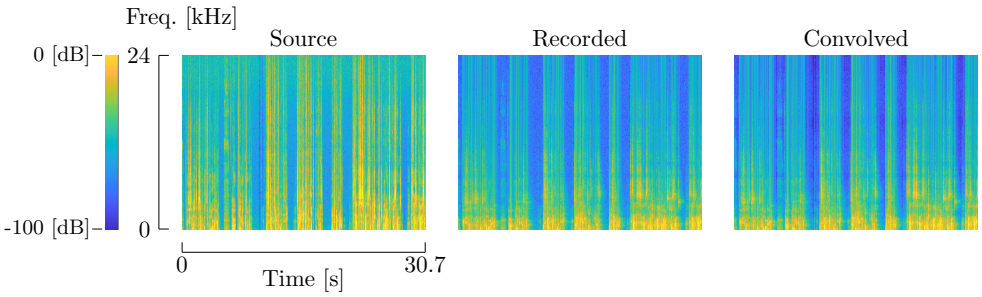
microphone at the same side mDHEDR (mDHEDL), and between the lRICR (lRICL) loudspeaker and the BTE hearing instrument microphones at the same side mBTERF and mBTERB (mBTELF and mBTELB). All domes and reverberation times are considered simultaneously, yet the closed dome was excluded when considering the impulse response from the lRICR (lRICL) loudspeaker to mBTERF and mBTERB (mBTELF and mBTELB) microphones due the closed dome having low SNR (cfr. Section 6.1).

Each of the individual data points exceeds an SER of 12 dB, and each of the mean SER values exceeds 14 dB as shown in Fig. 10a. These mean values are in agreement with [45], where mean values of 12 dB were found for the SER for similar recordings.

Fig. 10b shows the LDR for the same scenarios. The LDR in Fig. 10b demonstrates the same patterns as the SER in Fig. 10a. Consequently, the impulse responses provide an adequate model for the responses from the loudspeakers to the microphones.

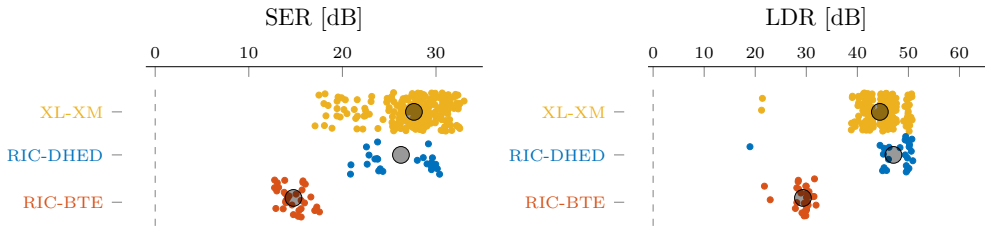


(a) Singing voice with $T_{30} = 0.09\,s$



(b) VCTK male speech $T_{30} = 1.48\,s$

Fig. 9: Spectrograms of source audio, recorded audio, and source audio convolved with the impulse response between lP000 and mXM1. The recorded and convolved audio visually match well. The source audio has been normalised to have the same root-mean-square (RMS) value as the recorded audio.



(a) SER (b) LDR

Fig. 10: The signal-to-error ratio (SER) for the Voice Cloning Toolkit (VCTK) male speech audio and the linear-to-distortion ratio (LDR). Both metrics are computed between each of the external loudspeakers and the external microphones (XL-XM), between the `lRICR` (`lRICL`) loudspeaker and the dummy head ear drum microphone at the same side `mDHEDR` (`mDHEDL`) (RIC-DHED), and between the `lRICR` (`lRICL`) loudspeaker and the behind-the-ear (BTE) hearing instrument microphones at the same side `mBTERF` and `mBTERB` (`mBTELF` and `mBTELB`) (RIC-BTE). All domes and reverberation times are considered simultaneously, although the closed dome is excluded from the RIC-BTE condition due to the low signal-to-noise ratio (SNR) of the feedback signal. Individual data points represent one recording. The mean across domes and rooms is represented as a black dot.

6 Example Use Cases

To illustrate the applicability of the dataset, three example use cases motivate the different components included in the dataset. To this end, the following example use cases are studied: a feedback use case exploring the effect of the dome and reverberation time on the feedback path from the RIC hearing instrument loudspeaker to the ipsilateral BTE hearing instrument microphone (Section 6.1), an assisted listening device use case exploring the DRR both in the external microphone close to the external sound source and in the BTE hearing instrument microphones (Section 6.2), and a leakage use case exploring the propagation of sound from an external source as played by an external loudspeaker to the dummy head ear drum microphones for different domes (Section 6.3). Consequently, the feedback use case motivates the inclusion in the dataset of the RIC hearing instrument loudspeakers and different domes; the assisted listening device use case motivates the inclusion of the external microphones; and the leakage use case motivates the inclusion of the different domes.

6.1 Feedback Use Case

Fig. 11 shows the impulse responses and magnitude spectra of the feedback paths between the `lRICR` loudspeaker and `lBTERF` microphone, across domes and reverberation times.

Across domes, the open and semi-open domes result in a similar magnitude of the feedback path as the direct path and early reflections have a similar magnitude

T_{30}	open	semi-open	closed
0.47 s	−12.28	−8.97	−7.78
0.73 s	−2.76	−10.54	−6.98
1.48 s	−3.93	−11.06	−7.69

(a) `lRICR - mBTERF`

T_{30}	open	semi-open	closed
0.47 s	−20.05	−19.24	−8.90
0.73 s	−9.46	−9.02	−5.06
1.48 s	−9.19	−9.59	−1.72

(b) `lRICL - mBTELF`

Table 2: Normalised Euclidean system distance (NESD) between the impulse response recorded with $T_{30} = 0.09\,\text{s}$ and the impulse responses with the corresponding dome and loudspeaker - microphone combination at different reverberation times.

despite the open dome having more venting (cfr. Fig. 4c). The closed dome, having no venting, blocks a larger proportion of the sound leaking from the ear canal, resulting in a smaller magnitude of the feedback path as shown in Fig. 11.

Across reverberation times, the overall shape and magnitude of the feedback paths also remains similar for each particular dome due to the path from `lRICR` loudspeaker to `mBTERF` microphone being short (in the cm order of magnitude) and being mainly affected by the dummy head geometry rather than the room itself, indicating only a minor effect of the reverberation time. This is further quantified by the normalised Euclidean system distance (NESD), calculated as [46]

$$\text{NESD} = 20\log_{10}\left(\frac{\sqrt{\frac{1}{L}\sum_{k=0}^{L-1}\left(\mathbf{h}_{\text{T,d}}[k] - \mathbf{h}_{0.09\,\text{s,d}}[k]\right)^2}}{\sqrt{\frac{1}{L}\sum_{k=0}^{L-1}\mathbf{h}_{0.09\,\text{s,d}}[k]^2}}\right). \tag{3}$$

To this end, the NESD quantifies the distance between the impulse response for each reverberation time $\text{T} \in \{0.47\,\text{s}, 0.73\,\text{s}, 1.48\,\text{s}\}$ and dome $\text{d} \in \{\text{open, semi-open, closed}\}$ with respect to the impulse response recorded with the reverberation time 0.09 s and the same dome d. In Table 2, the NESD is shown for the `lRICR - mBTERF` and `lRICL - mBTELF` loudspeaker - microphone combinations. Before calculating the NESD, the impulse responses were time aligned. While the NESD generally increases with the reverberation time, the NESD remains significantly negative, indicating only a minor effect of the reverberation time.

6.2 Assisted Listening Device Use Case

Fig. 12 shows the DRR for the impulse responses from the frontal `lP000` loudspeaker to both the `mXM1` external microphone close to the loudspeaker (0.17 m) and the BTE hearing instrument microphones far away from the loudspeaker ($\approx 1\,\text{m}$). These DRR measurements were repeated four times, as there are four dome types and these domes do not affect the DRR in the `mXM1` external microphone or the BTE hearing instrument microphones. This DRR compares the energy of the direct sound to the energy of the

reverberation, and is defined as [28]

$$\mathrm{DRR} = 10\log_{10}\left(\frac{\sum_{k=k_0-\delta}^{k_0+\delta} h[k]^2}{\sum_{k=k_0+\delta+1}^{L-1} h[k]^2}\right). \tag{4}$$

Herein, k_0 represents the arrival time in samples of the direct sound, δ defines a temporal region in samples around this arrival time, thereby partially incorporating some early reflections, and L is the length of the impulse response $h[k]$. The DRR was calculated using the IoSR MATLAB Toolbox [44] with $\frac{\delta}{f_s} = 2.5\,\mathrm{ms}$ and f_s the sampling frequency. Higher DRR values have been shown to correlate well with increased speech intelligibility [28]. Consequently, assisted listening devices aim at utilising a microphone placed close to the speaker of interest, in order to stream the recordings directly to the hearing instrument [47].

As shown in Fig. 12, the DRR in the `mXM1` external microphone is larger than the DRR in the BTE hearing instrument microphones due to the smaller distance from `mXM1` to `lP000`. Indeed, while the energy of the direct sound follows an inverse square law related to the distance between loudspeaker and microphone, the energy of the reverberation is unaffected by the distance [28]. The unexpected small DRR increase from `mXM1` to `mBTELF` and `mBTERF` for T_{30}=0.09 s can be attributed to the spread of the estimates and to the usage of a temporal region defined by δ, such that some early reflections are incorporated. The DRR is, additionally, larger in the frontal than in the rear BTE hearing instrument microphones due to the increased distance from the `lP000` loudspeaker. Finally, the difference between the DRR in the `mXM1` external microphone and the DRR in the BTE hearing instrument microphones increases as the reverberation time increases as the energy of the reverberation increases with the reverberation time [28]. Consequently, the effectiveness of using an external microphone in assisted listening devices is demonstrated, especially when the reverberation time is large.

6.3 Leakage Use Case

Fig. 13 shows the power ratio P_{ratio} of the audio signals recorded in the dummy head ear drum microphones when there is a dome inserted (open, semi-open, and closed) and the audio signals recorded in the dummy head ear drum microphones when there is no dome inserted (no-RIC). This power ratio P_{ratio} quantifies the amount of leakage blocked by each of the domes compared to the no-RIC case. Indeed, ideally, only the sound processed by the hearing instrument as played back by the RIC hearing instrument loudspeaker should reach the eardrum, while the unprocessed sound from the external sound source should be blocked. Consequently, this P_{ratio} between audio signal $a_{\mathrm{d}}[k]$ at the `mDHEDR` (`mDHEDL`) microphone for dome $\mathrm{d} \in \{\mathrm{open}, \mathrm{semi\text{-}open}, \mathrm{closed}\}$ and audio signal $a_{\mathrm{NR}}[k]$ at the `mDHEDR` (`mDHEDL`) microphone for the no-RIC case, defined as

$$P_{\mathrm{ratio}} = 20\log_{10}\left(\frac{\sqrt{P_{\mathrm{d}}}}{\sqrt{P_{\mathrm{NR}}}},\right) \tag{5}$$

with $P_{\mathrm{d}} = \frac{1}{K}\sum_{k=0}^{K-1} a_{\mathrm{d}}[k]^2$ and $P_{\mathrm{NR}} = \frac{1}{K}\sum_{k=0}^{K-1} a_{\mathrm{NR}}[k]^2$, should be as small as possible. This P_{ratio} is assessed using the frontal loudspeaker 1P000.

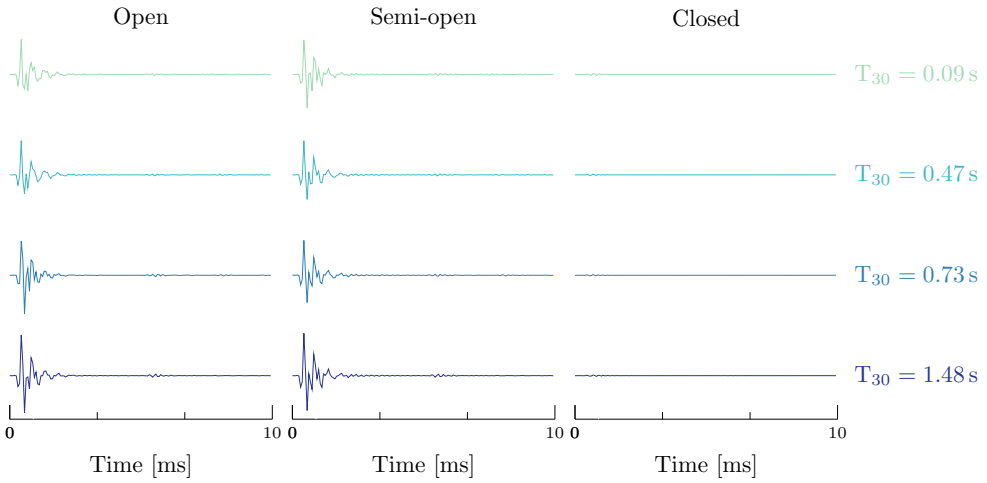
Similar to the observations in Section 6.1, the open and semi-open domes do not significantly block the external sound source leaking into the ear canal due to their venting, such that P_{ratio} in Fig. 13 is close to 0 dB and similar for both dome types. On the contrary, the closed dome has no venting and blocks the sound leakage from the `1P000` loudspeaker into the ear canal, leading to a mean P_{ratio} of -33.1 dB.

7 Conclusion

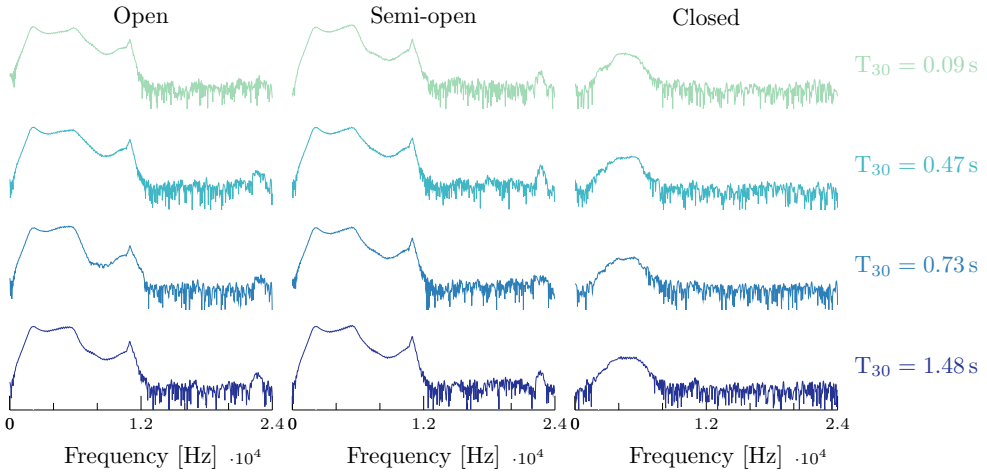
In this paper, the Hearing Instrument Dataset in Various Acoustical Scenarios (HIDVAS) has been introduced. This dataset consists of both impulse responses and audio recordings (male and female speech, speech shaped noise, singing voice, stringed instrument, wind instrument, and percussion instrument), which were recorded using eight external loudspeakers, two external microphones, and a dummy head on which behind-the-ear (BTE) hearing instrument shells were mounted with two microphones per shell, and in which receiver-in-canal (RIC) hearing instrument loudspeakers were inserted. The dummy head also has microphones, located at the ear drum, which were used to record microphone signals at the eardrum of the dummy head. These recordings were repeated for four hearing instrument domes (open, semi-open, closed, and no-RIC) in three reverberation conditions ($T_{30} = 0.09$ s, $T_{30} = 0.47$ s, and $T_{30} = 0.73$ s) in the ExpORL Audio Laboratory (EAL), and in one reverberation condition ($T_{30} = 1.44$ s) in the SONORA Audio Laboratory (SAL).

The domes affected the feedback paths and leakage, with the closed dome blocking more of the feedback and leakage than the open an semi-open domes. These open and semi-open domes, additionally, resulted in similar feedback paths and leakage. The reverberation time only had a minor effect on the feedback paths. The external microphones, placed close to the external loudspeakers resulted in an increase in direct-to-reverberant ratio (DRR), which is correlated with increased speech intelligibility, thereby motivating the use of the dataset for assisted listening device use cases as well.

The dataset is available at [1, 2], and the code for playback, recording, and analysis is available at [48].



(a) Impulse response (Linear scale)



(b) Magnitude spectrum [dB]

Fig. 11: Comparison of the ipsilateral feedback paths between `lRICR` loudspeaker and `mBTERF` microphone for the different reverberation times and domes. All paths are shown with the same magnitude, for which the numeric values are omitted due to this magnitude being of arbitrary unit. Across domes, the open and semi-open feedback paths are similar in magnitude, while the closed feedback paths are smaller in magnitude. Across reverberation times, the feedback paths are similar.

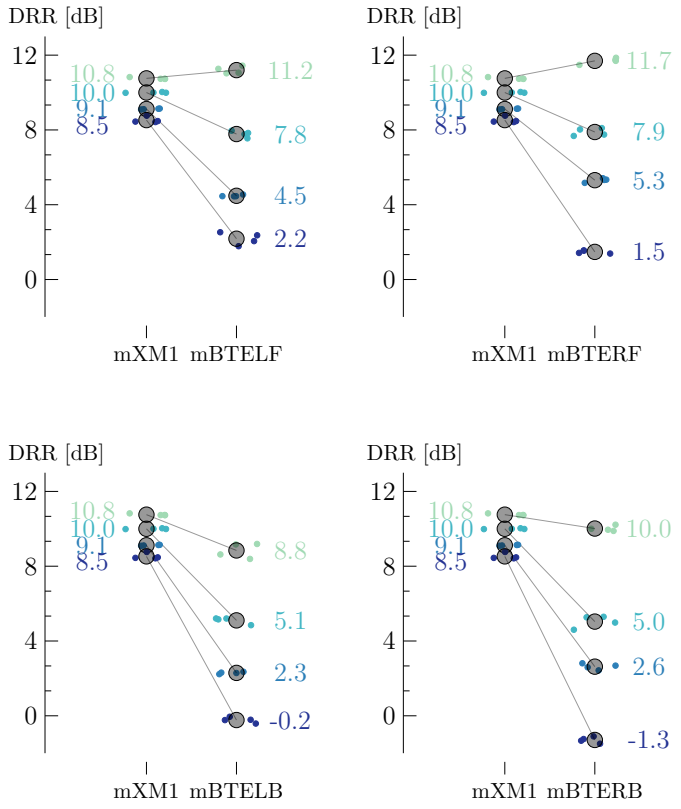


Fig. 12: The direct-to-reverberant ratio (DRR) between the `1P000` loudspeaker and both the `mXM1` external microphone and behind-the-ear (BTE) hearing instrument microphones and reverberation times. Each DRR was remeasured four times, as the measurement procedure was repeated for four dome types and the dome does not affect the DRR in the `mXM1` external microphone or the BTE hearing instrument microphones. The individual data points represent different domes, the black dots represent the average across domes, and the numbers display the numeric value of this average. The DRR is larger in the `mXM1` external microphone, which is closer to the `1P000` loudspeaker than the BTE hearing instrument microphones. The DRR increase from the `mXM1` external microphone to the BTE hearing instrument microphones is larger as the reverberation time increases.

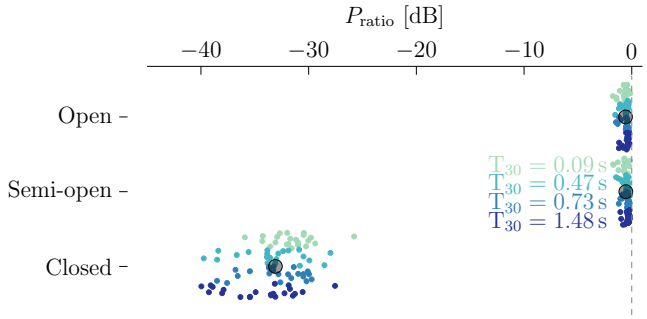


Fig. 13: The ratio of the power of the audio recordings played back by the `1P000` loudspeaker as measured at the `mDHEDR` (`mDHEDL`) microphone for each dome (d $\in$ {open, semi-open, closed}) compared to the power of the audio recordings as measured at the `mDHEDR` (`mDHEDL`) microphone when there is no dome inserted (no-receiver-in-canal (RIC)). The individual data points represent different audio recordings, and the black dots represent the average across audio recordings and reverberation times. The open and semi-open domes do not reduce this leakage from the external sound source into the ear canal well, while the closed dome does so.

List of abbreviations

BTE behind-the-ear.

CSTR Centre for Speech Technology Research.

DRR direct-to-reverberant ratio.

EAL ExpORL Audio Laboratory.

HIDVAS Hearing Instrument Dataset in Various Acoustical Scenarios.

ITE in-the-ear.

LDR linear-to-distortion ratio.
LISTf Leuven intelligibility sentences test spoken by a female speaker.
LISTm Leuven intelligibility sentences test spoken by a male speaker.
LUFS loudness units full scale.

NESD normalised Euclidean system distance.

PCB printed circuit board.

RIC receiver-in-canal.
RMS root-mean-square.

SAL SONORA Audio Laboratory.
SER signal-to-error ratio.
SNR signal-to-noise ratio.
SPL sound pressure level.
SSN speech shaped noise.

VCTK Voice Cloning Toolkit.

Declarations

Availability of data and materials. The HIDVAS dataset is openly accessible at [1], except for the inclusion of the Leuven intelligibility sentences test spoken by a female speaker (LISTf) and Leuven intelligibility sentences test spoken by a male speaker (LISTm) material and accompanying speech shaped noise (SSN) recordings due to license restrictions. At [1], the dataset is available either in full, thus including both audio recordings and impulse responses, or in an economic version containing only impulse responses. The dataset including the LISTf and LISTm material and

accompanying SSN recordings is available at [2] together with the information on how to comply with the LISTf and LISTm material licenses.

The code used for playback and recording is available in a GitHub repository [48]. In the same GitHub repository [48], the code used to generate the results in Section 3.2 (Fig. 6), in Section 5 (Fig. 8-Fig. 10), and in Section 6 (Table 2 and Fig. 11-Fig. 13) is available as well.

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

Funding. This research was carried out at the ESAT Laboratory of KU Leuven in the frame of Research Council KU Leuven C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices", and Aspirant Grant 11PDH24N (for AR) from the Research Foundation - Flanders (FWO), and at the ExpORL laboratories of the Dept. Of Neurosciences, with support of the EU-project MSCA-DN-CherISH-101120054 "Cochlear implants and spatial hearing: Enabling access to the next dimension of hearing" (JW) and VLAIO-O&O-PreENI-HBC.2024.0184 "Precision Electrode Neural Interface for Cochlear Implants" (JW).

Authors' contributions. AR, GB, JW, TVW, and MM designed the setup of the dataset and the recording chain. AR and GB carried out the recordings, and post-processed and analysed the data. JW, TVW, and MM contributed to the analysis of the data. AR and GB wrote the draft of the manuscript. All authors read and revised the final manuscript.

Acknowledgements. The authors would also like to thank Jan P. Koopman from Sonion for providing the RIC hearing instrument loudspeakers, Lieselot Van Deun from UZ Leuven for providing the different hearing instrument domes, and Thomas Dietzen from KU Leuven for assisting with the preparation of the recordings. Additionally, the authors would like to thank the Central Services for Electronics and Mechanics of the Department of Electrical Engineering (ESAT) of KU Leuven, including Joren Van Cleynenbreugel, for facilitating the manufacturing of the printed circuit boards (PCBs) on which the RIC hearing instrument loudspeakers were mounted.

References

[1] Roebben, A., Bernardi, G., Wouters, J., van Waterschoot, T., Moonen, M.: HIDVAS: A Hearing Instrument Dataset in Various Acoustical Scenarios for algorithm evaluation and training. Zenodo. https://doi.org/10.5281/zenodo.15773862 (2026)

[2] Roebben, A., Bernardi, G., Wouters, J., van Waterschoot, T., Moonen, M.: HIDVAS: A Hearing Instrument Dataset in Various Acoustical Scenarios for algorithm evaluation and training [Full version including license-restricted material]. Zenodo. https://doi.org/10.5281/zenodo.15774050 (2026)

[3] Kates, J.M.: Digital Hearing Aids, 1st edn. Plural Pub., San Diego (2008)

[4] Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., Weinzierl, S.: A round robin on room acoustical simulation and auralization. J. Acoust. Soc. Amer. **145**(4), 2746–2760 (2019)

[5] De Sena, E., Antonello, N., Moonen, M., van Waterschoot, T.: On the modeling of rectangular geometries in room acoustic simulations. IEEE/ACM Trans. Audio Speech Lang. Process. **23**(4), 774–786 (2015)

[6] Denk, F., Kollmeier, B.: The Hearpiece database of individual transfer functions of an openly available in-the-ear earpiece for hearing device research. Acta Acust. **5**(2), 2 (2021)

[7] Sankowsky-Rothe, T., Blau, M., Schepker, H., Doclo, S.: Reciprocal measurement of acoustic feedback paths in hearing aids. J. Acoust. Soc. Amer. **138**(4), 399–404 (2015)

[8] Hettler, F., Denk, F., Jürgens, T., Husstedt, H.: Acoustic feedback tendency in hearing aids for different types and couplings in relation to insertion gain. Zenodo (2024). https://doi.org/10.5281/zenodo.10835180

[9] Fejgin, D., Middelberg, W., Doclo, S.: BRUDEX database: Binaural Room Impulse Responses with Uniformly Distributed External Microphones. In: Proc. Speech Communication; 15th ITG Conf., pp. 126–130. IEEE, Aachen, Germany (2023)

[10] Dietzen, T., Ali, R., Taseska, M., van Waterschoot, T.: MYRiAD: A Multi-Array Room Acoustic Database. EURASIP J. Audio, Speech, Music Process. **2023**(17), 17 (2023)

[11] Delebecque, L., Serizel, R.: BinauRec: A dataset to test the influence of the use of room impulse responses on binaural speech enhancement. In: Proc. 31st European Signal Process. Conf. (EUSIPCO '23), pp. 126–130. EURASIP, Helsinki, Finland (2023)

[12] Woods, W.S., Hadad, E., Merks, I., Xu, B., Gannot, S., Zhang, T.: A real-world recording database for ad hoc microphone arrays. In: Proc. 2015 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '15), pp. 1–5. IEEE, New Platz, NY, USA (2015)

[13] Jeub, M., Schafer, M., Vary, P.: A binaural room impulse response database for the evaluation of dereverberation algorithms. In: Proc. 2009 IEEE Int. Conf. Digital Signal Process. (DSP '09), pp. 1–5. IEEE, Santorini, Greece (2009)

[14] Kayser, H., Ewert, S.D., Anemüller, J., Rohdenburg, T., Hohmann, V., Kollmeier, B.: Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. EURASIP J. Adv. Signal Process. **2009**(1), 298605 (2009)

[15] Maazaoui, M., Grenier, Y.: Romeo-HRTF: A multimicrophone Head Related Transfer Functions database (2011). https://adasp.telecom-paris.fr/resources/2011-03-31-romeo-hrtf

[16] Algazi, V.R., Duda, R.O., Thompson, D.M., Avendano, C.: The CIPIC HRTF database. In: Proc. 2001 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '01), pp. 99–102. IEEE, New Platz, NY, USA (2001)

[17] Gardner, B., Martin, K.: 3D SOUND: HRTF measurements (1994). https://www.lim.di.unimi.it/IEEE/BEGAULT/INDEX.HTM

[18] Lamba, M., Ohlmann, K., Ewert, S.D., Kollmeier, B.: Dummy head HRTFs with different head-above-torso orientations. In: Proc. DAS—DAGA 2024, pp. 1607–1609. DEGA e.V., Hannover, Germany (2024)

[19] Thiemann, J., Escher, A.: Multiple model high-spatial resolution HRTF measurements. In: Proc. DAS—DAGA 2015, pp. 797–798. DEGA e.V., Nuremberg, Germany (2015)

[20] Sass, R., Werner, S., Siegel, A.: Comparison of recording methods for measurements of individualized head-related transfer functions. In: Proc. 2010 Tonmeistertagung-VDT Int. Conv., Leipzig, Germany, pp. 721–726 (2011)

[21] Moore, A.H., Haan, J.M., Pedersen, M.S., Naylor, P.A., Brookes, M., Jensen, J.: OTIMP: The Oticon-Imperial hearing aid impulse response database. Zenodo (2019). https://doi.org/10.5281/zenodo.5040465

[22] Oreinos, C., Buchholz, J.M.: Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a Head and Torso Simulator (HATS). Acta Acust. United Ac. **99**(5), 836–844 (2013)

[23] Denk, F., Ernst, S.M.A., Ewert, S.D., Kollmeier, B.: Adapting hearing devices to the individual ear acoustics: database and target response correction functions for various device styles. Trends Hearing **22**, 2331216518779313–2331216518779313 (2018)

[24] Spriet, A., Rombouts, G., Moonen, M., Wouters, J.: Combined feedback and noise suppression in hearing aids. IEEE Trans. Audio Speech Lang. Process. **15**(6), 1777–1790 (2007)

[25] Damiano, S., MacWilliam, K., Lorenzoni, V., Dietzen, T., Waterschoot, T.: The trajectorir database: Room acoustic recordings along a trajectory of moving microphones. EURASIP J. Audio, Speech, Music Process., 1–45 (2025)

[26] Li, C., Wang, W., Sach, M., Zhang, W., Saijo, K., Cornell, S., Fu, Y., Ni, Z., Fingscheidt, T., Watanabe, S., Qian, Y.: Icassp 2026 urgent speech enhancement challenge. In: Proc. 2026 IEEE Int. Conf. Acoust., Speech, Signal Process.

(ICASSP '26), Barcelona, Spain, pp. 21919–21921 (2026)

[27] van Waterschoot, T.: KU Leuven ESAT-STADIUS Audio Research Labs (2022). https://lirias.kuleuven.be/3940173

[28] Naylor, P.A., Gaubitch, N.D.: Speech Dereverberation, 1st edn. Signals and communication technology. Springer, New York (2010)

[29] Farina, A.: Simultaneous measurement of impulse Response and distortion with a swept-sine technique. In: Preprints AES 108th Convention, pp. 1–24. AES, Paris, France (2000)

[30] Farina, A.: Advancements in impulse response measurements by sine sweeps. In: Preprints AES 123rd Convention, pp. 1–21. AES, Vienna, Austria (2007)

[31] Novak, A., Simon, L., Lotton, P.: Synchronized swept-sine: Theory, application, and implementation. J. Audio Eng. Soc. **63**(10), 786–798 (2015)

[32] van Wieringen, A., Wouters, J.: List and lint: Sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. Int. J. audiology **47**(6), 348–355 (2008)

[33] Jansen, S., Koning, R., Wouters, J., van Wieringen, A.: Development and validation of the Leuven intelligibility sentence test with male speaker (list-m). Int. J. audiology **53**(1), 55–59 (2014)

[34] Yamagishi, J., Veaux, C., MacDonald, K., et al.: CSTR VCTK Corpus: English multi-speaker corpus for CSTR Voice Cloning Toolkit (version 0.92) (2019). https://datashare.ed.ac.uk/handle/10283/3443

[35] Dubey, H., Gopal, V., Cutler, R., Aazami, A., Matusevych, S., Braun, S., Eskimez, S.E., Thakker, M., Yoshioka, T., Gamper, H., Aichner, R.: ICASSP 2022 Deep Noise Suppression Challenge (2022). https://arxiv.org/abs/2202.13288

[36] Francart, T., Van Wieringen, A., Wouters, J.: Comparison of fluctuating maskers for speech recognition tests. Int. J. audiology **50**(1), 2–13 (2011)

[37] Roa-Dabike, G., Akeroyd, M.A., Bannister, S., Barker, J.P., Cox, T.J., Fazenda, B., Firth, J., Graetzer, S., Greasley, A., Vos, R.R., *et al.*: The first cadenza challenges: using machine learning competitions to improve music for listeners with a hearing loss. IEEE Open J. Signal Process. **6**, 722–734 (2025)

[38] D'Orazio, D.: Anechoic recordings of Italian opera. Zenodo (2020). https://doi.org/10.5281/zenodo.3628247

[39] Bang, Olufsen: Music for Archimedes. CD B&O 101 (1992)

[40] Hansen, V., Munch, G.: Making recordings for simulation tests in the archimedes

project. J. Audio Eng. Soc. **39**(10), 768–774 (1991)

[41] ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level. International Telecommunication Union, Geneva, Switzerland (2015)

[42] Steinmet, C.J., Reiss, J.D.: pyloudnorm. GitHub repository (2019). https://github.com/csteinmetz1/pyloudnorm

[43] ISO 3382-1:2009: Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces. International Organization for Standardization, Geneva, Switzerland (2009)

[44] Hummersone, C., Prätzlich, T.: IoSR Matlab toolbox. GitHub repository (2017). https://github.com/IoSR-Surrey/MatlabToolbox

[45] Trainor, L., Sonnadara, R., Wiklund, K., Bondy, J., Gupta, S., Becker, S., Bruce, I.C., Haykin, S.: Development of a flexible, realistic hearing in noise test environment (r-hint-e). Signal Process. **84**(2), 299–309 (2004)

[46] Lydaki, E., Tan, Z.-H., Jensen, J., Guo, M.: Deep feedback cancellation for hearing aids with improved system stability and sound quality. In: Proc. 2025 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '25), Hyderabad, India, pp. 1–5 (2025)

[47] Dillon, H.: Hearing Aids, 2nd edition edn. Thieme, Sydney, Australia (2012)

[48] Roebben, A., Bernardi, G.: HIDVAS: A Hearing Instrument Dataset in Various Acoustical Scenarios for algorithm evaluation and training. GitHub repository (2025). https://github.com/Arnout-Roebben/HIDVAS