Title

OHHR – The Oldenburg Hearing Health Repository

Authors

Sumbul Jafri[1,2], Daniel Berg[2,3], Mareike Buhl[4], Matthias Vormann [2, 5], Samira Saak[2,3], Kirsten C. Wagener [2, 5], Christiane M. Thiel[1,2], Andrea Hildebrandt [1,2], & Birger Kollmeier[2,3]

Affiliations

1. Department of Psychology, School of Medicine and Health Sciences, Carl von Ossietzky Universität Oldenburg, Germany

2. Cluster of Excellence "Hearing4All", Carl von Ossietzky Universität Oldenburg, Germany

3. Department of Medical Physics and Acoustics, School of Medicine and Health Sciences, Carl von Ossietzky Universität Oldenburg, Germany

4. Université Paris Cité, Institut Pasteur, AP-HP, Inserm, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France

5. Hörzentrum Oldenburg gGmbH, Germany

Corresponding authors

Correspondence should be addressed to Andrea Hildebrandt or Birger Kollmeier.

Email: andrea.hildebrandt@uni-oldenburg.de, birger.kollmeier@uni-oldenburg.de

Abstract

Hearing health is shaped by both measurable auditory function and the perceived ability to navigate daily life. To fully understand its complexities, it is essential to integrate objective assessments—behavioural tests that quantify hearing acuity and functional performance—with subjective reports on how individuals navigate and manage their hearing in everyday life. The Oldenburg Hearing Health Repository (OHHR) has been developed to unite these perspectives, providing a comprehensive dataset on hearing health. Collected between 2013 and 2015 at the Hörzentrum Oldenburg in collaboration with Hearing4all, OHHR includes data from 581 individuals (aged 18–86 years; 255 females; mean age = 67.31 years; *SD* = 11.93) with varying degrees of hearing loss. This publicly accessible dataset combines audiometric tests (Pure Tone Audiometry, Loudness Scaling, Speech in Noise tests) with self-reports on hearing difficulties, lifestyle, technology use, and cognitive assessments (DemTect, Vocabulary size test). By integrating subjective experience with objective measures, OHHR will enable researchers to explore the links between hearing ability, cognition, and quality of life, providing valuable insights to advance precision medicine.

Keywords

Hearing health dataset, pure-tone audiometry, speech in noise, cognition, quality of life, hearing impairment

Background

Hearing loss is a major global health concern[1,2,3]. More than 1.5 billion people experience some degree of hearing loss during their lifetime, with at least 430 million requiring hearing health care. In Europe alone, approximately 190 million individuals (20% of the population) are affected to varying degrees by hearing loss or deafness[1]. Beyond its sheer prevalence, hearing loss has a profound impact on an individual's health and quality of life. It has been associated with an elevated risk of cognitive decline, increased perception of cognitive load, and is a contributing factor to accelerated brain atrophy, particularly in regions critical for language and memory[4,5]. In addition, the psychosocial effects of hearing loss, including isolation, depression, and reduced social engagement, exacerbate its impact on well-being[6,7]. Indeed, hearing loss disrupts every aspect of daily life, from difficulties understanding speech in noisy environments to concerns about safety due to impaired sound localization. The constant effort required to listen can be exhausting, often resulting in fatigue, and leading to social withdrawal. Also, the social stigma associated with wearing hearing aids, and the added burden of comorbidity with tinnitus in many cases can make coping with hearing loss even more challenging. In workplace and educational settings, hearing loss can impede performance and achievement, further diminishing quality of life[8].

All these multifaceted associations between hearing loss, psychological well-being, and health underscore the need for a data-driven investigation and simultaneous modelling of these multivariate associations. Such an approach requires larger, shared databases that incorporate subjective reports, multiple objective auditory measures (e.g., speech-in-noise perception, loudness tolerance), assessment of cognitive performance, comorbid health conditions and psychosocial factors to improve and promote an individualized understanding of the impact of hearing loss. The promotion of comprehensive hearing health datasets will facilitate the extension of knowledge beyond the confines of traditional and small-scale audiometric assessment and provide insight into complex multivariate associations.

An overview of existing hearing data repositories was recently published by Sethukumar and colleagues (2023)[9]. Their catalogue brings together a wide range of international publications, including studies on hearing loss, ageing, neurodegenerative disorders, healthcare and disease registries, birth cohorts and population-based studies. Each study comprises one or a few limited types of hearing data, which taken together include pure-tone audiometry, speech-in-noise measurements, otoacoustic emissions, auditory brainstem response, speech discrimination, loudness recruitment, and other clinical and self-reported hearing measures. Notably, sixty percent of the studies included in the catalogue refer to objective hearing data, and in some cases, even genetic testing. The catalogue also highlights datasets with extensive audiological testing, such as the Japanese Intractable Inner Ear Databank[10], the Swedish National Databank on Sudden Sensorineural Hearing Loss[11], and the Childhood Development after Cochlear Implant Dataset from Johns Hopkins University, USA[12]. Some cohort studies also incorporate broader biopsychosocial data along with hearing related measures from larger populations, thus providing potential for valuable insights into hearing health.

Despite these substantial contributions to the study of hearing loss, the advancement of hearing research is hampered by the lack of comprehensive datasets that integrate extensive audiological test data with cognitive, social, lifestyle, and medical variables in a single

repository that is openly available for scientific discovery. Many of the studies catalogued by Sethukumar and colleagues (2023) do not have access to the participants' lifestyle and medical history. Although the UK Biobank[13] for example offers an extensive biomedical dataset, it lacks pure tone audiometry data and uses only the Digit Triplet Test as a screening tool for speech in noise testing, limiting its utility in evaluating functional hearing abilities. Other studies, such as the British 1970 Cohort[14] and the AGES Reykjavik Study[15], provide audiometric and self-reported data but do not include any speech-in-noise measurements, which are crucial for understanding real-world communication challenges. The Rotterdam Study[16] is a longitudinal investigation of the incidence and determinants of chronic disease and disability in older adults. It combines data on hearing, cognition, physical and mental health, genetic testing, neuroimaging, and epidemiological data. However, the primary focus of this study in the field of otolaryngology is alone age-related hearing loss. This focus limits its applicability to other forms of hearing impairment, such as noise induced or ototoxic hearing loss and congenital conditions. The ongoing Rheinland study[17] also has a special focus, namely on aging and neurodegenerative disorders. Both studies employ the Mini-Mental State Examination for cognitive screening, which cannot detect subtle cognitive challenges faced by individuals with hearing problems[18]. Although the standard Mini-Mental State Examination remains widely used, research indicates that the Montreal Cognitive Assessment and the Dementia Detection test are more sensitive in detecting early cognitive decline in diverse populations[19,20]. Importantly, access to many of these datasets requires formal and complex authorization, some of which are financially burdensome, and others have institutional affiliation restrictions, all of which limit their availability for widespread and accelerated research. It is worth noting, however, that some of these studies are ongoing and their policies and access options may update in the future.

To address the above limitations of current hearing health datasets and the issue of limited access to them, we describe and share here a comprehensive open access dataset – the Oldenburg Hearing Health Data Repository (OHHR) – collected at the Hörzentrum Oldenburg gGmbH and the Cluster of Excellence “Hearing4all” in Germany between 2013 and 2015. OHHR is focused on individuals with well-defined hearing loss and comprises a hearing loss cohort of 442 out of a total of 581 participants. Having a well-defined cohort of individuals who are deaf or hard of hearing, and those who are unaffected allows researchers to isolate the effects of hearing impairment from potential confounding variables. Although other datasets, such as birth cohorts or disease-specific studies, offer valuable insights, they may not capture the same level of detail on hearing loss and associated health and well-being. The current data resource offers an extensive array of self-reported and examiner-driven measures in conjunction with pure tone audiometry, thereby providing researchers with a multifaceted perspective on the nature of hearing loss and associated factors. It also includes a precise speech-in-noise test, i.e. the Göttingen Sentence Test[21] to characterize the individual´s functional hearing deficit (to be treated with a hearing device) and the Digit Triplet Test, which is widely used as a hearing screening tool for assessing speech recognition in noisy environments[22,23,24]. The Dementia Detection test was used to identify cognitive impairment, and a vocabulary size test was also applied as a proxy for verbal crystalized intelligence. To achieve a holistic assessment, the database considers not only clinical evaluations but also draws from self-reported information using standardized questionnaires. Thus, OHHR captures a broad spectrum of data, including socio-demographic characteristics, hearing health history and impact on daily life, noise exposure, hearing aid use and

satisfaction, mental and physical health status, and potential comorbidities of hearing loss. Its increased granularity facilitates the exploration of the interplay between hearing loss, functional hearing deficit, cognitive decline, and overall quality of life and health.

Potential lines of research that OHHR could enable

Precision medicine in audiology undoubtedly benefits from such extensive datasets as the OHHR. For instance, Gieseler et al. (2017)[25] utilized this dataset to identify key individual predictors of speech-in-noise performance, including pure tone audiometry and verbal intelligence, among older adults who are deaf or hard of hearing. The comprehensive test battery provided the authors with a rich view of the participants' hearing characteristics, revealing how cognitive abilities and hearing aids impact speech recognition. Tahden et al. (2018)[26] compared hearing aid non-users with their aided counterparts across auditory and non-auditory measures in the dataset to explore factors influencing hearing aid adoption among older adults. They found that self-perceived hearing difficulties, technology commitment, and socioeconomic status play key roles, highlighting the dataset's value beyond statistical audiology in supporting rehabilitative strategies and public health policies. Furthermore, Buhl et al. (2020)[27] leveraged the same dataset to develop the so-called Common Audiological Functional Parameters (CAFPAs). These parameters serve as abstract representations of functional aspects of the auditory system, independent of specific audiological measures. By creating an abstract intermediate layer that connects audiological data with diagnostic insights, their approach aims to improve the interpretability and independence of predictions from specific measurement tools within a clinical decision-support system for audiology[28,29]. The development of CAFPAs clearly benefited from the comprehensive database and the dataset's specialized cohort of individuals who are deaf or hard of hearing [27]. To automatically predict the CAFPAs from individual patient data, Saak et al. (2020)[30] trained several machine learning models. This automated approach allows for data-driven CAFPA estimation without solely relying on expert input, thus enhancing scalability and interpretability. The OHHR was also used as a medical data example to illustrate a novel machine learning approach in the framework of the proposed maximal causes generative model with Beta observables, specifically to learn symptom-cause relationships, i.e. between CAFPAs and hearing loss types[31]. In addition, Saak et al. (2022)[32] stratified patient groups into distinct auditory profiles based on a wide array of hearing health measures available in the dataset. This stratification effectively summarizes patient information for clinical decisions, thereby advancing personalized care in audiology. Taken together, these studies demonstrate the dataset's versatility and pivotal role in supporting innovative approaches to improve clinical outcomes in audiology and hearing healthcare.

In summary, OHHR provides a comprehensive resource for developing predictive models that inform decisions for personalized rehabilitation approaches and guide public health initiatives. It will allow researchers to identify high-risk populations, gain comparative data to evaluate prevention strategies, and optimize hearing aid utilization. The inclusion of psychosocial and demographic information can help to further enhance the understanding of the aetiology of hearing loss. Additionally, OHHR offers the opportunity to investigate a wide variety of factors that influence the perception and management of hearing difficulties. The integration of data on hearing aid adoption and usage patterns with patient-reported outcomes provides a foundation for developing patient-centric care models and evaluating

assistive technologies. By making the OHHR openly available for research, the intention is to support discoveries that enhance the quality of life for people with hearing loss through early detection, effective interventions, and tailored rehabilitation services. The dataset is of interest for researchers in the field of audiology, auditory modelling, computational audiology, cognitive and health psychology, and public health with potential to advance global hearing health.

Methods

Participants

Participants aged 18 years and older were recruited from an existing database of volunteers at Hörzentrum Oldenburg gGmbH (abbreviated as HZO in the following). This non-profit organization is affiliated with the Carl von Ossietzky Universität Oldenburg and partners with the "Hearing4all" Cluster of Excellence funded by the German Research Foundation (https://hearing4all.de/en/). It works with a global network of scientific and industrial collaborators to develop innovative audiological methods and relies on the support of about 2,000 local volunteers. After screening for completeness, the final dataset prepared for public release via Zenodo (https://zenodo.org/records/14808650)[33] includes 581 individuals, aged 18 to 86 years (*n* = 255 female; median age = 70.0 years; mean age = 67.31 years; standard deviation (*SD*) = 11.93 years). The data was anonymized in accordance with the General Data Protection Regulation (GDPR; Regulation (EU) 2016/679). Each participant was assigned a unique identifier to maintain anonymity while enabling multivariate analyses. Data protection approval for the preparation and open release of the dataset was granted by the Data protection and Information Security Management Office at the Carl von Ossietzky University of Oldenburg, Germany.

Procedure

The data collection process began with an 11-page questionnaire, referred to as the "home questionnaire", which volunteers completed at home. It was sent via postal mail to potential participants along with study information and consent forms, with instructions to complete everything using pencil and paper and return it via prepaid mail. Participants who returned the completed home questionnaire were then invited to a one-hour lab-based test session. The test session included a face-to-face interview to discuss the individual's medical history, as well as several assessments. Besides a clinically standard pure tone audiogram, supra-threshold loudness perception was determined with the Adaptive Categorical Loudness Scaling[34]. In addition, two speech-recognition-in-noise tests (the Göttingen Sentence Test - GÖSA[21] and the German Digit Triple Test - DTT[23]), and two cognitive tests (Vocabulary size test or Wortschatztest - WST[35] and the Dementia Detection test - DemTect[20]) were performed. All measures were undertaken by trained professional staff members. Hearing assessments, including Pure tone Audiometry, Adaptive Categorical Loudness Scaling, the GÖSA, and the German DTT, were performed without hearing aids, even for regular users. However, cognitive assessments and the questionnaires, where optimal hearing was required were conducted with hearing aids for those who typically used them[25].

The local ethics committee at the Carl von Ossietzky Universität Oldenburg reviewed and authorized the data collection, ensuring compliance with the ethical standards described in

the Declaration of Helsinki[36], except for the requirement for pre-registration of the study. To protect the privacy of the individuals involved, the entire dataset was anonymized using the k-anonymity rule, ensuring that all cases remain unidentifiable in the shared database (see details below). Participants were also compensated monetarily for their involvement.

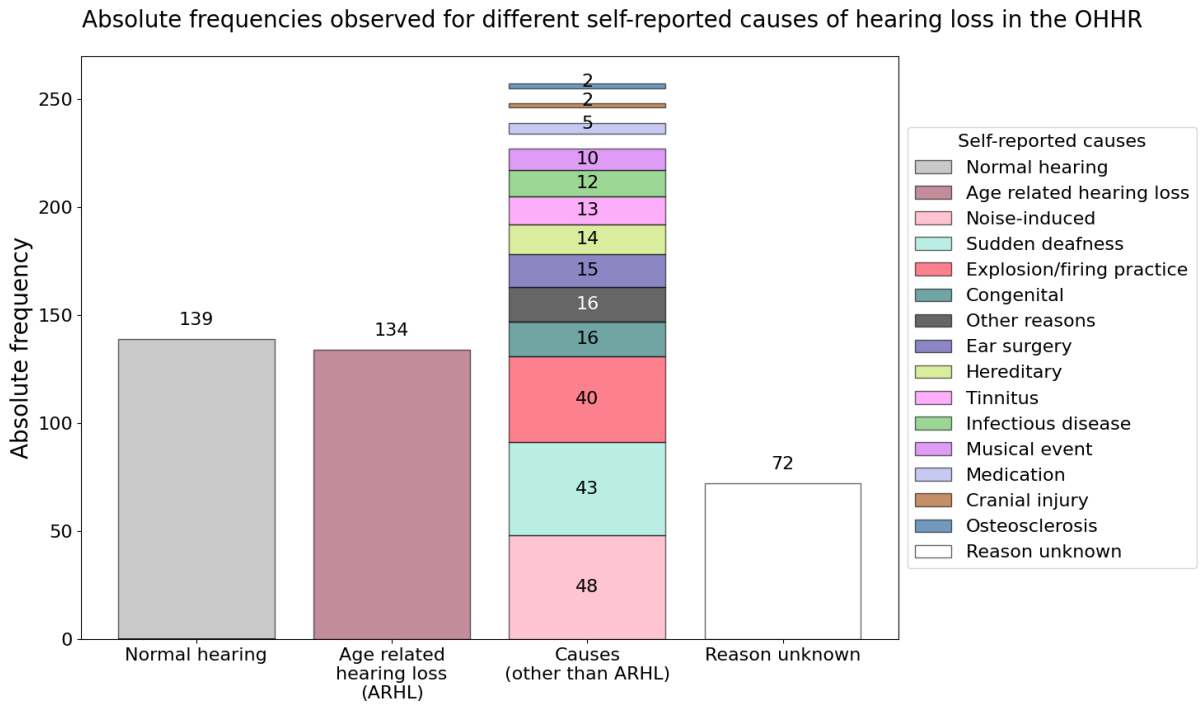
Self-reports included in the home questionnaire (HQ)

Subjective experience of hearing loss related consequences reflects the functional impact of hearing loss on an individual's life, an aspect that cannot be captured by audiometric measures alone. There has been much discussion in the literature about the low association between examiner-driven ("objective") and self-reported ("subjective") measures of hearing loss. However, it is crucial to also characterize individuals with hearing loss in terms of their subjective hearing-related experiences, along with their subjective health status (chronic diseases), self-reported demographics (age, gender), personality traits (self-efficacy, self-esteem), mood (depressive symptoms) and social context (social network size)[37]. Research by Wang and colleagues (2020)[38] shows that underreporting of hearing problems is associated with age, ear problems, lifestyle, and subjective beliefs about hearing health. Notably, older adults tend to underreport their hearing difficulties[39]. A combined approach of self-reports and audiometric measures is also effective in detecting high-frequency hearing loss. Furthermore, self-reports are particularly sensitive in contributing to the identification of mid-frequency hearing loss[40]. Understanding the factors that influence self-reported hearing problems, including cognitive abilities, education, and lifestyle, enables researchers and practitioners to interpret these reports more precisely and use them as complementary measures to audiometry for prediction and for deriving individualized treatment recommendations[41]. The HQ covered several key assessments, which are described in detail below.

Hearing anamnesis

The HQ collected detailed information on the individual's hearing history, including diagnosis and duration of hearing impairment, subjective ratings of hearing problems in both quiet and noisy environments, and noise exposure. Individuals were asked about the causes of their hearing difficulties (see Figure 1), hearing aid use (past and present), duration of use, and any history of sudden sensorineural hearing loss, for example due to middle ear infections. Ear noise/tinnitus was also addressed. This included its occurrence, duration and the physical or mental discomfort it causes. Furthermore, to evaluate the comfort and experience of hearing aid use, the questionnaire included three items from the seven-item International Outcome Inventory for Hearing Aids[42].

254 **Figure 1. Self-reported causes of hearing difficulties**



255
256 *Note*. A total of 441 individuals reported on the causes or possible contributing factors for their
257 hearing problems.

258 *General health and lifestyle*

259 Individuals provided information about their overall health, including any limitations in daily
260 activities due to health problems, how their physical and mental health had affected task
261 performance in the four weeks prior to data collection, pain impacting daily activities,
262 emotional well-being, and perceived memory issues. They also reported any chronic
263 conditions diagnosed in their lifetime and in the past 12 months, such as respiratory disorders,
264 cardiovascular diseases, diabetes, musculoskeletal, or psychiatric conditions (see Table 1).
265 These data are valuable, for example, for assessing the burden of comorbid conditions and
266 for calculating a comorbidity index according to the German National Health Examination
267 Survey[43]. This index quantifies the impact of chronic conditions on the use of health care
268 services, on quality of life, and on other health-related outcomes.

269 **Table 1. Prevalence of health problems and hearing difficulties, overall and by gender**

Health problems	*n*	% in the sample	% of men	% of women
Hearing Difficulties	443	75.0	46.87	28.09
Musculoskeletal conditions	325	55.0	28.43	26.57
Respiratory conditions	79	13.4	6.09	7.28
Cardiovascular conditions	77	13.0	10.49	2.54
High blood pressure/cholesterol	317	53.6	32.49	21.15
Diabetes mellitus	52	8.8	5.41	3.38
Neurological disease, e.g., Parkinson's disease	33	5.6	4.57	1.02
Sight disorder/blindness	83	14.0	8.46	5.58
Risk of falling/dizziness	75	12.7	5.08	7.61
Chronic gastric-intestinal disease	65	11.0	5.41	5.58
Cancer	61	10.3	5.92	4.4
Psychiatric conditions	71	12.0	5.58	6.43
Other chronic diseases (kidney, bladder, skin, etc.)	172	29.1	17.77	11.34
Physical handicap (arms/legs)	108	18.3	9.31	8.97

270 *Short form 12 health survey* (SF-12)[44]

271 The health status section of the HQ also included the SF-12 health survey, which is a concise
272 version of the 36-item health survey, but comprehensive enough to provide an overview of
273 an individual's physical and mental health. The SF-12 contains 12 questions, yielding two
274 summary scores. The physical component score (PCS) quantifies limitations in physical
275 functioning, role limitations due to physical health problems, bodily pain, general health
276 perceptions, and social functioning limitations due to physical problems. The mental
277 component score (MCS) indicates vitality, emotional well-being, and social functioning
278 limitations due to emotional problems, as well as interference with daily activities caused by
279 mental health problems. The PCS and MCS raw scores are separately transformed into
280 standardized scores and included in the OHHR as separate scores (*T*-values, with *M = 50, SD*
281 *= 10)*.

Media consumption and device usage

A specific section of the HQ aimed to assess the participant's media consumption habits and preferences. It included questions about headphone use for TV and radio, detailing the types of hardware and software used (e.g., radio, Bluetooth, infrared) at the time of data collection and how often they were used. Individuals also reported their use of technology for work and personal activities, both with and without hearing aids. In addition, they were asked about their experiences with sound levels, including situations where sounds were perceived as too loud with or without hearing aids. The survey also addressed the demand and perceived sound quality of hearing aids, the frequency of increasing TV volume without them, and individual musical preferences and expectations for sound quality.

Technology readiness

The HQ also included a 12-item questionnaire to measure the use of adaptive technology, particularly among older adults[45]. The measurement instrument is based on the theoretical concepts outlined in the technology acceptance model[46], which postulates how perceived usefulness and ease of use influence technology adoption. Furthermore, it also incorporates concepts from Ajzen's (1991)[47] work on personal beliefs about competence and control. Neyer and colleagues (2012)[45] extended these concepts by developing a comprehensive model of technology readiness that integrates both attitude-oriented and personality-theoretical perspectives. The main construct measured by the 12 items is technology readiness, or the willingness to invest time and effort in learning and using modern technology. Additionally, there are three subscales that can be measured with this questionnaire. Technology competence reflects the perceived ability and confidence in using technology effectively[48]. Technology acceptance quantifies an individual's relationship with modern technologies, primarily in terms of interest in technical innovations[46] (Davis, 1989). Technology control reflects the degree to which individuals feel that they have control over their technology use and can overcome challenges[48].

Demographic and socio-economic status

The demographic information was collected in both the clinical interview (age, gender, language status, etc.) and the HQ (household composition and living situation). Participants also reported socio-economic details about themselves, such as their educational degree, profession, occupation, and household income. The Scheuch-Winkler Index (SWI) was used to quantify socio-economic status (SES) by summing up education, occupation, and net income, resulting in a score ranging from 3 to 21[49]. This index is crucial for describing health disparities by investigating how socio-economic factors influence health outcomes. These scores were categorized into three SES groups—low, medium, and high—to represent different socio-economic strata within the German population.

Overall, the HQ collected extensive data that allows for a comprehensive characterization of the impact of hearing loss on an individual's quality of life and their biopsychosocial characteristics, making this dataset highly valuable for informing precision audiology. By including validated instruments along with newly designed questions, the HQ provided a thorough dataset for analysing the complex, multivariate interactions between hearing health, general health, technology use, and socio-economic factors. All of this is

complemented by further objective measures of hearing and cognition, which are described in the following.

Clinical interview for anamnesis

To gain a detailed overview about the participants' hearing health, beyond the survey data recorded with the HQ, an expert interviewer conducted a structured face-to-face interview with the participants. The interview was designed to record further demographic information and questions pertaining to the onset, progression, and specific characteristics of the individual's hearing loss, along with details on cochlear implants or any other implant, hearing aid use (access and frequency), family history of hearing loss, and history of ear infections.

Audiological tests

Pure Tone Audiometry

Pure tone audiometry is a standard tool used in clinical settings to identify hearing impairments. The resulting audiogram provides a comprehensive assessment of hearing sensitivity, conductive hearing loss, cochlear function, and discomfort levels. The audiometry procedure for OHHR was performed in the laboratory, with each ear assessed independently. Air conduction hearing thresholds were measured using a Siemens Unity II audiometer with Sennheiser HDA200 headphones, while bone conduction thresholds were evaluated with a RadioEar B71 bone conduction transducer. This was done with sinusoids for frequencies ranging from 125 Hz to 8000 Hz in an acoustically shielded audiometry booth. Bone conduction measurements were limited to 500 to 4000 Hz, following standard clinical practice. Below 500 Hz, the vibration threshold is too close to the hearing threshold, making measurements unreliable. Frequencies above 4000 Hz are excluded due to excessive distortion produced by the transducer. A standard clinical threshold method was used with a step size of 5 dB, where a threshold was accepted if the probe tone was not heard twice at lower levels and detected twice at higher levels. To summarize overall hearing sensitivity, the Pure Tone Average (PTA) was calculated as the mean hearing threshold across 500, 1000, 2000, and 4000 Hz. Additionally, measurements of uncomfortable loudness levels (UCL) were performed at 500, 1000, 2000, and 4000 Hz. Tone levels were increased in 5 dB increments and participants were asked to indicate when the sound became uncomfortably loud by pressing a button. The respective presentation level was recorded, and the presentation sequence was continued for the next test frequency at a presumably comfortable level. The complete assessment took approximately 13 minutes.

Adaptive Categorical Loudness Scaling

Adaptive Categorical Loudness Scaling (ACALOS) is a method used in clinical audiology to measure an individual's subjective perception of loudness according to Brand and Hohmann (2002)[34]. It is particularly adept at diagnosing loudness recruitment, which is commonly encountered among individuals with sensorineural hearing loss. This phenomenon is characterized by an excessive increase in perceived loudness as the sound level increases.

The stimulus signals used were one-third-octave bands of noise (duration 2 seconds) presented through Sennheiser HDA 200 headphones for 20 trials in a pseudo-random order.

Participants were asked to rate their perceived loudness of each stimulus level using an 11-category response scale, ranging from "not heard" to "too loud". The categories corresponded to 0, 5, 15, 25, 35, 45 and 50 categorical units, with four unnamed categories situated between them (10, 20, 30, 40 cu). The presentation levels were adaptively adjusted based on the participants' previous responses, thereby ensuring that the test assessed the whole individual range of loudness perception[34]. Measurements were recorded successively for 1500 and 4000 Hz narrow-band noise stimuli for the left and right ear separately. A loudness function can be fitted to the responses of each of the ACALOS measurements using either the Brand and Kollmeier (2002)[50] staircase method or the BTUX fitting method[51], depending on the researcher's preference[25].

The ACALOS procedure was designed to provide a reliable and efficient estimation of loudness functions with a small number of trials. It has good test-retest reliability, with intraindividual standard deviations for loudness levels ranging from 4-5 dB, comparable to or slightly better than other procedures that require more trials. The adaptive approach allows for efficient use of trials and better coverage of the auditory dynamic range, particularly benefiting people who are hard of hearing. Additionally, ACALOS eliminates the need for prior measurements of hearing thresholds, simplifying the testing procedure[34].

While Pure tone audiometry and ACALOS are effective in assessing hearing sensitivity, everyday communication depends on the ability to recognize and understand speech, particularly in noisy environments. Individuals with similar ages and audiogram thresholds can still exhibit substantial differences in speech comprehension in noisy settings. This underscores the importance of including speech-in-noise measures as part of a comprehensive audiological assessment such as the OHHR intends to provide. These additional measures are described below.

Digit Triplet Test

The Digit Triplet Test (DTT) employs a closed-set response format that facilitates convenient self-testing via telephone or the internet, primarily for hearing screening purposes. It consists of lists of twenty-seven-digit triplets spoken in background noise used to adaptively determine the speech recognition threshold (SRT). During the adaptive procedure, the presentation SNR of a trial depends on the previous individual performance: The starting SNR was set to 0 dB, the noise presentation level was kept constant, and the speech presentation level varied, i.e., it was lowered subsequent to correct identifications and increased if the listener had difficulty understanding any digit[23]. For the current database, the test was administered in the presence of an expert in a controlled laboratory environment. The DTT lasts approximately three minutes to complete. The background noise presentation level was set at 65 dB SPL and was changed to 80 dB SPL only for those with severe hearing loss, if 65 dB was not sufficiently audible[52]. The purpose of this test is to estimate the Speech Recognition Threshold (SRT), which is defined as the speech level at which a person can hear 50% of the numbers correctly. The DTT offers several advantages, including its brevity, familiarity with the stimuli employed, efficiency, online accessibility, and availability in multiple languages. It has also been shown to be an effective tool for both adults and children with cochlear implants, with of reliability and resilience to learning effects, linguistic abilities, and personal factors such as educational background[53].

Göttingen Sentence Test

The Göttingen Sentence Test (GÖSA) is a sentence matrix test in which listeners hear a set of 20 meaningful sentences, each consisting of three to seven words, delivered one at a time in background noise. This list is randomly selected from a pool of 10 predefined sets. It provides a more realistic representation of communication situations than tests that have single words or sounds presented in noise. Each sentence is meticulously designed to maintain perceptual equivalence and consistent difficulty levels across different sets, ensuring reliable and comparable results[21,54].

During the testing session, the sentences were presented with test-specific speech-shaped noise via a free-field loudspeaker in a sound-attenuated room. Participants were instructed to repeat as many parts of the sentence as they could after each sentence presentation. The GÖSA takes about 5 minutes to complete. The noise level during testing is set at 65 dB sound pressure level (SPL) and increased to 80 dB SPL for participants with severe hearing loss. The GÖSA also employs an adaptive procedure to determine the SRT with precision in the range of ±1 dB. The noise level is kept constant, and speech levels are adjusted based on the listener's responses, facilitating accurate assessment of speech perception abilities. In summary, the test offers a comprehensive assessment of speech intelligibility, using ecologically valid materials and carefully calibrated background noise to simulate real-world listening conditions[50].

Cognitive measures

Screening tests for dementia and cognitive assessment in general require the participant to hear and comprehend the test items sufficiently well. Similarly, audiological tests, except visual inspection of the ears, demand cognitive abilities such as attention, working memory, and semantic knowledge. These abilities are essential for attending to, comprehending, remembering, executing instructions, and communicating with the examiner. Therefore, an accurate diagnosis of cognitive and audiological conditions is contingent upon a thorough examination of both[55].

Dementia Detection Test

The Dementia Detection test (DemTect) is a neuropsychological screening tool employed to assess cognitive impairment. It is recognized for its high sensitivity and time efficiency, given that it can be administered in 8 to 10 minutes. DemTect comprises five subtests that cover a wide range of cognitive abilities, including immediate and delayed recall of verbal information, working memory, language and number processing, and executive functioning. In the Word List subtest, participants are presented with a 10-item word list over two trials to assess both immediate and delayed recall, thus targeting memory function, which is an essential domain for detecting mild cognitive impairment (MCI) and dementia[56,57,58,20]. The Number Conversion task assesses executive function and language processing by requiring participants to switch between different representations of numbers, such as words and numerals. It captures a range of errors, including those related to language impairment, lexical and syntactic processing, and literacy difficulties. Such impairments have been frequently observed in individuals with dementia[59,60,20]. In the Semantic Fluency Task ('Supermarket'), participants are asked to generate and name items belonging to a specific

category, such as supermarket items, within a limited time. It assesses a range of cognitive abilities including attention, working memory, cognitive flexibility, problem solving, semantic memory, language production and processing speed. Studies show that verbal fluency is often impaired in the early stages of dementia[61,62,63,64,20]. The Digit Span task requires participants to repeat sequences of numbers in reverse order and is a measure of working memory. Deficits in this area are considered one of the earliest signs of dementia[65,66,20]. Finally, in the Delayed Recall task, participants are asked to recall the 10-item word list presented at the beginning. This task assesses long-term memory.

The overall score is calculated by summing the scores of each subtest, with each subtest incorporating age-based scoring adjustments for participants under and over 60 years of age to account for age-related differences in cognitive abilities. DemTect has demonstrated robust construct validity and high reliability in both test-retest and inter-rater assessments. Notably, it excels in detecting mild Alzheimer's disease (AD) with a sensitivity of 100% and mild cognitive impairment (MCI) with a sensitivity of 80%. In contrast, the widely used Mini-Mental State Examination (MMSE) shows limited efficacy in identifying MCI, with a sensitivity of only 69%[20,18]. This tool enables the identification of potential cognitive decline or impairment of participants in OHHR.

Vocabulary size test

The vocabulary size test (in German Wortschatztest-WST[35]) was used to assess the verbal intelligence and language comprehension of the participants included in the OHHR. Factor analytic validation studies have revealed that indicators derived from the vocabulary tests load highly on a general cognition factor (commonly called the *g*-factor), meaning that the scores from WST can be interpreted as effective indicators of *g*[35,25]. Therefore, this assessment has the additional purpose of estimating premorbid intelligence levels in individuals with mild to moderate cognitive impairment and tracking the progression of dementia.

The WST is a 10-minute test consisting of 40-word recognition tasks. In each task, participants must identify the real word presented alongside five similar non-words. The tasks are arranged in a line-by-line format, with increasing difficulty. The raw score is determined by the number of correctly identified words. This test demonstrates high reliability, as evidenced by a split-half reliability coefficient (r = .95) and Cronbach's Alpha (α = .94). The score is largely independent of age, exhibiting a very low correlation with age (r = .08), but it is positively correlated with educational and vocational qualifications (r = .60). It has been normed for a wide age range (20-90 years) and has been standardized using a representative sample (N = 572) and Rasch scaling to ensure equivalent measurement of abilities across items.

Data records

The OHHR is accessible via Zenodo and is arranged into three principal folders, including **data**, **metadata**, and **tools**. These folders provide all the essential resources needed to access, interpret, and use the dataset easily and effectively. The tools folder contains conversion scripts that read and join the audiogram and loudness scaling JSON tables for potential use in Excel, MATLAB, Python, and SQL, thereby facilitating data management and visualization. The metadata folder includes detailed description files that define all the variables within

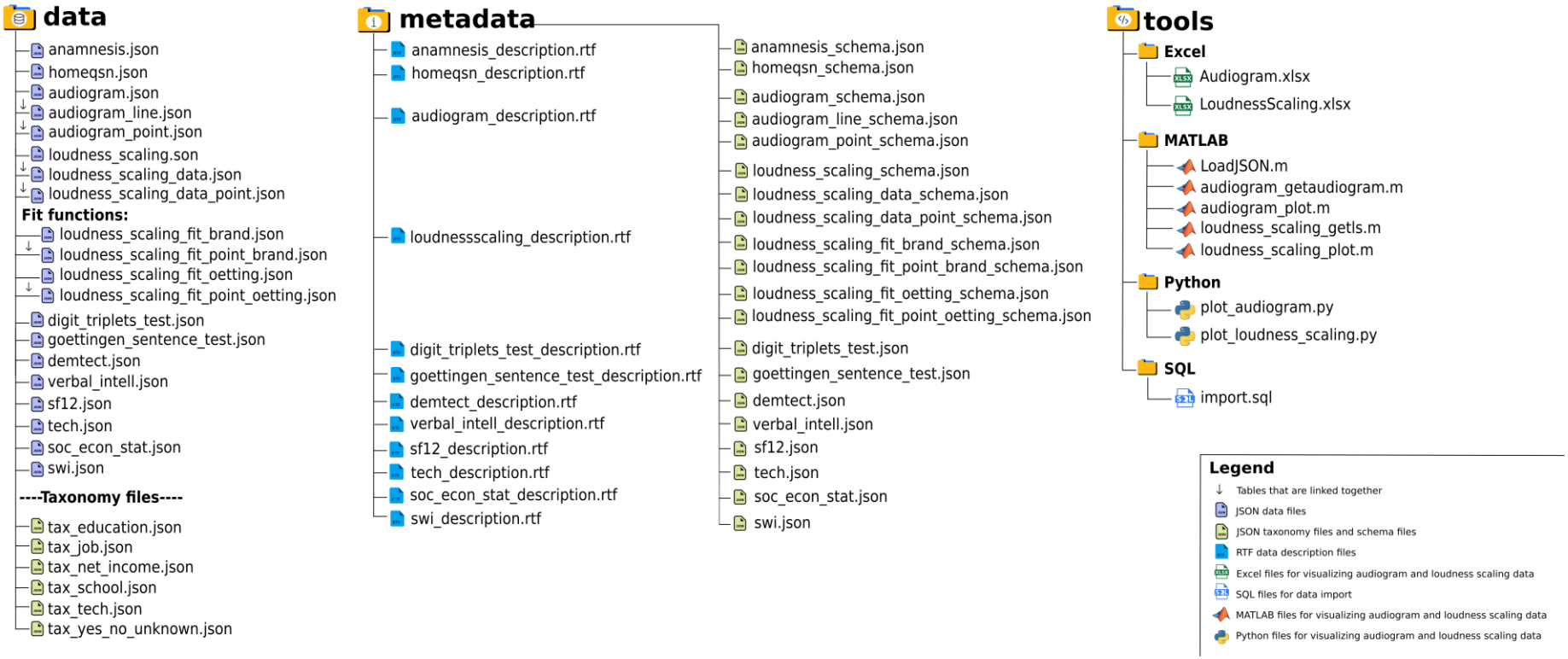
each .json file, while comprehensive schema files ensure data consistency and enable accurate interpretation across varying platforms. Below is an overview of the principal folders and their structure (see Figure 2), as well as the file names within Table 2.

All the data tables can be found (.json files) inside the data folder and their corresponding schema files in the metadata folder. There are also description files (.rtf format) in the metadata folder that explain the various columns inside each data table. In addition to the primary dataset, the data folder contains six taxonomical .json files that provide verbal representations of different categorical data. The most relevant files are listed and explained in Table 2.

Table 2. Overview of the files providing verbal explanations for levels in categorical data

File name	Representation
tax_education.json	Contains verbal representations of education levels
tax_job.json	Contains verbal representations of job levels
tax_net_income.json	Contains verbal representations of income levels
tax_school.json	Contains verbal representations of school levels
tax_tech.json	Contains verbal representations of responses from the technical survey, extracted from the original home questionnaire
tax_yes_no_unknown.json	Contains verbal representations for yes, no, and unknown responses

Figure 2. Overview of the data folders and tools of OHHR



Note. The arrow between the audiogram and loudness scaling files in the data folder represents linked tables that need to be joined together for use.

Data anonymity and identification risk assessment

As described above, participants were recruited from a volunteer database. The recruitment strategy involved mailing study information, consent forms, and the HQ to potential participants, who completed and returned these forms via post. This initial process ensured that participants understood the nature of the data they would be sharing prior to any in-person testing. To protect participants' privacy, all data were anonymized and additional consent for data publication, even for anonymized data, was confirmed by approximately half of the sample prior to inclusion in the OHHR. We applied *k*-anonymity (*k*=4), ensuring that each participant's quasi-identifiers were indistinguishable from at least three others, thereby minimizing the risk of re-identification. Additionally, further measures were taken, such as grouping categories with fewer than four cases into broader categories to further safeguard anonymity. A data protection officer reviewed the entire anonymization process, confirming that all privacy protocols were appropriately implemented. Documentation detailing the anonymization procedures and risk assessments are available upon request, demonstrating compliance with data protection regulations and ensuring participant confidentiality.

Technical validation

Standardized testing protocol

A standardized testing protocol was followed to ensure consistency and reliability in data collection. First, the order and timing of the assessments was standardized. Before any laboratory testing, the HQ was duly completed and returned. The subsequent interviews and assessments were primarily conducted on the same day. However, in certain instances, the assessments were done on alternative days when necessary or to obtain any missing information. To ensure adherence to the established protocol, the dates for each test were recorded. Second, standardized questionnaires and measurements were used wherever possible – see descriptions of the instruments above.

Medical Device Certification

The audiometric testing software used was certified as a medical device according to EU regulations and bears the CE mark. Classified as a legacy device under the EU directive, it is suitable for diagnostics in clinical settings and by hearing care professionals. This certification guarantees compliance with stringent safety and performance standards. More details can be found on the Hörzentrum Oldenburg website.

Calibration procedures

The audiometers were calibrated according to established standards and guidelines (ANSI S3.6-2010; ISO 389-8-2004), ensuring the precision and consistency of our auditory measurements. The speech tests and ACALOS were conducted using the Research and Development version of the Oldenburg Measurement Applications (OMA), designed for scientific research. This software operates on a commercially available PC with a high-quality sound card, supporting multi-channel capability, speech tests in various languages, and customizable noise signals. More details can be found on the Hörzentrum Oldenburg website.

Data completeness

During the validation process, we thoroughly inspected missing data across both auditory and non-auditory variables. All critical measures and questionnaires were complete, except for SF-12, which had 13 missing entries. The SF-12 was part of the HQ filled out by participants without supervision. Some entries are missing, especially in free response sections if the questions were not relevant to the participant.

Data quality assessment

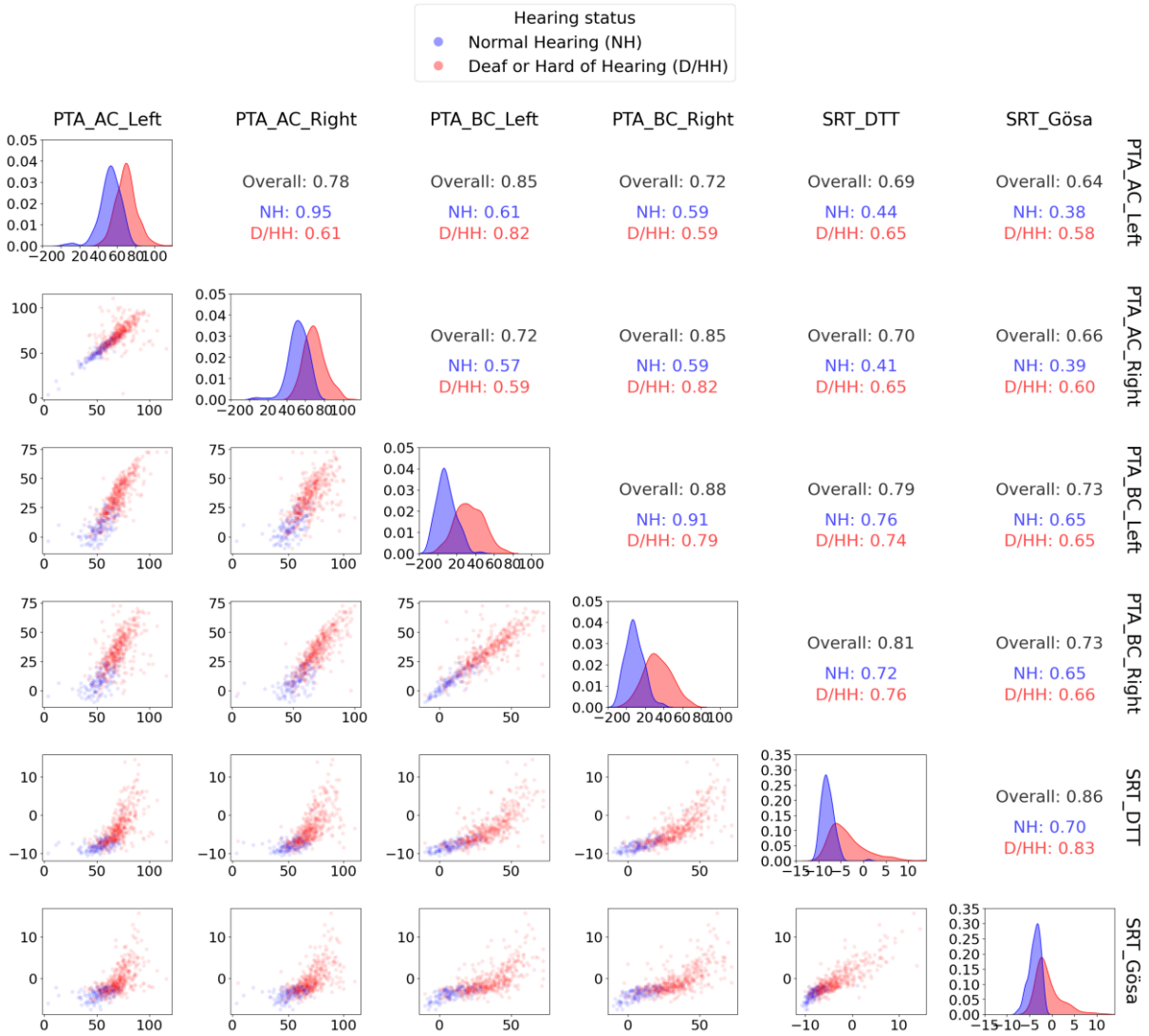
We have implemented a comprehensive technical validation process to ensure the highest standards of data quality and reliability. This section details the applied methods and validation steps, emphasizing compliance with prior established knowledge.

First, where applicable, comparisons with normative data were completed to contextualize the study results within established benchmarks or reference points. For instance, the noise level was calibrated to 80 dB for individuals with severe hearing loss during speech in noise measurements, and different scoring criteria were used in the DemTect assessment to account for age-related changes in cognition. Supplementary Table 1 presents the descriptive statistics for the standardized measurements in the dataset, including the number of items, measurement intention, mean (±SD), and range of values. Normative reference data are primarily available from normal-hearing population samples. The measurements are sensitive enough to differentiate between normal hearing and impairment. However, normative data for specific classes of hearing impairment in the German population were not consistently available in the literature; such data, wherever found, were included in the table[67,68,69].

Second, the univariate and bivariate distributions of the audiological measures were analysed according to self-reported normal hearing status or a professional diagnosis of hearing loss. This analysis was conducted to validate that the audiological measures in the OHHR effectively capture group differences. Figure 3 illustrates these distributions, showing clear differences in mean values between groups. Additionally, the figure demonstrates approximately Gaussian distributions within groups and as expected, strong bivariate associations between different measures of hearing ability, supporting the reliability and validity of the audiological data. It is important to note that this figure is intended solely to validate the consistency of hearing-related measures, rather than to assess differences in other audiological or non-audiological variables between groups.

Third, we explored whether the expected low, but well-established association between examiner-assessed and self-reported measures of hearing loss, as reported in the literature, was also present for the variables included in the OHHR. To this end, Figure 4 illustrates the distribution of Pure Tone Average (PTA) thresholds, as well as Speech Recognition Thresholds (SRT) from the Göttingen Sentence Test (GÖSA) and Digit Triplet Test (DTT), across varying levels of self-reported hearing problems. These measures are presented for both quiet and noise conditions. The figure demonstrates a small but consistent positive association between self-reported hearing difficulties and objective measures, including bone and air conduction thresholds, as well as speech-in-noise performance. These findings validate the OHHR hearing-related variables against the established literature.

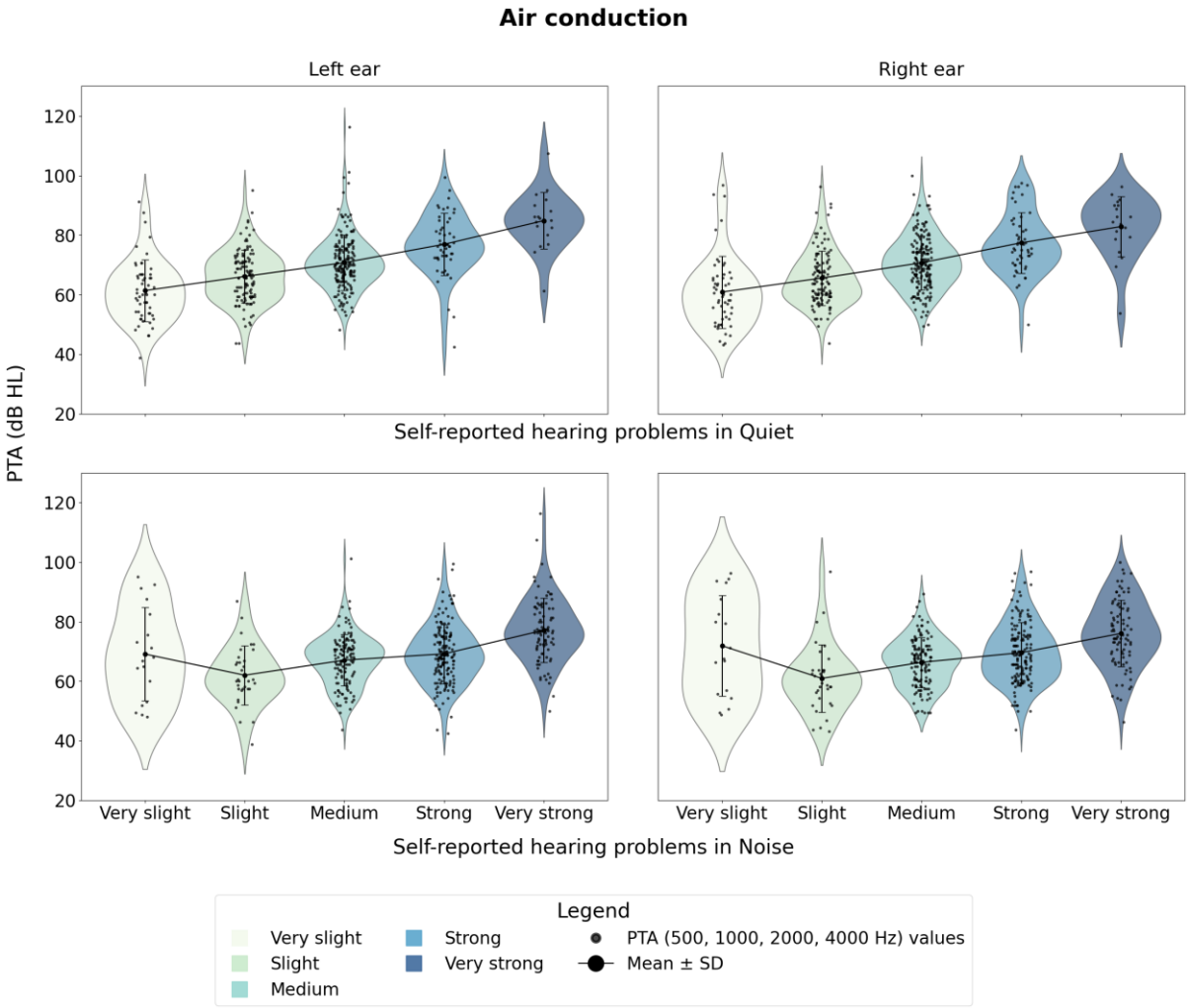
Figure 3. Univariate and bivariate distributions of audiological measures stratified by hearing status



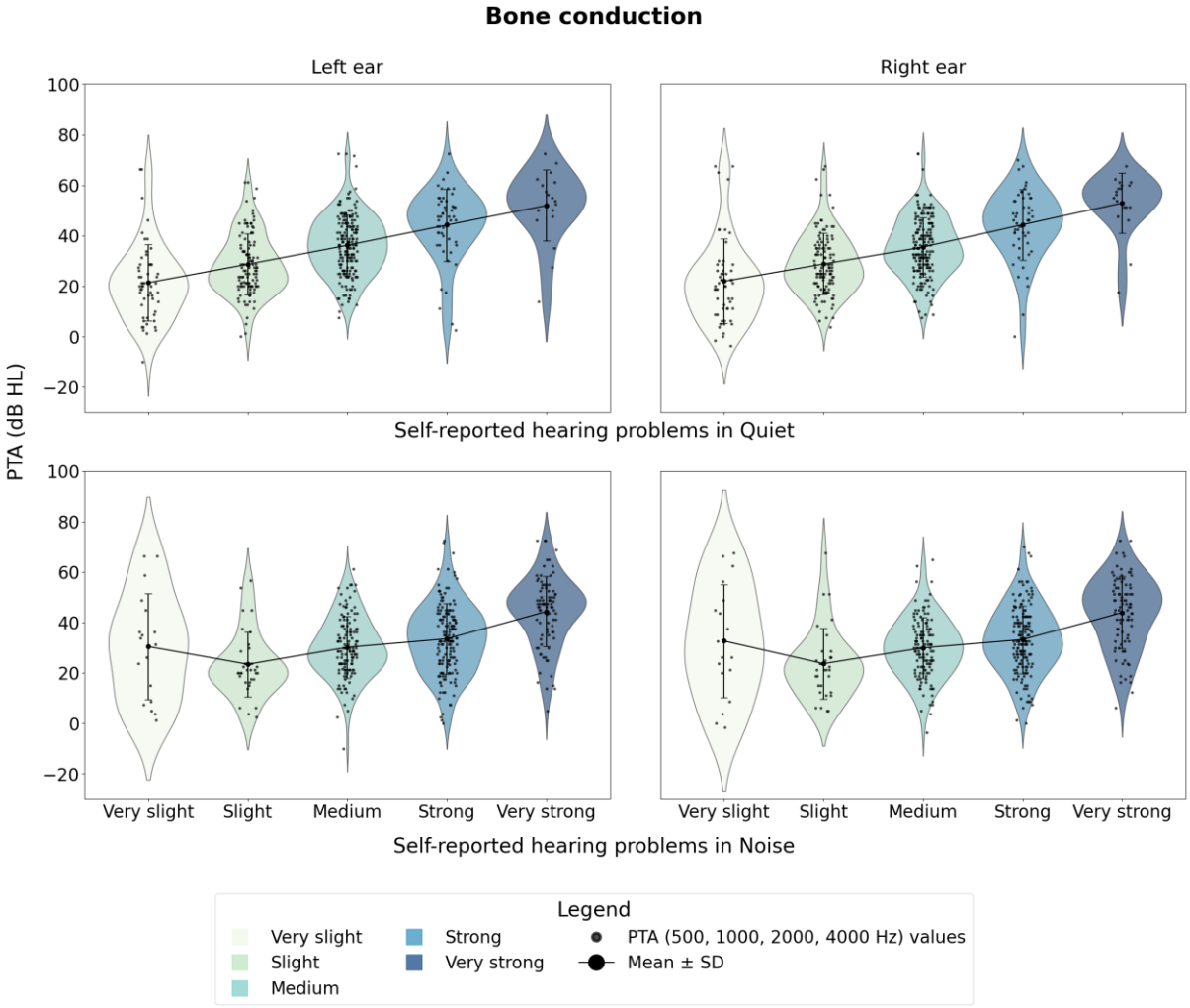
Note: PTA (Pure Tone Average): Measures hearing sensitivity across frequencies (0.5, 1, 2, 4 kHz) using air (AC) and bone conduction (BC) thresholds for both left and right ears, expressed in decibels Hearing Level (dB HL). SRT_DTT (Speech Reception Threshold – Digit Triplet Test): Assesses the ability to understand digit sequences in background noise at 50% accuracy, expressed in decibels Signal-to-Noise Ratio (dB SNR). SRT_Gösa (Speech Reception Threshold – Göttingen Sentence Test): Evaluates speech recognition at 50% accuracy with sentences presented in noise, expressed in dB SNR. Scatter plots display the relationships between each pair of measures for normal-hearing (blue) and deaf or hard of hearing individuals (red) individuals. Both axes represent the respective measurement values in dB HL for PTA and dB SNR for DTT and Gösa. Diagonal plots show the distribution of each measure by hearing status, with the x-axis representing measurement values (in dB HL or dB SNR) and the y-axis representing probability density. Correlation coefficients (Overall, NH, and D/HH) are indicated in each cell, illustrating relationships within and across groups.

Figure 4. Bivariate distributions of PTA air and bone conductance measures, Speech in noise measures and self-reported hearing problems

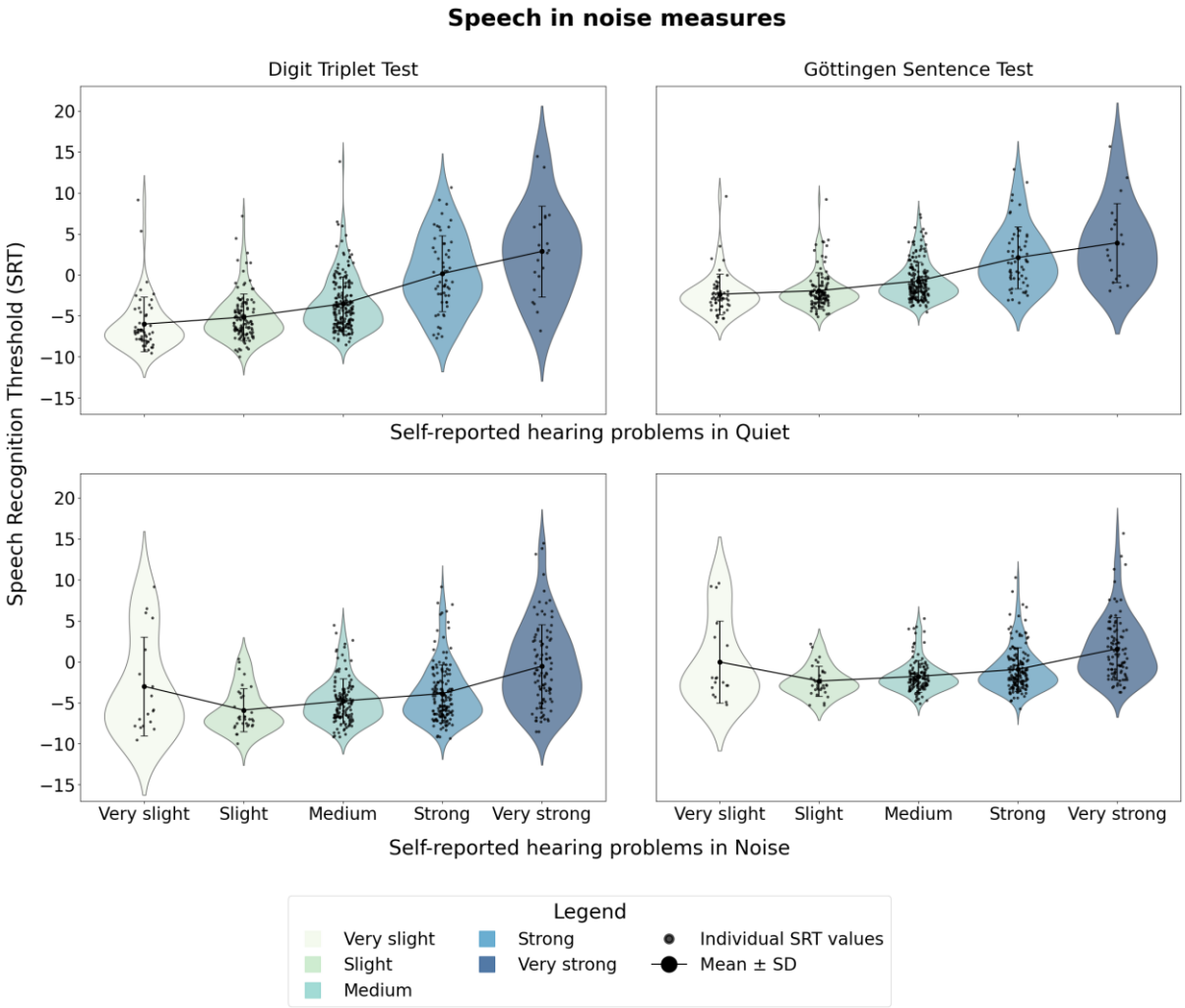
a



b


Bone conduction
Left ear
Right ear
PTA (dB HL)
100
80
60
40
20
0
−20
Self-reported hearing problems in Quiet
Very slight
Slight
Medium
Strong
Very strong
Self-reported hearing problems in Noise
Legend
Very slight
Slight
Medium
Strong
Very strong
PTA (500, 1000, 2000, 4000 Hz) values
Mean ± SD

c



Note. Comparison of self-reported hearing difficulties and objective auditory measures across three domains: Air Conduction (a), Bone Conduction (b), and Speech-in-Noise Measures (c). The violin plots display the distribution of Pure Tone Average (PTA) thresholds at key frequencies (500, 1000, 2000, and 4000 Hz) for both ears, measured through bone and air conduction, as well as Speech Recognition Thresholds (SRT in dB SNR) from the Digit Triplet Test (DTT) and Göttingen Sentence Test (GÖSA). Self-reported hearing problems are categorized by severity in quiet and noise conditions. Black dots represent individual data points, and solid lines indicate the Mean ± Standard deviation. These plots illustrate how subjective perceptions of hearing difficulties align with objective auditory thresholds and speech-in-noise performance across different testing modalities.

Usage notes

We recommend verifying the completeness of any columns of interest before analysis. If missing data is identified, it is advisable to either impute the missing cases[70] or exclude them to ensure accuracy.

Documentation and version control

Comprehensive documentation of data collection methods, transformations, and version control procedures has been prepared. This ensures transparency and accuracy in handling the dataset throughout its lifecycle. Users can access the latest version of the dataset on Zenodo (see below for access information), where all version histories are recorded.

Compliance with the FAIR principles

The dataset adheres to the FAIR principles (Findable, Accessible, Interoperable, and Reusable), ensuring usability for a wide range of research. Findability is achieved through its deposition in a public repository such as the "Hearing4all" Zenodo Community, with the present description as a supporting reference. Accessibility is ensured through its open-access status (under CC BY 4.0 license), allowing unrestricted access to all data files, which are stored in machine-readable JSON format and accompanied by comprehensive documentation. Interoperability is supported by storing all data in JSON format including JSON schema files and taxonomies, providing standardized data structures and definitions to facilitate integration with various analytical systems and tools. Reusability is ensured by rich metadata, clear licensing, and custom tools that support data access across popular analysis platforms (Python, MATLAB, Excel, SQL), supporting easy adaptation for future research.

Code availability

The dataset is available in a machine-readable JSON format, ensuring easy accessibility across different platforms. Custom example scripts are provided to access individual audiogram and loudness scaling files from Python, MATLAB, and Excel, allowing users to analyse and visualize the data effortlessly. Additionally, an SQL script is available for importing the entire dataset into an SQL database for advanced querying. These tools support smooth integration into a wide range of analysis environments.

Excel scripts

- **Excel/Audiogram.xlsx**: An example for using audiograms in Excel. Reads the JSON tables (live data connection), joins the tables as needed, and shows audiogram plots with the possibility to select an observation by its identification number (ID). The relative path of the .json files can be configured on the 'Setup' sheet.
- **Excel/LoudnessScaling.xlsx**: An example for using loudness scaling measurements in Excel. Reads the JSON tables (live data connection), joins the tables as needed, and shows loudness function plots with raw data and fits with the possibility to select an observation by ID.

MATLAB scripts

- **MATLAB/audiogram_getaudiogram.m:** Returns an audiogram for a user specified observation by joining relevant tables.
- **MATLAB/audiogram_plot.m**: Example script for calling 'audiogram_getaudiogram.m' and plotting the audiogram.
- **MATLAB/loudness_scaling_getls.m**: Returns a loudness scaling measurement for a passed observation by joining relevant tables.
- **MATLAB/loudness_scaling_plot.m**: Example script for calling 'loudness_scaling_getls.m' and plotting up to four loudness functions with up to two fits and the raw data points.

Python scripts

- **Python/plot_audiogram.py**: A script for plotting audiograms for an observation by joining relevant tables.
- **Python/plot_loudness_scaling.py**: A script for plotting loudness scaling for an observation by joining relevant tables.

SQL script

- **SQL/import.sql**: A script including documentation for importing the database into an SQL Database.

The full dataset, metadata and the above-mentioned scripts are available under CC BY 4.0 license at https://zenodo.org/records/14808650.

Acknowledgements

We extend our gratitude to all participants who generously contributed their data for the development of OHHR. We also acknowledge Anja Gieseler and Maike Tahden for their efforts in consolidating the specification document detailing the data codes in the previous version of this dataset, which significantly facilitated the description and refinement of the database. This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 - Project ID 390895286.

Author Contributions

Conception, design, and preparation of the manuscript: Sumbul Jafri, Daniel Berg, Christiane Thiel, Andrea Hildebrandt

Contribution to study design: Kirsten C. Wagener, Matthias Vormann, Birger Kollmeier

Recruiting participants and data acquisition: Matthias Vormann, Kirsten C. Wagener

Database management: Daniel Berg, Matthias Vormann, Mareike Buhl, Samira Saak

All authors critically reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Rights and permissions

Open Access. This article is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This license allows for unrestricted use, sharing, adaptation, distribution, and reproduction in any medium or format, provided that appropriate credit is given to the original author(s) and the source, a link to the license is provided, and any changes made are clearly indicated. All figures included in this article are licensed under the same terms and can be credited with a single line. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. World Health Organization. World report on hearing. (2021). Available at: https://www.who.int/publications/i/item/9789240020481

2. Döge, J. *et al.* The Prevalence of Hearing Loss and Provision with Hearing Aids in the Gutenberg Health Study. *Dtsch. Ärztebl. Int.* **120**, 99, https://doi.org/10.3238/arztebl.m2022.0385 (2023).

3. Hackenberg, B. *et al.* Hearing loss and its burden of disease in a large German cohort—hearing loss in Germany. *The Laryngoscope* **132**, 1843–1849, https://doi.org/10.1002/lary.29980 (2022).

4. Ray, J., Popli, G. & Fell, G. Association of cognition and age-related hearing impairment in the English longitudinal study of ageing. *JAMA Otolaryngol. Neck Surg.* **144**, 876–882, https://doi.org/10.1001/jamaoto.2018.1656 (2018).

5. Wang, H. F. *et al.* Hearing impairment is associated with cognitive decline, brain atrophy and tau pathology. *EBioMedicine* **86,** https://doi.org/10.1016/j.ebiom.2022.104336 (2022).

6. Shukla, A. *et al.* Hearing loss, loneliness, and social isolation: a systematic review. *Otolaryngol. Neck Surg.* **162**, 622–633, https://doi.org/10.1177/0194599820910377 (2020).

7. Nordvik, Ø. *et al.* Generic quality of life in persons with hearing loss: a systematic literature review. *BMC Ear Nose Throat Disord.* **18**, 1–13, https://doi.org/10.1186/s12901-018-0051-6 (2018).

8. Dixon, P. R. *et al.* Health-Related Quality of Life Changes Associated With Hearing Loss. *JAMA Otolaryngol. Neck Surg.* **146**, 630–638, https://doi.org/10.1001/jamaoto.2020.0674 (2020).

9. Sethukumar, P. *et al.* Cataloging Existing Hearing Loss Cohort Data to Guide the Development of Precision Medicine for Sensorineural Hearing Loss: A Systematic Review of Hearing Repositories. *J. Int. Adv. Otol.* **19**, 420, https://doi.org/10.5152/iao.2023.22690 (2023).

10. Kanatani, Y. *et al.* National registry of designated intractable diseases in Japan: present status and future prospects. *Neurol. Med. Chir. (Tokyo)* **57**, 1–7, https://doi.org/10.2176/nmc.st.2016-0135 (2017).

11. Nosrati-Zarenoe, R., Hansson, M. & Hultcrantz, E. Assessment of diagnostic approaches to idiopathic sudden sensorineural hearing loss and their influence on treatment and outcome. *Acta Otolaryngol. (Stockh.)* **130**, 384–391, https://doi.org/10.3109/00016480903161541 (2010).

12. Fink, N. E. *et al.* Childhood development after cochlear implantation (CDaCI) study: Design and baseline characteristics. *Cochlear Implants Int.* **8**, 92–116, https://doi.org/10.1002/cii.333 (2007).

13. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779, https://doi.org/10.1371/journal.pmed.1001779 (2015).

14. Brown, M. 1970 British cohort study. *Open Health Data* **2**, e6 (2014).

15. Fisher, D. E. *et al.* Sex-specific predictors of hearing-aid use in older persons: The age, gene/environment susceptibility-Reykjavik study. *Int. J. Audiol.* **54**, 634–641, https://doi.org/10.3109/14992027.2015.1024889 (2015).

16. Ikram, M. A. *et al.* The Rotterdam Study. Design update and major findings between 2020 and 2024. *Eur. J. Epidemiol.* **39**, 183–206, https://doi.org/10.1007/s10654-023-01094-1 (2024).

17. Xu, P. *et al.* Hypothalamic substructural integrity is associated with age, sex and cognitive function across lifespan: A comparative analysis of two large population-based cohort studies. *Alzheimers Dement.* **20**, e084672, https://doi.org/10.1002/alz.084672 (2024).

18. Pinto, T. C. *et al.* Is the Montreal Cognitive Assessment (MoCA) screening superior to the Mini-Mental State Examination (MMSE) in the detection of mild cognitive impairment (MCI) and Alzheimer's Disease (AD) in the elderly? *Int. Psychogeriatr.* **31**, 491–504, https://doi.org/10.1017/s1041610218001370 (2019).

19. Tang-Wai, D. F. *et al.* CCCDTD5 recommendations on early and timely assessment of neurocognitive disorders using cognitive, behavioral, and functional scales. *Alzheimers Dement. NY* **6,** https://doi.org/10.1002/trc2.12057 (2020).

20. Kalbe, E. *et al.* DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int. J. Geriatr. Psychiatry* **19**, 136–143, https://doi.org/10.1002/gps.1042 (2004).

21. Kollmeier, B. & Wesselkamp, M. Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J. Acoust. Soc. Am.* **102**, 2412–2421, https://doi.org/10.1121/1.419624 (1997).

22. Smits, C., Kapteyn, T. S. & Houtgast, T. Development and validation of an automatic speech-in-noise screening test by telephone. *Int. J. Audiol.* **43**, 15–28, https://doi.org/10.1080/14992020400050004 (2004).

23. Zokoll, M. A., Wagener, K. C., Brand, T., Buschermöhle, M. & Kollmeier, B. Internationally

comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype. *Int. J. Audiol.* **51**, 697–705, https://doi.org/10.3109/14992027.2012.690078 (2012).

24. Van den Borre, E., Denys, S., van Wieringen, A. & Wouters, J. The digit triplet test: a scoping review. *Int. J. Audiol.* **60**, 946–963, https://doi.org/10.1080/14992027.2021.1902579 (2021).

25. Gieseler, A. *et al.* Auditory and Non-Auditory Contributions for Unaided Speech Recognition in Noise as a Function of Hearing Aid Use. *Front. Psychol.* **8**, 219, https://doi.org/10.3389/fpsyg.2017.00219 (2017).

26. Tahden, M. A., Gieseler, A., Meis, M., Wagener, K. C. & Colonius, H. What keeps older adults with hearing impairment from adopting hearing aids? *Trends Hear.* **22**, 2331216518809737, https://doi.org/10.1177/2331216518809737 (2018).

27. Buhl, M., Warzybok, A., Schädler, M. R., Majdani, O. & Kollmeier, B. Common Audiological Functional Parameters (CAFPAs) for single patient cases: Deriving statistical models from an expert-labelled data set. *Int. J. Audiol.* **59**, 534–547, https://doi.org/10.1080/14992027.2020.1728401 (2020).

28. Buhl, M., Warzybok, A., Schädler, M. R. & Kollmeier, B. Sensitivity and Specificity of Automatic Audiological Classification Using Expert-Labelled Audiological Data and Common Audiological Functional Parameters. *Int. J. Audiol.* **60**, 16–26, https://doi.org/10.1080/14992027.2020.1817581 (2021).

29. Buhl, M. Interpretable clinical decision support system for audiology based on predicted Common Audiological Functional Parameters (CAFPAs). *Diagnostics* **12**, 463, https://doi.org/10.3390/diagnostics12020463 (2022).

30. Saak, S. K., Hildebrandt, A., Kollmeier, B. & Buhl, M. Predicting Common Audiological Functional Parameters (CAFPAs) as Interpretable Intermediate Representation in a Clinical Decision-Support System for Audiology. *Front. Digit. Health* **2**, 596433, https://doi.org/10.3389/fdgth.2020.596433 (2020).

31. Mousavi, H., Buhl, M., Guiraud, E., Drefs, J. & Lücke, J. Inference and learning in a latent variable model for beta distributed interval data. *Entropy* **23,** https://doi.org/10.3390/e23050552 (2021).

32. Saak, S., Huelsmeier, D., Kollmeier, B. & Buhl, M. A flexible data-driven audiological patient stratification method for deriving auditory profiles. *Front. Neurol.* **13**, 959582, https://doi.org/10.3389/fneur.2022.959582 (2022).

33. Jafri, Sumbul *et al.* OHHR – The Oldenburg Hearing Health Repository [Dataset]. Zenodo, https://doi.org/10.5281/zenodo.14177902 (2025).

34. Brand, T. & Hohmann, V. An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.* **112**, 1597–1604, https://doi.org/10.1121/1.1502902 (2002).

35. Schmidt, K.-H. & Metzler, P. *Wortschatztest*. (Beltz, 1992).

36. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191–2194, https://doi.org/10.1001/jama.2013.281053 (2013).

37. Pronk, M., Deeg, D. J. & Kramer, S. E. Explaining discrepancies between the digit triplet speech-in-noise test score and self-reported hearing problems in older adults. *J. Speech Lang. Hear. Res.* **61**, 986–999, https://doi.org/10.1044/2018_jslhr-h-17-0124 (2018).

38. Wang, D. *et al.* Analysis of influential factors of self-reported hearing loss deviation in young adults. *J. Public Health* **28**, 455–461 (2020).

39. Kamil, R. J., Genther, D. J. & Lin, F. R. Factors associated with the accuracy of subjective assessments of hearing impairment. *Ear Hear.* **36**, 164–167, https://doi.org/10.1097/aud.0000000000000075 (2015).

40. Louw, C., Swanepoel, D. W. & Eikelboom, R. H. Self-Reported Hearing Loss and Pure Tone Audiometry for Screening in Primary Health Care Clinics. *J. Prim. Care Community Health* **9**, 2150132718803156, https://doi.org/10.1177/2150132718803156 (2018).

41. Kenny Gibson, W., Cronin, H., Kenny, R. A. & Setti, A. Validation of the self-reported hearing questions in the Irish Longitudinal Study on Ageing against the Whispered Voice Test. *BMC Res. Notes* **7**, 361, https://doi.org/10.1186/1756-0500-7-361 (2014).

42. Cox, R. M. & Alexander, G. C. The International Outcome Inventory for Hearing Aids (IOI-HA): Psychometric Properties of the English Version. *Int. J. Audiol.* **41**, 30-35, https://doi.org/10.3109/14992020209101309 (2002).

43. Bellach, B. M., Knopf, H. & Thefeld, W. Der Bundes-Gesundheitssurvey 1997/98. *Gesundheitswesen Sonderh. Stuttg. Thieme* **60**, S59–S68 (1998).

44. Ware, J. E., Kosinski, M. & Keller, S. D. A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. Med. Care 34, 220–233, https://doi.org/10.1097/00005650-199603000-00003 (1996).

45. Neyer, F. J., Felber, J. & Gebhardt, C. Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft. *Diagnostica* **58**, 87–99, https://doi.org/10.1026/0012-1924/a000067 (2012).

46. Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **13**, 319–340, https://doi.org/10.2307/249008 (1989).

47. Ajzen, I. The Theory of Planned Behavior. *Organ. Behav. Hum. Decis. Process.* (1991).

48. Krampen, G. *Fragebogen Zu Kompetenz-Und Kontrollüberzeugungen:(FKK)*. (Hogrefe, Verlag für Psychologie, 1991).

49. Winkler, J. & Stolzenberg, H. Adjustierung des Sozialen-Schicht-Index für die Anwendung im Kinder- und Jugendgesundheitssurvey (KiGGS). *Wismar Fachbereich Wirtsch.* (2009).

50. Brand, T. & Kollmeier, B. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.* **111**, 2801–2810, https://doi.org/10.1121/1.1479152 (2002).

51. Oetting, D., Brand, T. & Ewert, S. D. Optimized loudness-function estimation for categorical loudness scaling data. *Hear. Res.* **316**, 16–27, https://doi.org/10.1016/j.heares.2014.07.003 (2014).

52. Wagener, K. C. & Brand, T. Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. Int. J. Audiol. 44, 144-156, https://doi.org/10.1080/14992020500057517 (2005).

53. Cullington, H. E. & Aidi, T. Is the Digit Triplet Test an Effective and Acceptable Way to Assess Speech Recognition in Adults Using Cochlear Implants in a Home Environment. *Cochlear Implants Int.* **18**, 97–105, https://doi.org/10.1080/14670100.2016.1273435 (2017).

54. Kuehnel, V., Kollmeier, B. & Wagener, K. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Z. Für Audiol.* **38**, 4–15 (1999).

55. Lemke, U. Hearing Impairment in Dementia – How to Reconcile Two Intertwined Challenges in Diagnostic Screening. *Audiol. Res.* **1**, https://doi.org/10.4081/audiores.2011.e15 (2011).

56. Petersen, R. C. *et al.* Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* **56**, 303–308 (1999).

57. Petersen, R. C. *et al.* Current concepts in mild cognitive impairment. *Arch. Neurol.* **58**, 1985–1992 (2001).

58. Petersen, R. C. *et al.* Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) [RETIRED]. *Neurology* **56**, 1133–1142, https://doi.org/10.1212/wnl.56.9.1133 (2001).

59. Deloche, G. & Seron, X. From Three to 3: A Differential Analysis of Skills in Transcoding Quantities Between Patients with Broca's and Wernicke's Aphasia. *Brain* **105**, 719–733, https://doi.org/10.1093/brain/105.4.719 (1982).

60. Seron, X. & Deloche, G. From 4 to four: a supplement to 'From three to 3'. *Brain* **106**, 735–744, https://doi.org/10.1093/brain/106.3.735 (1983).

61. Tröster, A. I., Salmon, D. P., McCullough, D. & Butters, N. A comparison of the category fluency deficits associated with Alzheimer's and Huntington's disease. *Brain Lang.* **37**, 500–513, https://doi.org/10.1016/0093-934x(89)90032-1 (1989).

62. Chertkow, H. & Bub, D. Semantic Memory Loss in Dementia of Alzheimer's Type: What Do Various Measures Measure? *Brain* **113**, 397–417, https://doi.org/10.1093/brain/113.2.397 (1990).

63. Monsch, A. U. *et al.* Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Arch. Neurol.* **49**, 1253–1258, https://doi.org/10.1001/archneur.1992.00530360051017 (1992).

64. Cerhan, J. H. *et al.* Diagnostic Utility of Letter Fluency, Category Fluency, and Fluency Difference Scores in Alzheimer's Disease. *Clin. Neuropsychol.* **16**, 35–42, https://doi.org/10.1076/clin.16.1.35.8326 (2002).

65. Belleville, S., Peretz, I. & Malenfant, D. Examination of the working memory components in normal aging and in dementia of the Alzheimer type. *Neuropsychologia* **34**, 195–207, https://doi.org/10.1016/0028-3932(95)00097-6 (1996).

66. Baddeley, A. D., Baddeley, H. A., Bucks, R. S. & Wilcock, G. K. Attentional control in Alzheimer's disease. *Brain* **124**, 1492–1508, https://doi.org/10.1093/brain/124.8.1492 (2001).

67. von Gablenz, P. & Holube, I. Hearing loss and speech recognition in the elderly. Laryngo-Rhino-Otologie 96, 759–764, https://doi.org/10.1055/s-0043-119388 (2017).

68. Thiele, C., Sukowksi, H., Lenarz, T. & Lesinski-Schiedat, A. Göttinger Satztest im Störgeräusch für verschiedene Gruppen von Schwerhörigkeit. *Laryngo-Rhino-Otologie* **91**, 782–788, https://doi.org/10.1055/s-0031-1295419 (2012).

69. Drixler, K., Morfeld, M., Glaesmer, H., Brähler, E. & Wirtz, M. Validierung der messung gesundheitsbezogener lebensqualität mittels des short-form-health-survey-12 (SF-12 Version 2.0) in einer deutschen normstichprobe. *Zeitschrift für Psychosomatische Medizin und Psychotherapie* **66**, 272–286, https://doi.org/10.13109/zptm.2020.66.3.272 (2020).

70. Jäger, S., Allhorn, A. & Bießmann, F. A benchmark for data imputation methods. *Front. Big Data* **4**, 693674, https://doi.org/10.3389/fdata.2021.693674 (2021).