frontiers | Research Topics

From the ear to the brain – new data analytics techniques for a better understanding of human hearing

Edited by
Alessia Paglialonga, Jan Wouters and Norbert Dillier

Published in
Frontiers in Neurology
Frontiers in Systems Neuroscience
Frontiers in Neuroinformatics

ISSN 1664-8714
ISBN 978-2-8325-2722-1
DOI 10.3389/978-2-8325-2722-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

From the ear to the brain – new data analytics techniques for a better understanding of human hearing

Topic editors

Alessia Paglialonga — National Research Council (CNR), Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

Jan Wouters — KU Leuven, Belgium

Norbert Dillier — University of Zurich, Switzerland

Citation

Paglialonga, A., Wouters, J., Dillier, N., eds. (2023). *From the ear to the brain – new data analytics techniques for a better understanding of human hearing*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2722-1

Table of contents

frontiers in Neurology

ORIGINAL RESEARCH
published: 01 March 2022
doi: 10.3389/fneur.2022.836683

Bottom-Up and Top-Down Attention Impairment Induced by Long-Term Exposure to Noise in the Absence of Threshold Shifts

Ying Wang[1,2,3†], *Xuan Huang*[1,2,3†], *Jiajia Zhang*[1,2,3†], *Shujian Huang*[1,2,3], *Jiping Wang*[1,2,3], *Yanmei Feng*[1,2,3], *Zhuang Jiang*[4*], *Hui Wang*[1,2,3*] *and Shankai Yin*[1,2,3]

[1] *Department of Otolaryngology-Head and Neck Surgery, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China,* [2] *Otolaryngology Institute of Shanghai Jiao Tong University, Shanghai, China,* [3] *Shanghai Key Laboratory of Sleep Disordered Breathing, Shanghai, China,* [4] *Department of Otolaryngology, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China*

OPEN ACCESS

Edited by:
Norbert Dillier,
University of Zurich, Switzerland

Reviewed by:
Katrien Vermeire,
Long Island University-Brooklyn, United States
Antonio Greco,
Sapienza University of Rome, Italy

***Correspondence:**
Hui Wang
wangh2005@alumni.sjtu.edu.cn
Zhuang Jiang
jiangzhuang0908@163.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

Received: *15 December 2021*
Accepted: *31 January 2022*
Published: *01 March 2022*

Citation:
Wang Y, Huang X, Zhang J, Huang S, Wang J, Feng Y, Jiang Z, Wang H and Yin S (2022) Bottom-Up and Top-Down Attention Impairment Induced by Long-Term Exposure to Noise in the Absence of Threshold Shifts. Front. Neurol. 13:836683. doi: 10.3389/fneur.2022.836683

Objective: We aimed to assess the effect of noise exposure on bottom-up and top-down attention functions in industrial workers based on behavioral and brain responses recorded by the multichannel electroencephalogram (EEG).

Method: In this cross-sectional study, 563 shipyard noise-exposed workers with clinical normal hearing were recruited for cognitive testing. Personal cumulative noise exposure (CNE) was calculated with the long-term equivalent noise level and employment duration. The performance of cognitive tests was compared between the high CNE group (H-CNE, >92.2) and the low CNE group; additionally, brain responses were recorded with a 256-channel EEG from a subgroup of 20 noise-exposed (NG) workers, who were selected from the cohort with a pure tone threshold <25 dB HL from 0.25 to 16 kHz and 20 healthy controls matched for age, sex, and education. P300 and mismatch negativity (MMN) evoked by auditory stimuli were obtained to evaluate the top-down and bottom-up attention functions. The sources of P300 and MMN were investigated using GeoSource.

Results: The total score of the cognitive test (24.55 ± 3.71 vs. 25.32 ± 2.62, $p < 0.01$) and the subscale of attention score (5.43 ± 1.02 vs. 5.62 ± 0.67, $p < 0.001$) were significantly lower in the H-CNE group than in the L-CNE group. The attention score has the fastest decline of all the cognitive domain dimensions (slope = −0.03 in individuals under 40 years old, $p < 0.001$; slope = −0.06 in individuals older than 40 years old, $p < 0.001$). When NG was compared with controls, the P300 amplitude was significantly decreased in NG at Cz (3.9 ± 2.1 vs. 6.7 ± 2.3 μV, $p < 0.001$). In addition, the latency of P300 (390.7 ± 12.1 vs. 369.4 ± 7.5 ms, $p < 0.001$) and MMN (172.8 ± 15.5 vs. 157.8 ± 10.5 ms, $p < 0.01$) was significantly prolonged in NG compared with controls. The source for MMN for controls was in the left BA11, whereas the noise exposure group's source was lateralized to the BA20.

Conclusion: Long-term exposure to noise deteriorated the bottom-up and top-down attention functions even in the absence of threshold shifts, as evidenced by behavioral and brain responses.

Keywords: noise, attention function, P300, mismatch negativity, bottom-up, top-down

INTRODUCTION

Noise is one of the most common types of pollution in both occupational and non-occupational environments (1). Long-term noise exposure that exceeds certain levels can harm the auditory system, resulting in progressive hearing loss and an increase in hearing sensitivity threshold (2, 3). Meanwhile, evidence of the non-auditory effects related to noise exposure is growing (4, 5), such as, annoyance (6), disturbed sleep (7), cardiovascular disease (8), and anxiety (9). In addition to these effects, noise exposure affects a variety of cognitive processes, such as reaction time, memory, perception, and attention (10). Human error and, in some cases, increased accidents may result from the alteration of attention performance (11). A previous study demonstrated that noise exposure could impair performance on the focused attention task (12), while some studies found that noise could increase arousal levels and accuracy in computerized attention tests (13). The effect of noise exposure on attention performance remain rather inconclusive (14, 15).

One of the influential parameters in the effect of noise on attention performance could be noise characteristics. Jafari et al. (10) discovered the decreased attention in low-frequency noise-exposed subjects (16) and a significant reduction of visual and auditory attention when noise intensity was at 95 dBA level. Smith and Miles (17) found that subjects who were exposed to noise for 5 h made more errors than those who were exposed for 2 h in a reaction time task. Pawlaczyk-Łuszczyńska et al. (18) discovered that the low-frequency noise might affect the concentration and attention function. Furthermore, exposure duration, intensity, education years, gender, age, hearing level, and even basic diseases could all be influential parameters regarding the effect of noise on attention performance and might lead to these apparently contradictory results.

Attention is not a monolithic process, and two types of attention are commonly distinguished: top-down and bottom-up attention (19, 20). The voluntary allocation of attention to certain features or objects is referred to as top-down attention (21). Attention, on the other hand, is not only voluntarily directed. Salient stimuli can attract attention, even though the subject has no intention of focusing on these stimuli (22). Bottom-up attention refers to solely being guided by externally driven factors to stimuli (22). The attention process can be modulated by "top-down" specific task goals and expectations as well as "bottom-up" external-driving factors (23). "Bottom-up" attention plays a critical role during auditory processing in noisy environments (24), which is capable of tracking certain auditory stimuli in noisy environments without paying attention voluntarily to the auditory modality. In tasks with several components, noise may cause an increase in concentration on the dominant or high-probability component at the expense of other features (12). However, there is still a scarcity of solid evidence from people who have documented the effects of noise exposure on top-down and bottom-up attention performance.

In this study, we aimed to evaluate the effect of noise exposure on bottom-up and top-down attention functions in industrial workers in the absence of peripheral hearing loss based on behavioral and brain responses recorded by the multichannel electroencephalogram (EEG). First, we utilized the Montreal Cognitive Assessment (MoCA) cognitive test to assess the cognitive performance, particularly attention, in a large cohort of shipyard workers with long-term noise exposure. In addition, we measured the P300 and the mismatch negativity (MMN), which reflect the brain's sound encoding, in a subgroup of 20 noise-exposed workers with pure tone thresholds <25 dB HL from 0.25 to 16 kHz, selected from the cohort and 20 healthy controls matched for age, gender, and education; their hearing functions were further evaluated by a comprehensive test battery containing both subjective and objective measures (25).

METHODS

Participants and Study Design

A large-scale epidemiological survey was conducted from June to July 2019 (25). A questionnaire was used to collect the cross-sectional physical examination data from 807 sanding, welding, metal, and cutting workers, such as demographics, noise exposure duration, type of work, history of major diseases, including genetic and drug-related hearing loss, diabetes, hypertension, smoking, and alcohol consumption, and use of hearing protection devices. Audiologic evaluations and personal cumulative noise exposure (CNE) estimates were conducted, as described in our previous study (25). By the median (92.2 dBA-year) of CNE, all participants were divided into two groups: high CNE (H-CNE) and low CNE (L-CNE). Then, recruited participants completed cognitive tests to assess the cognitive function by professional physicals (26). The procedures and criteria for participant inclusion and exclusion are outlined in **Figure 1.** Inclusion criteria include: (1) age < 50 years; (2) air conduction thresholds < 25 dB HL at 0.25–8 kHz in bilateral ear; (3) employment duration > 2 years; (4) right-handed; and (5) native Mandarin speaker. Exclusion criteria include abnormal tympanograms, a history of otological diseases, or reading or language difficulties.

Furthermore, 20 participants were selected at random from L-CNE group as the noise-exposed group (NG) based on the following criteria: (1) under the age of 40 years; (2) pure-tone

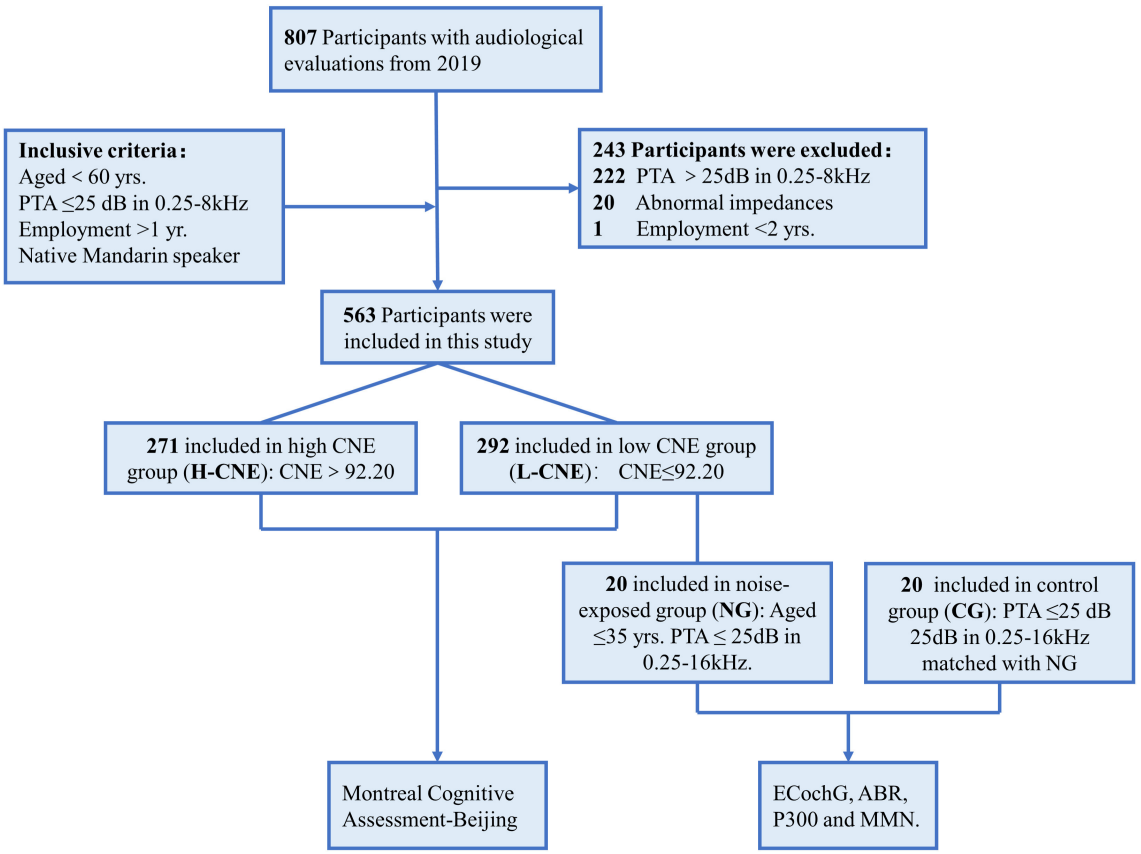


FIGURE 1 | The flowchart illustrates the study design and participants.

average (PTA) < 25 dB hearing level at any frequency between 0.25 and 16 kHz; (3) right-handedness; and (4) native Mandarin speakers. The NG group underwent more extensive auditory processing tests, such as an electrocochleogram (ECochG) and auditory brainstem responses (ABR). A control group (CG) of 20 healthy subjects without a history of occupational noise exposure was matched for age, gender, education level, and hearing thresholds. On-site measurements of ECochG and ABR were taken. The high-density EEG was performed during a routine visit to our hospital.

This study was approved by the Institutional Ethics Review Board of the Shanghai Sixth People's Hospital affiliated with Shanghai Jiao Tong University and was registered in the Chinese Clinical Trial Registry (http://www.chictr.org.cn/index.aspx, registration number: ChiCTR-RPC-17012580). Potential consequences and benefits of the study were explained, and a written informed consent was obtained from every subject before this study.

Cognitive Test

The MoCA Beijing Version (MoCA-BJ) was administered by professional geriatricians (26), which is considered as an acceptable tool for lower education level groups in both urban and rural areas (27). The MoCA-BJ scale contained seven cognitive domains (5 points-visuospatial and executive function, 3 points-naming, 6 points-attention, 2 points-abstraction, 3 points-language, 5 points-delayed memory, and 6 points-orientation) ranging from 0 to 30, with a higher number indicating better performance. One point was used for education adjustment, in which an additional point can be added to the total score if the individual education years ≤12 years.

ECochG and ABR

The SmartEP auditory evoked potential system (Intelligent Hearing Systems; Miami, FL) was used to measure the ECochG and ABR in a soundproof room. The acoustic stimulation was delivered *via* ER-3A insertable earphones (Etymotic Research; Elk Grove Village, IL). The recording electrode was placed near the tympanic membrane for ECochG or the hairline in the middle of the forehead for ABR, and the reference electrode was on the mastoid. The amplitude and latency of the compound action potential (CAP) in ECochG and waves I and V in ABR were measured in the response to 80 dB HL clicks. The stimulating rate was 13.1 Hz, and the electrical resistance was <3 kΩ. The responses were band-pass filtered between 200 and 2,000 Hz and averaged 1,024 times in each trial.

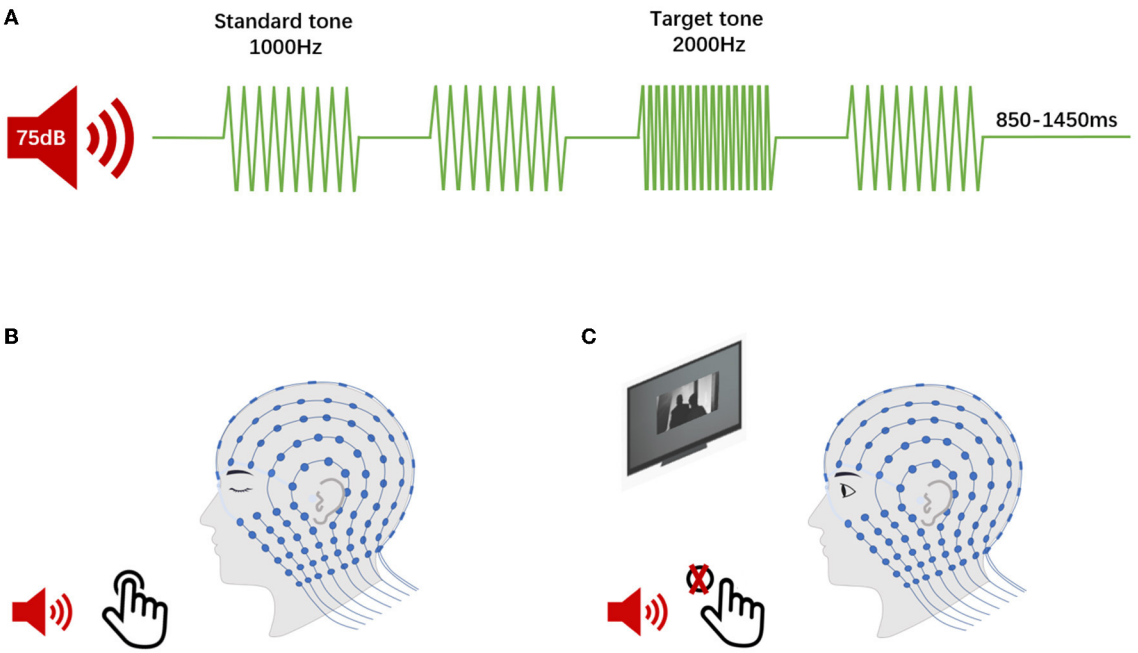


FIGURE 2 | Event-related potentials (ERPs) procedure. **(A)** The continuous auditory stimulus comprised both rarely presented target sounds and frequently presented standard sounds in two tasks. **(B)** The 2-tone auditory oddball task (P300, for top-down analyses). The participants were required to discriminate the target stimulus from the standard tone by pressing a button. **(C)** The passive listening task (MMN, for bottom-up analysis). The participants were instructed to watch the silent movie without responding to the presented auditory stimulus.

Event-Related Potential

EEG Acquisition

Electroencephalogram signals were collected in a soundproof room using the Geodesic EEG System (GES 300, Electrical Geodesics; Eugene, OR). A 256-channel HydroCel Geodesic Sensor Net was used to place all the electrodes, and all electrode-skin impedance values were kept below 50 kΩ during the recording. Responses were recorded online relative to a vertex reference electrode (Cz) at a sampling rate of 1,000 Hz and then digitally filtered (0.3–70 Hz). Participants were instructed to keep awake and avoid moving their eyes or changing their posture, and the EEG data were monitored for signs of drowsiness.

Event-Related Potential Procedure

The auditory oddball task required participants' responses based on a cognitive decision regarding the auditory stimulus types. The results of this oddball task were interpreted as auditory "top-down" effects, principally (28). Afterwards, in a passive listening task, participants would hear the same stream of auditory stimuli as in the oddball task, and this passive listening task could reflect the "bottom-up" attention effect (28). Therefore, participants engaged in the following two auditory tasks during EEG acquisition (**Figure 2**): (1) a 2-tone auditory oddball task. The oddball task consisted of two stimuli that were presented in a random order. One stimulus is the quasi-random sequence of frequent standard tones (1,000 Hz, an 85% occurrence probability), while another stimulus is infrequent deviant (target) tones (2,000 Hz, a 15% occurrence probability). The whole task consisted of a total of 1,000 auditory stimuli with random interstimulus intervals (ISIs) ranging from 850 to 1,450 ms. In the oddball paradigm, all stimuli (75-dB sound pressure level with 50-ms duration shaped by a 5-ms rise/fall time window) were delivered through a loudspeaker (Micro-DSP, Sichuan, China) placed 100 cm from the subject at an 180 degrees azimuth. The participants were required to discriminate the target stimulus from the standard tone by pressing a button with their eyes closed to minimize any destructive effects due to alterations in visual attention. (2) A passive listening task used the same series of stimuli in the auditory oddball task. During this task, we showed a silent movie to the participants to divert their attention away from the presented auditory stimuli. They were instructed to watch the movie and not respond to the simultaneously presented target auditory stimuli.

ERP Analysis

Event-related potential (ERP) data were analyzed offline with the Net Station 4.3 software (EGI). The continuous EEG signals were digitally filtered between 0.1 and 40 Hz, and then segmented using the event stimulus timestamp. All epochs were calculated 100 ms before and 700 ms after stimulus onset. After segmentation, artifact detection was performed using the Net Station artifact detection tool, which automatically detects eye blinks and eye movements and marks bad channels. Data were baseline-corrected using a 100 ms pre-stimulus period. A single-trial examination was performed for each participant, and artifacts were rejected before grand averages were computed. The

TABLE 1 | Demographic characteristics of subjects in the high-cumulative noise exposure (H-CNE) and low-CNE (L-CNE) groups.

Variable	H-CNE group			L-CNE group			*P*-value[#]
	≤40 yrs. (*n* = 216)	>40 yrs. (*n* = 55)	Overall (*n* = 271)	≤40 yrs. (*n* = 245)	>40 yrs. (*n* = 47)	Overall (*n* = 292)	
Age, mean (±SD), *yrs.*	32.5 ± 4.4	45.7 ± 4.0	35.2 ± 6.8	31.7 ± 4.6	44.5 ± 3.0	33.8 ± 6.4	0.012
Sex, male, (%)	202 (93.5)	51 (92.7)	253 (93.4)	228 (93.1)	41 (87.2)	269 (91.8)	0.483
Education years, mean (±SD), *yrs.*	10.2 ± 2.1	9.4 ± 2.0	10.1 ± 2.1	10.5 ± 2.1	9.8 ± 2.2	10.4 ± 2.1	0.0 76
Exposure duration, mean (±SD), *yrs.*	8.9 ± 4.1***	12.0 ± 5.5**	9.5 ± 4.6	6.6 ± 3.7	8.7 ± 4.3	7.0 ± 4.0	<0.001
CNE, median (IQR), *dBA-year*	94.8 (92.5–105.4)***	96.4 (92.9–106.4)***	95.2 (92.5–106.4)	90.4 (76.0–92.2)	90.1 (77.8–92.2)	90.4 (76.0.9–92.2)	<0.001
Diabetes, *n* (%)	2 (0.9)	2 (3.6)	4 (1.5)	2 (0.8)	0 (0)	2 (0.7)	0.362
Hypertension, *n* (%)	191 (88.4)	43 (78.2)	234 (86.3)	203 (82.9)	38 (80.9)	240 (82.2)	0.176
Smoking, *n* (%)	105 (48.6)	23 (41.8)	128 (47.2)	116 (47.7)	17 (36.2)	133 (45.9)	0.745
Drinking, *n* (%)	96 (44.4)	24 (43.6)	120 (44.3)	103 (42.4)	19 (40.0)	122 (42.1)	0.597
PTA, mean (±SD), *dB*							
0.25–8 kHz	17.0 ± 4.4***	18.0 ± 4.0	17.16 ± 4.3	15.4 ± 5.0	17.2 ± 4.3	15.67 ± 4.9	<0.001
10–16 kHz	31.2 ± 14.0*	39.2 ± 12.6	32.8 ± 14.1	28.4 ± 13.3	38.7 ± 10.0	30.0 ± 13.4	0.016

[#]*Indicates statistical significance between the H-CNE and L-CNE groups. The number of asterisks indicates statistical significance against the L-CNE in the same age group (*, <0.05; **, <0.01; ***, p < 0.001). H-CNE, high cumulative noise exposure group; L-CNE, low cumulative noise exposure group; PTA, pure-tone average (dB HL); yrs, years.*

P300 elicited by the target in this task is a large, positive-going potential that peaks around 300 ms post-stimulus in normal young adults. The MMN was quantified from the deviant-standard difference waveforms. Peak latency or peak amplitude was determined as the most negative (for MMN) or positive (for P300) point. The amplitude was measured from the baseline, defined as the mean voltage of the pre-stimulus interval, while the latency was measured from the point in time when the deviance occurred (100 ms). We analyzed three (Fz, Cz, and Pz) electrodes to observe the distribution of the P300 and MMN components. Furthermore, the ERP data were input to the GeoSource module of the Net Station software (version 4.5.7) to compute the standardized low-resolution brain electromagnetic tomography (sLORETA) for the purpose of source localization (29, 30).

Statistics

For parametric data, the results were presented as a mean (SD) or median [interquartile range (IQR)], and for categorical data, as a number (percentage). Depending on the data type, Pearson's 2 test, independent samples *t*-test, and Mann–Whitney *U*-test were used to determine intergroup differences. A linear regression line was fitted to the data to determine the decline rate of cognitive test scores (slope) from 70 to 110 dBA-year of CNE, which was compared using the Mann–Whitney *U*–test. The independent samples *t*-test or the Mann–Whitney *U*–test were used to compare the latencies and amplitudes of AEPs and ERPs between the NG and CG. The 2-tailed $p < 0.05$ was considered to indicate statistical significance, and data analysis was performed using the SPSS 24.0 (IBM, Armonk, NY) and Prism version 9 (GraphPad Software).

RESULTS

Baseline Characteristics of Participants

The overall median CNE was ∼92.20 dBA-year approximately. In the H-CNE group ($n = 271$), the mean age was 35.2 ± 4.4 years old and the median CNE was 95.2 (92.5–106.4) dBA-year, whereas the mean age of the L-CNE group ($n = 292$) was 33.8 ± 6.4 years and the median CNE was 90.4 (76.0.9–92.2) dBA-year. The subjects in the H-CNE and L-CNE groups were matched well in terms of age, gender, education years, smoking and alcohol drinking habits, and basic diseases. Furthermore, there were no significant differences regarding the terms mentioned above in the same age group (≤40 years and >40 years) between the H-CNE and L-CNE groups. An overview of the demographic and clinical characteristics is shown in **Table 1**.

Cognitive Test Results

Figure 3A presents the results of the MoCA-BJ education adjustment scores and cognitive domain scores in H-CNE and L-CHE subjects. The H-CNE group performed significantly worse than the L-CNE group in the education adjustment scores (24.55 ± 3.71 vs. 25.32 ± 2.62) and domains of attention, visual spatial/executive (5.34 ± 1.02 vs. 5.62 ± 0.67; 3.37 ± 1.37 vs. 3.60 ± 1.13). For subjects under 40 years old, almost all cognitive test scores in the H-CNE group were similar to those in the L-CNE group. Only attention subscales differed significantly between the L-CNE (5.64 ± 0.67) and H-CNE groups (5.40 ± 1.00) ($t = -3.071$, $p = 0.002$). For subjects aged over 40 years, attention scores, visual spatial/executive scores, and education adjustment scores in the H-CNE group were 5.11 ± 1.07, 2.71 ± 1.32, and 22.73 ± 3.72, respectively, while in the L-CNE group, scores were 5.48 ± 0.68, 3.33 ± 1.28, and 24.13±2.83, respectively. There were significant differences in

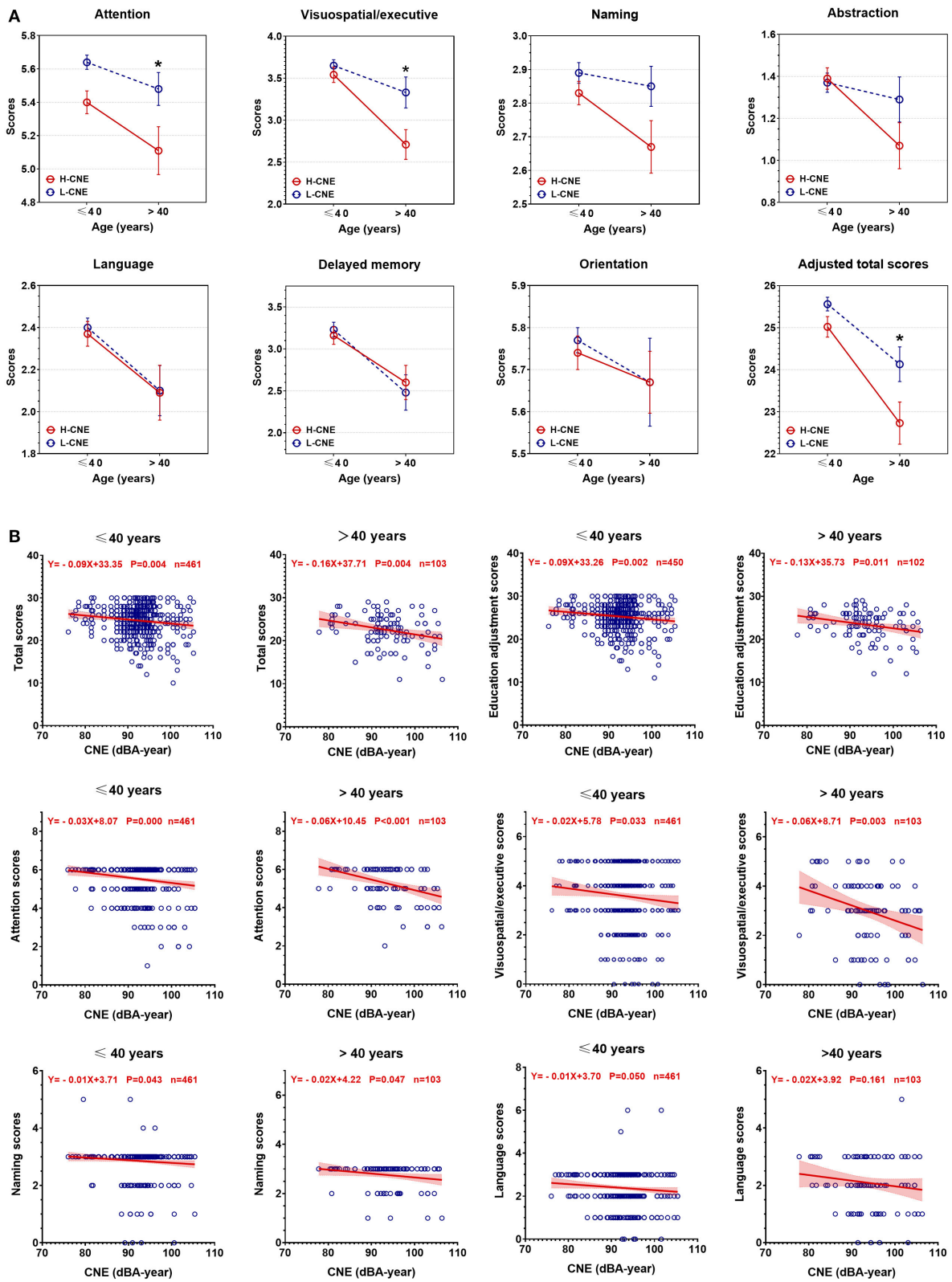


FIGURE 3 | The between-group differences in Montreal Cognitive Assessment Beijing Version (MoCA-BJ) scores. **(A)** Group analysis of MoCA-BJ scores between high-cumulative noise exposure (H-CNE) and low-CNE (L-CNE) groups. For subjects aged under 40 years old, attention function scores were significantly higher in the *(Continued)*

FIGURE 3 | L-CNE group compared with the H-CNE group. For subjects aged over 40 years old, attention, visuospatial and executive, and education adjustment scores showed a difference between H-CNE and L-CNE. **(B)** The scatter plot depicted the decrease of MoCA-BJ scores with the increase of CNE among participants aged over 40 years or younger. For educational adjusted scores, attention, visuospatial/executive, naming, and language scores, there were significant differences in the rate of decrease in scores with CNE. The asterisks indicates statistical significance between the L-CNE and the H-CNE group in the same age group (*, <0.05; **, <0.01; ***, $p < 0.001$).

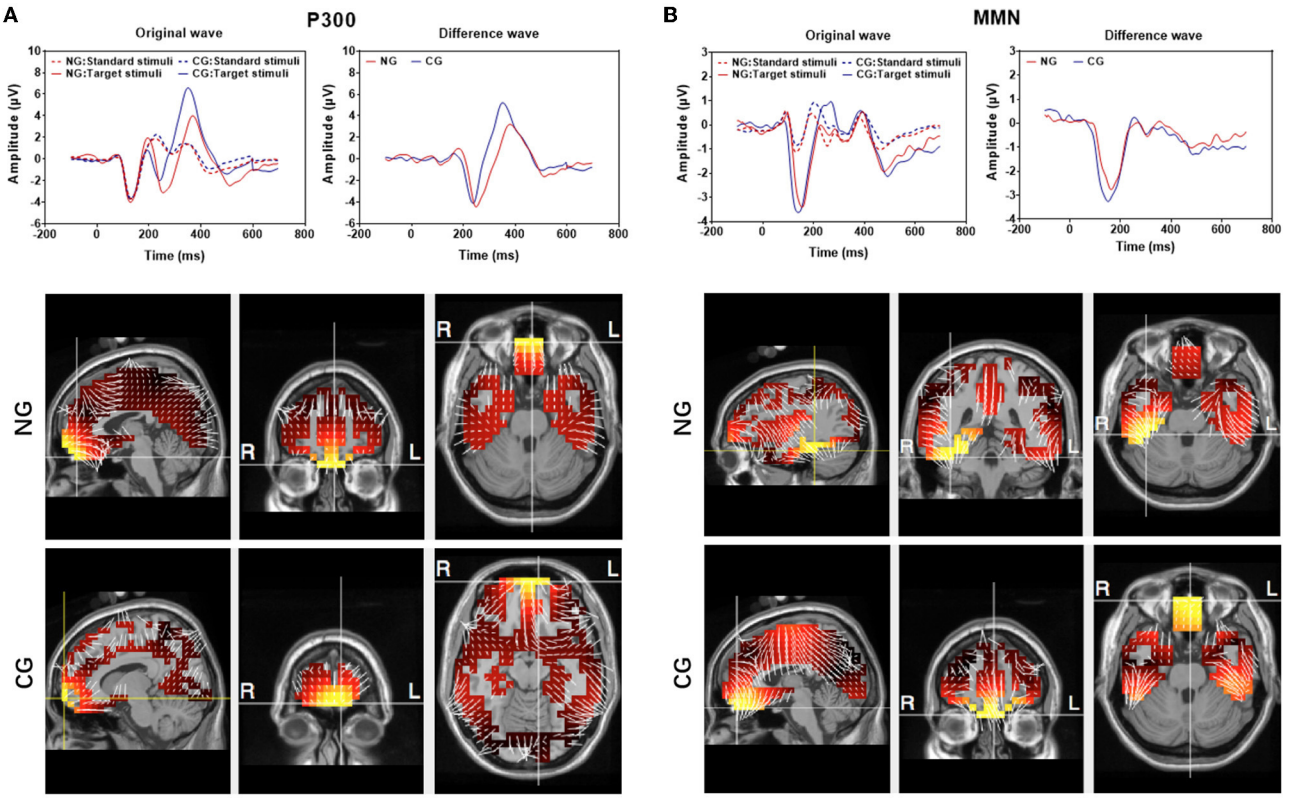


FIGURE 4 | Averaged P300 **(A)** and mismatch negativity (MMN) **(B)** recorded at Cz electrode. Top: Original responses to the standard and deviant stimuli from the 20 subjects in the NG and CG groups. There is a difference between the responses to two types of stimuli. (Dotted lines reflected the response evoked by target stimuli while solid lines reflected the response evoked by standard stimuli; red lines presented the response in NG while blue lines were in CG). Button: sLORETA images of the MMN and P300 components of the two groups at the sagittal, coronal, and axial slices of the maximum current density.

attention scores, visual spatial/executive scores, and education adjustment scores between these two groups ($t = -2.123$, $p = 0.036$; $t = -2.436$, $p = 0.017$; and $t = -2.436$, $p = 0.017$).

Scatterplots revealed a negative relationship between cognitive test scores and CNE, as the values of CNE increased, the corresponding cognitive total scores and subscale scores decreased **(Figure 3B)**. There were significant differences in the rates of decrease in scores among all individuals for educational adjusted scores ($Z = 1.903$, $p = 0.05$), attention scores ($Z = 2.984$, $p = 0.003$), and naming scores ($Z = 2.131$, $p = 0.033$). Among all dimensions of cognitive domains, attention scores were the ones with the fastest decline (slope = −0.03 point/dBA-year, $p < 0.001$ in individuals under 40 years old; slope = −0.06 point/dBA-year, $p < 0.001$ in individuals over 40 years old).

MMN and P300

Demographic and clinical characteristics of the NG and CG subgroups are compared in **Supplementary Table 1.** The NG subjects ($n = 20$) were exposed for 8 h/day for an average of 6.9 years, with a mean PTA at 0.25–8 kHz of 9.3 ± 3.1 and 9.8 ± 4.3 dB at 10–16 kHz. Subjects in the CG group ($n = 20$) worked in silent conditions and the mean PTA at 0.25–8 kHz was 10.4 ± 2.7 dB and at 10–16 kHz was 13.1 ± 6.8 dB. There were no significant differences in the amplitude and latency of ABR waves Iand V, as well as the ECochG wave AP between the NG and CG groups (all $p > 0.05$). The other clinical characteristics, such as age, gender, years of education, and cognitive test scores, were not significantly different between the two groups (all $p > 0.05$).

The group-averaged waveforms at Cz are presented in **Figure 4** and group-averaged latency and amplitude at Cz, Pz, and Fz are shown in **Supplementary Table 2**. Overall, deviant stimuli elicited much larger responses from both subgroups in both P300 and MMN measurements. The peak latencies for both P300 and MMN were longer in the responses of NG subjects. In the NG group, subjects' responses had slightly smaller P300 and MMN amplitudes. The P300 latency and amplitude at Cz were

390 ± 12.1 ms and 3.9 ± 2.1 μV, respectively, and the MMN latency and amplitude at Cz were 172.8 ± 15.5 ms and −2.7 ± 0.6 μV. In the CG group, the P300 latency and amplitude at Cz were 369 ± 7.5 ms and 6.7 ± 2.3 μV, respectively, and the MMN latency and amplitude at Cz were 157.8 ± 10.5 ms and −3.2 ± 0.7 μV. The peak latency of MMN from all three sites differed significantly between NG and CG groups (all $p < 0.01$), while there was no significant between-group difference in the amplitudes of MMN ($p > 0.05$).

The source localization was performed in both MMN and P300 by using group-averaged EEG data from the 20 subjects in each group (**Figure 4).** The maximum current strength of MMN in CG was identified in the front lobe close to the left BA 11 (orbitofrontal area, voxel locations: −3, 52, −27), whereas the maximum current strength of NG was considerably lateralized to the right BA20 (inferior temporal gyrus, voxel locations: 39, −39, −27). The source localization for the maximum current of P300 was in the left BA11, and there was not a significant difference between the NG (locations: −3, 52, −27) and CG (locations: −10, 66, −13) groups.

DISCUSSION

The present study demonstrated that long-term noise exposure impairs bottom-up and top-down attention functions in the absence of threshold shifts, as evidenced by behavioral and brain responses. The alterations of MMN and P300 suggested impairments in bottom-up and top-down attention functions in participants under long-term noise exposure. In the NG subgroup, significantly lower MMN amplitudes were observed, and the peak latencies of both MMN and P300 were considerably longer. Furthermore, we found a shift of MMN source localization in the right temporal lobe of the noise exposure group, indicating a reorganization of the auditory cortex and alterations of hemisphere dominance. In addition, CNE was a significant factor in the impairment of cognitive function, suggesting that the low-level noise was not as effective compared with high levels of noise.

The association of ambient noise with attention function was less investigated (31, 32), and nearly all early field studies of noise exposure and cognitive performance had some weaknesses, such as small sample sizes, inadequate noise measurement data, and auditory evaluation of each subject accurately. On the other hand, solid evidence from prospective and epidemiological studies (33) revealed that hearing loss was an independent risk factor for cognitive decline, containing the attenuated attention functions, while the mechanism of this association has yet to be elucidated (34). There was likely overlap among the peripheral auditory, central auditory, and cognitive function (35). Animal studies showed that even under a brief exposure to noise, there would be a significant loss of cochlear afferent synapses (36–44). It remained a concern whether such synapse loss could occur in humans and lead to attention function deterioration. Further, noise altered neuronal dendrites (45) and induced peroxidation in specific areas of the lemniscal ascending auditory pathway in mice (46). Noise exposure would result in the substantial impairment of the auditory cortex function and behavioral consequences in mice, regardless of the intensity and duration of noise exposure (47). In the present study, the noise exposure of each subject was documented by their employment duration in the industrial environment, and by the noise survey in the workplaces. All subjects were exposed to industrial noise for 8 h/day for more than 300 days/year. In addition, all individuals maintained good hearing sensitivity over the frequency range from 0.25 to 8 kHz (the hearing thresholds of NG subjects were <25 dB from 0.25 to 16 kHz). The attention deficits observed in this study could be attributable to hard-to-detect cochlea damage and related central plasticity, as there was no interference from hearing threshold or other confounders.

Besides top-down and bottom-up attention, attention could be divided into arousal, sustained attention, selective attention, and divided attention according to hierarchical models from Sohlberg and Mateer (48). Selective attention might be a crucial component of cognitive function (10). The altered amplitude and latency of MMN and P300 could indicate a decrease in not only bottom-up and top-down attention but also selective attention, sustained attention, and divided function (49, 50). On the one hand, the bottom-up and top-down attention models claim that, although distinct processes mediate the attention guidance based on bottom-up and top-down factors, both types of attentional processes require a common neural apparatus, the frontoparietal network (21). On the other hand, the anterior attentional system (AAS), also known as the executive network, oversees selective attention, sustained attention, and divided attention. This system is related to the prefrontal dorsolateral cortex, the orbitofrontal cortex, and the anterior cingulate cortex (48), according to the Posner and Petersen neuroanatomical model (48). The frontoparietal network is clearly the core area of various attention models. Previous animal studies showed that noise exposure could increase oxidative stress, decrease brain-derived neurotrophic factor and synapse-associated protein (51), and cause neuronal dendritic alteration and free radical imbalance in the prefrontal cortex and hippocampus (45). In the present study, we found a significant difference between the NG and CG subgroups in the auditory oddball and the passive listening tasks, indicating a decreased top-down and bottom-up attention process as well as decreased selective, sustained, and divided attention function. In addition, we found that the source localization for maximal MMN was lateralized to the right BA20 (inferior temporal gyrus) in NG subjects, while it was the left BA11 (orbitofrontal area) in CG subjects. These findings were consistent with previous studies, which discovered that the frontal area was the source of MMN in subjects who had not been exposed to noise, and the right temporal lobe appeared to be more susceptible to functional reorganization in subjects who had been exposed to noise (52, 53). Our findings were consistent with that the speech-discrimination-induced ERP was dominant in the right hemisphere in individuals exposed to occupational noise, in contrast to the left hemisphere dominance in control subjects (54). While there was no distinct difference for the P300 source, the underlying mechanisms might be that in noisy environments, bottom-up driven attention is more important during auditory processing (24), and long-term

noise exposure might deteriorate bottom-up driven attention function first. Noise exposure induced the reorganization of tonotopic areas (55), as well as structural and molecular changes in human auditory (temporal gyrus) and non-auditory areas (frontal area) (56). However, it was not clear whether similar central plasticity occurs in association with difficult-to-test cochlear damage, which could also reduce the auditory input from cochleae to the auditory brain, although the threshold might not be increased.

Our study has some limitations that should be taken into consideration. We only compare the cognitive performances between different levels of CNE and lack a set of data from the control group of healthy subjects without noise exposure. Our sample size for the EEG measurements remains small, and we cannot completely rule out the existence of peripheral damage in these subjects that requires more sensitive and reliable tests. Due to the large sample size, no further cognitive assessments, such as the Stroop test were performed to evaluate the attention function.

CONCLUSIONS

In conclusion, we found that noise exposure deteriorated both bottom-up and top-down attention functions, as evidenced by the behavioral and brain responses. Behavioral test results revealed that the higher cumulative noise exposure could result in more severe damage to attention function, which was also confirmed by the reduced ERP amplitude and latency. The difficult-to-test cochlear damage, reorganization of auditory and non-auditory areas, and hemisphere dominance alteration might contribute to the significant attention deficits.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Ethics Review Board of the Shanghai Sixth People's Hospital affiliated with Shanghai Jiao Tong University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SY and HW: study conception and design. ZJ, HW, JW, and SH: acquisition of data. YW, ZJ, JZ, and YF: analysis and interpretation of data. YW, ZJ, XH, and HW: drafting of manuscript. HW: critical revision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (82071041/H1304), Innovative research team of high-level local universities in Shanghai (SHSMU-ZLCX20211702), Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 82101220), the First Grant (2020YFC2005201) of Chinese National Key Research and Development Program (2020YFC2005200).

ACKNOWLEDGMENTS

The authors would like to acknowledge all the participants and institutions in this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2022.836683/full#supplementary-material

REFERENCES

1. Hammer MS, Swinburn TK, Neitzel RL. Environmental noise pollution in the United States: developing an effective public health response. *Environ Health Perspect.* (2014) 122:115–9. doi: 10.1289/ehp.1307272
2. Willis S, Moore BCJ, Galvin JJ3rd, Fu QJ. Effects of noise on integration of acoustic and electric hearing within and across ears. *PLoS ONE.* (2020) 15:e0240752. doi: 10.1371/journal.pone.0240752
3. Sha SH, Schacht J. Emerging therapeutic interventions against noise-induced hearing loss. *Expert Opin Investig Drugs.* (2017) 26:85–96. doi: 10.1080/13543784.2017.1269171
4. Basner M, Babisch W, Davis A, Brink M, Clark C, Janssen S, et al. Auditory and non-auditory effects of noise on health. *Lancet.* (2014) 383:1325–32. doi: 10.1016/S0140-6736(13)61613-X
5. Stansfeld SA, Matheson MP. Noise pollution: non-auditory effects on health. *Br Med Bull.* (2003) 68:243–57. doi: 10.1093/bmb/ldg033
6. Beutel ME, Jünger C, Klein EM, Wild P, Lackner K, Blettner M, et al. Noise annoyance is associated with depression and anxiety in the general population- the contribution of aircraft noise. *PLoS ONE.* (2016) 11:e0155357. doi: 10.1371/journal.pone.0155357
7. Muzet A. Environmental noise, sleep and health. *Sleep Med Rev.* (2007) 11:135–42. doi: 10.1016/j.smrv.2006.09.001
8. Sørensen M, Andersen ZJ, Nordsborg RB, Jensen SS, Lillelund KG, Beelen R, et al. Road traffic noise and incident myocardial infarction: a prospective cohort study. *PLoS ONE.* (2012) 7:e39283. doi: 10.1371/journal.pone.0039283
9. Miedema HM, Oudshoorn CG. Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals. *Environ Health Perspect.* (2001) 109:409–16. doi: 10.1289/ehp.01109409
10. Jafari MJ, Khosrowabadi R, Khodakarim S, Mohammadian F. The effect of noise exposure on cognitive performance and brain activity patterns. *Open Access Maced J Med Sci.* (2019) 7:2924–31. doi: 10.3889/oamjms.2019.742
11. Wilkins PA, Action WI. Noise and accidents–a review. *Ann Occup Hyg.* (1982) 25:249–60.
12. Smith AP. Noise and aspects of attention. *Br J Psychol.* (1991) 82:313–24. doi: 10.1111/j.2044-8295.1991.tb02402.x

13. Alimohammadi I, Sandrock S, Gohari MR. The effects of low frequency noise on mental performance and annoyance. *Environ Monit Assess.* (2013) 185:7043–51. doi: 10.1007/s10661-013-3084-8
14. Tzivian L, Winkler A, Dlugaj M, Schikowski T, Vossoughi M, Fuks K, et al. Effect of long-term outdoor air pollution and noise on cognitive and psychological functions in adults. *Int J Hyg Environ Health.* (2015) 218:1–11. doi: 10.1016/j.ijheh.2014.08.002
15. Gawron VJ. Performance effects of noise intensity, psychological set, and task type and complexity. *Hum Factors.* (1982) 24:225–43. doi: 10.1177/001872088202400208
16. Pawlaczyk-Luszczyńska M, Szymczak W, Dudarewicz A, Sliwińska-Kowalska M. Proposed criteria for assessing low frequency noise annoyance in occupational settings. *Int J Occup Med Environ Health.* (2006) 19:185–97. doi: 10.2478/v10001-006-0022-9
17. Smith AP, Miles C. The combined effects of occupational health hazards: an experimental investigation of the effects of noise, nightwork and meals. *Int Arch Occup Environ Health.* (1987) 59:83–9. doi: 10.1007/BF00377682
18. Pawlaczyk-Luszczyńiska M, Dudarewicz A, Waszkowska M, Szymczak W, Sliwińska-Kowalska M. The impact of low-frequency noise on human mental performance. *Int J Occup Med Environ Health.* (2005) 18:185–98.
19. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci.* (1995) 18:193–222. doi: 10.1146/annurev.ne.18.030195.001205
20. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci.* (2002) 3:201–15. doi: 10.1038/nrn755
21. Katsuki F, Constantinidis C. Bottom-up and top-down attention: different processes and overlapping neural systems. *Neuroscientist.* (2014) 20:509–21. doi: 10.1177/1073858413514136
22. Pinto Y, van der Leij AR, Sligte IG, Lamme VA, Scholte HS. Bottom-up and top-down attention are independent. *J Vis.* (2013) 13:16. doi: 10.1167/13.3.16
23. Kaya EM, Elhilali M. Investigating bottom-up auditory attention. *Front Hum Neurosci.* (2014) 8:327. doi: 10.3389/fnhum.2014.00327
24. Lagemann L, Okamoto H, Teismann H, Pantev C. Bottom-up driven involuntary attention modulates auditory signal in noise processing. *BMC Neurosci.* (2010) 11:156. doi: 10.1186/1471-2202-11-156
25. Jiang Z, Wang J, Feng Y, Sun D, Zhang X, Shi H, et al. Analysis of early biomarkers associated with noise-induced hearing loss among shipyard workers. *JAMA Netw Open.* (2021) 4:e2124100. doi: 10.1001/jamanetworkopen.2021.24100
26. Huang YY, Qian SX, Guan QB, Chen KL, Zhao QH, Lu JH, et al. Comparative study of two Chinese versions of montreal cognitive assessment for screening of mild cognitive impairment. *Appl Neuropsychol Adult.* (2021) 28:88–93. doi: 10.1080/23279095.2019.1602530
27. Yu J, Li J, Huang X. The Beijing version of the montreal cognitive assessment as a brief screening tool for mild cognitive impairment: a community-based study. *BMC Psychiatry.* (2012) 12:156. doi: 10.1186/1471-244X-12-156
28. Hong SK, Park S, Ahn MH, Min BK. Top-down and bottom-up neurodynamic evidence in patients with tinnitus. *Hear Res.* (2016) 342:86–100. doi: 10.1016/j.heares.2016.10.002
29. Ghumare EG, Schrooten M, Vandenberghe R, Dupont P. A time-varying connectivity analysis from distributed EEG sources: a simulation study. *Brain Topogr.* (2018) 31:721–37. doi: 10.1007/s10548-018-0621-3
30. Wojcik GM, Masiak J, Kawiak A, Schneider P, Kwasniewicz L, Polak N, et al. New protocol for quantitative analysis of brain cortex electroencephalographic activity in patients with psychiatric disorders. *Front Neuroinform.* (2018) 12:27. doi: 10.3389/fninf.2018.00027
31. Elmenhorst EM, Elmenhorst D, Wenzel J, Quehl J, Mueller U, Maass H, et al. Effects of nocturnal aircraft noise on cognitive performance in the following morning: dose-response relationships in laboratory and field. *Int Arch Occup Environ Health.* (2010) 83:743–51. doi: 10.1007/s00420-010-0515-5
32. Schapkin SA, Falkenstein M, Marks A, Griefahn B. Executive brain functions after exposure to nocturnal traffic noise: effects of task difficulty and sleep quality. *Eur J Appl Physiol.* (2006) 96:693–702. doi: 10.1007/s00421-005-0049-9
33. Lin FR, Yaffe K, Xia J, Xue QL, Harris TB, Purchase-Helzner E, et al. Hearing loss and cognitive decline in older adults. *JAMA Intern Med.* (2013) 173:293–9. doi: 10.1001/jamainternmed.2013.1868
34. Gurgel RK, Ward PD, Schwartz S, Norton MC, Foster NL, Tschanz JT. Relationship of hearing loss and dementia: a prospective, population-based study. *Otol Neurotol.* (2014) 35:775–81. doi: 10.1097/MAO.0000000000000313
35. Humes LE. Speech understanding in the elderly. *J Am Acad Audiol.* (1996) 7:161–7.
36. Kujawa SG, Liberman MC. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J Neurosci.* (2009) 29:14077–85. doi: 10.1523/JNEUROSCI.2845-09.2009
37. Kim KX, Payne S, Yang-Hood A, Li SZ, Davis B, Carlquist J, et al. Vesicular glutamatergic transmission in noise-induced loss and repair of cochlear ribbon synapses. *J Neurosci.* (2019) 39:4434–47. doi: 10.1523/JNEUROSCI.2228-18.2019
38. Kaur T, Clayman AC, Nash AJ, Schrader AD, Warchol ME, Ohlemiller KK. Lack of fractalkine receptor on macrophages impairs spontaneous recovery of ribbon synapses after moderate noise trauma in C57BL/6 mice. *Front Neurosci.* (2019) 13:620. doi: 10.3389/fnins.2019.00620
39. Song Q, Shen P, Li X, Shi L, Liu L, Wang J, et al. Coding deficits in hidden hearing loss induced by noise: the nature and impacts. *Sci Rep.* (2016) 6:25200. doi: 10.1038/srep25200
40. Shi L, Chang Y, Li X, Aiken SJ, Liu L, Wang J. Coding deficits in noise-induced hidden hearing loss may stem from incomplete repair of ribbon synapses in the cochlea. *Front Neurosci.* (2016) 10:231. doi: 10.3389/fnins.2016.00231
41. Shi L, Liu K, Wang H, Zhang Y, Hong Z, Wang M, et al. Noise induced reversible changes of cochlear ribbon synapses contribute to temporary hearing loss in mice. *Acta Otolaryngol.* (2015) 135:1093–102. doi: 10.3109/00016489.2015.1061699
42. Shi L, Guo X, Shen P, Liu L, Tao S, Li X, et al. Noise-induced damage to ribbon synapses without permanent threshold shifts in neonatal mice. *Neuroscience.* (2015) 304:368–77. doi: 10.1016/j.neuroscience.2015.07.066
43. Shi L, Liu L, He T, Guo X, Yu Z, Yin S, et al. Ribbon synapse plasticity in the cochleae of Guinea pigs after noise-induced silent damage. *PLoS ONE.* (2013) 8:e81566. doi: 10.1371/journal.pone.0081566
44. Liu L, Wang H, Shi L, Almuklass A, He T, Aiken S, et al. Silent damage of noise on cochlear afferent innervation in guinea pigs and the impact on temporal processing. *PLoS ONE.* (2012) 7:e49550. doi: 10.1371/journal.pone.0049550
45. Manikandan S, Padma MK, Srikumar R, Jeya Parthasarathy N, Muthuvel A, Sheela Devi R. Effects of chronic noise stress on spatial memory of rats in relation to neuronal dendritic alteration and free radical-imbalance in hippocampus and medial prefrontal cortex. *Neurosci Lett.* (2006) 399:17–22. doi: 10.1016/j.neulet.2006.01.037
46. Cheng L, Wang SH, Chen QC, Liao XM. Moderate noise induced cognition impairment of mice and its underlying mechanisms. *Physiol Behav.* (2011) 104:981–8. doi: 10.1016/j.physbeh.2011.06.018
47. Zhou X, Merzenich MM. Environmental noise exposure degrades normal listening processes. *Nat Commun.* (2012) 3:843. doi: 10.1038/ncomms1849
48. Posner MI, Petersen SE. The attention system of the human brain. *Annu Rev Neurosci.* (1990) 13:25–42. doi: 10.1146/annurev.ne.13.030190.000325
49. Light GA, Braff DL. Mismatch negativity deficits are associated with poor functioning in schizophrenia patients. *Arch Gen Psychiatry.* (2005) 62:127–36. doi: 10.1001/archpsyc.62.2.127
50. Pratt N, Willoughby A, Swick D. Effects of working memory load on visual selective attention: behavioral and electrophysiological evidence. *Front Hum Neurosc.* (2011) 5:57. doi: 10.3389/fnhum.2011.00057
51. Wang S, Yu Y, Feng Y, Zou F, Zhang X, Huang J, et al. Protective effect of the orientin on noise-induced cognitive impairments in mice. *Behav Brain Res.* (2016) 296:290–300. doi: 10.1016/j.bbr.2015.09.024
52. Shiell MM, Champoux F, Zatorre RJ. The right hemisphere planum temporale supports enhanced visual motion detection ability in deaf people: evidence from cortical thickness. *Neural Plast.* (2016) 2016:7217630. doi: 10.1155/2016/7217630
53. Lee YS, Min NE, Wingfield A, Grossman M, Peelle JE. Acoustic richness modulates the neural networks supporting intelligible speech processing. *Hear Res.* (2016) 333:108–17. doi: 10.1016/j.heares.2015.12.008

54. Brattico E, Kujala T, Tervaniemi M, Alku P, Ambrosi L, Monitillo V. Long-term exposure to occupational noise alters the cortical organization of sound processing. *Clin Neurophysiol.* (2005) 116:190–203. doi: 10.1016/j.clinph.2004.07.030
55. Dietrich V, Nieschalk M, Stoll W, Rajan R, Pantev C. Cortical reorganization in patients with high frequency cochlear hearing loss. *Hear Res.* (2001) 158:95–101. doi: 10.1016/S0378-5955(01)00282-9
56. Husain FT, Medina RE, Davis CW, Szymko-Bennett Y, Simonyan K, Pajor NM, et al. Neuroanatomical changes due to hearing loss and chronic tinnitus: a combined VBM and DTI study. *Brain Res.* (2011) 1369:74–88. doi: 10.1016/j.brainres.2010.10.095

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Systems Neuroscience

BRIEF RESEARCH REPORT
published: 14 July 2022
doi: 10.3389/fnsys.2022.892198

Deconvolution of Ears' Activity (DEA): A New Experimental Paradigm to Investigate Central Auditory Processing

Fabrice Bardy [1,2,3,4]*

[1] HEARing Co-operative Research Center, Carlton, VIC, Australia, [2] Department of Linguistics, Macquarie University, Sydney, NSW, Australia, [3] School of Psychology, University of Auckland, Auckland, New Zealand, [4] Eisdell Moore Centre for Hearing and Balance Research, University of Auckland, Auckland, New Zealand

A novel experimental paradigm, "deconvolution of ears' activity" (DEA), is presented which allows to disentangle overlapping neural activity from both auditory cortices when two auditory stimuli are presented closely together in time in each ear. Pairs of multi-tone complexes were presented either binaurally, or sequentially by alternating presentation order in each ear (i.e., first tone complex of the pair presented to one ear and second tone complex to the other ear), using stimulus onset asynchronies (SOAs) shorter than the neural response length. This timing strategy creates overlapping responses, which can be mathematically separated using least-squares deconvolution. The DEA paradigm allowed the evaluation of the neural representation in the auditory cortex of responses to stimuli presented at syllabic rates (i.e., SOAs between 120 and 260 ms). Analysis of the neuromagnetic responses in each cortex offered a sensitive technique to study hemispheric lateralization, ear representation (right vs. left), pathway advantage (contra- vs. ipsi-lateral) and cortical binaural interaction. To provide a proof-of-concept of the DEA paradigm, data was recorded from three normal-hearing adults. Results showed good test-retest reliability, and indicated that the difference score between hemispheres can potentially be used to assess central auditory processing. This suggests that the method could be a potentially valuable tool for generating an objective "auditory profile" by assessing individual fine-grained auditory processing using a non-invasive recording method.

Keywords: auditory cortical responses, overlapping neural responses, auditory stimulation, least-squares deconvolution, rapid acoustic stimulation

OPEN ACCESS

Edited by:
Norbert Dillier,
University of Zurich, Switzerland

Reviewed by:
Lina Reiss,
Oregon Health and Science University, United States
Pavel Zahorik,
University of Louisville, United States

***Correspondence:**
Fabrice Bardy
fabrice.bardy@auckland.ac.nz

Received: *08 March 2022*
Accepted: *16 June 2022*
Published: *14 July 2022*

Citation:
Bardy F (2022) Deconvolution of Ears' Activity (DEA): A New Experimental Paradigm to Investigate Central Auditory Processing. Front. Syst. Neurosci. 16:892198. doi: 10.3389/fnsys.2022.892198

INTRODUCTION

The auditory system is a binaural system. Auditory cortices in right and left hemispheres receive ascending projections originating from each ear. The resulting activity in one cortex is a mixture of signals from both ears. The effects of monaural and binaural stimulation on cortical responses have been studied considerably in humans, using techniques such as magnetoencephalography (MEG) (Pantev et al., 1986). MEG is well suited to study hemispheric processing differences given the low dispersion of the magnetic field and the location of the cerebral auditory cortical centers in the temporal lobe of each hemisphere. For monaural sound presentation, there is evidence of a predominant contra-lateral pathway in the human auditory system (Pantev et al., 1986, 1998; Mäkel et al., 1993). The contra-lateral advantage is characterized by shorter latencies and larger amplitudes

of the N100m. These measures reflect anatomical differences, especially the larger number of neurons projecting on the contra-lateral compared to the ipsi-lateral side of the ascending auditory pathways. For binaural presentation at the cortical level, MEG frequency-tagging of cortical steady-state responses can be employed (Fujiki et al., 2002). Here, stimuli receive a marker, or tag, using a specific modulation frequency. This makes it possible to identify which stimulus evoked the observed cortical response.

The auditory system is a temporally fast system. It can process acoustic stimuli presented with short temporal disparities between the ears. Processing rapidly changing sounds encompasses several levels of transformation from one cochlea to the auditory cortex of both hemispheres. Unfortunately, a non-invasive objective measure of binaural interaction in the auditory cortex during rapid stimulation with temporally restricted sounds is not yet available. However, if such a method were to be available, research on the interaction and/or integration of signals in the auditory cortex for stimuli presented at syllabic rates (i.e., between 4 and 10 Hz) could provide new insights into normally developed and disordered central auditory processing systems.

This report describes a novel experimental paradigm, named "deconvolution of ears' activity" (DEA), which makes use of the least-squares (LS) deconvolution technique to allow separation of left and right ear activity in each hemisphere to rapidly presented stimuli (Bardy et al., 2014a,b). The LS deconvolution technique is a mathematical algorithm designed to disentangle temporally overlapping brain responses. The technique, described in Bardy et al. (2014a), relies on the timing characteristics of the stimulus sequence to be unequally spaced. This specific property is called "jitter". The LS deconvolution has been validated in a pair paradigm using EEG data (Bardy et al., 2014b). In the DEA paradigm, LS deconvolution is applied to a sequence of stimuli presented in pairs either binaurally or sequentially, using stimulus onset asynchronies (SOAs) shorter than the duration of the cortical. Right and left ear activity is extracted from the mixture of signals in both auditory cortices such that, using this method, the signal propagation from each ear to each auditory cortex can be tracked. The DEA paradigm is introduced in this paper, and is evaluated on three normal hearing adults as a proof-of-concept.

Two hypotheses were investigated: (1) the LS deconvolution technique can disentangle temporally overlapping brain responses in each auditory cortex originating from both ears with a high test-retest reliability; and (2) an auditory profile can be generated based on measures of the auditory pathway lateralization, hemispheric advantage, ear advantage and binaural cortical interaction.

METHODS

Subjects

Test and retest MEG data were obtained from 3 right-handed adult subjects (3 males, age: 37, 32, 29) on two separate occasions. Subjects had no history of neurological or audiological problems and had pure tone audiometric thresholds $\leq$20 dB HL in all octave frequencies between 250 to 8,000 Hz. This study was approved by and conducted under oversight of the Macquarie University Human Research Ethics Committee. All subjects gave written informed consent to participate in this study.

Stimulation

Two multi-tone (MT) stimuli, selected to optimize the amplitude of the cortical response (Bardy et al., 2015), were obtained by amplitude-modulated tone-bursts composed of carrier frequencies of 2 and 1 kHz with modulation frequencies 800 and 400 Hz respectively. Changing the frequency of the stimuli was used to minimize the habituation of the cortical neural response. The stimuli were presented through custom insert earphones, using pneumatic tubes to deliver sound to the subject, with a frequency response relatively flat between 500 and 8 kHz and an approximate 10 dB/octave roll-off for frequencies below 500 Hz (Raicevich et al., 2010). The two MTs were presented in pairs, using jittered SOAs with means of 120, 190, or 260 ms. The jitter distribution, permitting the deconvolution, was rectangular with a width of 70 ms and a step size of 13.3 ms. The inter-pair interval (IPI), representing the time interval between the onset of two successive pairs of stimuli, was jittered with 400 ms around an average of 1,400 ms. The MTs had a rise and decay time of 10 ms, a duration of 50 ms and an rms intensity of 70 dB SPL. They were presented through shielded transducers (Oldfield, 1971). The stimuli were presented in three presentation conditions. The first presentation condition was binaural (both stimuli of the pair presented simultaneously to the right and left ears). In the two other presentation conditions, stimuli were alternated sequentially in each ear (i.e., when the left ear received the first tone, the right ear received the second tone of the pair, and vice versa). All 9 conditions (3 SOAs x 3 presentation conditions) were randomly presented in a 25-min-long stimulus sequence.

In conditions where the cortical response was longer than the SOA, brain responses overlapped in time, and LS deconvolution described by Bardy et al. (2014a) was employed to disentangle the occurring overlapping responses. Thus, for example, in the alternating sequential condition, it was possible within each auditory cortex to separate the activity elicited by the stimulus to the right and left ears respectively from the overlapping cortical response (**Figure 1**).

Procedure

MEG data were continuously recorded using a whole-head MEG system (Model PQ1160R-N2, KIT, Kanazawa, Japan) consisting of 160 coaxial first-order gradiometers with a 50 mm baseline (Kado et al., 1999; Uehara et al., 2003). MEG data were acquired in a magnetically shielded room using a sampling rate of 1,000 Hz with a bandpass filter of 0.1–200 Hz and a 50 Hz notch filter. For co-registration, the location of five indicator coils placed on the participant's head were digitized. A pen digitizer (Polhemus Fastrack, Colchester, VT) was used to measure the shape of each participant's head which was then carefully centered in the MEG dewar (position error $<$10 mm for each subject). Artifact removal from MEG data included signals exceeding amplitude ($>$2,700 fT/cm) and magnetic gradient ($>$800 fT/cm/sample) criteria (Yetkin et al., 2004). Averaging and band-pass filtering between 3 Hz (6 dB/octave, forward) and 30 Hz (48 dB/octave, zero-phase) was performed for each trigger

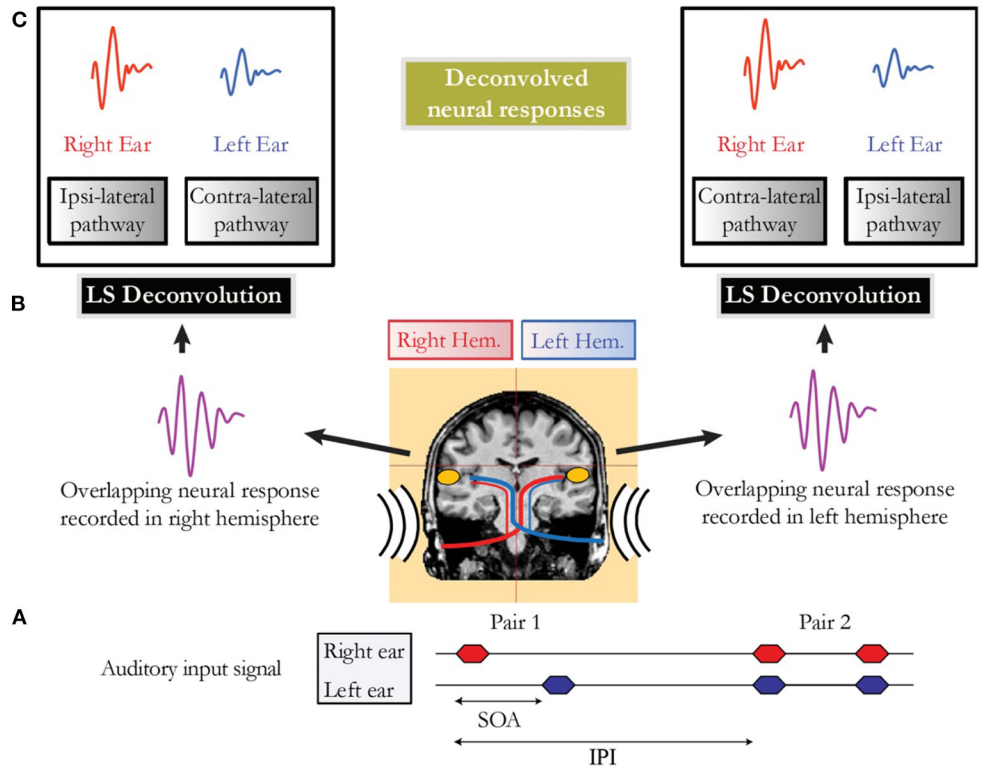


FIGURE 1 | (A) Schematic representation of the auditory input signal for a sequential presentation condition (e.g., Pair 1) followed by a binaural presentation condition (e.g., Pair 2). The stimulus onset asynchrony (SOA) that represents the timing between the onset of the two auditory stimuli of a pair is jittered to allow deconvolution. The time interval between the onset of two pairs is referred as the inter-pair interval (IPI). **(B)** Representation of the overlapping neural activity of Pair 1. The MEG signal is recorded in both hemispheres in response to the first auditory stimulus presented to the right ear, followed closely by a stimulus presented in the left ear. **(C)** Deconvolved neural responses in each auditory cortex obtained using the least-squares (LS) deconvolution technique.

condition using the non-contaminated epochs. The accepted epochs after artifact rejection were exported from BESA 5.3 into MATLAB (MathWorks, Natick, MA) and downsampled to 100 Hz. Deconvolution was performed for each of the 160 channels to disentangle overlapping responses. For each condition, recovered responses were defined by epochs of 100 ms pre-stimulus to 380 ms post-stimulus.

Statistical Analysis

Amplitudes and latencies were defined by peak measures of magnetic global field power (mGFP) calculated on 40 sensors located over the temporal lobe in each hemisphere. For each subject and each condition, the N100m was defined as the most positive peak in the 80–150 ms following the sound onset. The selected time window for the P200m was 120–200 ms. A repeated measures ANOVAs was performed. Greenhouse-Geisser corrections for sphericity were applied, as indicated by the cited ε value (Greenhouse and Geisser, 1959). Bonferroni corrections were applied for *post hoc* analysis.

Individual laterality indices (LIs) for hemisphere, pathway, ear and cortical binaural interaction were calculated. For each subject, LIs were calculated based on the relevant mGFP response amplitudes, time-averaged over a 200-ms window post-onset. **Figure 2** displays an example of auditory cortical responses elicited by pairs of auditory stimuli presented binaurally or alternated sequentially for an individual subject with SOAs jittered around 190 ms. For hemispheric lateralization, the LI was calculated as the difference between left and right mGFP response amplitudes (bottom vs. top 6 panels in **Figure 2B**) normalized by the sum of left and right mGFP responses (i.e. $LI = \frac{mGFP(left)-mGFP(right)}{mGFP(left)+mGFP(right)}$). The LI was +1 for a response geared completely asymmetrical toward the left hemisphere, zero for a symmetrical response, and −1 for a response geared completely asymmetrical toward the right hemisphere. For pathway advantage, the LI was calculated employing the same method using the responses associated with the contra- (panels labeled 3R, 4L, 5L and 6R in **Figure 2B**) and the ipsi-lateral pathways (panels labeled 3L, 4R, 5R, 6L in **Figure 2B**). The ear LI was calculated by comparing mGFP responses from the left ear (3rd and 6th columns in **Figure 2B**) to the responses from the right ear (4th and 5th columns in **Figure 2B**). Finally, the binaural interaction LI was computed by comparing binaural stimulation (first 2 columns in **Figure 2B**) and monaural stimulation responses (last 4 columns in **Figure 2B**). The

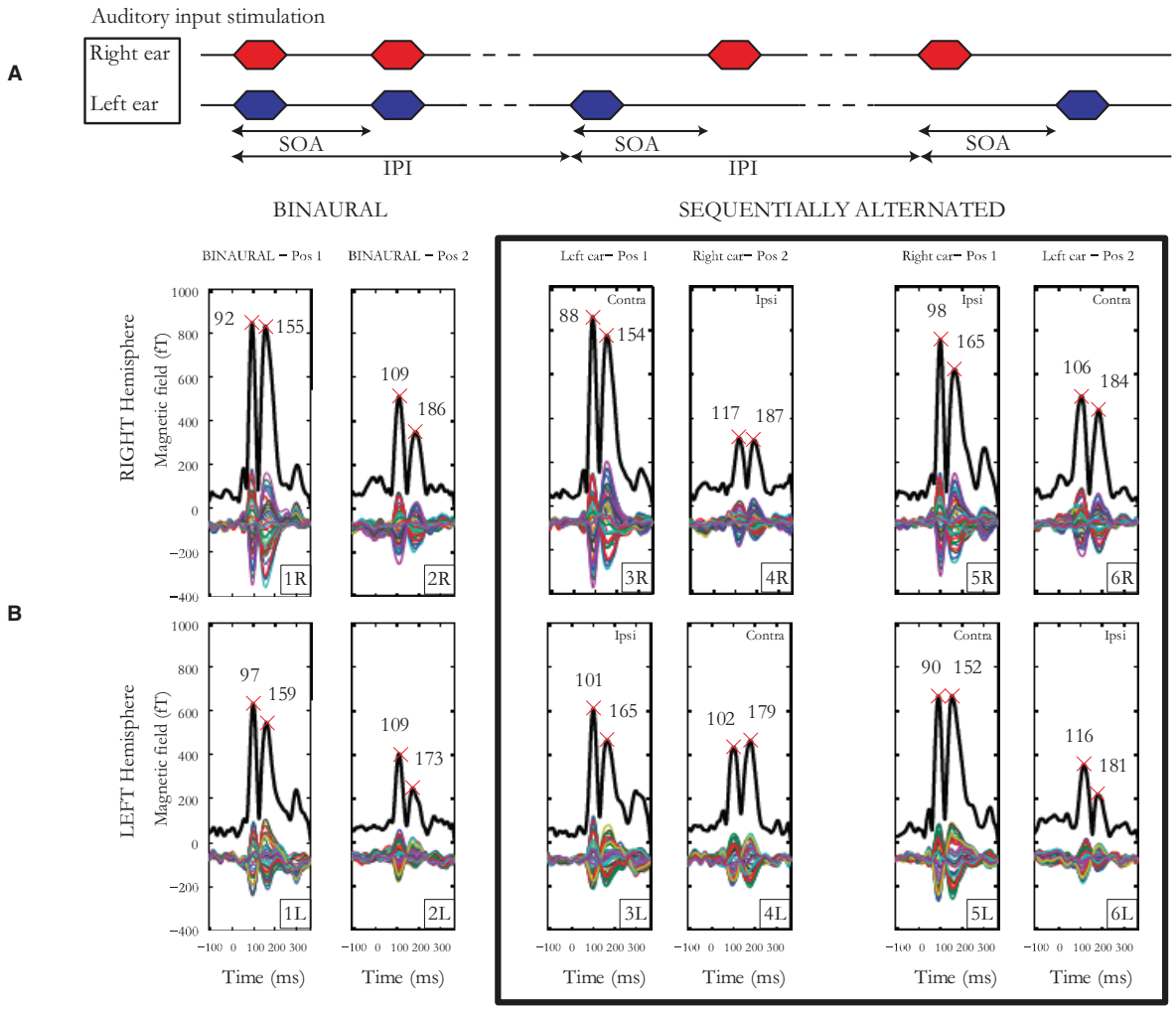


FIGURE 2 | (A) Auditory input stimulation representing the binaural condition to the left, the sequentially alternated conditions "left ear" followed by "right ear" in the middle and then the sequentially alternated "right ear" followed by "left ear" on the right. The stimulus onset asynchrony (SOA) represents the time between the start of the two stimuli of a pair, while the inter-pair interval (IPI) represents the time interval between the onset of two successive pairs of stimuli. **(B)** Cortical responses from subject 1 for SOAs jittered around 190 ms. Multiple thin waveforms represent activity recorded by each of the 40 sensors located over the temporal lobe, in each hemisphere, after LS deconvolution, from −100 to 380 ms after stimulus onset. mGFP waveforms are represented with a thick black line, provided for both right and left hemispheres, the 3 presentation conditions (1 x binaural, 2 x sequentially alternated) and both first and second tone-bursts. Latencies of the N100m and P200m are indicated by crosses.

binaural interaction LI was computed for both hemispheres and for each pathway (i.e., ipsi- and contra-lateral). For each subject, the difference between the means for each LI was checked by the Student's *t*-test. The threshold for significance after Bonferroni correction was $p < 0.0041$. Test-retest reliability indices were obtained using the mean squared error for each measure of LI as well as the intra-class correlation coefficients (ICCs) on mGFP waveforms.

RESULTS

Cortical Responses to Rapidly Presented Stimuli

Figure 3 presents means and standard deviations of N100m and P200m amplitudes and latencies for ear, stimulus, pathway, and hemisphere. Data analysis was conducted on the amplitude and latency of N100m and P200m in response to the second stimulus

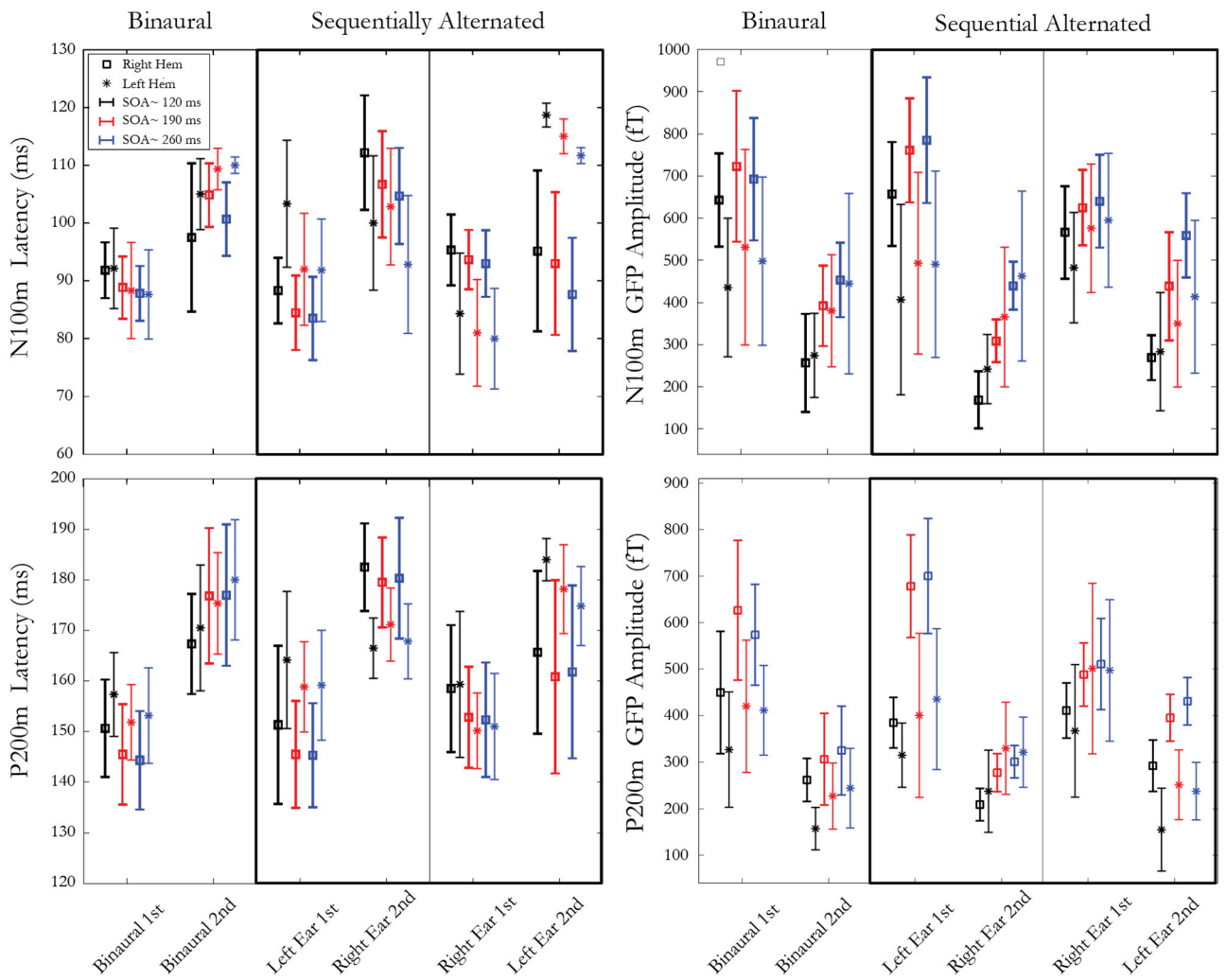


FIGURE 3 | Latencies (**left**) and amplitudes (**right**) of mGFP N100m (**top**) and P200m (**bottom**) components. Each panel represents 3 presentation conditions: 1 binaural condition and 2 sequentially alternated conditions (stimulus presented first at either left or right ear). Within each presentation condition, three SOAs (~120 ms, ~190 ms and ~ 260 ms) are used, resulting in two responses to both stimuli of the pair, recorded from both right and left hemispheres. Error bars denote standard deviations between participants.

of the pair. A repeated measure ANOVA was computed with these factors: hemisphere (right, left), presentation condition (binaural, sequentially alternated left ear first, sequentially alternated right ear first), and SOA (~120, ~190, ~260 ms). The effect of SOA was found to be significant for both amplitudes and latencies of N100m (Amp. $F(2,10)=46.48$, $p = 0.000009$, $\varepsilon = 0.58$; Lat. $F(2,10) = 7.30$, $p = 0.03$, $\varepsilon = 0.54$) and for P200m amplitude (Amp. $F(2,10) = 53.95$, $p = 0.000004$, $\varepsilon = 0.78$). *Post hoc* analysis for N100m and P200m Amp showed a significant increase in amplitude from SOA ~120 to SOA ~190 ms. The amplitude increased between SOA ~190 to SOA ~260 ms was only significant for N100m. A significant interaction was present between SOA and presentation condition for both N100m ($F(4,20) = 10.07$; $p = 0.001$, $\varepsilon = 0.60$) and P200m [$F(4,20) = 8.29$; $p = 0.004$, $\varepsilon = 0.60$] latencies. *Post hoc* analysis revealed a decrease in N100m response latency when SOA increased from ~120 to ~190ms ($p < 0.003$) and from ~120 to ~ 260 ms ($p < 0.02$) for both sequentially alternated presentation conditions, while this trend was absent in the binaural presentation conditions. For P200m, the only significant difference was between binaural presentation and right-left sequential for SOA ~260 ms. A significant interaction was observed between hemisphere and presentation condition for N100m [Lat. $F(2,10) = 41.78$, $p = 0.00001$, $\varepsilon = 0.75$] and for P200m [Amp. $F(2,10)=16.18$, $p = 0.0007$, $\varepsilon = 0.87$; Lat. $F(2,10)=14.60$, $p=0.001$, $\varepsilon = 0.82$]. For N100m latencies, *post hoc* analysis revealed shorter latencies in the right hemisphere compared to the left hemisphere when stimuli were presented binaurally ($p<0.04$). Moreover, pairwise comparisons revealed longer latency for the ipsilateral pathway compared to the contralateral pathway in the sequential stimulation mode for both N100m and P200m when the second stimulus of the pair

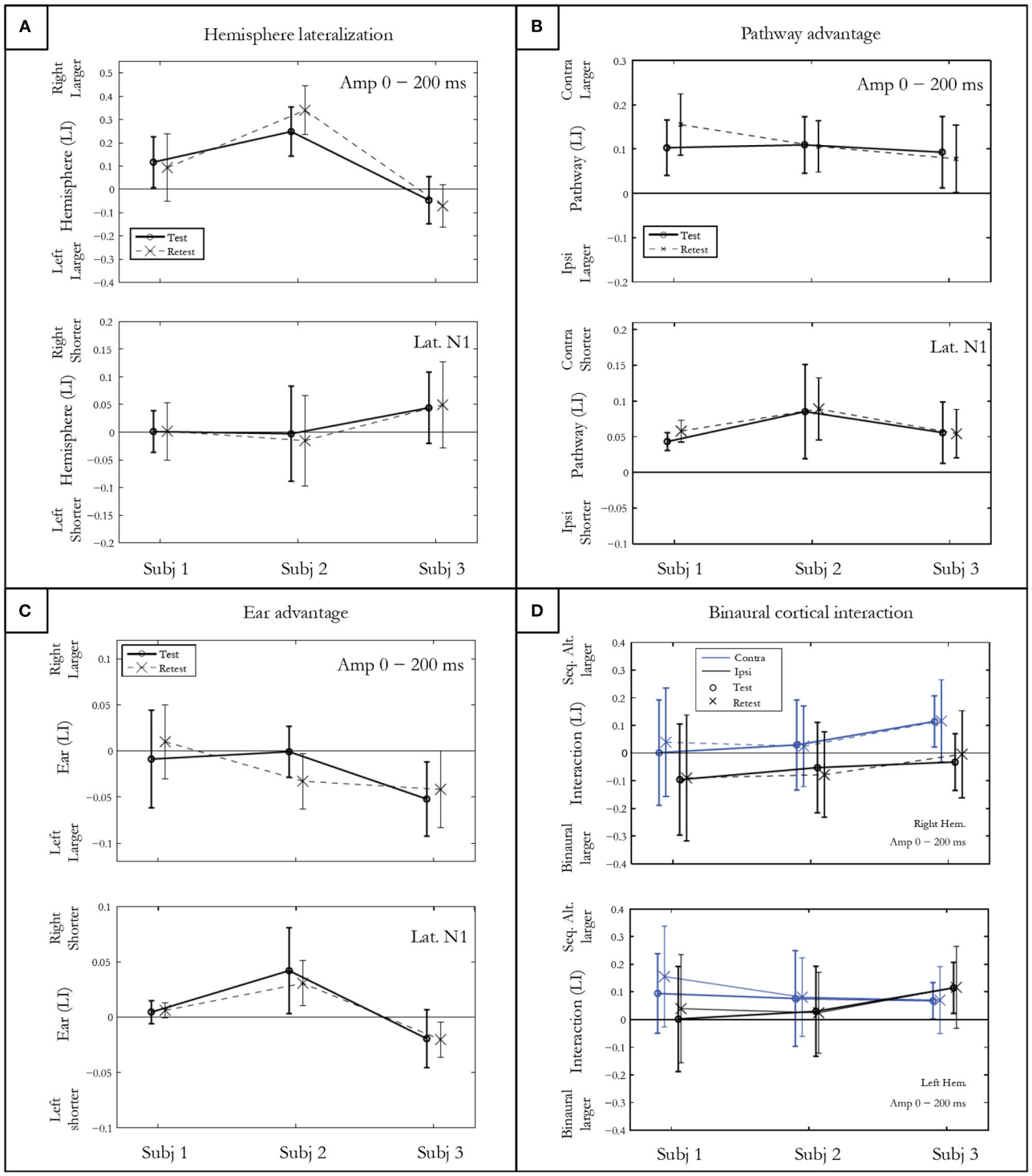


FIGURE 4 | Indices of hemispheric lateralization **(A)**, pathway advantage **(B)**, and ear advantage **(C)** for mGFP amplitudes in a 200 ms post onset window, and for N100m latency. Both test and retest conditions are shown. The binaural cortical interaction **(D)** is represented for the mean mGFP amplitude in the right and left hemisphere for the contra- and ipsi-lateral pathways. Error bars denote standard deviations between conditions for each participant.

was presented to the left ear ($p < 0.009$) while this difference was significant only for P200m ($p < 0.037$) when the second stimulus was presented to the right ear. The amplitude of P200m was also significantly larger in the left hemisphere when the second stimulus of the pair was presented to the right ear. Lastly, an interaction between hemisphere, SOA and presentation

condition was significant for N100m [$F(4,20) = 3.34$, $p = 0.02$, $\varepsilon = 0.68$]. *Post hoc* analysis revealed a significant difference between hemispheres for all SOAs in sequential presentation condition when the second stimulus of the pair was presented to the right ear ($p < 0.008$). The SOA $\sim$ 260 ms for the binaural presentation was the only other condition that showed a significant hemispheric difference.

Hemispheric Lateralization

The hemispheric lateralization index (LI) for response amplitude presented in **Figure 4A** shows intra-subject differences on the vertical abscissa, and inter-subject differences on the horizontal abscissa. Subject 1 presented a rightward, subject 2 a large rightward, and subject 3 a slightly leftward lateralization. The *t*-test, which allows comparing the hemispheric LI to 0, was significant for each subject ($p<0.001$) after Bonferroni correction. No differences in symmetrical activation were found for the latencies either for subject 1 ($p = 0.86$) or subject 2 ($p = 0.51$). However, significantly earlier latencies were found in the right hemisphere for subject 3 ($p = 0.0003$).

Pathway Advantage

The pathway LI calculated by contrasting contra- vs. ipsi-lateral pathway responses in the sequential conditions is represented in **Figure 4B**. After Bonferroni correction, significantly larger amplitudes and shorter latencies for the N100m and P200m were measured in the contra-lateral pathway for all subjects *($p < 0.0001$)*.

Ear Advantage

The statistical results of ear LI presented in **Figure 4C** indicated no significant amplitude difference between the activity elicited by the right and the left ear for subject 1 ($p = 0.96$) and for subject 2 ($p = 0.01$). A left ear advantage was observed for subject 3 for both amplitude ($p = 0.002$) and latency ($p = 0.002$).

Cortical Binaural Interaction (CBI)

Figure 4D shows the CBI for the three subjects in both hemispheres for contra- and ipsi-lateral pathways. The finding of a positive CBI LI indicates that the response recorded in the sequentially alternated condition is larger compared to the response in the ipsi-lateral pathway. CBI of different natures are observed for each subject. When collapsed across hemispheres, the *t*-test showed that CBI was close to significance only for subject 3 (subject 1: $p = 0.02$; subject 2: $p = 0.10$; subject 3: $p = 0.006$).

Test-Retest Reliability

Two different test-retest reliability measures were computed. First, the mGFP waveforms were compared for test and retest conditions by computing the intra-class correlation coefficients (ICCs) for the three subjects in a 250 ms window post onset. A mean ICC value larger than 0.75 for each subject (i.e., subject 1 = 0.78, subject 2 = 0.79; subject 3 = 0.84) demonstrated a good test-retest reliability.

Second, a test-retest index was calculated using the mean squared error (mean = 0.057; SD = 0.026) of all four indices presented in **Figure 4** (i.e., hemispheric lateralization, pathway advantage, ear advantage and CBI).

DISCUSSION

The central aim of this paper was to introduce the deconvolution of ears' activity (DEA) paradigm which disentangles the activity in both auditory cortices elicited by stimuli presented to both ears simultaneously or separately. In this paradigm, the LS deconvolution technique was applied to MEG data recorded using pairs of stimuli presented either binaurally or alternating sequentially (i.e., right-left and left-right). The DEA paradigm allowed the investigation of auditory information transfer from one specific ear to both auditory cortices. It could also be used to explore response lateralization, the strength of crossed auditory pathways and the response adaptation properties to auditory stimuli closely separated in time. Furthermore, it allowed for the investigation of non-linear processing in the brain and CBI, mainly caused by inhibition mechanisms (Imig and Brugge, 1978; Imig and Reale, 1981; Reite et al., 1981; Papanicolaou et al., 1990).

We demonstrated the feasibility and test–retest reproducibility of this non-invasive measure on 3 right-handed normal-hearing subjects. The case studies provided examples of different auditory processing characteristics at the cortical level, identifiable at the individual level. The inter-individual differences were detectable by assessment of the difference in response between experimental conditions. For example, hemispheric lateralization was assessed by computation of the LI calculated from the responses in each hemisphere. The CBI was investigated by contrasting binaural and monaural stimulation both in contra- and ipsi-lateral pathways. The results collected using the DEA paradigm allows an objective auditory processing characterization and the generation of an individual "auditory profile" in a relatively quick time (i.e., 25 min).

Experimental Results

The data recorded from three normal-hearing subjects confirmed that both ears were represented in each cortical hemisphere. However, differences in latency and amplitude were observed for each response to various conditions.

Beyond the idea proposed by Poeppel (2003) that sound processing in the brain is a bilateral phenomenon, the present study revealed inter-individual differences in the hemispheric lateralization of the cortical response. While two subjects showed a rightward hemisphere lateralization for response amplitude, the third subject had a leftward lateralization. These hemispheric asymmetries and specializations for processing auditory stimuli were also reported previously by Mäkel et al. (1993) and Jamison et al. (2006). The cerebral lateralization of the auditory cortical area however is still highly debated (Bishop, 2013; Scott and McGettigan, 2013).

For all subjects tested, the N100m was larger and approximately 10 ms shorter for the contra-lateral compared to the ipsi-lateral auditory pathway in the sequentially alternated conditions. These results are in agreement with several studies showing a contra-lateral dominance based on lateralization of

the N100m component (Pantev et al., 1986, 1998; Tiihonen et al., 1989; Woldorff et al., 1999).

Individual differences were also observed when comparing ear activity. Further research will need to investigate whether this objective measure of ear advantage is correlated with behavioral performance on a dichotic listening task such as the Dichotic Digits Test (Musiek, 1983).

The DEA paradigm allowed to investigate the suppression-type interaction and neural mechanisms underlying the processing of rapidly presented signals. As shown in **Figure 4D**, different binaural interactions were observed. Amplitudes of responses elicited in the sequentially alternated presentation condition were found to be either slightly larger, slightly smaller or of similar amplitude compared to the binaural presentation condition. Inter-subject differences were observed with different interactions depending on hemisphere and pathway involved. A MEG study using complex tones showed that responses to ipsi-lateral stimuli over the right auditory cortex are inhibited by the stimuli presented in the contra-lateral (left) ear (Brancucci et al., 2004).

Lastly, cortical responses to stimulus pairs separated by short SOAs allowed the study of the representation in the auditory cortex of stimuli presented closely together. The significant interactions between hemisphere, presentation condition, and SOA revealed by ANOVA indicate the complex binaural interactions occurring in the brain when processing rapidly presented stimuli.

We conclude that the DEA paradigm could represent a technique to study interesting properties of the central auditory system. Individual differences are of special interest as they provide an alternative characterization of the hearing profile of a person which could potentially be useful to for example objectively identify auditory processing disorder (APD) subjects. Using the LS deconvolution technique to separate overlapping ear activity in both auditory cortices, recorded MEG data can provide a measure for rapid temporal processing, response lateralization, auditory pathway and ear advantage, and CBI for rapidly presented sound stimuli. Such a test would allow studying the temporal acuity of the human auditory system when processing rapid changes in the acoustic signal. Moreover, it could provide insights concerning the flow of neural signals from the cochlea to the cerebral cortex. From a clinical perspective, tests are needed to better evaluate and understand the neurological characteristics of binaural processing occurring in the auditory system. Such tests could contribute to the diagnosis of neurodevelopment disorders, such as specific language impairment (SLI) or dyslexia where abnormal crossing pathways or the disability to process rapid auditory stimuli has been identified (Lamminmäki et al., 2012). However, further studies are needed to record normative data on normal hearing subjects, that can then be used as a benchmark to characterize other populations. Moreover, other complex sounds, such as speech syllables (using carefully selected jitter parameters), could be used in the future to investigate the influence of stimuli on binaural interaction mechanisms and lateralization of the response.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Macquarie University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FB confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was supported in part by: the HEARing CRC, established and supported under the Australian Cooperative Research Centers Program, an Australian Government Initiative, by the Australian Government Department of Health and by the Oticon Foundation.

ACKNOWLEDGMENTS

The author gratefully thank Bram Van Dun, Harvey Dillon, Catherine McMahon, Robert Cowan, Mark Seeto and Ramesh Rajan for their suggestions during the preparation of this manuscript.

REFERENCES

Bardy, F., Dillon, H., and Van Dun, B. (2014a). Least-squares deconvolution of evoked potentials and sequence optimization for multiple stimuli under low-jitter conditions. *Clin. Neurophysiol.* 125, 727–737. doi: 10.1016/j.clinph.2013.09.030

Bardy, F., Van Dun, B., and Dillon, H. (2014b). Mc,Mahon CM: Deconvolution of overlapping cortical auditory evoked potentials (caeps) recorded using short stimulus onset-asynchrony (soa) ranges. *Clin. Neurophysiol.* 125, 814–826. doi: 10.1016/j.clinph.2013.09.031

Bardy, F., Van Dun, B., and Dillon, H. (2015). "Bigger is better: Increasing cortical response amplitude via stimulus spectral complexity". *Ear Hear.* 36, 677–687. doi: 10.1097/AUD.0000000000000183

Bishop, D. V. (2013). Cerebral asymmetry and language development: cause, correlate, or consequence? *Science.* 2013, 340. doi: 10.1126/science.1230531

Brancucci, A., Babiloni, C., Babiloni, F., Galderisi, S., Mucci, A., Tecchio, F., et al. (2004). Inhibition of auditory cortical responses to ipsilateral stimuli during dichotic listening: Evidence from magnetoencephalography. *Eur. J. Neurosci.* 19, 2329–2336. doi: 10.1111/j.0953-816X.2004.03302.x

Fujiki, N., Jousmaki, V., and Hari, R. (2002). Neuromagnetic responses to frequency-tagged sounds: A new method to follow inputs from each ear to the human auditory cortex during binaural hearing. *J. Neurosc.i* 22, 201–204. doi: 10.1523/JNEUROSCI.22-03-j0003.2002

Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*. 24, 95–112. doi: 10.1007/BF02289823

Imig, T. J., and Brugge, J. F. (1978). Sources and terminations of callosal axons related to binaural and frequency maps in primary auditory cortex of the cat. *J. Comp. Neurol.* 182, 637–660. doi: 10.1002/cne.901820406

Imig, T. J., and Reale, R. A. (1981). Ipsilateral corticocortical projections related to binaural columns in cat primary auditory cortex. *J. Comp. Neurol.* 203, 1–14. doi: 10.1002/cne.902030102

Jamison, H. L., Watkins, K. E., Bishop, D. V., and Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex.* 16, 1266–1275. doi: 10.1093/cercor/bhj068

Kado, H., Higuchi, M., Shimogawara, M., Haruta, Y., Adachi, Y., Kawai, J., et al. (1999). Magnetoencephalogram systems developed at kit. *IEEE Trans. Appl. Supercond.* 9, 4057–4062. doi: 10.1109/77.783918

Lamminmäki, S., Massinen, S., Nopola-Hemmi, J., Kere, J., and Hari, R. (2012). Human robo1 regulates interaural interaction in auditory pathways. *J. Neurosci.* 32, 966–971. doi: 10.1523/JNEUROSCI.4007-11.2012

Mäkel,ä, J. P., Ahonen, A., Hämäläinen, M., Hari, R., Llmoniemi, R., Kajola, M., et al. (1993). McEvoy L, Salmelin R: Functional differences between auditory cortices of the two hemispheres revealed by whole-head neuromagnetic recordings. *Hum. Brain Mapp.* 1, 48–56. doi: 10.1002/hbm.460010106

Musiek, F. E. (1983). Assessment of central auditory dysfunction: the dichotic digit test revisited. *Ear Hear.* 4, 79–83. doi: 10.1097/00003446-198303000-00002

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia.* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Pantev, C., Lütkenhöner, B., Hoke, M., and Lehnertz, K. (1986). Comparison between simultaneously recorded auditory-evoked magnetic fields and potentials elicited by ipsilateral, contralateral and binaural tone burst stimulation. *Int. J. Audiol.* 25, 54–61. doi: 10.3109/00206098609078369

Pantev, C., Ross, B., Berg, P., Elbert, T., and Rockstroh, B. (1998). Study of the human auditory cortices using a whole-head magnetometer: left vs. right hemisphere and ipsilateral vs. Contralateral stimulation. *Audiol. Neurootol.* 3, 183–190. doi: 10.1159/000013789

Papanicolaou, A. C., Baumann, S., Rogers, R. L., Saydjari, C., Amparo, E. G., and Eisenberg, H. M. (1990). Localization of auditory response sources using magnetoencephalography and magnetic resonance imaging. *Arch. Neurol.* 47, 33–37. doi: 10.1001/archneur.1990.00530010041016

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3

Raicevich, G., Burwood, E., Dillon, H., Johnson, B. W., and Crain, S. (2010). Wide band pneumatic sound system for MEG. 20th International Congress on Acoustics. *Sydney: ICA* 2010, 1–5. Available online at: http://www.acoustics.asn.au/conference_proceedings/ICA2010/cdrom-ICA2010/papers/p570.pdf

Reite, M., Zimmerman, J. T., and Zimmerman, J. E. (1981). Magnetic auditory evoked fields: Interhemispheric asymmetry. *Electroencephalogr Clin. Neurophysiol.* 51, 388–392. doi: 10.1016/0013-4694(81)90102-4

Scott, S. K., and McGettigan, C. (2013). Do temporal processes underlie left hemisphere dominance in speech perception? *Brain Lang.* 127, 36–45. doi: 10.1016/j.bandl.2013.07.006

Tiihonen, J., Hari, R., Kaukoranta, E., and Kajola, M. (1989). Interaural interaction in the human auditory cortex. *Int. J. Audiol.* 28, 37–48. doi: 10.3109/00206098909081609

Uehara, G., Adachi, Y., Kawai, J., Shimogawara, M., Higuchi, M., Haruta, Y., et al. (2003). Multi-channel squid systems for biomagnetic measurement. *IEICE Trans. Electron.* 86, 43–54. Available online at: https://search.ieice.org/bin/summary.php?id=e86-c_1_43

Woldorff, M. G., Tempelmann, C., Fell, J., Tegeler, C., Gaschler-Markefski, B., Hinrichs, H., et al. (1999). Lateralized auditory spatial perception and the contralaterality of cortical processing as studied with functional magnetic resonance imaging and magnetoencephalography. *Hum. Brain Mapp.* 7, 49–66. doi: 10.1002/(SICI)1097-0193(1999)7:1<49::AID-HBM5>3.0.CO;2-J

Yetkin, F. Z., Roland, P. S., Christensen, W. F., and Purdy, P. D. (2004). Silent functional magnetic resonance imaging (fmri) of tonotopicity and stimulus intensity coding in human primary auditory cortex. *Laryngoscope.* 114, 512–518. doi: 10.1097/00005537-200403000-00024

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 03 August 2022
DOI 10.3389/fneur.2022.928158

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Sraboni Chaudhury,
University of Michigan, United States
Christoffer Hatlestad-Hall,
Oslo University Hospital, Norway

*CORRESPONDENCE
Axelle Calcus
axelle.calcus@ulb.be

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 25 April 2022
ACCEPTED 13 July 2022
PUBLISHED 03 August 2022

CITATION
Calcus A, Undurraga JA and Vickers D (2022) Simultaneous subcortical and cortical electrophysiological recordings of spectro-temporal processing in humans.
Front. Neurol. 13:928158.
doi: 10.3389/fneur.2022.928158

Simultaneous subcortical and cortical electrophysiological recordings of spectro-temporal processing in humans

Axelle Calcus[1,2,3]*, Jaime A. Undurraga[4,5] and Deborah Vickers[1,6]

[1]Department of Speech, Hearing and Phonetic Sciences, University College London, London, United Kingdom, [2]Laboratoire des Systèmes Perceptifs, Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL University, CNRS, Paris, France, [3]Center for Research in Cognitive Neuroscience, Université Libre de Bruxelles (ULB), Brussels, Belgium, [4]Department of Linguistics, Macquarie University, Sydney, NSW, Australia, [5]Interacoustics Research Unit, Technical University of Denmark, Lyngby, Denmark, [6]SOUND Lab, Cambridge Hearing Group, Department of Clinical Neurosciences, Herchel Smith Building for Brain and Mind Sciences, Cambridge, United Kingdom

Objective assessment of auditory discrimination has often been measured using the Auditory Change Complex (ACC), which is a cortically generated potential elicited by a change occurring within an ongoing, long-duration auditory stimulus. In cochlear implant users, the electrically-evoked ACC has been used to measure electrode discrimination by changing the stimulating electrode during stimulus presentation. In addition to this cortical component, subcortical measures provide further information about early auditory processing in both normal hearing listeners and cochlear implant users. In particular, the frequency-following response (FFR) is thought to reflect the auditory encoding at the level of the brainstem. Interestingly, recent research suggests that it is possible to simultaneously measure both subcortical and cortical physiological activity. The aim of this research was twofold: first, to understand the scope for simultaneously recording both the FFR (subcortical) and ACC (cortical) responses in normal hearing adults. Second, to determine the best recording parameters for optimizing the simultaneous capture of both responses with clinical applications in mind. Electrophysiological responses were recorded in 10 normally-hearing adults while they listened to 16-second-long pure tone sequences. The carrier frequency of these sequences was either steady or alternating periodically throughout the sequence, generating an ACC response to each alternation—the alternating ACC paradigm. In the "alternating" sequences, both the alternating rate and the carrier frequency varied parametrically. We investigated three alternating rates (1, 2.5, and 6.5 Hz) and seven frequency pairs covering the low-, mid-, and high-frequency range, including narrow and wide frequency separations. Our results indicate that both the slowest (1 Hz) and medium (2.5 Hz) alternation rates led to significant FFR and ACC responses in most frequency ranges tested. Low carrier frequencies led to larger FFR amplitudes, larger P1 amplitudes, and N1-P2 amplitude difference at slow alternation rates. No significant relationship was found between subcortical and cortical response amplitudes, in line with different generators and processing levels across the auditory pathway. Overall, the alternating ACC paradigm can be used to measure sub-cortical and cortical responses as

indicators of auditory early neural encoding (FFR) and sound discrimination (ACC) in the pathway, and these are best obtained at slow alternation rates (1 Hz) in the low-frequency range (300–1200 Hz).

KEYWORDS

auditory change complex, frequency following response (FFR), cortical auditory evoked potential (CAEP), brainstem, auditory processing (AP)

Introduction

Auditory evoked potentials are electrophysiological responses providing information on underlying neurophysiological function of structures in the auditory pathways. They are useful in audiological diagnostic assessment and for populations who cannot provide reliable responses to sounds. Electrophysiological responses are routinely used to explore the viability of different stages of the auditory pathway, from otoacoustic emissions, recording responses from the organ of Corti, through to cortical auditory evoked potentials, showing responsiveness of higher brain centers [e.g., (1–3)]. However, measurements can be time consuming particularly if responses to multiple stimulus parameters are required, for example, when recording responses to different sound frequencies. Measurement of responses at different stages in the auditory pathway allow for identification of site of lesion or loci of sound transmission difficulties for individuals with atypical sound processing abilities. The best approach to understanding sound processing at different stages of the auditory pathway is to measure concurrent responses at different sites.

Knebel et al. (4) have suggested that the combination of speech auditory brainstem responses (ABRs) and cortical responses to the same stimuli can be used to understand the inter-relationship between the generators of the different potentials and also the interaction between different brain regions. Musacchia et al. (5) recorded simultaneous speech ABRs and cortical onset responses (ORs) to /da/ stimuli to determine if musicians compared to non-musicians exhibited differences in ABRs and associated cortical ORs. They found that stronger ABRs to periodicity was associated with shorter latency of the OR and that musicians showed larger ABR amplitudes and shorter OR latencies than non-musicians.

Krishnan et al. (6) reported an approach for simultaneously acquiring the brainstem frequency following response (FFR) and cortical evoked pitch responses. The FFR is a sustained response evoked by the neurons in the brainstem able to track, on a cycle-by-cycle basis, the frequency of the periodic stimuli—phase locking. Pitch salience was varied by adapting the number of stimulus periodicity in an iterated rippled noise. The cortical responses to pitch were measured for stimulus onset (OR) and in response to a change in the pitch salience of the stimulus [auditory change complex, ACC (7)]. The ACC is a cortical response evoked by a change in an ongoing stimulus, with a fronto-central topographic distribution when referenced to the mastoid (7, 8). Morphologically, the ACC is characterized by a series of peaks usually within 50 and 250 ms after the stimulus onset – P1-N1-P2 response – and is measured using EEG electrodes typically placed in fronto-central regions. The latency, amplitude and morphology of the peaks (P1, P2) and trough (N1) are used as indicators of neural synchrony and maturation of the auditory pathways. Contrary to the OR, in which response characteristics have not been associated with pitch salience, the magnitude and latency of the ACC show a clear relation with pitch perception. For example, Mathew et al. (9) observed strong associations between the ACC and the ability to discriminate between stimulating electrodes in cochlear implant (CI) users. There is evidence that ACC responses to change in stimulus characteristics relate to speech perception abilities: Han and Dimitrijevic (10) showed a relationship between the N1 latency for the ACC to modulation detection and speech perception. However, behavioral discrimination seems to be best predicted by combining both subcortical (brainstem FFRs) and cortical (ACC) responses (6) to improve understanding of the processing in different auditory regions.

This approach for simultaneous measurement of the brainstem FFR and the ACC is of interest here. By means of a modified ACC paradigm, in which the fundamental frequency (F0) of an otherwise continuous stimulus, is periodically alternated—the alternating ACC (8, 11) - we investigate spectro-temporal processing in subcortical and cortical regions. The goals of the current research were to determine if brainstem FFRs and cortical ACC responses could be evoked and recorded simultaneously to periodic frequency alternations in a stimulus, allowing multiple measurements across the auditory pathway to investigate F0 processing. We varied parameters to understand the optimal approach for maximizing responses. This research is directed at developing electrophysiological measures that can help to understand perceptual capabilities in normal hearing, hearing impairment, and listening with a CI. In particular, we aim to develop electrophysiological paradigms that can be efficiently used to objectively measure discrimination and temporal processing abilities, hence allowing for identification of spectral regions where signal transmission/processing might

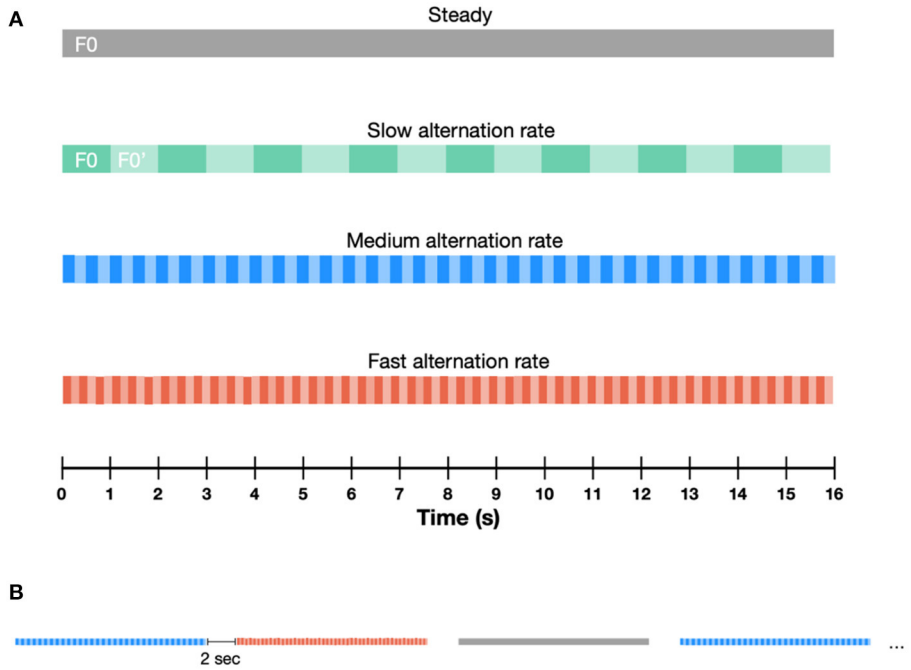


FIGURE 1
Schematic illustration of the paradigm containing different types of auditory sequences. **(A)** The F0 of these sequences was either steady or alternating between F0 and F0′, throughout the sequence. Three alternation rates were presented: slow (1 Hz), medium (2.5 Hz) and fast (6.5 Hz). **(B)** Sequence duration was fixed at 16 s. Sequences were separated by 2 s-long pauses.

be impaired. Such measures can also be used to evaluate phase locking and adaptation in the auditory system (8).

Here, we investigate subcortical and cortical functional integrity to periodic changes in F0 occurring at several alternating rates. F0s were chosen to correspond to center frequencies of CI electrodes, ranging from 300 to 3,000 Hz, for future application with CI users (using Advanced Bionics frequency allocation table). Alternation rates varied from 1 to 6.5 Hz, hence being close to the syllabic rate. This paradigm aimed to identify the condition that would provide the most information in a minimum amount of time, in the objective of developing a reliable, fast clinical tool.

Methods

Participants

Ten young (21–27 years old, mean 23.66 years, 2 males) English speakers participated in this study. All participants had normal hearing defined as air-conducted pure-tone thresholds of 25 dB HL or better at octave frequencies from 0.25 to 8 kHz in both ears. None of the participants reported a history of neurological disorders. All participants provided written consent as approved by the UCL Research Ethics Committee (SHaPS-2018-DV-028) and were compensated for their time.

Stimuli

Participants were presented with 16-second-long pure tone sequences. The fundamental frequency (F0) of these sequences was either steady or alternating throughout the sequence. In the steady sequences, F0 was set to 320 Hz. In the alternating sequences, both the alternation rate and the F0 varied parametrically. A schematic illustration of the paradigm is provided in Figure 1. We investigated three alternation rates (1, 2.5 and 6.5 Hz) and seven F0 changes, covering the low- (300–1,320 Hz), mid- (1,320–3,120 Hz) and high- (2,620–3,120 Hz) frequency range, with varying separations between the lower and higher F0 within each frequency range (F0 and F0′, respectively). Each F0 alternating condition consisted of two frequency pairs alternating periodically at a given alternating rate. In the low frequency range (300–1,320 Hz) F0 alternated between 320–340 Hz, 320–480 Hz, 320–720 Hz, and 320 −1,320 Hz. In the mid-frequency range (1,320–3,120 Hz), F0 alternated between 1,320–1,520 Hz and 1,320–3,120 Hz, whilst in the high-frequency range, F0 alternated between 2,620–3,120 Hz. The range of F0 were selected to cover important speech frequency range. Stimuli were presented at 75 dB (A), with alternating polarities to minimize stimulus artifacts. Sound calibration was performed separately for the low-, mid- and high-frequency ranges, as an intensity average over the whole duration of the sequences.

TABLE 1 Summary of the stimulation metrics for all three alternation rate, at one F0 change.

Alternation rate	Number of sequences	Number of F0/F0′ iterations	Duration
6.5 Hz	5	520	1.3 min
2.5 Hz	12	480	3.2 min
1 Hz	30	480	8 min

Given that we presented seven F0 changes (see Stimuli), number of sequences and duration must be multiplied by 7 to provide total duration.

Sequences were presented in random order, separated by a 2 second inter-stimulus interval. Participants were presented with a total of 336 sequences (total recording time: 100 min), over two sessions that were scheduled no more than 2 weeks apart. Note that there was no significant difference in the number of rejected epochs during the first and second recording session [subcortical data: $t_{(9)} = -1.58$, $p = 0.148$; cortical data: $t_{(9)} = -0.58$, $p = 0.574$].

The number of sequences in each F0 condition was equalized across alternation rates in order to generate approximately the same number of iterations of the F0 and F0′ tones constituting sequences (see Table 1).

Recording parameters

Participants watched a muted movie with subtitles while seated comfortably in a double-walled, electrically shielded soundproof booth.

Stimuli and trigger signals were generated using a custom interface programmed in MATLAB, and delivered diotically using a external soundcard (RME FireFace UC, 44.1 kHz) connected to a custom-made trigger box which separated the two channels and simultaneously sent the trigger to the BioSemi system and the stimuli to electrically shielded ER-2 insert earphones (Intelligent Hearing Systems, Miami, FL).

Electrophysiological responses were collected using a BioSemi ActiveTwo system at a sampling rate of 8,192 Hz from 32 scalp electrodes positioned in the standard 10/20 configuration. Additional electrodes were placed on each mastoid; recordings were re-referenced offline to the average of activity at the mastoid electrodes.

Subcortical analyses

Epochs used to analyse subcortical FFRs were obtained by applying a band-pass filter (200–4,000 Hz) to the EEG data recorded at Cz, epoching the data 0–16 s relative to target onset, and averaging across epochs. Averaged mastoids to vertex (Cz) is a commonly used electrode montage (12). The average response was transformed to the frequency-domain (FFT of 131072 points) at a resolution of 0.0625 Hz. Trials exceeding ± 100 μ at Cz or Fz were excluded, leading to an average of 2% rejected trials.

The frequency peak was computed as the highest amplitude within 1 Hz centered around the target frequencies of a given sequence. Spectral noise floor was computed as the mean amplitude within 10 Hz surrounding the target frequencies (5 Hz on each side, excluding 5 immediately adjacent bins).

Cortical analyses

Evoked potentials of cortical origin were obtained by band-pass filtering (0.5–35 Hz) the EEG waveforms recorded at electrode C3, C4, Cz (vertex of the head), F3, F4 and Fz at 35 Hz, and creating epochs lasting −0.5 to 16 s relative to each target tone onset time. Fronto-central electrodes were chosen because they are thought to provide the most reliable estimates of both FFR and ACC measures (7, 13). Epochs were baseline corrected using the mean value from −100 to 0 ms. Trials exceeding ± 100 μ at Cz or Fz were excluded, leading to an average of 18.14% rejected trials.

To obtain the transient response, the magnitude of the auditory-evoked P1, N1 and P2 for each participants' set of data was computed as the mean amplitude in a fixed time window of 30–90, 75–150, and 150–290 ms respectively, after each alternation of frequency within every sequence type. The time windows have been selected based on visual inspection of the individual ERP responses, and are coherent with the typical latencies for each peak (14). To obtain the frequency response, data were epoched using a time window of 0 to 16 s relative to each sequence onset time.

Statistical analyses

The aim of the first subcortical analysis was to determine whether the FFR responses were significantly above the noise floor. One outlier whose EEG responses were more than 3 S.D. above the interquartile range was excluded from the analyses of the subcortical measures, and has also been removed from the grand average plots (Figure 2) and boxplots (Figure 3). A linear mixed-effects (LME) model was used [lme4 package of R; (15)] to determine whether overall measurement type (i.e., target frequency peak or spectral noise floor), alternation rate (1, 2.5 or 6.5 Hz), condition (320 vs. 340 Hz, 320 vs. 480 Hz, 320 vs. 720 Hz, 320 vs. 1,320 Hz, 1,320 vs. 1,520 Hz, 1,320 vs. 3,120 Hz, and 2,620 vs. 3120 Hz), and F0 category (F0 or F0′), or any of their four-, three- and two-way interactions predicted the amplitude of the response. Subsequently, a LME was conducted to determine whether alternation rate, condition, and F0 category

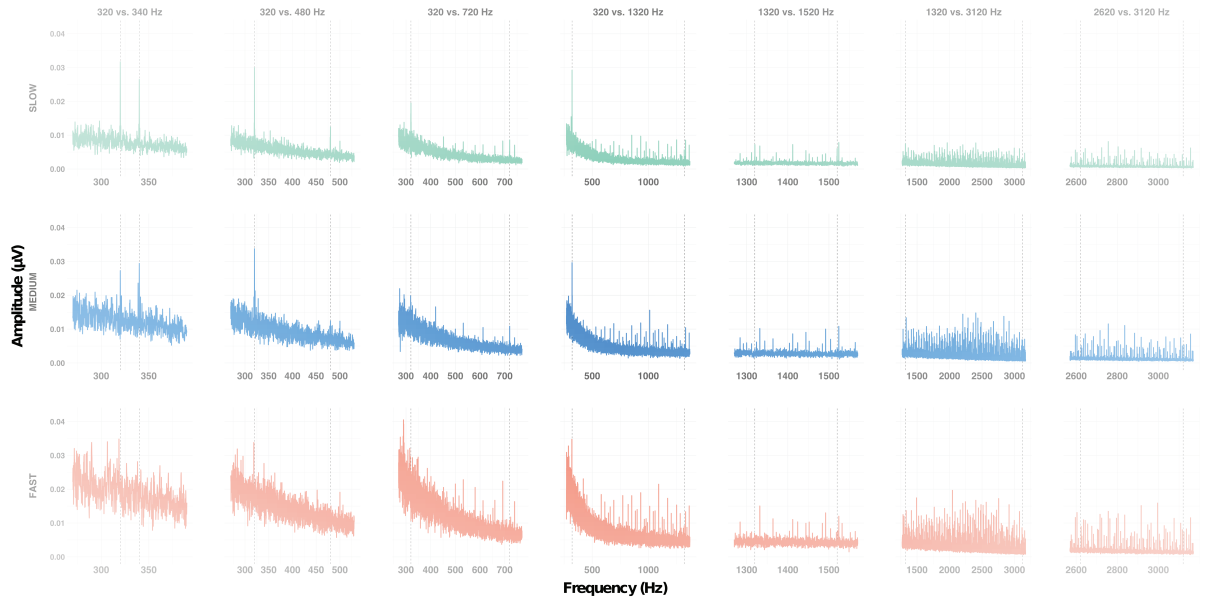


FIGURE 2
Frequency response of the subcortical grand average responses at Cz, for each condition (columns) at each alternation rate (rows). Vertical dotted lines indicate the expected frequencies for each condition.

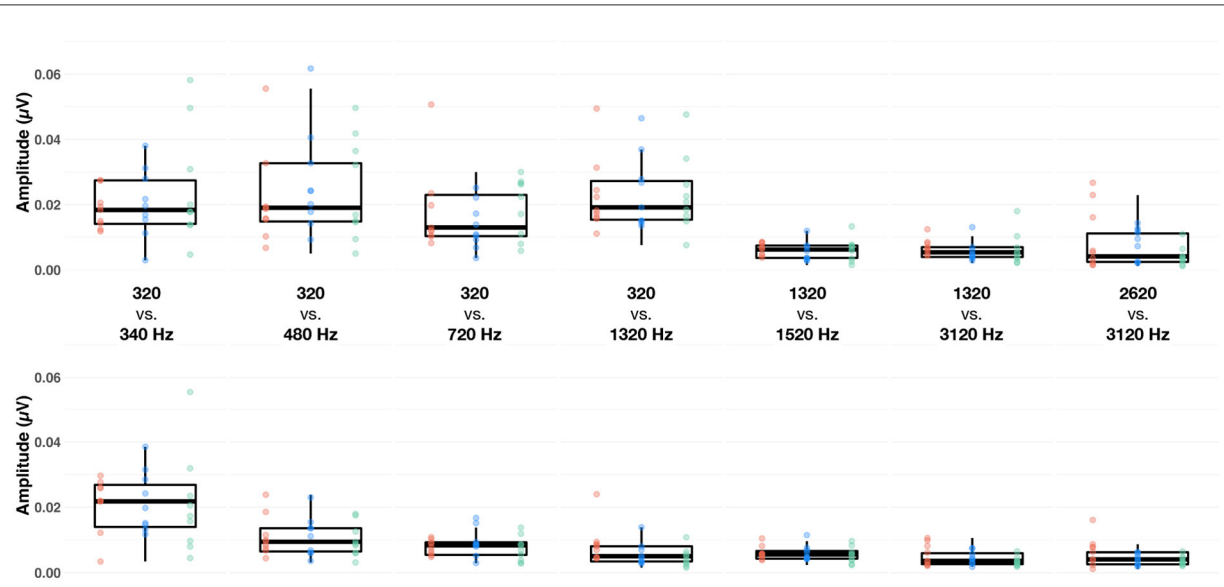


FIGURE 3
Boxplots of baseline-corrected amplitude (μV) of the subcortical response evoked by the F0 (upper row) and F0′ (lower row) in each of the seven conditions (columns), at fast (red dots), medium (blue dots) and slow (green dots) alternating rates, recorded at Cz. The whiskers indicate values that fall within 1.5 times the interquartile range. Dots falling outside the whiskers are outliers.

or any of their three- or two-way interactions significantly predicted the amplitude of the FFR at the target peak. In all models, the factor listener was used as a random intercept. Only the significant predictors are reported in the results section.

Visual inspection of the cortical measures (Figure 4, top panel) suggested that, as the alternation rate increased, only the P1 remained visible. This is due to the fact that, in the fast alternation condition (6.5 Hz), the ACC evoked by the new F0 started 150 ms after the previous F0, hence leading to an

overlap between the P1 elicited by the new F0 and the N1-P2 of the previous sound. Therefore, statistical analyses of the cortical measures were run in two steps. First, we used an LME model to determine whether alternation rate (1, 2.5 or 6.5 Hz), condition, recording electrode and F0 range (F0 or F0′) significantly predicted the amplitude of P1. Next, we fed a LME with the same factors to determine if these could predict the N1-P2 amplitude.

The correlation between brainstem and cortical measures was investigated using Pearson correlation coefficient (r).

Results

Subcortical measure (FFR)

First, we set out to determine whether the amplitude of the FFR evoked by both F0 and F0′ within a sequence was significantly above the noise floor (Figure 2). The LME model including the interaction between F0 category × condition × measurement type interaction [F (6, 706) = 6.63, $p < 0.001$, $\eta_p^2 = 0.05$] was significant. Overall, the amplitude of the target frequency peak was always larger than amplitude of spectral noise floor, i.e., positive signal-to-noise ratio [SNR; $F(1, 706) = 559.49$, $p < 0.001$, $\eta_p^2 = 0.44$]. However, the magnitude of this effect was variable across conditions. As shown in Supplementary Figure 1, the SNR was larger for F0s in conditions 320 vs. 340 Hz, 320 vs. 480 Hz, 320 vs. 720 Hz, 320 vs. 1320 Hz than in the remaining conditions (1,320 vs. 1,520 Hz, 1,320 vs. 3,120 Hz, and 2,620 vs. 3,120 Hz). The SNR was larger for high F0s in condition 320 vs. 340 Hz than in all remaining conditions. To account for the differences in SNR in further analyses, we computed the baseline-corrected amplitude as the difference between target frequency peak and spectral noise floor (Figure 3).

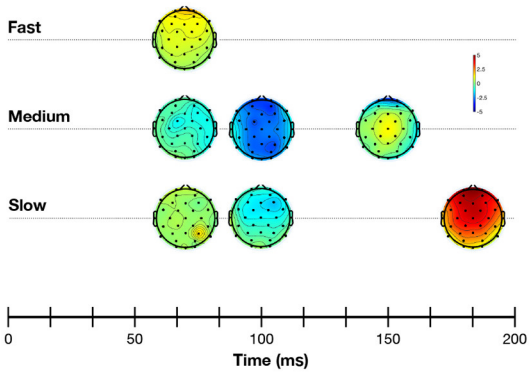


FIGURE 5
Voltage maps showing mean cortical activity during the 50–200 ms post-stimulus time window. Negative values are shown in blue, and positive values in red.

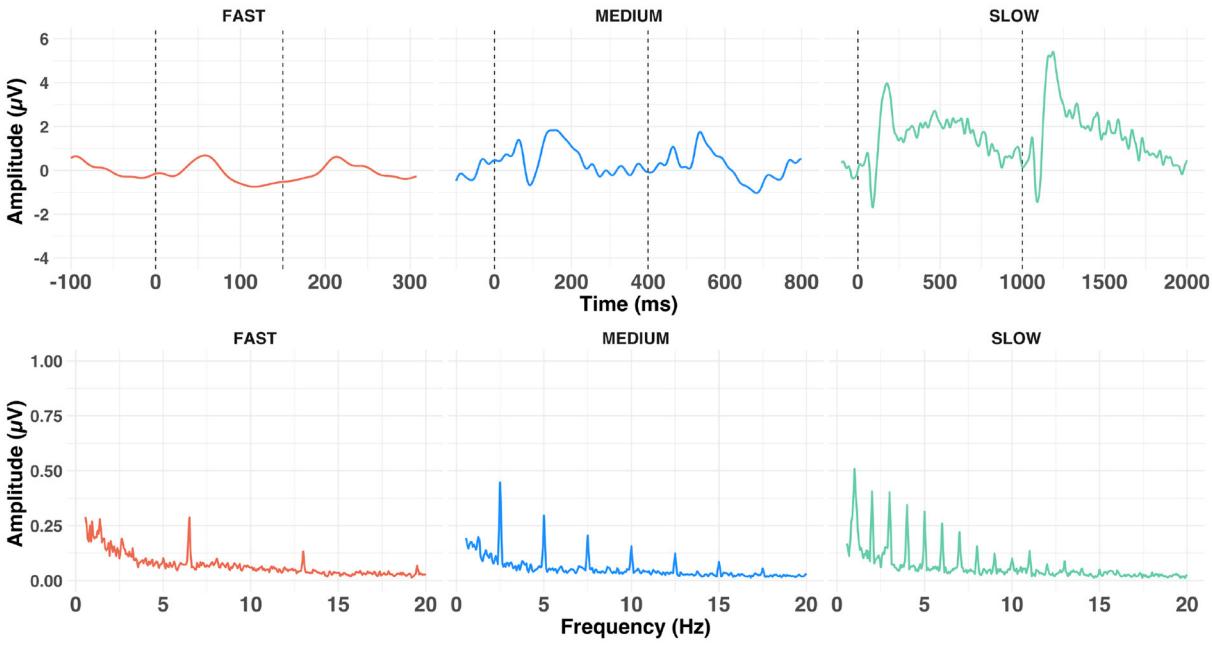


FIGURE 4
Average time- (upper row) and frequency- (lower row) domain representation of the grand average waveforms (at Cz) of the cortical response to a-ACC stimuli aggregated across conditions, presented at fast (left), medium (middle) and slow (right) alternating rates.

Next, we sought to identify factors that influenced the amplitude of the FFR. A LME model indicated that only the factor *condition* was significant [$F_{(6,404)} = 3.66$, $p = 0.001$, $\eta_p^2 = 0.05$]. Bonferroni-corrected *post-hoc t*-tests indicated that amplitude of the FFR evoked in both 320 vs. 340 Hz and 320 vs. 480 Hz conditions was significantly larger than that evoked in the 2,620 vs. 3,120 Hz condition (both *ps* < 0.05). This suggests that, irrespective of the alternation rate, FFR amplitude is larger for low- than mid- or high- frequency range (Figure 3).

Cortical measures

Figure 4 shows the grand average response evoked at Cz at each of the alternation rates in the time- and frequency-domain. Time-domain traces show the morphology of the response transitioned from a transient P1-N1-P2 waveform (Figure 4, slow condition) to a steady-state cortical response (Figure 4, fast condition). Voltage maps are illustrated in Figure 5. This is evident in the frequency domain plots where the spectrum transitioned from having multiple peaks (integer number of the slow and medium alternating rates) to an almost unimodal frequency peak at the fast alternating rate. Note that the morphology of the fast alternating may arise from the overlap of ACC responses leading to the steady-state sinusoidal morphology. Bearing this in mind, we will refer to P1 and N1 as the maximum and minimum of the time-domain response in the fast condition, respectively.

An LME model applied to time-domain responses indicated that P1 amplitude was significantly affected by factors alternation rate and condition, as well as their two-way interaction [respectively: $F_{(2,2,490)} = 41.02$, $p < 0.001$, $\eta_p^2 = 0.03$; $F_{(2,2,490)} = 5.93$, $p < 0.001$, $\eta_p^2 = 0.00$; $F_{(12,2,490)} = 3.84$, $p < 0.001$, $\eta_p^2 = 0.02$]. Bonferroni-corrected *post-hoc t*-tests were used to decompose the alternation rate × condition interaction (Figure 6). P1 amplitude did not vary with condition at fast alternation rates (all *ps* > 0.10). At medium alternation rates, it was significantly larger at 320 vs. 340 Hz than any other condition (all *ps* < 0.05). At slow alternation rates, P1 amplitude was significantly smaller in 320 vs. 480 Hz than any other condition (all *ps* < 0.05). Note that overall, P1 amplitude was significantly larger at slow than medium ($p = 0.020$) alternation rate, and at medium than fast alternation rate ($p < 0.001$). This suggest that a slow alternation rate is optimal to elicit a large P1, except in the 320 vs. 480 Hz condition.

Similarly, we investigated the effect of different parameters on N1-P2 amplitudes. A LME model revealed that alternation rate, condition and EEG recording electrode, as well as the alternation rate × condition interaction were significant [respectively: $F_{(1,1,648)} = 834.68$, $p < 0.001$, $\eta_p^2 = 0.34$; $F_{(6,1,648)} = 14.26$, $p < 0.001$, $\eta_p^2 = 0.05$; $F_{(5,1,648)} = 7.56$, $p < 0.001$, $\eta_p^2 = 0.02$; $F_{(6,1,648)} = 8.61$, $p < 0.001$, $\eta_p^2 = 0.03$]. N1-P2 was significantly smaller at C3 and C4 than at F3, F4 and Fz (all *ps* <0.05). Bonferroni-corrected *post-hoc t*-tests were used to decompose the alternation rate × condition interaction. At medium alternation rates, N1-P2 amplitude observed in conditions 320 vs. 1,320 Hz and 1,320 vs. 3,120 Hz were significantly smaller than observed in conditions 1,320 vs. 1,520 Hz and 2,620 vs. 3,120 Hz, respectively (both *ps* $<$ 0.05). At slow alternation rates, N1-P2 amplitude was smaller in condition 320 vs. 340 Hz than in all other conditions (all ps $<$ 0.01) except in 1,320 vs. 3,120 Hz ($p = 0.445$). On the contrary, N1-P2 amplitude was larger in condition 320 vs. 480 Hz than both conditions 1,320 vs. 1,520 Hz and 1,320 vs. 3,120 Hz (both *ps* < 0.05). Last, N1-P2 amplitude was larger in condition 2,620 vs. 3,120 Hz than all other conditions (all *ps* < 0.05), except 320 vs. 480 Hz ($p = 0.085$). Note that, similarly to P1 amplitude, N1-P2 amplitude was significantly larger at slow than medium

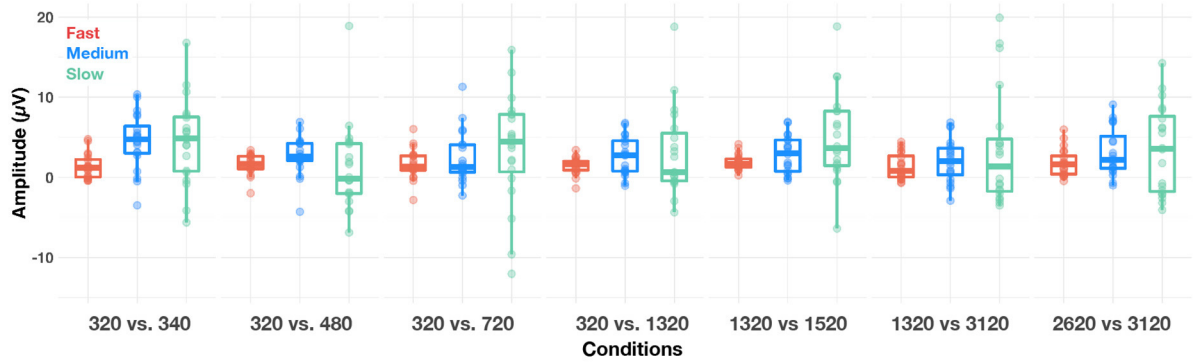


FIGURE 6
Boxplots of amplitude of the P1 (cortical) response evoked in each condition, at fast (red dots), medium (blue dots) and slow (green dots) alternating rates, recorded at Cz. The whiskers indicate values that fall within 1.5 times the interquartile range. Dots falling outside the whiskers are outliers.

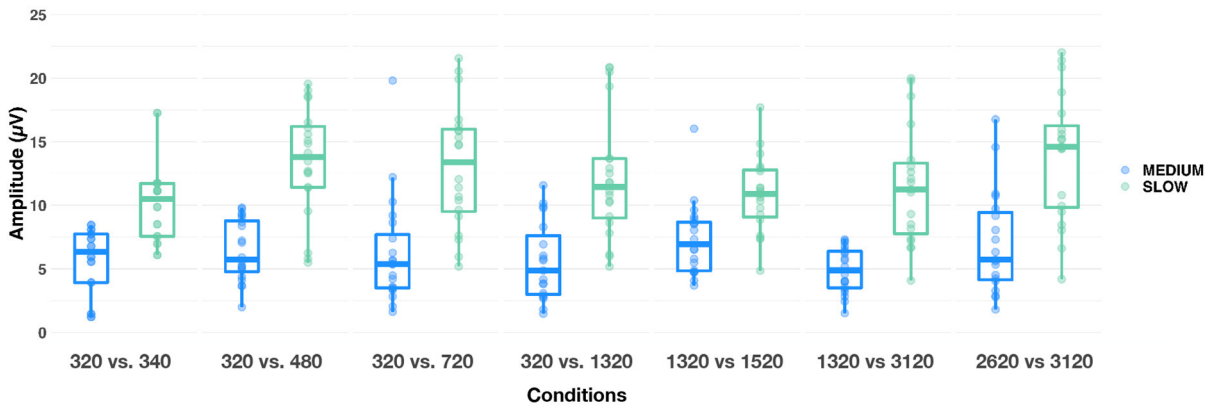


FIGURE 7
Boxplots showing N1-P2 amplitude responses evoked in each condition, at medium (blue) and slow (green) alternating rates. The whiskers indicate values that fall within 1.5 times the interquartile range. Dots falling outside the whiskers are outliers.

alternation rate ($p < 0.0001$). Together, this suggest that a slow alternation rate might not influence the amplitude of the subcortical response (see above), but appears to be the optimal candidate to elicit large transient cortical responses.

As an exploratory follow-up, we sought to determine whether increasing frequency separation between F0 and F0′ led to a larger N1-P2 amplitude difference. This analysis was only conducted on the four conditions where F0 = 320 Hz. A LME model revealed that alternation rate, condition and alternation rate × condition interaction were significant [respectively: $F_{(1,453)} = 250.9$, $p < 0.001$, $\eta_p^2 = 0.36$; $F_{(3,453)} = 7.48$, $p < 0.001$, $\eta_p^2 = 0.05$; $F_{(3,453)} = 7.53$, $p < 0.001$, $\eta_p^2 = 0.05$]. Bonferroni-corrected *post-hoc* comparisons failed to show significant amplitude differences between conditions at the medium alternation rate (all *p*s >0.50, see first 4 Conditions in Figure 7). However, at the slow alternation rate, the 320 vs. 480 Hz condition led to significantly larger N1-P2 amplitude differences than all three other conditions (all *p*s ≤ 0.01). N1-P2 was also significantly larger in the 320 vs. 720 Hz condition than in the 320 vs. 340 Hz ($p < 0.01$). No other comparisons were statistically significant (*p*s > 0.10).

Relationship between subcortical and cortical measures

To investigate the relationship between brainstem and cortical responses we computed the correlation between P1 amplitude and FFR amplitude, as well as between N1-P2 amplitude difference and FFR amplitude. After aggregating conditions for each of the three alternation rates, none of the correlations were found to be significant (all *p*s >0.10), see Figure 8. Similarly, there was no significant correlation between the amplitude of either F0 or F0′ subcortical response and amplitude of the ACC (all *p*s >0.10).

Discussion

The aim of this study was to identify stimulus parameters that would maximize simultaneous recording of subcortical (FFR) and cortical (OR) responses to the alternating ACC. Using this paradigm, we were able to measure significant cortical ACC and subcortical FFRs using the same stimuli. The alternating ACC maximizes data collection efficiency because each stimulus change produces a response for averaging and time is not wasted in dead periods between stimulus presentation.

The cortical and subcortical responses demonstrated different patterns across frequency range (conditions), frequency differences and alternation rate. Using a repeated-measures design (n = 10), it appears that the optimal condition for simultaneous subcortical and transient cortical recording was slowly alternating (1 Hz) between either 320 and 340 Hz or 320 and 480 Hz (see Figure 4 upper row, Figures 6, 7). Subcortical FFRs were overall larger in the low frequency range, and for F0 than F0′, consistent with more robust phase-locking at lower than higher frequencies (see Figures 2, 3). All transient cortical measures were larger at slower alternation rates, consistent with adaptation to repeating stimuli in the human auditory cortex (16). The choice of F0 conditions might depend upon the ACC response of interest. To maximize P1 amplitude, one might prefer the 320 vs. 340 Hz condition rather than 320 vs. 480 Hz condition, which elicited the smallest P1 response. However, 320 vs. 480 sequences elicited the largest N1-P2 difference. To our knowledge, this is the first study that parametrically explored auditory stimulation for optimizing recording parameters. Further studies replicating this finding

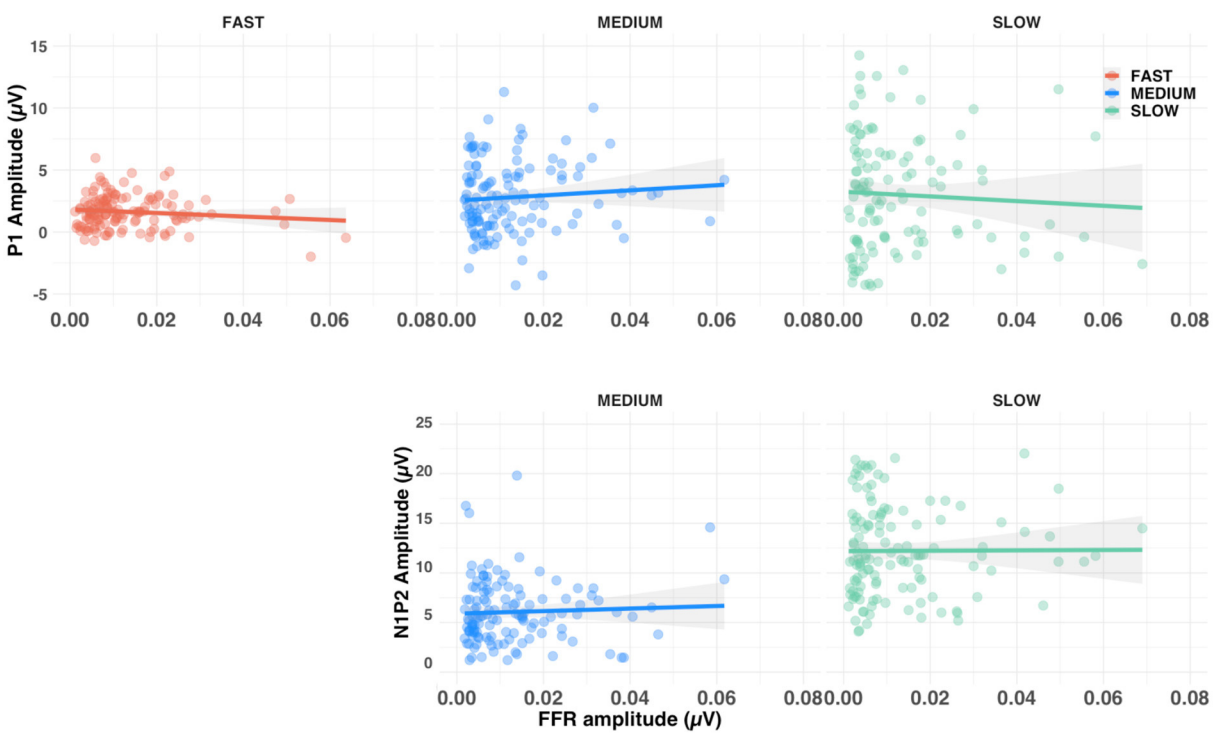


FIGURE 8
Scatterplots of the relationship between FFR amplitude and both P1 amplitude (upper row) and N1-P2 amplitude difference (lower row). Shaded lines represent the 95% confidence interval around the linear regression. Note that the fast-alternating rate did not allow enough time to elicit an N1-P2, which are not shown.

on larger sample sizes would be useful both for researchers and clinicians.

A previous study investigated the use of several presentation schemes to measure the ACC to frequency changes (17). In their study, the maximum time interval between alternations was 500 ms and the reported RMS amplitudes for the ACC were in the range of 0.5 to 1 μV in adult listeners. This is smaller (roughly 3 μV if we estimate the peak-to-peak amplitude from the RMS scaling by sqrt(2) to obtain the peak amplitude and assuming that positive and negative peaks have the same peak amplitude) but comparable to our medium condition, where we observed N1-P2 amplitudes in the order of 5 μV. However, this was significantly smaller than in the slow alternating rate, where the average ACC amplitude was on average 12 μV, both of which were obtained with a similar number of epochs and presumably a similar amount of background noise. Recording time for any condition of the slow (or medium) alternating rate was 8 min, making it considerably faster than previous studies using short, broadband stimuli [e.g., (6)]. Interestingly, a similar alternation rate was successfully used to elicit electrically-evoked FFR and ACC in cochlear implant users (8, 18).

Whilst the slow alternating rate seems to be optimal for the detection of transient ACC in the time-domain, we did not investigate whether frequency-domain analysis will lead to improved detection of the ACC. A visual inspection of Figure 4 shows that use of a periodic alternation rate leads to a spectrum with peaks at the alternation rate and its harmonics. Therefore, the detection of the ACC could be performed in the frequency-domain by taking the energy of the frequency bin corresponding to the alternation rate and its harmonics and comparing those to unrelated frequencies. It remains unclear whether this approach will lead to better results than in the time-domain but it could be a promising method for detecting the ACC. Further studies could investigate if this approach can indeed improve the detection of the ACC for clinical applications.

There was no significant relationship between amplitude for subcortical and either (cortical) P1 or N1-P2 response, suggesting that they are measuring different aspects of perception. This might appear to contrast with the literature showing significant brainstem-cortical relationships (6, 19). However, previous reports showed correlations between subcortical FFR responses and late (> 500 ms), cortical pitch responses; or with N1 and P2 latency (19). Our results do not indicate a clear relationship between the subcortical FFR amplitude and cortical P1 or N1-P2 amplitudes most likely due

to the different generators of the responses and the nature of their behaviors [for reviews, see (20, 21)].

We anticipate that these measures will be useful for objectively studying auditory processing in populations such as children with dyslexia or auditory processing disorders (22–25). Indeed, simultaneously acquired FFR and OR ACC would be able to inform personalized auditory training programs, enable teachers to position children in classroom locations with good signal-to-noise ratios and provide clinicians with information to optimally set up hearing aids, CIs or a combination of both.

Conclusion

We believe that the alternating ACC paradigm can be used to measure sub-cortical and cortical responses that provide complimentary information regarding auditory processing. For probing auditory discrimination we recommend the use of slow alternation rates (<3 Hz) in the low-frequency range (300–1,200 Hz) to strike a balance between the sub-cortical and cortical levels of processing. Future work is required to evaluate how this can be used to inform clinical interventions for people with CIs or other auditory processing difficulties.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by UCL Research Ethics Committee. The participants provided their written informed consent to participate in this study.

Author contributions

CA and VD designed experiments. CA collected and processed EEG data and drafted the manuscript. CA and UJ conducted statistical analysis. All authors contributed to the interpretation of data and to revising the manuscript. All authors contributed to the article and approved the submitted version.

Funding

We gratefully acknowledge funding from the People Programme (Marie Curie Actions) of the European Union H2020 grant agreement no. 798093 (EAR-DNA). Debi Vickers was funded by a Medical Research Council (MRC) Senior Fellowship in Hearing (MR/S002537/1) and a National Institute Health and Care Research programme grant for applied research (201608).

Acknowledgments

The authors would like to thank all the participants who took part in this study. Thanks go to Andrew Clark for his technical help with this project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2022.928158/full#supplementary-material

References

1. Gibson WP. The clinical uses of electrocochleography. *Front Neurosci.* (2017) 11:274. doi: 10.3389/fnins.2017.00274

2. Lucchetti F, Nonclercq A, Avan P, Giraudet F, Fan X, Deltenre P. Subcortical neural generators of the envelope-following response in sleeping children: a transfer function analysis. *Hearing Res.* (2021) 401:108157. doi: 10.1016/j.heares.2020.108157

3. Mehta K, Watkin P, Baldwin M, Marriage J, Mahon M, Vickers D. Role of cortical auditory evoked potentials in reducing the age at hearing aid fitting in

children with hearing loss identified by newborn hearing screening. *Trends Hear.* (2017) 21:2331216517744094. doi: 10.1177/2331216517744094

4. Knebel JF, Jeanvoine A, Guignard F, Vesin JM, Richard C. Differences in click and speech auditory brainstem responses and cortical response patterns: a pilot study. *J Neurology Neurophysiology.* (2018) 9:1–18.

5. Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Res.* (2008) 241:34–42. doi: 10.1016/j.heares.2008.04.013

6. Krishnan A, Bidelman GM, Smalt CJ. Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia.* (2012) 50:2849–59. doi: 10.1016/j.neuropsychologia.2012.08.013

7. Ostroff JM, Martin BA, Boothroyd A. Cortical evoked response to acoustic change within a syllable. *Ear Hearing.* (1998) 19:290–7. doi: 10.1097/00003446-199808000-00004

8. Undurraga JA, Yper LV, Bance M, McAlpine D, Vickers D. Neural encoding of spectro-temporal cues at slow and near speech-rate in cochlear implant users. *Hearing Res.* (2021) 403:108160. doi: 10.1016/j.heares.2020.108160

9. Mathew R, Undurraga J, Li G, Meerton L, Boyle P, Shaida A, et al. Objective assessment of electrode discrimination with the auditory change complex in adult cochlear implant users. *Hearing Res.* (2017) 354:86–101. doi: 10.1016/j.heares.2017.07.008

10. Han JH, Dimitrijevic A. Acoustic change responses to amplitude modulation: a method to quantify cortical temporal processing and hemispheric asymmetry. *Front Neurosci.* (2015) 9:38. doi: 10.3389/fnins.2015.00038

11. Vickers D, Moore BCJ, Boyle P, Schlittenlacher J, Yper LP, Undurraga J. Electrophysiological and psychophysical measures of amplitude modulation. In: *Proceedings of the 23rd International Congress on Acoustics*. Aachen (2019). Available online at: https://research-management.mq.edu.au/ws/portalfiles/portal/139792801/139698416.pdf

12. Skoe E, Kraus N. Auditory brainstem response to complex sounds: a tutorial. *Ear Hearing.* (2010) 31:302–24. doi: 10.1097/AUD.0b013e3181cdb272

13. Krizman J, Kraus N. Analyzing the FFR: a tutorial for decoding the richness of auditory function. *Hearing Res.* (2019) 382:107779. doi: 10.1016/j.heares.2019.107779

14. Sussman E, Steinschneider M, Gumenyuk V, Grushko J, Lawson K. The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Res.* (2008) 236:61–79. doi: 10.1016/j.heares.2007.12.001

15. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* (2015) 67:1–48. doi: 10.18637/jss.v067.i01

16. Lanting CP, Briley PM, Sumner CJ, Krumbholz K. Mechanisms of adaptation in human auditory cortex. *J Neurophysiol.* (2013) 110:973–83. doi: 10.1152/jn.00547.2012

17. Martin BA, Boothroyd A, Ali D, Leach-berth t. stimulus presentation strategies for eliciting the acoustic change complex: increasing efficiency. *Ear Hearing.* (2010) 31:356–66. doi: 10.1097/AUD.0b013e3181ce6355

18. Gransier R, Guérit F, Carlyon RP, Wouters J. Frequency following responses and rate change complexes in cochlear implant users. *Hearing Res.* (2021) 404:108200. doi: 10.1016/j.heares.2021.108200

19. Parbery-Clark A, Marmel F, Bair J, Kraus N. What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci.* (2011) 33:549–57. doi: 10.1111/j.1460-9568.2010.07546.x

20. Coffey EBJ, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, et al. Evolving perspectives on the sources of the frequency-following response. *Nat Commun.* (2019) 10:5036. doi: 10.1038/s41467-019-13003-w

21. Alain C, Tremblay K. The role of event-related brain potentials in assessing central auditory processing. *J Am Acad Audiol.* (2007) 18:573–89. doi: 10.3766/jaaa.18.7.5

22. Calcus A, Deltenre P, Colin C, Kolinsky R. Peripheral and central contribution to the difficulty of speech in noise perception in dyslexic children. *Dev Sci.* (2017) 51:e12558–13. doi: 10.1111/desc.12558

23. Hornickel J, Kraus N. Unstable representation of sound: a biological marker of dyslexia. *J Neurosci.* (2013) 33:3500–4. doi: 10.1523/JNEUROSCI.4205-12.2013

24. Koravand A, Jutras B, Lassonde M. Abnormalities in cortical auditory responses in children with central auditory processing disorder. *Neuroscience.* (2017) 346:135–48. doi: 10.1016/j.neuroscience.2017.01.011

25. Sharma M, Purdy S, Humburg P. Cluster analyses reveals subgroups of children with suspected auditory processing disorders. *Front Psychol.* (2019) 10:1–14. doi: 10.3389/fpsyg.2019.02481

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 03 August 2022
DOI 10.3389/fneur.2022.852030

OPEN ACCESS

EDITED BY
Edmund C. Lalor,
University of Rochester, United States

REVIEWED BY
Xiangbin Teng,
Max Planck Institute for Empirical Aesthetics, MPG, Germany
Nathaniel J. Zuk,
University of Rochester, United States

*CORRESPONDENCE
Wouter David
wouter.david@kuleuven.be

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 10 January 2022
ACCEPTED 29 June 2022
PUBLISHED 03 August 2022

CITATION
David W, Gransier R and Wouters J (2022) Evaluation of phase-locking to parameterized speech envelopes. *Front. Neurol.* 13:852030. doi: 10.3389/fneur.2022.852030

Evaluation of phase-locking to parameterized speech envelopes

Wouter David*, Robin Gransier and Jan Wouters

ExpORL, Department of Neurosciences, KU Leuven, Leuven, Belgium

Humans rely on the temporal processing ability of the auditory system to perceive speech during everyday communication. The temporal envelope of speech is essential for speech perception, particularly envelope modulations below 20 Hz. In the literature, the neural representation of this speech envelope is usually investigated by recording neural phase-locked responses to speech stimuli. However, these phase-locked responses are not only associated with envelope modulation processing, but also with processing of linguistic information at a higher-order level when speech is comprehended. It is thus difficult to disentangle the responses into components from the acoustic envelope itself and the linguistic structures in speech (such as words, phrases and sentences). Another way to investigate neural modulation processing is to use sinusoidal amplitude-modulated stimuli at different modulation frequencies to obtain the temporal modulation transfer function. However, these transfer functions are considerably variable across modulation frequencies and individual listeners. To tackle the issues of both speech and sinusoidal amplitude-modulated stimuli, the recently introduced Temporal Speech Envelope Tracking (TEMPEST) framework proposed the use of stimuli with a distribution of envelope modulations. The framework aims to assess the brain's capability to process temporal envelopes in different frequency bands using stimuli with speech-like envelope modulations. In this study, we provide a proof-of-concept of the framework using stimuli with modulation frequency bands around the syllable and phoneme rate in natural speech. We evaluated whether the evoked phase-locked neural activity correlates with the speech-weighted modulation transfer function measured using sinusoidal amplitude-modulated stimuli in normal-hearing listeners. Since many studies on modulation processing employ different metrics and comparing their results is difficult, we included different power- and phase-based metrics and investigate how these metrics relate to each other. Results reveal a strong correspondence across listeners between the neural activity evoked by the speech-like stimuli and the activity evoked by the sinusoidal amplitude-modulated stimuli. Furthermore, strong correspondence was also apparent between each metric, facilitating comparisons between studies using different metrics. These findings indicate the potential of the TEMPEST framework to efficiently assess the neural capability to process temporal envelope modulations within a frequency band that is important for speech perception.

KEYWORDS

temporal processing, envelope modulations, envelope encoding, auditory steady-state responses (ASSR), speech processing

Introduction

Natural speech is a complex and dynamic signal. One prominent component of the speech signal is the temporal envelope. The speech envelope contains slow modulations that are related to linguistic information at different timescales such as phrases, words, syllables, and phonemes (1, 2). The modulation spectrum of the speech envelope exhibits a prominent peak for slow modulations of 4–5 Hz (3, 4), which corresponds to the syllable rate in speech (1, 5–7). Since the timescales of these slow modulations coincide with spoken syllables, access to these envelope modulations and their representation in the neural signal traveling through the auditory pathway is essential for speech perception, especially when access to spectral information is limited (8–12).

Two main electrophysiological paradigms are often used to investigate the neural representation of these slow envelope modulations throughout the auditory pathway. One paradigm involves neural entrainment to speech, which refers to cortical responses that consistently phase-lock to slow modulations of the speech envelope (13). The relation between neural responses and the speech envelope through phase-locking has been established with magneto- and electroencephalography (MEG/EEG) (14–16). While listening to speech, the phase pattern of the neural response is consistent with the speech envelope modulations of 4–8 Hz (17, 18). Interestingly, several studies suggested that speech perception performance is associated with the degree of phase-locking to the speech envelope (19–21). In other words, neural phase-locked patterns that are less consistent with the speech envelope are associated with degraded speech perception. For example, higher disruption of neural phase-locking during listening with electrical transcranial stimulation has been shown to result in more degraded speech perception (22). These findings suggest that phase-locking to the speech envelope in the auditory pathway plays an important role in speech perception. Moreover, hierarchical linguistic structures – such as words, phrases, and sentences – are differentiated by input acoustical cues and linguistic higher-order comprehension processes (23–25). The phase-locked responses to speech from the auditory pathway consist of cortical activity at different timescales (or modulation frequency bands) that concurrently track different linguistic structures at different hierarchical levels.

Analyses of phase-locked responses to speech have pointed to distinct functional roles of the delta (1–4 Hz) and theta (4–8 Hz) bands. On the one hand, phase-locking in the delta band is largely associated with the amount of linguistic information in the speech signal (26, 27) and with the listener's proficiency in the language (28–30). By manipulating the different levels of linguistic structure in the speech signals, this can be studied. When listening to a stream of synthesized Chinese sentences, in which the sentence rate was not present in the envelope but was encoded in the linguistic structure, native Chinese listeners did show phase-locking at the sentence rate while native English listeners did not (29). Neural phase-locking is also associated with lexical, syntactic, and/or semantic changes in the linguistic content when the speech is comprehended. The theta band (4–8 Hz), on the other hand, seems to be more dependent on the saliency of the perceived acoustic envelope. To assess how envelope modulations at these low frequencies are processed by the auditory system, one can use techniques that alter the linguistic content of speech. Distortions to the speech signal can consequently also affect the linguistic message conveyed (31, 32). These findings show that the envelope and the linguistic content of speech are interdependent (13, 33–35). However, the relative contributions to neural phase-locked responses of the speech envelope on the one hand and the linguistic content of speech, on the other hand, are difficult to disentangle from each other. Several studies have shown the applicability to use amplitude-modulated (AM) stimuli to assess phase-locked responses to envelope modulations (36–39).

Sinusoidally amplitude-modulated (SAM) stimuli are at the basis of the other paradigm to investigate the neural representation of envelope modulations. These stimuli evoke auditory steady-state responses (ASSR) (40) of which the strength reflects the ability of the auditory pathway to phase-lock to the stimulus' modulation frequency (i.e., the response is synchronized to the envelope fluctuations). ASSRs evoked by stimuli with modulations below 20 Hz originate predominately from the auditory cortex, while those evoked with higher frequencies originate from subcortical and brainstem regions (41–44). Studies have indicated that speech perception performance in noise is correlated with 40-Hz ASSRs (45–47) and 80-Hz ASSRs (47–49). In addition, ASSRs elicited by 20-Hz and 4-Hz modulations are associated with phoneme and sentence scores, respectively (48–50). To obtain a sense of the overall capacity of neural modulation processing, ASSRs are measured over a wide range of modulation frequencies. The ASSR amplitude as a function of modulation frequency is the temporal modulation transfer function (TMTF). The TMTF shows a broad peak around 80 and 40 Hz (36–39), and also around 20 Hz (36). Interestingly, the TMTF shows large variations in ASSR evoked by modulation frequencies below 20 Hz and across listeners (36). Therefore, to gain insight into the overall processing capacity of these slow modulations, one would have to measure several ASSRs within this range to evaluate the overall capability to process speech-relevant modulations. However, this approach is time-consuming and could potentially be performed more efficiently using a speech-like stimulus that contains the modulation frequencies of interest.

To overcome the issues that are encountered with speech and SAM stimuli, Gransier and Wouters (51) developed the Temporal Envelope Speech Tracking (TEMPEST) framework. The TEMPEST framework enables the creation of stimuli with parameterized envelopes which can be used to assess the effect

of specific characteristics of the speech envelope on neural processing (e.g., envelopes that contain the same modulations as natural speech). In the present study, we investigate whether TEMPEST-based stimuli that consist of syllabic-like and phonemic-like modulations—as present in natural speech—can be used to gain insight into the speech-weighted electrophysiological TMTF of normal-hearing listeners. To this end, we elicited responses with TEMPEST stimuli based on distributions of modulation frequencies close the syllable (∼4 Hz) and phoneme (∼20 Hz) rates in speech. Furthermore, we also recorded ASSRs, which are normally used to assess the electrophysiological TMTF, with modulation frequencies that covered the same range as those in the TEMPEST stimuli. We compared the overall activity of the TEMPEST neural responses and that of the ASSRs. We expect that the overall TEMPEST neural activity corresponds to the speech-weighted overall activity within the ASSR TMTF and that the TEMPEST framework can be used to efficiently probe the speech-weighted electrophysiological TMTF in normal-hearing listeners. To this end, we used different power- and phase-based electrophysiological metrics that are widely used in the literature. Many studies make use of various electrophysiological metrics (or terminologies) to characterize phase-locked responses to AM stimuli. Some of the studies made use of power-based metrics [e.g., in Gransier et al. (36), Purcell et al. (37), Poulsen et al. (38)] while other studies applied phase-based metrics [e.g., in Luo and Poeppel (17), Howard and Poeppel (18)]. Due to the use of different metrics, comparing results across studies is difficult. Therefore, we included different power- and phase-based metrics and investigated how they relate to each other in order to facilitate these comparisons across studies.

Materials and methods

TEMPEST framework

Gransier and Wouters (51) introduced the TEMPEST framework in which amplitude-modulated stimuli are created based on an a-priori distribution of modulation frequencies that are relevant for speech. The purpose of the TEMPEST framework is to evaluate the overall envelope encoding ability of the auditory system with stimuli containing a range of envelope modulations. TEMPEST stimuli have a quasi-regular envelope which is generated by concatenating windows over time (Figure 1). Each window in the envelope can represent the occurrence of an acoustic unit in natural speech. The duration of each window depends on random sampling from a probability distribution of modulation frequencies. Each randomly sampled modulation frequency (f_m) is inverted to determine the duration ($T_{window} = 1/f_m$) of subsequent windows (Figure 1). Furthermore, each window can have some fixed or variable parameters, such as peak amplitude, onset time, etc. A simple example is the SAM stimulus, which can be created within the TEMPEST framework using sinusoidal windows with a fixed peak amplitude and only one modulation frequency. The next examples are two TEMPEST stimuli used in this

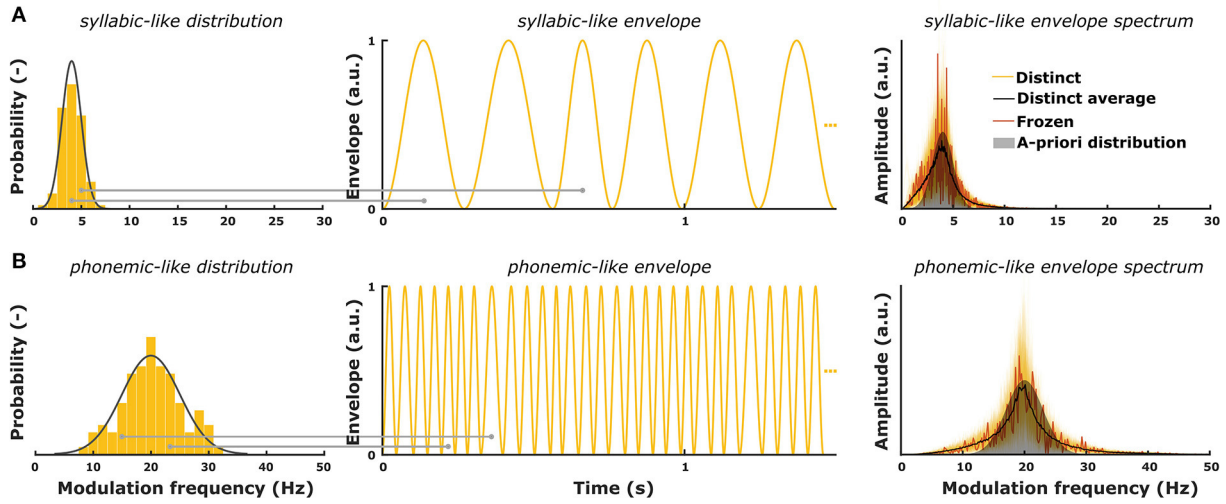


FIGURE 1
(A) A-priori modulation frequency distribution for syllabic-like envelopes (mean = 4 Hz; standard deviation = 1 Hz) and an exemplary envelope. The right panel shows modulation spectra of 200 syllabic-like TEMPEST envelopes (yellow) along with the averaged distinct spectrum (black) and the spectrum of the frozen envelope (red). **(B)** A-priori modulation frequency distribution for phonemic-like envelopes (mean = 20 Hz; standard deviation = 3 Hz) and an exemplary envelope. The right panel shows modulation spectra of 200 phonemic-like TEMPEST envelopes (yellow) along with the averaged distinct spectrum (black) and the spectrum of the frozen envelope (red). Histograms show the sampled modulation frequencies of the exemplary envelopes. Horizontal gray lines depict the sampling of modulation frequency for two envelope windows with corresponding window length (= $1/f_m$).

study (Figure 1, right). These stimuli have different modulation frequency distributions: one centered around 4 Hz (syllable rate) and one around 20 Hz (phoneme rate) (Figure 1, left). Due to sampling of the distributions, the envelope modulation spectrum will also contain these modulation frequencies with a peak at the center frequency. After its generation, the envelope is used to modulate a carrier signal to finalize the creation of the TEMPEST stimulus.

The main goal of this study is to validate whether the TEMPEST framework can be used to assess the speech-weighted electrophysiological TMTF in normal-hearing listeners. The TEMPEST framework would be a useful tool to investigate the overall neural capability to process envelope modulations which can potentially be related to speech perception performance. To this end, we generated "basic" TEMPEST stimuli using a Gaussian probability function of low modulation frequencies that are apparent in the speech envelope.

Participants

Ten normal-hearing native-Dutch young adults (ages from 19 to 27 years; 3 males and 7 females) participated in this study. No participants had neurological deficits. All participants had normal hearing (pure tone thresholds $\leq$ 25 dB HL for all octave frequencies between 250 and 8,000 Hz). This study was approved by the Medical Ethical Committee of the UZ Leuven hospital (study number: B322201524931). All participants gave written informed consent before participation.

Stimuli

SAM stimuli

ASSRs with different modulation frequencies were recorded to obtain individual electrophysiological TMTFs within the modulation frequency ranges of the TEMPEST stimuli. Modulation frequencies of the SAM stimuli were chosen to sample the modulation bands of the TEMPEST stimuli (Figure 1, left). Syllabic-like SAM stimuli with modulation frequencies of 2–6 Hz and phonemic-like SAM stimuli with modulation frequencies of 17–23 Hz were included. All SAM stimuli were created in a custom stimulation software (52). Modulation frequencies were adjusted such that there is an integer number of cycles within one trial of 1.024 s. However, we will further report using rounded modulation frequencies for readability. Modulation depth was set at a maximum of 100% in order to elicit as large ASSRs as possible. The carrier was speech-weighted noise which was generated from the long-term average spectrum of 730 Dutch sentences of the LIST corpus (53). Blocks of 2.56 min were recorded in each measurement session so that 300 trials in total were recorded for each modulation frequency.

TEMPEST stimuli

TEMPEST envelopes for this study were generated in Matlab R2016b using Hann windows. Hann windows were used because they have a start- and endpoint at zero to prevent discontinuities in the envelope. The peak amplitude of the windows was always at a maximum of 1 such that the effective modulation depth of the TEMPEST stimuli was 100%. We generated two types of TEMPEST stimuli: syllabic-like and phonemic-like stimuli (Figure 1). Modulation distributions of the TEMPEST stimuli were based on modulation rates that are particularly important for speech, i.e., the natural rates of syllables and phonemes (2, 7). The modulation distribution of syllabic-like TEMPEST envelopes closely matched the low envelope modulation spectrum of speech, which shows a peak around 4 Hz (3, 4). The phonemic-like modulation distribution was based on phoneme length statistics in speech from which the mean duration was found to be around 50 ms (54), which corresponds to a center modulation frequency of 20 Hz. The standard deviations of the distributions were 1 Hz and 3 Hz the envelopes of the syllabic-like and phonemic-like TEMPEST stimuli, respectively (Figure 1).

The duration of the syllabic-like and phonemic-like stimuli were 5.12 s and 25.6 s long in order to reach a similar number of envelope windows and to sufficiently sample the modulation distributions. The envelopes were tested for sufficient statistical similarity to the modulation distribution using the Kolmogorov-Smirnov test with a significance level of $\alpha = 0.05$. Additionally, we applied criteria to ensure that the envelope modulation sample mean and standard deviation did not deviate too far from those of the a-priori distribution. We used $\Delta\mu \leq 0.05$ Hz and $\Delta\sigma \leq 0.05$ Hz for syllabic-like envelopes, and $\Delta\mu \leq 0.25$ Hz and $\Delta\sigma \leq 0.1$ Hz for phonemic-like envelopes, with $\Delta\mu$ the difference between the means and $\Delta\sigma$ the difference between standard deviations of the sample and a-priori distributions. Envelopes that did not meet these criteria were discarded and new ones were generated instead until they met the criteria. This procedure was continued until 200 syllabic-like and phonemic-like TEMPEST envelopes were obtained. Only 20% of the total amount of generated envelopes passed the test and both criteria. Finally, these envelopes were used to modulate segments of speech-weighted noise based on Dutch LIST sentences (53).

In the main experiment, one single syllabic-like and one phonemic-like stimulus were presented repeatedly to the listener. These stimuli are referred to as frozen stimuli since the same temporal pattern was used over again. The goal of the frozen stimuli was to test robust neural phase-locking and evoked power in the modulation distribution frequency range and to compare this neural activity with ASSRs. Additionally, the remaining syllabic-like and phonemic-like stimuli were presented only once to the listener. Since these stimuli were temporally different from each other, they are referred to as distinct stimuli. Distinct stimuli were used as a baseline measurement with respect to the frozen stimuli (17, 55–57).

The number of distinct stimuli equaled the number of repeated presentations of the frozen stimulus so that bias by differences in the number of trials is minimized (58). Stimuli were presented in blocks of 5.12 min, in which either frozen stimuli were repeated or distinct stimuli were presented in random order. In total, there were 156 frozen and distinct syllabic-like trials (12 presentations in 13 blocks), and 180 frozen and distinct phonemic-like trials (60 presentations in 3 blocks). Each block was preceded with a short 2.56-s TEMPEST segment generated with the same parameters. The evoked neural activity to this segment contains an onset response that would interfere with the main analysis. Therefore, the EEG recordings corresponding to this segment were immediately discarded.

Equipment

Calibration and presentation setup

Presentation of all stimuli was done using custom-built software interfacing with an RME-Hammerfall DSP Multiface II soundcard and delivered monaurally through an Etymotic ER-3A insert earphone to the right ear. All stimuli were calibrated using a 2-cc coupler of an artificial ear (Brüel & Kjær, type 4,152) and presented at 70 dB sound pressure level (SPL) at a sampling rate of 32 kHz. Two measurement sessions were conducted whereby each session started with a set of ASSR stimuli in a pseudo-random order which was followed by a set of phonemic-like and syllabic-like TEMPEST stimuli in a pseudo-random order as well.

EEG recording setup

EEG was recorded using a 64-channel BioSemi ActiveTwo recording system with a sampling rate of 8,192 Hz and a recording bandwidth of 0 to 1,683 Hz. A head cap with 64 Ag/AgCl recording electrodes was placed on the scalp of every participant. The electrode positions were placed across the scalp according to the international standard 10–20 system (59). All recordings were made in a double-walled soundproof booth that is equipped with a Faraday cage to avoid signal interference as much as possible. Participants watched a silent movie by choice while seated in a relaxing chair. They were offered a head pillow and asked to move as little as possible to minimize head movement/muscle artifacts.

Signal processing and response quantification

Preprocessing

Offline signal processing was done in Matlab R2016b. EEG recordings were high-pass filtered using a 1st order Butterworth filter with a cut-off frequency of 0.5 Hz to remove any DC component and slow drifts. Recordings were referenced to electrode Cz by subtracting the recording of Cz from those of the other channels. 5% of the trials were discarded from the analysis based on the highest peak-to-peak amplitudes, as they were assumed to contain muscle and other recording artifacts. Due to measurement errors, not all trials could be obtained from each participant. Only 108 frozen phonemic-like trials could be retained from Participant 7, while 162 phonemic-like trials could be retained in all other cases. The minimum number of retained syllabic-like trials is 115 and the maximum number is 136 across all participants. Time signals of the parieto-occipital recording electrodes were averaged into a left and a right hemispheric channel. Recording electrodes O1, PO3, PO7, P9, P7, P5, CP5, and TP7 formed the left hemispheric channel, while recording electrodes O2, PO4, PO8, P10, P8, P6, CP6, and TP8 formed the right hemispheric channel. See Figure 2 for a visualization of the selected electrodes.

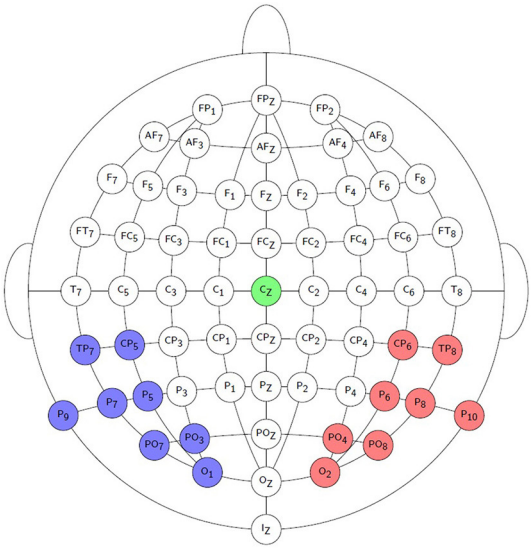


FIGURE 2
Visualization of the EEG recording electrodes used to form the left channel (blue) and the right channel (red). The reference EEG electrode Cz is indicated by the green color.

In the case of ASSRs, all 300 trials of each modulation frequency were successfully recorded. Syllabic ASSR (f_{mod} = 2–6 Hz) recordings were grouped into sweeps of 5 trials, while phonemic-like ASSR (f_{mod} = 17–23 Hz) recordings were grouped into sweeps of 1 trial. Syllabic-like and phonemic-like ASSR sweep lengths were thus 5.12 s and 1.024 s, respectively. Consequently, the number of cycles in each sweep is similar for both syllabic-like and phonemic-like ASSRs in order to have similar phase estimation during analysis. The rest of the preprocessing procedure is the same as for the TEMPEST recordings.

Neural response analyses

Amplitude and phase for each modulation frequency were extracted from the individual or averaged response trials after transforming into the spectral domain. ASSR sweeps were transformed using the discrete Fourier transform. TEMPEST response trials were transformed into Fourier spectrograms with Hanning windows in which the window length and window overlap were tuned such that phase estimation is similar to that for ASSRs. The window length was equal to the length of the corresponding syllabic-like or phonemic-like ASSR sweep. The window overlap corresponded to three times the reciprocal of the mean modulation frequency in each TEMPEST stimulus such that subsequent windows are, on average, one cycle from each other. Thus, for syllabic-like TEMPEST, spectrograms were computed with 5.12 s window length and 0.25 s window step, whereas for phonemic-like stimuli, a window length of 1.024 s and a window step of 0.05 s were used. Since different spectrogram parameter values were used, the frequency resolution differed between syllabic-like and phonemic-like stimuli. Response bins are 0.195 Hz/bin and 0.977 Hz/bin, respectively. Amplitude and phase were extracted from each time-frequency bin in the spectrogram. These values were used to compute several electrophysiological metrics listed below.

To gain insight into the characteristics and robustness of the recorded neural responses and to compare the TEMPEST responses with ASSRs, four electrophysiological metrics were employed in our analysis. A small selection of metrics have been employed because many different metrics are being used in the literature and this makes comparisons and conclusions across studies more difficult. In order to investigate how different metrics relate to each other and to facilitate comparisons between studies, the metrics used in our analyses represent some of the most widely used ones in the power and phase domain. Two of them are power-based metrics, namely power and signal-to-noise ratio (SNR) of the averaged response. Power is computed after obtaining the amplitude spectrum of the averaged neural response and squaring the amplitude in each frequency bin This metric reflects the overall neural activity evoked by the stimulus (60). The SNR is taken as the power of the averaged neural response divided by power of the neural background noise. Power of the averaged neural response is computed as the mean power across stimulus trials in each frequency bin. Power of the neural background noise is computed as the variance of power across stimulus trials divided by the number of trials in each frequency bin. This estimation of neural background noise is more viable for TEMPEST responses than the estimation from neighboring noise bins which is commonly used in case of ASSRs (40). This is because TEMPEST responses are expected to contain evoked power within a certain frequency band whereas ASSRs only have evoked power in the modulation frequency bin. Additionally, as the neural background noise typically exhibits a 1/f spectrum, noise power at the lower frequency side is higher than at the higher frequency side. Evoked responses to a repeated stimulus expected to be consistent in power and phase across trials, while neural background noise adds a random amplitude and phase to that of the evoked response in each trial. Under this assumption, variance in power across trials divided by the number of trials reflects neural background noise power (61). The two power-based metrics are ubiquitously used in the neuroscience field to indicate the strength and quality of the measured averaged response. The other two metrics are solely based on the phase of the individual response trials: inter-trial phase coherence (ITPC) and pairwise phase consistency (PPC). The first metric, ITPC, indicates consistency of phase-locking to a stimulus based on the magnitude of the average of unit vectors rotated by extracted phases θ_n across N trials (17, 62).

$$ITPC = \left(\sum_{n=1}^{N} \frac{\cos\theta_n}{N}\right)^2 + \left(\sum_{n=1}^{N} \frac{\sin\theta_n}{N}\right)^2 \quad (1)$$

The ITPC is commonly used to investigate robustness of phase-locking with different stimulus parameters (17, 18, 55, 57, 62–64). However, despite its considerable presence in the literature, the ITPC is biased by the number of trials with fewer trials resulting in a larger positive bias in the outcome. This bias could hamper comparison between conditions and/or studies with different amounts of trials (58, 61, 65). In contrast to ITPC, the PPC is an unbiased estimate of phase-locking because it is based on the averaged dot product of all possible phase pairs θ_n and θ_m across N trials (66).

$$PPC = \frac{2}{N(N-1)} \sum_{n=1}^{N-1} \sum_{m=n+1}^{N} cos(\theta_n - \theta_m) \quad (2)$$

When phase consistency is high, then distances between phase pairs will become smaller and thus dot products will be larger. The advantage of the PPC is that it allows for comparison between studies and conditions even with different trial numbers. Both ITPC and PPC take up values between 0 and 1, with 0 indicating no phase-locking at all and 1 indicating perfect phase-locking across trials. Note that ITPC and PPC for TEMPEST responses are computed for each time and frequency bin. In order to obtain electrophysiological patterns as a function of modulation frequency in each participant, results of each metric were averaged in the time domain.

Responses were tested for significance against the neural background noise using the Hotelling T^2 test (52, 67). ASSRs were tested only at their modulation frequency bin while TEMPEST neural activity was tested in each modulation frequency bin of the spectral domain. To evaluate similarity between ASSR patterns and between TEMPEST patterns measured with different electrophysiological metrics and whether different metrics would reveal different characteristics

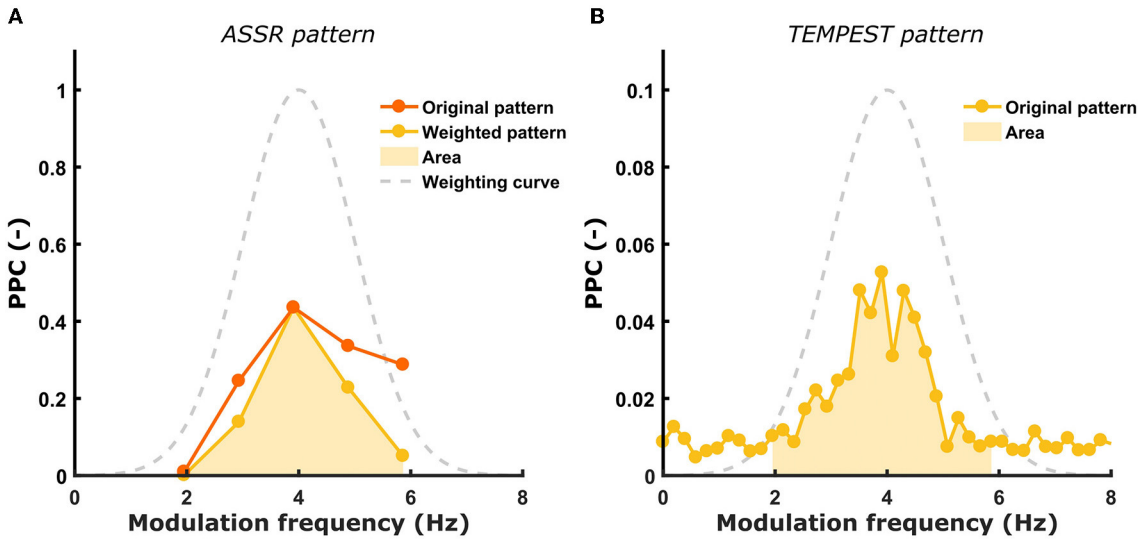


FIGURE 3
Illustration of ASSR pattern weighting and computation of the area under the curve. **(A)** the original ASSR pattern is weighted by the Gaussian modulation distribution of the corresponding TEMPEST stimuli. The computed area under the weighted ASSR pattern is depicted as the shaded area. **(B)** The computed area under the TEMPEST pattern is depicted as the shaded area. Syllabic-like pattern data came from the left hemispheric channel in Participant 3.

of the neural patterns, the patterns were subjected to correlation analyses. Only significant response bins were included in the analyses. Additionally, because ITPC and PPC are bounded between 0 and 1, their values were first transformed using the Fischer z-transformation. Pearson's correlation coefficients and corresponding p-values were then reported. The significance level was $\alpha = 0.05$ at all times and *post-hoc* Bonferroni correction was used to control for false discovery rate since multiple correlations were being tested simultaneously.

Comparison between TEMPEST and ASSR TMTF patterns

The main goal of this study is to investigate whether the neural activity evoked by TEMPEST stimuli is comparable to the overall ASSR activity (i.e., the TMTF) in the same frequency band. Usually, the TMTF is obtained by setting out ASSR amplitude as a function of modulation frequency (36–39). However, in this study not only ASSR patterns of power, but also of SNR, ITPC and PPC were used. When the ASSR TMTF shows a prominent peak, we hypothesize that the TEMPEST neural activity would also show a relatively large peak and vice versa for each metric. The presence of prominent peaks translates into a larger area under the TMTF pattern. To compare the TEMPEST patterns with those of ASSRs, areas under patterns of the same metric were computed and correlated with each other across all participants in the left and right hemispheres. Before computing the area of ASSR patterns, patterns were first weighted according to the corresponding TEMPEST modulation distribution in order to account for the relative contribution of each modulation frequency to the TEMPEST neural activity. Each modulation frequency of the SAM stimuli contribute equally to the ASSR patterns. However, these contributions are not equal anymore in case of TEMPEST due to the a-priori modulation frequency distribution used to generate the stimuli. To achieve this weighting of the ASSR pattern, it is multiplied with the Gaussian curve of the corresponding syllabic-like or phonemic-like TEMPEST modulation distribution. By doing this, the TEMPEST and ASSR neural evoked activity can be directly compared to each other after accounting for the modulation distribution shape. Areas under the patterns were computed between 2 and 6 Hz for syllabic-like responses, and between 17 and 23 Hz for phonemic-like responses (Figure 3).

The area was computed by summing up the values in each frequency bin within the restricted band. Finally, to test the relative correspondence between the ASSR and TEMPEST patterns, Pearson's correlations between the TEMPEST and ASSR areas across participants were computed. Only areas of the same metric from ASSR and TEMPEST analyses were correlated (e.g., the area of ASSR SNR was correlated with the area of TEMPEST SNR). Partial Pearson's correlations were computed between TEMPEST and ASSR power area in order to control for any potential effects of induced power area. Induced power is the power that appears in the EEG in any frequency band while listening to a stimulus. In order to investigate whether neural phase-locking to the TEMPEST and SAM stimuli correspond to each other, the correlation with induced power must be controlled for. The induced power

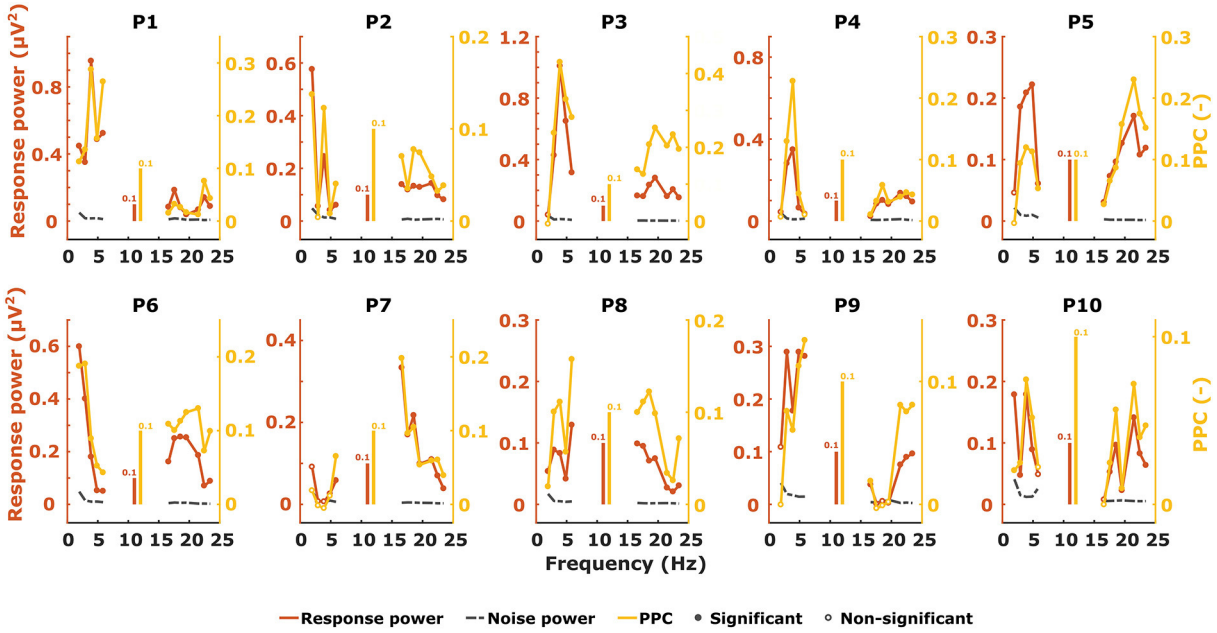


FIGURE 4
Unweighted ASSR power (red), noise (black), and PPC (yellow) patterns from the left hemispheric channel as a function of modulation frequency.

area in the syllabic-like frequency range was computed from the averaged power spectrum of the distinct phonemic-like TEMPEST stimuli, whereas the induced power area in the phonemic-like frequency range was computed from the distinct syllabic-like TEMPEST stimuli. The significance level for the correlations was $\alpha = 0.05$ and p-values were corrected with the Bonferroni procedure.

Results

Evaluation of electrophysiological metrics

ASSR

We measured ASSRs with 2–6 Hz (syllabic-like) and 17–23 Hz (phonemic-like) modulation frequencies and obtained the response pattern across modulation frequency for each participant and electrophysiological metric, which is very similar to how TMTFs are obtained elsewhere. Almost all ASSRs were found to be statistically significantly different from noise using the Hotelling T^2 test. Figure 4 shows the individual ASSR patterns measured with response power and PPC for syllabic ASSRs in the left hemispheric channel. In this case, the patterns of these two metrics are relatively similar to each other within each participant. The different shapes of the patterns demonstrate the large variability in ASSRs across modulation frequency and participants.

Patterns of the other electrophysiological metrics are not shown but their similarity in shape to each other was evaluated with correlation analyses. Examples of correlation scatterplots for the syllabic-like responses in the left hemispheric channel are shown in Figure 5. Table 1 summarizes all Pearson's correlation coefficients between the different electrophysiological metrics. Since the response power and PPC patterns were relatively similar, they were highly correlated with each other [r (38) = 0.85, $p < 0.0001$]. The ITPC and PPC showed an almost perfect linear correlation (Figure 5, bottom left) based on the fact that the PPC is an unbiased estimate of phase-locking compared to the biased ITPC due to the number of trials. Exchanging the ITPC for PPC would not virtually change the interpretation of the results. The next highest correlations were found between SNR and PPC, which are very high [from r (38) = 0.92 to r (64) = 0.99, $p < 0.0001$]. Comparing the ASSR power with SNR and PPC resulted in moderate to high correlation coefficients. Each correlation coefficient was found to be highly significant (Table 1).

TEMPEST

When characterizing neural responses to TEMPEST stimuli for each modulation frequency, all electrophysiological metrics showed variation across participants. Figures 6, 7 show only response power and PPC patterns layered over each other for syllabic-like and phonemic-like neural

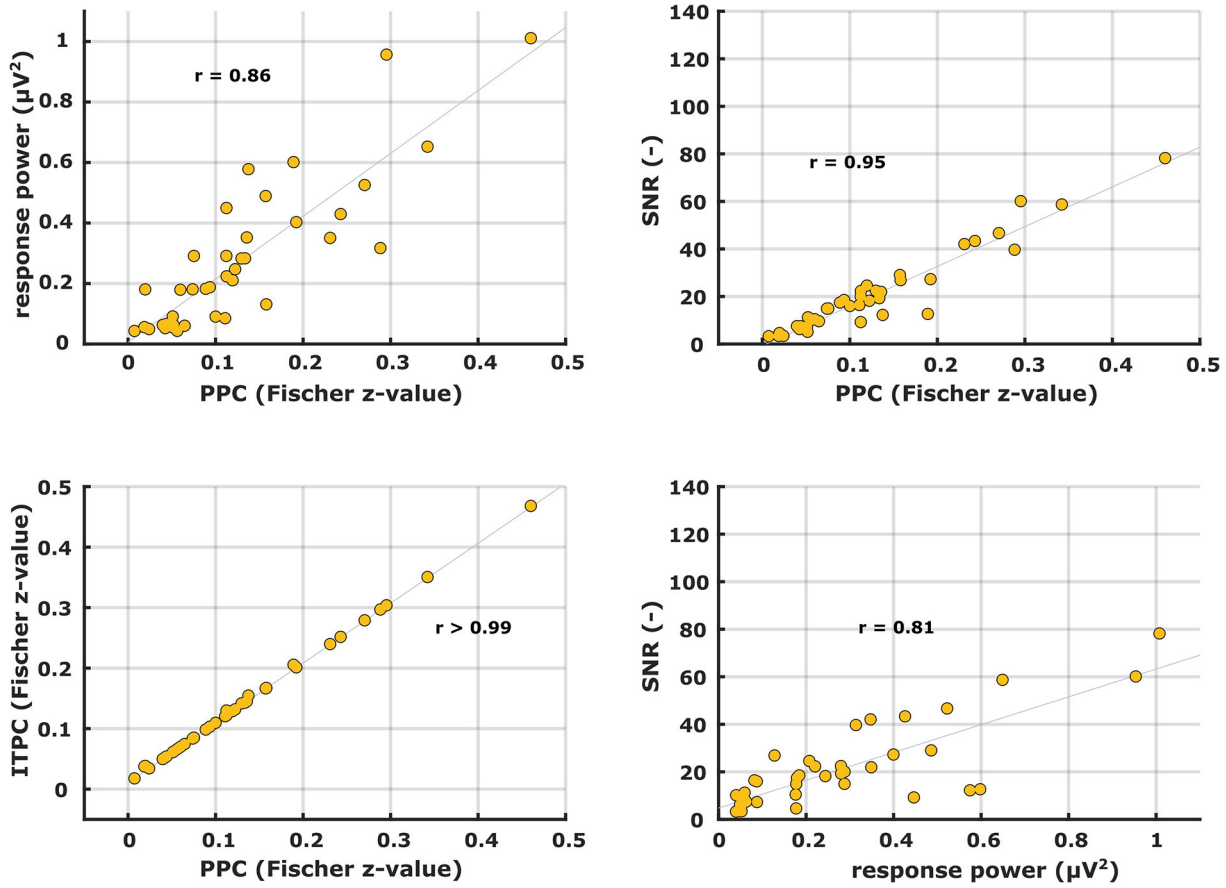


FIGURE 5
Scatter plots between different ASSR electrophysiological metrics for syllabic-like stimuli in the left hemispheric channel.

responses, respectively. Patterns of the other metrics are not shown, but pattern correlations between all four metrics are presented in Table 1. In some participants, distinctive peaks around the mean modulation frequency of the envelope were found in their patterns. For example, participants 1, 3, 5, and 9 showed increased activity around 4 Hz with syllabic-like stimuli. Interestingly, unlike these participants, participant 2 did not show peak activity around 4 Hz but a broader one around 7–8 Hz with syllabic-like stimuli, which corresponds to the range of second harmonic frequencies. With phonemic-like stimuli, participants 3, 5, 6, 7, and 8 showed highly prominent peaks around 20 Hz. As expected, responses to distinct stimuli did not show the increased averaged neural activity as with frozen stimuli.

Patterns of the other metrics are not shown but – like with the ASSRs – their similarity to the PPC patterns was evaluated with correlation analyses. Correlations between the different electrophysiological metric patterns are shown in Table 2. Again, unsurprisingly, the ITPC and PPC showed an almost perfect linear correlation ($r > 0.99$) (Figure 8, bottom left). Exchanging the ITPC for PPC would not virtually change the interpretation of the results as well in this case. Other metric comparisons resulted in moderate to high correlations except for power vs. PPC in the left hemisphere for phonemic-like stimuli. Correlations with PPC for power and SNR were not as strong as those for ASSRs. Each correlation coefficient was found to be highly significant, except for power vs. PPC in the left hemisphere for phonemic-like stimuli (Table 2).

Comparison of TEMPEST and ASSR neural activity

If an individual ASSR TMTF does not show a prominent peak, then we expect that the TEMPEST neural pattern would not show a peak as well and vice versa. To assess whether the overall activity of ASSR TMTFs corresponds to the activity of TEMPEST responses within participants, we computed the area under the patterns and performed correlation analyses. Only

TABLE 1 Pearson's correlation coefficients between different electrophysiological metrics of syllabic-like and phonemic-like ASSRs in the left and right hemispheric channels.

	Syllabic-like		Phonemic-like	
	Left ($df = 37$)	Right ($df = 40$)	Left ($df = 64$)	Right ($df = 63$)
Power vs. PPC	0.86* (<0.0001)	0.74* (<0.0001)	0.69* (<0.0001)	0.62* (<0.0001)
SNR vs. PPC	0.95* (<0.0001)	0.92* (<0.0001)	0.98* (<0.0001)	0.99* (<0.0001)
ITPC vs. PPC	> 0.99* (<0.0001)	> 0.99* (<0.0001)	> 0.99* (<0.0001)	> 0.99* (<0.0001)
SNR vs. power	0.81* (<0.0001)	0.63* (<0.0001)	0.68* (<0.0001)	0.63* (<0.0001)

Corresponding p-values are reported between parentheses. *Significant after post-hoc Bonferroni correction.

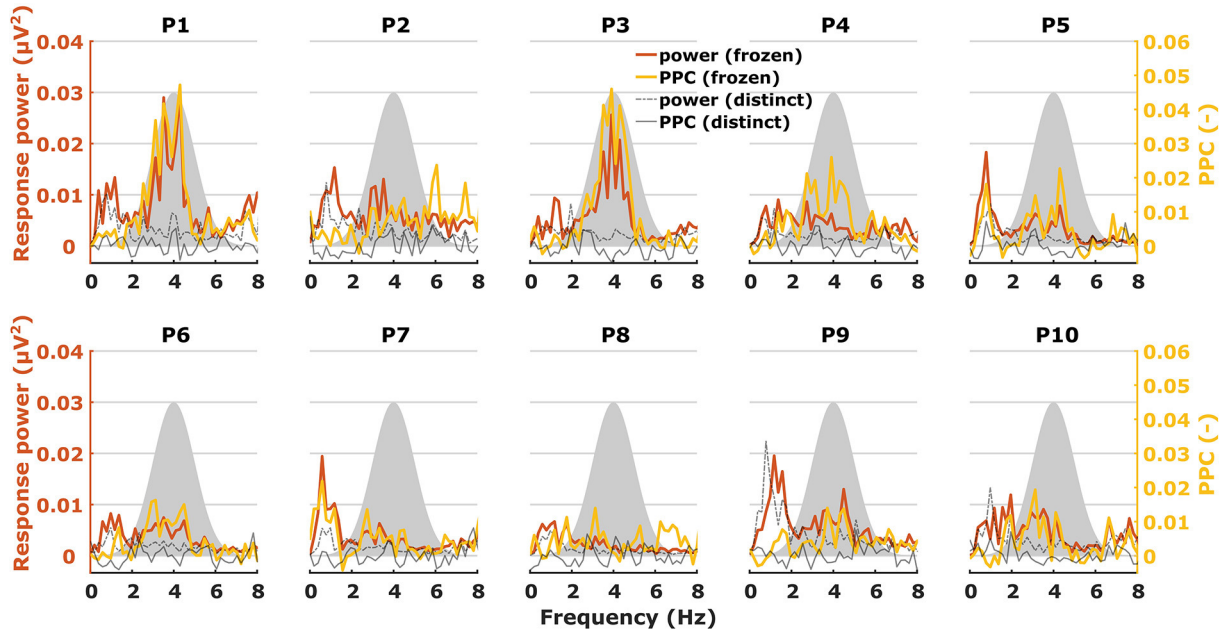


FIGURE 6
Two different patterns, one of power (red) and one of PPC (yellow), plotted over each other across modulation frequency. The patterns are from TEMPEST syllabic-like responses in the left hemispheric channel for each participant. Black patterns describe the baseline from distinct responses. The shaded Gaussian curve represents the modulation frequency distribution of the stimuli (scaled arbitrarily).

areas under the ASSR and the TEMPEST pattern of the same electrophysiological metric were used because comparing areas with different metrics would not be insightful (e.g., area under ASSR PPC pattern vs. area under TEMPEST SNR pattern). The computed areas were directly correlated across all ten participants for SNR and PPC of syllabic-like and phonemic-like responses in the left and right hemispheres separately (Figure 9, middle and right columns). Based on the almost perfect correlation between ITPC and PPC (Tables 1, 2), ITPC was left out because it would produce the same results as the PPC. All correlation coefficients were found to be strong [r (8) = 0.75–0.98] and highly significant after *post-hoc* Bonferroni correction ($p \leq 0.001$), except for the correlation coefficient between ASSR and TEMPEST SNR area for syllabic-like responses in the right hemispheric channel which was not significant anymore after *post-hoc* correction ($p = 0.013$). For the power metric, partial Pearson's correlations between TEMPEST power area and ASSR power area were computed in order to control for any potential effects of induced power area. Partial correlation coefficients were found to be strong [r (7) = 0.81–0.97] and highly significant ($p \leq 0.001$). These high correlations indicate that the overall activity of the TEMPEST responses corresponds to the speech-weighted overall activity of the ASSR TMTF.

Discussion

The TEMPEST framework was introduced by Gransier and Wouters (51) to provide an efficient method to investigate the

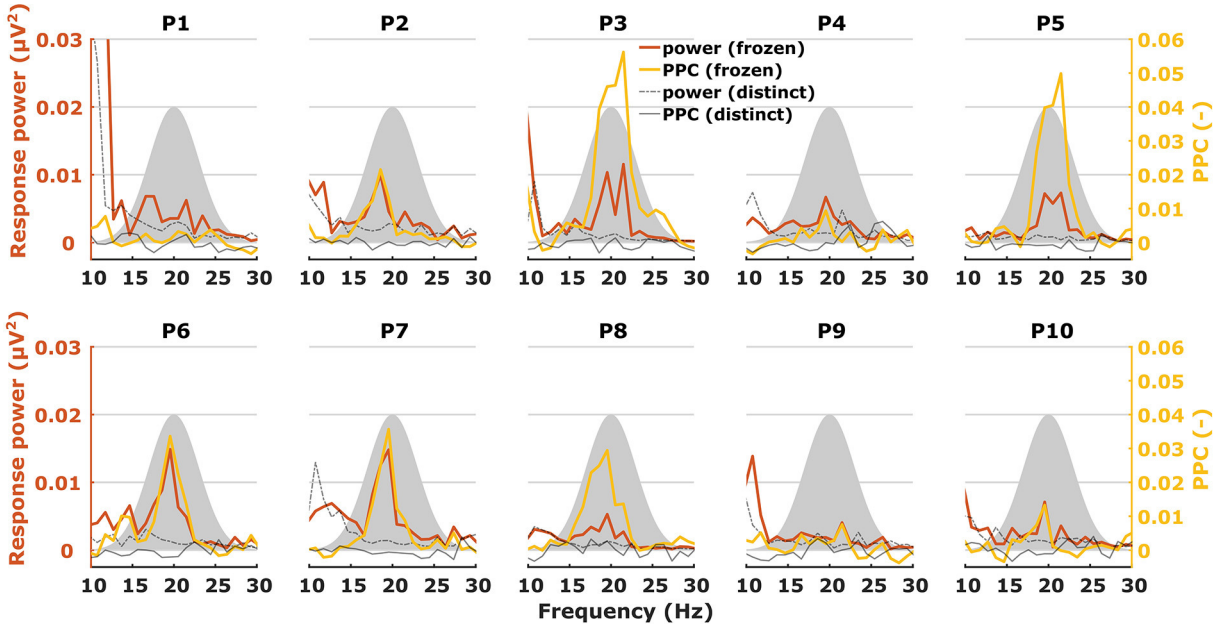


FIGURE 7
Two different patterns, one of power (red) and one of PPC (yellow), plotted over each other across modulation frequency. The patterns are from TEMPEST phonemic-like responses in the left hemispheric channel for each participant. Black patterns describe the baseline from distinct responses. The shaded Gaussian curve represents the modulation frequency distribution of the stimuli (scaled arbitrarily).

TABLE 2 Pearson's correlation coefficients between different electrophysiological metrics of syllabic-like and phonemic-like TEMPEST neural responses in the left and right hemispheric channels.

	Syllabic-like		Phonemic-like	
	Left ($df = 42$)	Right ($df = 39$)	Left ($df = 22$)	Right ($df = 27$)
Power vs. PPC	0.79* (<0.0001)	0.49* (0.0012)	0.31 (0.14)	0.44 (0.017)
SNR vs. PPC	0.85* (<0.0001)	0.70* (<0.0001)	0.86* (<0.0001)	0.91* (<0.0001)
ITPC vs. PPC	>0.99* (<0.0001)	>0.99* (<0.0001)	>0.99* (<0.0001)	>0.99* (<0.0001)
SNR vs. power	0.92* (<0.0001)	0.70* (<0.0001)	0.49 (0.015)	0.54* (0.0025)

Corresponding p-values are reported between parentheses. *Significant after post-hoc Bonferroni correction.

neural representation of the stimulus' envelope with speech-like modulation frequencies. In this study, we aimed to demonstrate a proof-of-concept of the TEMPEST framework to efficiently assess the overall capability of temporal envelope encoding in the auditory pathway. To this end, we investigated whether the neural activity evoked by TEMPEST stimuli corresponds to the speech-weighted electrophysiological TMTF, which is classically measured with ASSRs. We used four different electrophysiological metrics to characterize the neural responses. Two metrics were purely based on power (evoked power and SNR) and two other metrics were purely based on phase (ITPC and PPC) of the individual trials or the averaged trial of the neural response. These metrics were computed for each modulation frequency to obtain neural activity patterns as a function of modulation frequency. This approach is similar to how TMTFs were obtained in other studies using ASSR amplitude (36–39). Comparing the overall neural activity pattern obtained with TEMPEST to the speech-weighted TMTF obtained with ASSRs allowed us to investigate whether they correspond to each other across listeners.

First, we compared neural activity patterns of different metrics with each other for the ASSRs and the TEMPEST responses separately. This is to investigate whether different metrics would reveal different characteristics of the evoked neural activity. A notable case is the almost perfect linear correlation between the ITPC and PPC (Figures 5, 8, bottom left panel) because these two metrics are similar to each other except for a bias due to the number of trials in the ITPC

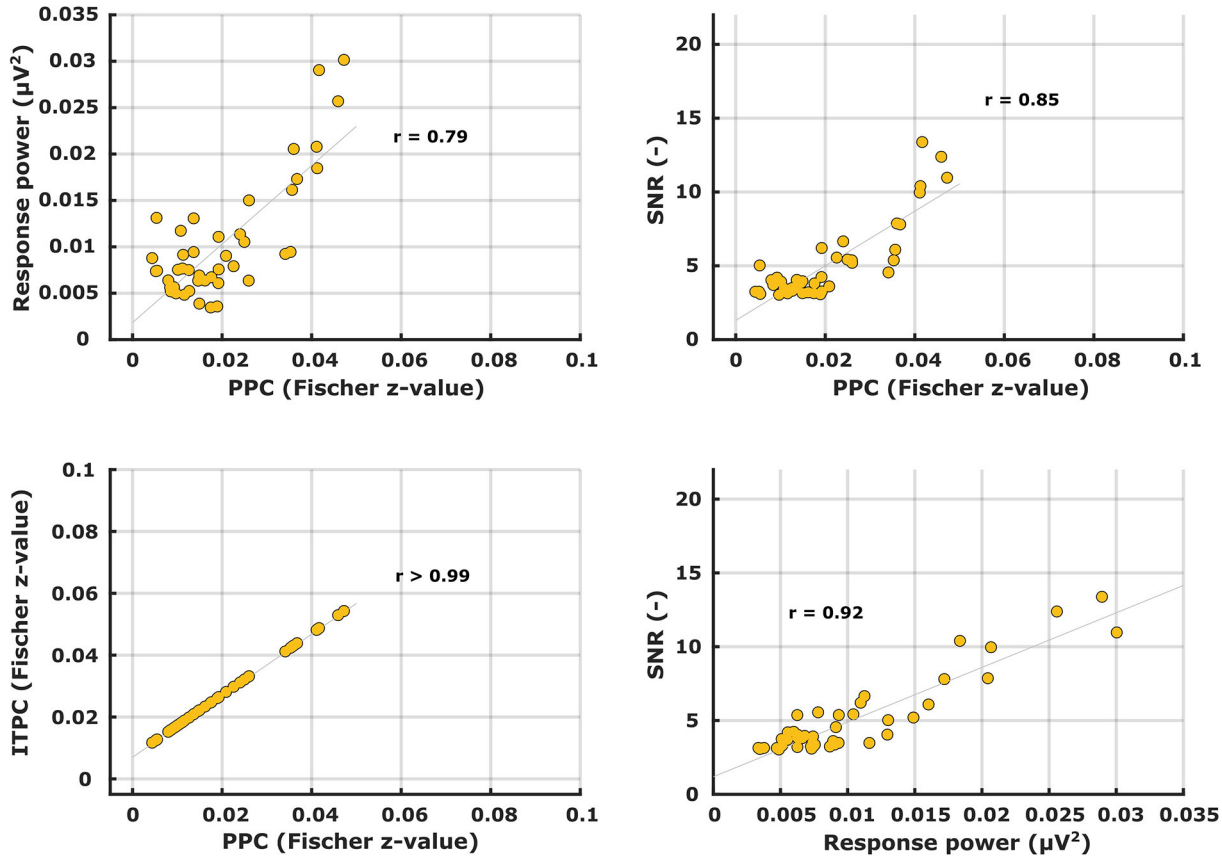


FIGURE 8
Scatter plots between different electrophysiological metrics for syllabic-like TEMPEST neural responses in the left hemispheric channel.

(66). Small deviations occurred because the number of trials was slightly different which resulted in a slightly different bias in ITPC. Furthermore, the bias was relatively small because considerable numbers of trials were used to compute the ITPC (58, 65). Consequently, ITPC results can be exchanged by PPC results without loss of interpretation. As indicated by the high and significant correlation coefficients (Table 1), all individual ASSR TMTF patterns had the same characteristics regardless of the metric used. In the case of TEMPEST activity, syllabic-like patterns of all different metrics significantly correlated with each other. However, phonemic-like patterns of power and PPC did not correlate significantly with each other in both hemispheric channels, and neither did the SNR and power patterns in the left channel (Table 2). Interestingly, the power-based SNR was highly correlated with the phase-based PPC for both ASSRs and TEMPEST responses. While SNR and PPC are based on two independent aspects of the response, i.e., power and phase, the high correlation might be explained by better representation of the phase pattern of the recorded responses due to higher SNR (55, 57). The power metric leads to mostly moderate correlations for both ASSR and TEMPEST stimuli. However, power by itself doesn't tell much about the presence of a response compared to the presence of background noise, which the SNR and PPC can do to a certain extent. This interaction might explain the smaller correlations between power and the other two metrics. Nevertheless, the high correlations also indicate a high similarity of the intersubject variability in envelope modulation processing across modulation frequency across metrics. For example, if a participant showed a large peak of SNR at a certain modulation frequency, then a large peak of PPC is also expected to appear at the same modulation frequency (Figures 5, 8). Therefore, patterns obtained with different electrophysiological metrics, are comparable.

Patterns obtained with TEMPEST stimuli differed across participants which is consistent with the notion that neural phase-locked activity varies considerably across individuals (36). Participants 1, 3, 5, and 9 had relatively large neural activity peaks around 4 Hz when listening to syllabic-like stimuli, while others showed less prominent or no peaks at all. Participants who had prominent syllabic-like neural activity do not necessarily have prominent phonemic-like neural activity as well (e.g., participant 1 in Figures 6, 7), demonstrating

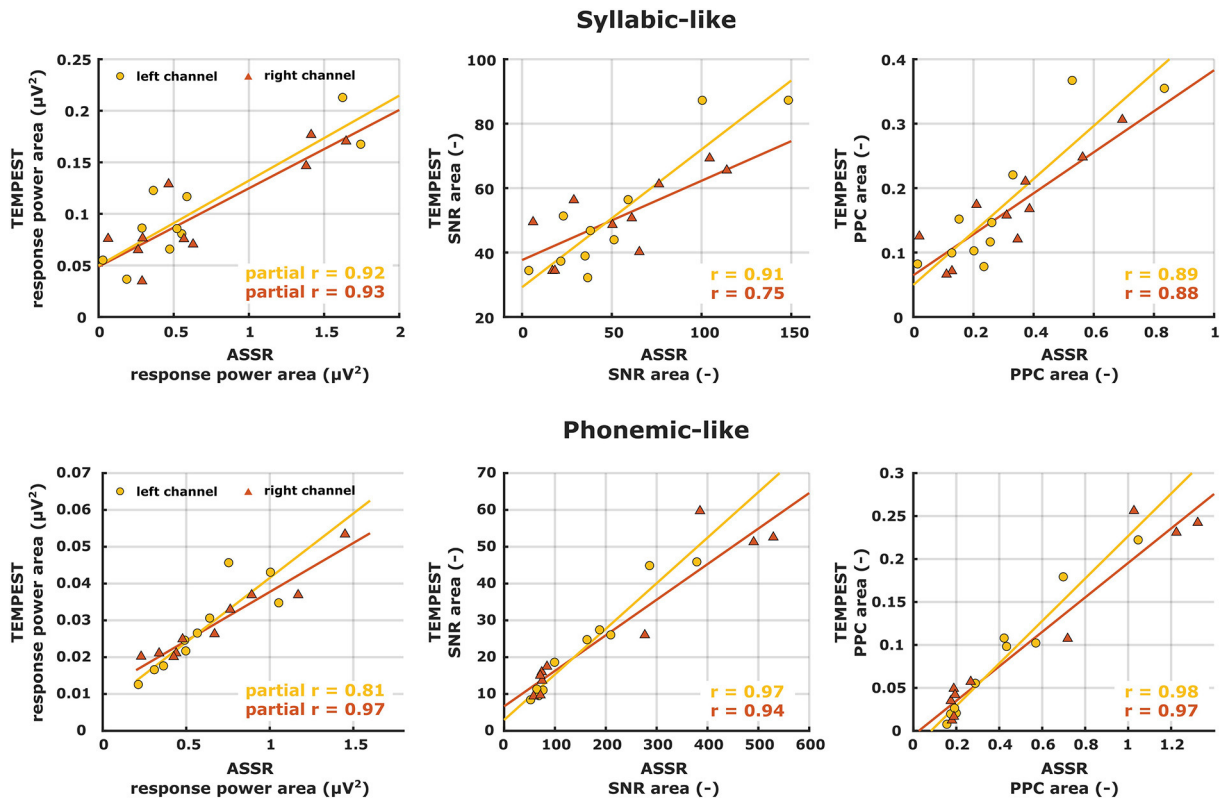


FIGURE 9
Correlation scatterplots between ASSR and TEMPEST area under patterns of the same electrophysiological metric for syllabic-like (first row) and phonemic-like (second row) neural responses. Pearson's partial correlation coefficients are shown only for the power metric whereas Pearson's correlation coefficients are shown for the other metrics in the right bottom corner.

the variability across modulation frequency as well (36). Interestingly, participant 2 had no prominent neural activity around 4 Hz when listening to syllabic-like stimuli, but it was instead shifted up to around 8 Hz. One likely explanation is that the higher harmonics of the envelope modulations were preferentially encoded and/or processed in the auditory system in this participant, which is more likely for such slow modulation frequencies (68, 69).

Some studies used non-speech stimuli with different irregular envelope characteristics and investigated their evoked response using phase coherence metrics (55, 57). Both studies of Teng and colleagues used stimuli with dynamic acoustic changes that occur at timescales similar to our stimuli. They used several different stimuli with dynamics at different timescales, some of which coincided with those of syllables and phonemes in speech. Two of those stimuli were the theta- and gamma-sounds. The theta-sound contained changes at a mean timescale of 190 ms (∼5 Hz modulation rate) which approximately corresponds to the syllable mean modulation frequency of our syllabic-like TEMPEST stimuli. Similarly, the gamma-sound was temporally related to the phoneme rate with a mean timescale of 27 ms (∼37 Hz modulation rate). The authors computed the ITPC of the brain's response for each modulation frequency [note that they used the formula from Lachaux et al. (70), not formula (1) in this study]. Responses evoked by theta sounds showed significantly increased ITPC around 4 Hz and those evoked by gamma sounds around 37 Hz. The peaks that we found in the neural patterns within the modulation frequency range of the TEMPEST stimuli are reminiscent of this finding. Teng and Poeppel (57) also included beta-sounds with mean timescales of 62 and 41 ms (modulation rates of ∼16 and ∼24 Hz, respectively), thus these stimuli are temporally more closely related to our phonemic-like stimuli. However, they reported a considerable decrease in ITPC with beta sounds compared to theta and gamma sounds. In contrast, we did not find a decrease in ITPC and PPC with phonemic-like TEMPEST stimuli compared to syllabic-like TEMPEST stimuli, and similar conclusions can also be made in the case of response power (Figures 6, 7). Another study by Teng and colleagues used complex stimuli with irregular 1/f modulation spectra (56). They investigated robustness of neural phase-locking by comparing ITPC results with frozen

and distinct stimuli ($n = 25$). To this end, the ITPC of the distinct stimuli was subtracted from the frozen ITPC. In a way, this is subtracting the bias from the frozen ITPC and this would be comparable to the PPC. The ITPC difference that they found was at approximately 0.06 in the delta and theta band, which is in line with our syllabic-like results (Figure 6).

Luo and colleagues have also looked at the difference in ITPC between responses evoked by the same (frozen) spoken sentence and responses evoked by different (distinct) sentences (17, 62). ITPC differences of responses to spoken sentences in the delta-theta band are comparable to our syllabic-like PPC results. Another study used mutual information to investigate how much the response phase in the theta band encodes information about the sentence stimulus (71). Peaks of mutual information in the theta band varied across participants, which is in line with the variability in ASSR TMTF for low frequencies (36) and with our results that show variable peaks of activity using syllabic-like stimuli. Additionally, small peaks of mutual information were present in the 22–27 Hz range in some participants and were slightly visible in the grand-average pattern. This frequency range is close to the modulation frequency range of our phonemic-like stimuli. Furthermore, the difference in order of magnitude in mutual information between the theta band and the 22–27 Hz range is similar to the difference that our results exhibit between the syllabic-like and phonemic-like responses. This similarity should be treated with caution because our metrics are not related to mutual information. One thing to keep in mind is that sentence stimuli contain a much wider range of modulation frequencies than our syllabic-like and phonemic-like TEMPEST stimuli.

The main goal of the study was to evaluate whether the global neural activity evoked by TEMPEST stimuli was qualitatively comparable to the speech-weighted overall activity in the electrophysiological TMTF measured with ASSRs. To this end, we computed the area under the patterns of power, SNR, and PPC as a function of modulation frequency of TEMPEST responses and area under the ASSR TMTFs by summing up the values at significant response frequency bins. Before the computation of the area, TMTFs were first weighted with the Gaussian curve of the modulation frequency distribution from the corresponding TEMPEST stimuli. We then computed same-metric correlation coefficients between these areas across participants. All correlations between ASSR and TEMPEST were found to be strong and significant except for the SNR in the right hemispheric channel (Figure 9). These significantly high correlations indicate that the neural activity evoked by TEMPEST stimuli is comparable to those of the speech-weighted TMTF measured with the classical ASSR paradigm. Furthermore, they also show that the variability in the global neural patterns across listeners as measured with TEMPEST stimuli is similar to that found with ASSR TMTFs, which is consistent with the findings by (36). Consequently, evoked TEMPEST responses characterized by any of the three metrics (power, SNR, or PPC) can be used as an indicator of individual neural temporal processing capability within the modulation frequency band of interest.

Although our approach of computing the area under the patterns of TEMPEST neural activity and the TMTF does not consider the exact pattern shapes, we found that the overall activity evoked by TEMPEST stimuli strongly corresponds to the overall activity found in the electrophysiological TMTF. This result is a clear indication that the TEMPEST framework has the potential to evaluate temporal envelope processing in the auditory pathway. Furthermore, since TEMPEST stimuli contain a range of envelope modulations as determined by an a-priori modulation frequency distribution, individual distribution-weighted electrophysiological TMTFs can be efficiently determined, which would otherwise be measured by multiple SAM stimuli, as is clear from Figures 6, 7. Further research on variations of TEMPEST stimuli and improvement of the neurophysiological analyses can potentially push the TEMPEST framework to more clinical usability. Moreover, the TEMPEST framework provides many possibilities to generate TEMPEST stimuli that are parameterized, for example, by a modulation frequency distribution, a modulation depth distribution, window shape with optionally varying parameters, etc. Furthermore, the framework also allows for more complex stimuli such as nesting of two or more TEMPEST envelopes (51), which combines multiple TEMPEST stimuli with different modulation frequency distributions into one stimulus. This approach would be comparable to combining multiple SAM stimuli at different carrier frequencies and is commonly used to electrophysiologically determine frequency-specific hearing thresholds in infants (72).

Conclusion

The TEMPEST framework (51) provides stimuli that evoke neural phase-locked activity with the same characteristics as the electrophysiological TMTF classically measured with ASSRs after weighting by the TEMPEST distribution. Since TEMPEST stimuli contain a range of envelope modulation frequencies in contrast to single-frequency SAM stimuli, they can be used to efficiently probe temporal envelope processing in the auditory pathway. Any of the four electrophysiological metrics (evoked power, SNR, ITPC, or PPC) can be used to evaluate the degree of neural tracking to amplitude-modulated stimuli. Moreover, TEMPEST stimuli that contain speech-like modulations (such as the syllable and the phoneme rate in speech) have the potential to provide a better understanding of the role of neural envelope processing in speech perception. Not only that, but they could also

potentially capture differences in temporal envelope processing in different listener groups with different types of auditory processing deficits. Future work would further investigate the potential of the TEMPEST framework using more complex stimuli by varying several other envelope parameters or combining different stimuli into one stimulus with multiple bands of modulation frequencies, and explore different analysis techniques to exploit its full potential in the neuroscientific and audiological fields.

Data availability statement

The data generated and/or analyzed during the current study are not publicly available for legal/ethical reasons, they can be obtained on reasonable request to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethical Committee of the University Hospitals and University of Leuven. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was partly funded by a research grant from Flanders Innovation & Entrepreneurship through the VLAIO research grant HBC.20192373, partly by a Wellcome Trust Collaborative Award in Science RG91976 to Robert P. Carlyon, John C. Middlebrooks, and JW, and partly by an SB Ph.D. grant 1S34121N from the Research Foundation Flanders (FWO) awarded to WD.

Acknowledgments

We are very grateful to all participants for their help in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Plomp R. The role of modulation in hearing. In: Klinke R, Hartmann R, editors. *HEARING—Physiological Bases and Psychophysics*. Berlin: Springer (1983).

2. Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci.* (1992) 336:367–73. doi: 10.1098/rstb.1992.0070

3. Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. *Neurosci Biobehav Rev.* (2017) 81:181–7. doi: 10.1016/j.neubiorev.2017.02.011

4. Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C. A cross-linguistic study of speech modulation spectra. *J Acoust Soc Am.* (2017) 142:1976–89. doi: 10.1121/1.5006179

5. Greenberg S, Carvey H, Hitchcock L, Chang S. Temporal properties of spontaneous speech - a syllable-centric perspective. *J Phon.* (2003) 31:465–85. doi: 10.1016/j.wocn.2003.09.005

6. Goswami U, Leong V. Speech rhythm and temporal structure: converging perspectives? *Lab Phonol.* (2013) 4:67–92. doi: 10.1515/lp-2013-0004

7. Greenberg S. Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Commun.* (1999) 29:159–76. doi: 10.1016/S0167-6393(99)00050-3

8. Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am.* (1994) 95:2670–80. doi: 10.1121/1.409836

9. Shannon R V, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science.* (1995) 270:303–4. doi: 10.1126/science.270.5234.303

10. Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature.* (2002) 416:87–90. doi: 10.1038/416087a

11. Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, et al. Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci U S A.* (2005) 102:2293–8. doi: 10.1073/pnas.0406460102

12. Friesen LM, Shannon R V, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am.* (2001) 110:1150–63. doi: 10.1121/1.1381538

13. Peelle JE, Davis MH. Neural oscillations carry speech rhythm through to comprehension. *Front Psychol.* (2012) 3:1–17. doi: 10.3389/fpsyg.2012.00320

14. Abrams DA, Nicol T, Zecker S, Kraus N. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci.* (2008) 28:3958–65. doi: 10.1523/JNEUROSCI.0187-08.2008

15. Aiken SJ, Picton TW. Human cortical responses to the speech envelope. *Ear Hear.* (2008) 29:139–57. doi: 10.1097/AUD.0b013e31816453dc

16. Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A.* (2001) 98:13367–72. doi: 10.1073/pnas.201400998

17. Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron.* (2007) 54:1001–10. doi: 10.1016/j.neuron.2007.06.004

18. Howard MF, Poeppel D. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol.* (2010) 104:2500–11. doi: 10.1152/jn.00251.2010

19. Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T. Speech intelligibility predicted from neural entrainment of the speech envelope. *J Assoc Res Otolaryngol.* (2018) 19:181–91. doi: 10.1007/s10162-018-0654-z

20. Ding N, Simon JZ. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J Neurosci.* (2013) 33:5728–35. doi: 10.1523/JNEUROSCI.5297-12.2013

21. Decruy L, Vanthornhout J, Francart T. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. *J Neurophysiol.* (2019) 122:601–15. doi: 10.1152/jn.00687.2018

22. Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E. Neural entrainment to speech modulates speech intelligibility. *Curr Biol.* (2018) 28:161–9.e5. doi: 10.1016/j.cub.2017.11.033

23. Di Liberto GM, O'Sullivan JA, Lalor EC. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr Biol.* (2015) 25:2457–65. doi: 10.1016/j.cub.2015.08.030

24. Peelle JE, Gross J, Davis MH. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb Cortex.* (2013) 23:1378–87. doi: 10.1093/cercor/bhs118

25. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* (2013) 11:e1001752. doi: 10.1371/journal.pbio.1001752

26. Molinaro N, Lizarazu M. Delta (but not theta)-band cortical entrainment involves speech-specific processing. *Eur J Neurosci.* (2018) 48:2642–50. doi: 10.1111/ejn.13811

27. Bonhage CE, Meyer L, Gruber T, Friederici AD, Mueller JL. Oscillatory EEG dynamics underlying automatic chunking during sentence processing. *Neuroimage.* (2017) 152:647–57. doi: 10.1016/j.neuroimage.2017.03.018

28. Etard O, Reichenbach T. Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *J Neurosci.* (2019) 39:5750–9. doi: 10.1523/JNEUROSCI.1828-18.2019

29. Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci.* (2016) 19:158–64. doi: 10.1038/nn.4186

30. Getz H, Ding N, Newport EL, Poeppel D. Cortical tracking of constituent structure in language acquisition. *Cognition.* (2018) 181:135–40. doi: 10.1016/j.cognition.2018.08.019

31. Houtgast T, Steeneken HJM. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J Acoust Soc Am.* (1985) 77:1069–77. doi: 10.1121/1.392224

32. Obleser J, Herrmann B, Henry MJ. Neural oscillations in speech: don't be enslaved by the envelope. *Front Hum Neurosci.* (2012) 6:2008–11. doi: 10.3389/fnhum.2012.00250

33. Ding N, Simon JZ. Cortical entrainment to continuous speech: functional roles and interpretations. *Front Hum Neurosci.* (2014) 8:1–7. doi: 10.3389/fnhum.2014.00311

34. Zoefel B, Vanrullen R. The role of high-level processes for oscillatory phase entrainment to speech sound. *Front Hum Neurosci.* (2015) 9:1–12. doi: 10.3389/fnhum.2015.00651

35. Ramus F, Nespor M, Mehler J. Correlates of linguistic rhythm in the speech signal. *Cognition.* (2000) 75:265–92. doi: 10.1016/S0010-0277(99)00058-X

36. Gransier R, Hofmann M, Wieringen A Van, Wouters J. Stimulus-evoked phase-locked activity along the human auditory pathway strongly varies across individuals. *Sci Rep.* (2021) 11:143. doi: 10.1038/s41598-020-80229-w

37. Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am.* (2004) 116:3581–93. doi: 10.1121/1.1798354

38. Poulsen C, Picton TW, Paus T. Age-related changes in transient and oscillatory brain responses to auditory stimulation during early adolescence. *Dev Sci.* (2009) 12:220–35. doi: 10.1111/j.1467-7687.2008.00760.x

39. Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C. A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J Acoust Soc Am.* (2000) 108:679–91. doi: 10.1121/1.429600

40. Picton TW, John MS, Dimitrijevic A, Purcell D. Human auditory steady-state responses. *Int J Audiol.* (2003) 42:177–219. doi: 10.3109/14992020309101316

41. Darestani EF, Goossens T, Wouters J, van Wieringen A. Spatiotemporal reconstruction of auditory steady-state responses to acoustic amplitude modulations: Potential sources beyond the auditory pathway. *Neuroimage.* (2017) 148:240–53. doi: 10.1016/j.neuroimage.2017.01.032

42. Herdman AT, Lins O, Van Roon P, Stapells DR, Scherg M, Picton TW. Intracerebral sources of human auditory steady-state responses. *Brain Topogr.* (2002) 15:69–86. doi: 10.1023/A:1021470822922

43. Luke R, De Vos A, Wouters J. Source analysis of auditory steady-state responses in acoustic and electric hearing. *Neuroimage.* (2017) 147:568–76. doi: 10.1016/j.neuroimage.2016.11.023

44. Bidelman GM. Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hear Res.* (2015) 323:68–80. doi: 10.1016/j.heares.2015.01.011

45. Gransier R, Luke R, Van Wieringen A, Wouters J. Neural modulation transmission is a marker for speech perception in noise in cochlear implant users. *Ear Hear.* (2020) 41:591–602. doi: 10.1097/AUD.0000000000000783

46. Leigh-Paffenroth ED, Fowler CG. Amplitude-modulated auditory steady-state responses in younger and older listeners. *J Am Acad Audiol.* (2006) 17:582–97. doi: 10.3766/jaaa.17.8.5

47. Dimitrijevic A, John MS, Picton TW. Auditory steady-state responses and word recognition scores in normal-hearing and hearing- impaired adults. *Ear Hear.* (2004) 25:68–84. doi: 10.1097/01.AUD.0000111545.71693.48

48. Goossens T, Vercammen C, Wouters J, van Wieringen A. Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. *Hear Res.* (2018) 370:189–200. doi: 10.1016/j.heares.2018.07.012

49. Poelmans H, Luts H, Vandermosten M, Boets B, Ghesquière P, Wouters J. Auditory steady state cortical responses indicate deviant phonemic-rate processing in adults with dyslexia. *Ear Hear.* (2012) 33:134–43. doi: 10.1097/AUD.0b013e31822c26b9

50. Alaerts J, Luts H, Hofmann M, Wouters J. Cortical auditory steady-state responses to low modulation rates. *Int J Audiol.* (2009) 48:582–93. doi: 10.1080/14992020902894558

51. Gransier R, Wouters J. Neural auditory processing of parameterized speech envelopes. *Hear Res.* (2021) 412:108374. doi: 10.1016/j.heares.2021.108374

52. Hofmann M, Wouters J. Improved electrically evoked auditory steady-state response thresholds in humans. *J Assoc Res Otolaryngol.* (2012) 13:573–89. doi: 10.1007/s10162-012-0321-8

53. 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 Audiol.* (2008) 47:348–55. doi: 10.1080/14992020801895144

54. Crystal TH, House AS. Segmental durations in connected-speech signals: current results. *J Acoust Soc Am.* (1988) 83:1553–73. doi: 10.1121/1.395911

55. Teng X, Tian X, Rowland J, Poeppel D. Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biol.* (2017) 15:1–29. doi: 10.1371/journal.pbio.2000812

56. Teng X, Tian X, Doelling K, Poeppel D. Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. *Eur J Neurosci.* (2018) 48:2770–82. doi: 10.1111/ejn.13742

57. Teng X, Poeppel D. Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cereb Cortex.* (2020) 30:2600–14. doi: 10.1093/cercor/bhz263

58. Bastos AM, Schoffelen JM. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front Syst Neurosci.* (2016) 9:1–23. doi: 10.3389/fnsys.2015.00175

59. Jasper HH. The ten twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol.* (1957) 10:371–5.

60. Gransier R, van Wieringen A, Wouters J. Binaural interaction effects of 30–50 Hz auditory steady state responses. *Ear Hear.* (2017) 38:e305–15. doi: 10.1097/AUD.0000000000000429

61. Cohen MX. *Analyzing Neural Time Series Data*. MIT Press (2014).

62. Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* (2010) 8:25–6. doi: 10.1371/journal.pbio.1000445

63. Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S. Rhythmic auditory cortex activity at multiple timescales shapes stimulus–response gain and background firing. *J Neurosci.* (2015) 35:7750–62. doi: 10.1523/JNEUROSCI.0268-15.2015

64. VanRullen R. How to evaluate phase differences between trial groups in ongoing electrophysiological signals. *Front Neurosci.* (2016) 10:1–22. doi: 10.3389/fnins.2016.00426

65. Van Diepen RM, Mazaheri A. The caveats of observing inter-trial phase-coherence in cognitive neuroscience. *Sci Rep.* (2018) 8:1–9. doi: 10.1038/s41598-018-20423-z

66. Vinck M, van Wingerden M, Womelsdorf T, Fries P, Pennartz CMA. The pairwise phase consistency: a bias-free measure of rhythmic neuronal synchronization. *Neuroimage.* (2010) 51:112–22. doi: 10.1016/j.neuroimage.2010.01.073

67. Hotelling H. The generalization of student's ratio. *Ann Math Stat.* (1931) 2:360–78. doi: 10.1214/aoms/1177732979

68. Tlumak AI, Durrant JD, Delgado RE, Boston JR. Steady-state analysis of auditory evoked potentials over a wide range of stimulus repetition rates: profile in adults. *Int J Audiol.* (2011) 50:448–58. doi: 10.3109/14992027.2011.560903

69. Tlumak AI, Durrant JD, Delgado RE, Boston JR. Steady-state analysis of auditory evoked potentials over a wide range of stimulus repetition rates: Profile in children vs. adults. *Int J Audiol.* (2012) 51:480–90. doi: 10.3109/14992027.2012.664289

70. Lachaux J-P, Rodriguez E, Martinerie J, Varela FJ. Measuring phase synchrony in brain signals. *Hum Brain Mapp.* (1999) 8:194–208. doi: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

71. Cogan GB, Poeppel D. A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J Neurophysiol.* (2011) 106:554–563. doi: 10.1152/jn.00075.2011

72. John MS, Picton TW, MASTER. A windows program for recording multiple auditory steady-state responses. *Comput Methods Programs Biomed.* (2000) 61:125–50. doi: 10.1016/S0169-2607(99)00035-8

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fneur.2022.905017

OPEN ACCESS

EDITED BY
Stephanie Clarke,
Centre Hospitalier Universitaire Vaudois (CHUV), Switzerland

REVIEWED BY
Alessandro Presacco,
University of Maryland, College Park, United States
Aravindakshan Parthasarathy,
University of Pittsburgh, United States
Hanin Karawani,
University of Haifa, Israel

*CORRESPONDENCE
Ehsan Darestani Farahani
e.darestani@gmail.com

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 26 March 2022
ACCEPTED 11 July 2022
PUBLISHED 05 August 2022

CITATION
Farahani ED, Wouters J and van Wieringen A (2022) Age-related hearing loss is associated with alterations in temporal envelope processing in different neural generators along the auditory pathway. *Front. Neurol.* 13:905017. doi: 10.3389/fneur.2022.905017

Age-related hearing loss is associated with alterations in temporal envelope processing in different neural generators along the auditory pathway

Ehsan Darestani Farahani*, Jan Wouters and Astrid van Wieringen

Research Group Experimental ORL, Department Neurosciences, KU Leuven, Leuven, Belgium

People with age-related hearing loss suffer from speech understanding difficulties, even after correcting for differences in hearing audibility. These problems are not only attributed to deficits in audibility but are also associated with changes in central temporal processing. The goal of this study is to obtain an understanding of potential alterations in temporal envelope processing for middle-aged and older persons with and without hearing impairment. The time series of activity of subcortical and cortical neural generators was reconstructed using a minimum-norm imaging technique. This novel technique allows for reconstructing a wide range of neural generators with minimal prior assumptions regarding the number and location of the generators. The results indicated that the response strength and phase coherence of middle-aged participants with hearing impairment (HI) were larger than for normal-hearing (NH) ones. In contrast, for the older participants, a significantly smaller response strength and phase coherence were observed in the participants with HI than the NH ones for most modulation frequencies. Hemispheric asymmetry in the response strength was also altered in middle-aged and older participants with hearing impairment and showed asymmetry toward the right hemisphere. Our brain source analyses show that age-related hearing loss is accompanied by changes in the temporal envelope processing, although the nature of these changes varies with age.

KEYWORDS

age-related hearing loss (ARHL), neural generators, auditory temporal processing, auditory steady-state response (ASSR), EEG

Introduction

Speech perception of individuals with hearing impairment (HI) is worse than that of persons with normal audiometric thresholds (NH), even after correcting for differences in hearing audibility (1–4). In addition to deficits in audibility, changes in central auditory processing, and in particular temporal processing, account for impaired speech perception of individuals with HI (5). Electrophysiological studies in animals have shown

that HI is associated with increased neural responses to amplitude-modulated stimuli in the auditory nerve fibers (6–8) and the midbrain (9). Similarly, human studies showed enhanced neural responses in the brainstem of adults around 60 years old with HI compared to NH ones in the same age range (10, 11).

The temporal envelope of speech (slow fluctuations of 2 to 50 Hz) is crucial for accurate speech understanding (12–14) and transmits both prosodic and linguistic information (15). Speech envelopes are encoded in the central auditory system through synchronized (phase-locked) neural activity (16, 17). Temporal envelope processing can be assessed through the auditory steady-state responses (ASSRs; 16). ASSRs are auditory-evoked responses to periodically varying acoustic stimuli and reflect the ability of the auditory system to follow the temporal envelope of sounds (18).

In our previous study (19), we investigated age-related changes in the activity of subcortical and cortical neural generators of ASSRs in middle-aged and older persons with normal audiometric thresholds (<25 dB HL). Analyses showed enhanced neural responses for older adults compared to younger ones for relatively slow modulations (<50 Hz). However, for faster modulations (i.e., 80 Hz), the neural responses were reduced for older adults compared to younger ones. While these age-related changes occur in persons with normal hearing, it remains unclear how HI affects temporal envelope processing. Aging is typically accompanied by decreasing audiometric thresholds in the high frequencies (presbycusis). These peripheral changes are accompanied by changes in the central auditory system (10, 20) and associated neural generators. The current study focuses on the potential aggravating role of HI on the activity of the neural generators for middle-aged and older adults.

Electrophysiological studies investigating how HI affects the processing of the temporal envelope demonstrated enhanced response strengths for middle-aged listeners with HI compared to middle-aged NH ones [∼60 years old; (10, 11, 21, 22)]. In contrast to middle-aged persons with HI, older adults with HI (∼75 years old) did not show enhanced responses to acoustic modulations (11). Note that stimulus audibility has been corrected in these studies. The absence of an enhanced response in older persons with HI could be because a significant neural enhancement had already been observed with NH older listeners and was, therefore, more a factor of aging than HI. However, how HI affects the temporal envelope processing in the different neural generators in middle-aged and older adults remains unclear. Sensor-level analysis (i.e., analysis based on the scalp's data) may not be sensitive enough to reveal all the dynamics of the neural generators underlying temporal envelope processing in persons with HI. This is because the recorded data at each sensor are a weighted average of the activity of several neural generators due to the volume conduction of the brain tissue.

On the other hand, brain source analysis estimates the original activity of each neural generator using computational modeling. In the current study, we use a source reconstruction approach based on minimum-norm imaging (MNI). In this approach, a large number of equivalent current dipoles in the brain are considered. Then, the amplitudes of all dipoles (for each time point) are estimated to reconstruct a source distribution map with minimum overall energy (23, 24).

The MNI approach imposes minimal restrictions about the number and location of the sources, contrary to more common methods like dipole source analysis, which makes prior assumptions regarding the number and location of the sources. Another advantage of the MNI approach is the ability to reconstruct a wide range of cortical and subcortical sources simultaneously (25). The beamforming method, another well-known method of brain source reconstruction, has more difficulty in reconstructing the cortical and subcortical sources. To reconstruct neural generators of ASSRs using beamforming methods, a supplementary preprocessing is necessary to suppress the correlated source from the other hemisphere (26–28). Additionally, the beamforming approaches cannot simultaneously reconstruct the cortical and subcortical sources.

Age-related hearing loss may also affect hemispheric asymmetry in temporal envelope processing. Previous data have shown that the pattern of neural synchronization in older adults with normal audiometric thresholds is symmetrical across hemispheres, while that of young NH adults is asymmetric (29, 30). With age, this altered hemispheric asymmetry is in line with the HAROLD model (31), which states that hemispheric asymmetry is reduced in older people compared to younger ones. Using brain source analyses, Farahani et al. (19) also showed that hemispheric asymmetry is reduced for NH older adults compared to younger normal hearing in response to the 20 and 80 Hz amplitude-modulated stimuli. However, age-related hearing loss may affect hemispheric asymmetry on top of age, as has been demonstrated for linguistic processing (32). In their sensor-level EEG study, Goossens et al. (11) observed a hemispheric asymmetry toward the right hemisphere for older participants with HI. The observed changes in hemispheric asymmetry in persons with HI are possibly due to anatomical changes related to presbycusis, such as reduced integrity of white matter tracts (33). However, it is also possible that the sensor-level analysis cannot capture changes related to HI in the other cohorts. It is expected that source-level analysis, due to the higher sensitivity explained before, might reflect more changes associated with HI concerning the hemispheric asymmetry than the sensor-level analysis.

The current study aims to investigate potential changes in temporal envelope processing for subcortical and cortical neural generators along the auditory pathway in middle-aged and older persons with age-related HI compared to normal-hearing ones. Different studies have shown that the diminished cochlear output of people with HI, due to hair cell loss and/or

synaptopathy, activates various mechanisms to increase central gain and preserve neural excitability (e.g., 32, 33). Hence, we hypothesize that the neural generators of ASSRs in middle-aged listeners with HI will show enhanced response strength compared to those with NH. However, we do not expect such an enhancement in older adults with HI because older adults with NH already exhibit compensatory mechanisms of increasing neural excitability and central gain (19, 34, 35). Concerning hemispheric asymmetry in temporal envelope processing, we hypothesize that the reconstructed activity at the auditory cortex reveals an altered pattern of hemispheric asymmetry in listeners with HI. However, these alterations may vary with age and stimulation conditions.

We investigate the potential alterations during temporal envelope processing of people with HI when stimulus audibility was corrected for. We look into ASSRs' cortical and subcortical neural generators along the auditory pathway in young, middle-aged, and older persons with and without HI. The activity of these neural generators is reconstructed using a minimum-norm imaging (MNI) approach (25). To investigate the response strength and the phase-locking to the stimulus, the ASSR amplitude and phase coherence are calculated for each neural generator. This is done for ASSRs in response to 4, 20, 40, and 80 Hz acoustic modulations presented separately to the left and right ears. The acoustic modulations at 4 and 20 Hz were presented as a model of the temporal envelope of syllables and phonemes, respectively. The modulation frequencies of 40 and 80 Hz were also selected because these modulations can activate more subcortical neural generators than cortical ones (26, 36). Potential alterations in hemispheric asymmetry are also investigated for the neural generators in the left and right auditory cortices (31, 37).

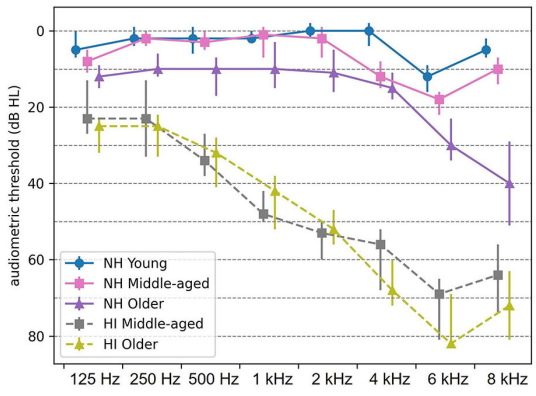


FIGURE 1
Median audiometric thresholds (dB HL) of normal-hearing (NH) and hearing-impaired (HI) participants, averaged across both ears. Thresholds are indicated by circles, squares, and triangles for young, middle-aged, and older persons, respectively. Error bars indicate the interquartile range. NH, normal hearing; HI, hearing impaired.

Materials and methods

Participants

The EEG data were adopted from Goossens et al. (29). Participants were either NH or with HI in three narrow age cohorts, including 19 young (20–30 years, nine men), 20 middle-aged (50–60 years, ten men), and 16 older adults (70–80 years, five men) in NH group and 14 middle-aged (50–60 years, four men) and 13 older adults (70–80 years, five men) with HI. Only individuals who showed symmetrical hearing based on the criteria of the audiogram classification system (38) were eligible for participation. The participants in the NH group had audiometric thresholds within normal limits [≤25 dB HL] at all octave frequencies from 125 Hz up to and including 4 kHz in both ears (Figure 1). However, the participants with HI had audiometric thresholds higher than 35 dB HL from 1 kHz onward (Figure 1). All middle-aged and older participants with HI were diagnosed with age-related hearing loss (i.e., presbycusis) and used hearing aids in both ears. To avoid cognitive impairment as a confounder, only adults who showed no indication of cognitive impairment were recruited. The participants were screened using the Montreal Cognitive Assessment Task (39), and the cutoff score was 26 out of 30. This screening with a stringent cutoff score ensured that all participants had cognitive capacities within the normal range. All participants were Dutch native speakers. They were right-handed based on the Edinburgh Handedness Inventory (40), and none of them had a medical history of brain injury, neurological disorders, or tinnitus.

Stimuli

The acoustic stimuli were amplitude-modulated (AM) noise at 4, 20, 40, and 80 Hz and generated in MATLAB (The MathWorks, Inc.). The white noise (bandwidth of 1 octave, centered at 1 kHz) was sinusoidally modulated with a modulation depth of 100%. The modulation frequencies were adjusted to ensure that there was an integer number of cycles in an epoch of 1.024 s (41).

Loudness balancing

The stimuli were presented *via* ER-3A insert phones to the left ear and the right ear. Each stimulus type was presented for 300 s continuously. For NH participants, the stimuli were presented at 70 dB SPL which they rated as comfortably loud. For participants with HI, no hearing aids were used during EEG recording. To correct for the audibility of listeners with

HI, each individual was asked to adjust the intensity level until he/she perceived it as comfortably loud, similar to the NH participants. This arrangement allowed us to present stimuli to all participants at equal loudness levels. There were two reasons for using equal loudness levels to correct for stimulus audibility instead of equal sensation levels. First, the equal sensation level for participants with HI reaches ∼108 dB SPL, which exceeds their uncomfortable loudness level (∼103 dB SPL). Second, it was shown that the magnitude of the ASSR was highly correlated with the perceived loudness of the acoustic modulations (42, 43). So, the equal loudness level is an effective way to control for differences in stimulus audibility between NH and HI.

Experiment protocol and EEG recordings

The experiment was conducted in a double-walled soundproof booth with a Faraday cage. The experiment procedure was arranged to ensure passive listening to acoustic stimuli during a wakeful state. During acoustic stimulation, the participants were asked to lay down on a bed and watch a muted movie with subtitles *via* a 21-inch LCD monitor with 60 Hz vertical refresh rate. All participants were encouraged to lie quietly and relaxed during the experiment to avoid movements and muscle artifacts caused by fatigue, especially in older adults. We used a large-size and very soft pillow to support the neck and backside of the head.

The EEG data were recorded using the BioSemi ActiveTwo system (BioSemi B.V., Amsterdam, the Netherlands, 2010) with 64 active electrodes. The electrodes were fixed in a head cap according to the 10–10 electrode system. The EEG signals were amplified and digitized at a sampling rate of 8,192 Hz with a gain of 32.25 nV/bit. The recording system used a built-in low-pass filter with a cutoff frequency of 1,638 Hz.

EEG source analysis

The activity of the neural generators of ASSRs along the auditory pathway was reconstructed using a method based on MNI, which was suggested for ASSR source analysis (25). An overview of this method is given below [for more details, see (25)]. The analyses were performed in MATLAB R2016b (MathWorks).

Preprocessing

To eliminate the low-frequency distortions and drift of the amplifier, the EEG data were filtered by a zero-phase high-pass filter with a cutoff frequency of 2 Hz (Butterworth, second order, 12 dB/octave). The filtered EEG data were split into epochs of 1.024 s. Subsequently, 10% of epochs with the highest peak-to-peak amplitude across channels were rejected for early noise reduction.

Afterward, the EEG data were re-referenced to a common average over all channels and epochs. To eliminate artifacts caused by eye movements, eye blinks, and heartbeats, we used independent component analysis (ICA) based on the Infomax algorithm implemented in the FieldTrip toolbox (44). The noisy components were identified with a visual inspection. In the end, the remaining artifacts not recognized by ICA were identified and eliminated using a threshold level of 70 μV for the maximum absolute amplitude of each epoch. To have a similar effect on the group-wise results, we kept the same number of epochs across participants. The first 192 artifact-free epochs (six sweeps of 32 epochs) were preserved for subsequent analyses to keep the same number of epochs across participants. We chose not to use a lower number of epochs in each sweep to keep our frequency resolution high enough (each frequency bin corresponds to 0.03 Hz). In case we could not find 192 epochs (six sweeps of 32 epochs) for a participant, then we gradually increased the threshold (step of 5 μV) up to 110 μV. These epochs were selected out of 300 epochs of each participant per condition. For the topographic map of ASSRs, see Farahani et al. (45).

Source reconstruction and developing ASSR map

Mixed head model

A mixed head model consisting of cortical and subcortical regions was generated to reconstruct the neural generators along the auditory pathway. This head model was generated using the boundary element method (BEM), as implemented in OpenMEEG (46). To this end, we used the template brain scan of ICBM152 (47) and the default channel location file in the Brainstorm application (48, 49).

Data averaging for group-wise analyses

Since the head model was generated based on a template brain scan, we used a group-wise framework in our source analyses instead of individual-level analyses to have a high localization accuracy (45). So, the preprocessed epochs of each participant were divided into sweeps of 32 concatenated epochs and averaged across all participants. The outcome grand-averaged sweep was used for source reconstruction.

Reconstruction source map of EEG in time domain

The distribution map of brain activity at each time point was estimated using dynamic statistical parametric mapping [dSPM; (50)] implemented in the Brainstorm application (48, 49). In the dSPM method, the standard minimum-norm solution is normalized with the estimated noise at each source (24). This noise normalization eliminates the bias toward superficial

sources, which is accompanied by the standard minimum-norm solution (24, 51).

Noise covariance matrix

The noise covariance matrix required for dSPM was calculated based on the EEG recorded in the absence of auditory stimulation. The silence EEG of participants was filtered by a zero-phase band-pass filter with a bandwidth of 4 Hz and modulation frequency as center frequency and concatenated before calculating the covariance matrix.

Regularization parameter

For each experimental condition (i.e., stimulation type, age group, and hearing status), the regularization parameter (λ^2) required for dSPM was specifically determined based on:

Equation 1

$$\lambda^2 = \frac{1}{\mathbf{SNR}^2_{\mathbf{scalp}}}$$

where SNR_{scalp} is the signal-to-noise ratio (based on the amplitude) of the whitened EEG data (52–54). The fast Fourier transform (FFT) was applied for each channel, and the magnitude of the spectrum at the modulation frequency was considered the ASSR strength. The highest response magnitude across channels was assigned to the signal of interest (19). The EEG background noise was estimated based on the average magnitude of 30 neighboring frequency bins on the left and the right sides of the response frequency bin. The median of the EEG background noise across channels was used as noise level for calculating SNR_{scalp} (25).

Generating ASSR map

ASSR map shows the magnitude of the response for different regions of the brain. To generate an ASSR map, the waveform of each dipole was transformed to the frequency domain using FFT. Then, for each dipole, the SNR of the ASSR was calculated according to Equation 2.

Equation 2

$$\mathbf{SNR(dB)} = \mathbf{10}\left(\frac{\mathbf{P_{S+N}}}{\mathbf{P_N}}\right)$$

where P_{S+N} is the power of the spectrum at the modulation frequency, which shows the power of the steady-state response plus neural background noise. P_N indicates the power of the neural background noise, which was estimated using the average power of 30 neighboring frequency bins (corresponding to 0.92 Hz) on each side of the modulation frequency bin.

The one-sample f-test based on SNR was employed to recognize the dipoles with significant ASSRs (43, 55). Results were corrected for multiple comparisons using the false discovery rate (FDR) method (56). Finally, the ASSR map illustrating ASSR amplitudes for dipoles with significant responses and zero for the dipoles with no significant responses was generated. The ASSR amplitude was calculated using Equation 3. For subcortical regions, the activity at each point was reconstructed using three orthogonal dipoles (across x, y, and z). The ASSR amplitude for subcortical regions was calculated based on Equation 4. A detailed explanation and a sample ASSR map can be retrieved from the study by Farahani et al. (25).

Equation 3

$$\mathbf{ASSR_{amp}} = \sqrt{\mathbf{P_{S+N}}} - \sqrt{\mathbf{P_N}}$$

Equation 4

$$\mathbf{Subcortical\ ASSR_{amp}} = \sqrt{ASSR^2_{amp\ x} + ASSR^2_{amp\ y} + ASSR^2_{amp\ z}}$$

Defining regions of interest

fMRI studies show that the main neural generators of the ASSRs along the auditory pathway are located in the cochlear nucleus (CN), the inferior colliculus (IC), the medial geniculate body (MGB), and the auditory cortex (AC) bilaterally (57–60). Therefore, we defined eight regions of interest (ROIs) for further analysis (Figure 2). At the subcortical level, the ROIs were defined bilaterally in the CN (recognized with reference to the medullary pontine junction; left CN: 0.49 cm^3; right CN: 0.47 cm^3), IC (identified with reference to the thalamus; left IC: 0.50 cm^3; right IC: 0.55 cm^3), and in the posterior thalamus (roughly the posterior third of the thalamus; left MGB: 1.24 cm^3; right MGB: 1.45 cm^3) (19, 60). The cortical ROIs of the AC were defined bilaterally in the Heschl's gyrus (left AC: 5.49 cm^2; right AC: 5.58 cm^2) with reference to the transverse temporal gyrus in the Desikan–Killiany atlas implemented in Brainstorm (48, 61).

Time series of ROIs and ASSR amplitude

A representative dipole in each ROI was selected for subsequent analysis using the algorithm suggested by Farahani et al. (25). First, inside each ROI, a patch with the highest mean ASSR amplitude was selected. Then, a dipole with the most similar response, regarding amplitude and phase, to the mean ASSR of the patch was selected as the representative dipole. The ASSR amplitudes of the representative dipoles in cortical and subcortical ROIs were obtained based on Eq. 3 and Eq. 4 and used for further analyses. The time series of the representative dipole was used for subsequent phase coherence analysis.

Phase coherence

Phase coherence (or intertrial phase coherence) shows the phase consistency of ASSRs across epochs (17, 62). It also explains the phase-locking capability of a neural generator to the acoustic stimulus and varies between 0 and 1 (45, 63). To calculate the phase coherence, the time series of each ROI with

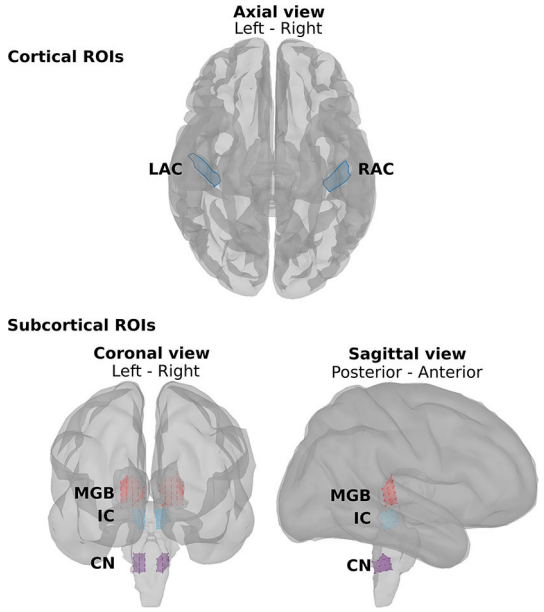


FIGURE 2
The regions of interest (ROIs) along the auditory pathway. The cortical ROIs are located bilaterally in the left auditory cortex (LAC) and right auditory cortex (RAC). The subcortical ROIs include the left and right medial geniculate body (LMGB, RMGB), the left and right inferior colliculus (LIC, RIC), and the left and right cochlear nucleus (LCN, RCN).

192 epochs were divided into 64 groups of three epochs. The phase of group i (θ_i, i = 1, 2,..., 64) was obtained from the complex responses averaged across the three epochs. Finally, phase coherence was calculated based on Equation 5 (62).

Equation 5

$$\textbf{PhaseCoherence}=\frac{1}{N}\sqrt{\left(\sum_{i=1}^{N}\cos\theta_i\right)^2+\left(\sum_{i=1}^{N}\sin\theta_i\right)^2}$$

For subcortical ROIs, the representative dipole had three time series (x, y, and z components). To reduce the dimension of this data, the optimal dipole direction representing most of the variance of the ASSR was estimated using singular value decomposition (SVD) (64). The three time series were projected in the optimal direction, and the outcome was used for calculating the phase coherence. It should be noted that before SVD, the three time series were filtered by a zero-phase band-pass filter with a bandwidth of 4 Hz and modulation frequency as the center frequency.

Hemispheric lateralization

To assess hemispheric asymmetry, we employed the laterality index (LI). The LI is a normalized index with the range of [-1, 1], where zero means symmetrical processing pattern and positive and negative values show lateralization to the right and left hemispheres, respectively. LI was calculated as:

Equation 6

$$\textbf{LI}=\frac{\textbf{ASSR}_{\textbf{amp R}}-\textbf{ASSR}_{\textbf{amp L}}}{\textbf{ASSR}_{\textbf{amp R}}+\textbf{ASSR}_{\textbf{amp L}}}$$

where $ASSR_{ampR}$ and $ASSR_{ampL}$ denote the ASSR amplitude (based on equations 3 and 4) of the neural generator located in the right and left hemispheres, respectively. To prevent inaccurate lateralization, the LI was only calculated when both neural generators had a significant ASSR.

Statistical analysis

Since we used a group-wise framework and the value of ASSR measures could not be obtained for each individual participant, the standard deviation could not be calculated in the traditional manner. The standard deviation was estimated based on the jackknife resampling method for each of the ASSR amplitude, phase coherence, and LI (65). The mean of ASSR amplitudes, phase coherence, and LI were obtained from all participants without resampling. The subsequent statistical analyses were performed based on the mean, estimated standard deviation, and the number of participants in each group, rather than on individual data points (66, 67) using custom scripts in MATLAB R2016b (MathWorks).

To investigate the overall effect of hearing impairment on ASSR amplitude, a factorial mixed analysis of variance (FM-ANOVA) with side of stimulation (two levels: left and right) and neural generators (eight levels: two cortical generators and six subcortical generators) as within-subject variables was separately carried out for middle-aged and older participants in response to 4, 20, 40, and 80 Hz acoustic modulations. *Post-hoc* comparisons were performed in cortical and subcortical categories of neural generators. The two-sample *t*-test was performed for each category based on the pooled mean and the pooled standard deviations across neural generators. The results were corrected for multiple comparisons using the FDR method (56). In the tests with neural generators as a within-subject variable, the sample size of the test has a high number, and in turn, the statistical tests often showed very small *p*-values. Thus, the effect sizes were also reported to measure significance independent of sample size (68). Cohen's d was used as a measure of effect size. The description of magnitudes of d was initially suggested by Cohen (69) and expanded by Sawilowsky (70). The magnitudes of 0.01, 0.2, 0.5, 0.8, and 1.2 were described as very small, small, medium, large, and very large effect sizes. Similar statistical analyses were also carried out for phase coherence.

For hemispheric lateralization, a one-sample *t*-test with FDR correction was employed to determine for which stimulation

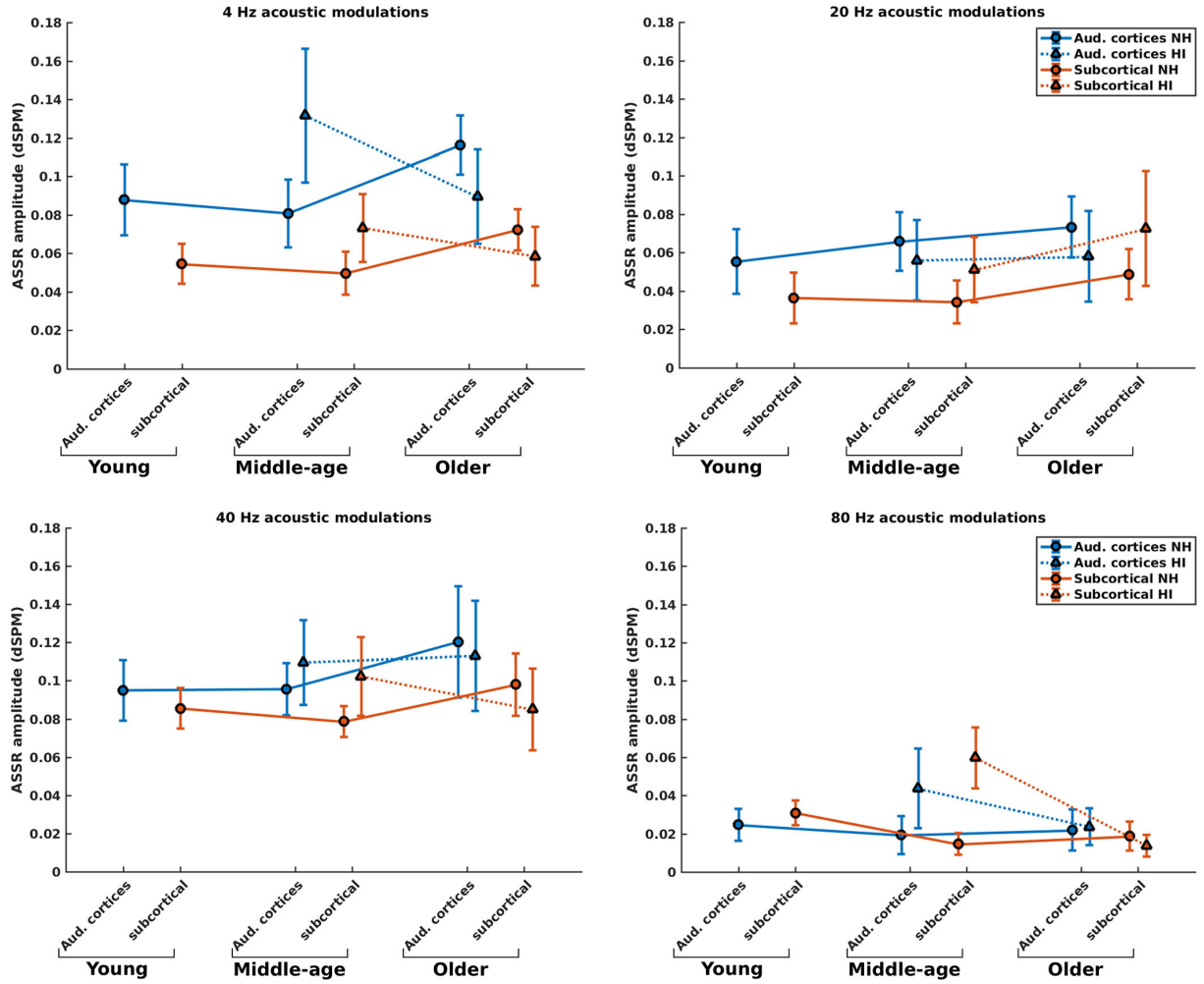


FIGURE 3
ASSR amplitudes of the neural generators in the auditory cortices and subcortical neural generators in NH and HI participants regardless of the side of stimulation across age and modulation frequency. The circle and triangle symbols indicate the pooled means (i.e., the weighted average of amplitudes across the side of stimulation and the side of generators; number of subjects as weights), and error bars represent the pooled standard deviations (69).

conditions the LI differed significantly from zero. A significant positive or negative LI shows lateralization to the right or left hemispheres, respectively. Finally, the potential effect of hearing impairment on hemispheric lateralization was investigated using a two-sample *t*-test per modulation frequency and side of stimulation.

Results

Effect of hearing impairment on the response strength of the neural generators

Figure 3 illustrates the mean response strengths for the cortical and subcortical neural generators (for anatomical locations, see Figure 2) for young, middle-aged, and older listeners for each of the four modulation frequencies. A significant main effect of HI was found in the middle-aged and older participants for 4, 20, 40, and 80 Hz modulations (see Table 1). However, the main effects in middle-aged participants were the opposite of those of older participants. For the middle-aged participants, the response strengths of listeners with HI were larger than those of listeners with NH. In contrast, for the older participants, a significantly smaller response strength was observed in the listeners with HI compared to the NH ones for 4, 40, and 80 Hz, yet not for 20 Hz acoustic modulations.

Post-hoc testing in middle-aged participants showed significantly larger response strengths for listeners with HI than NH listeners for both the cortical and subcortical neural generators and different modulation frequencies. The only

TABLE 1 The results of the main effect of hearing impairment and *post-hoc* testing for ASSR amplitude and phase coherence.

		ASSR amplitude		Phase coherence	
		Middle-aged NH, HI	**Older NH, HI**	**Middle-aged NH, HI**	**Older NH, HI**
4 Hz	Aud. cortices	d = −1.9	d = 1.3	d = −1.3	d = 0.7
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Subcortical	d = −1.6	d = 1.0	d = −0.6	d = -0.3
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.01$
	Main effect	d = −1.7	d = 1.1	d = −0.7	d = -0.1 n.s.
		$p < 0.001$	$p < 0.001$	$p < 0.001$	
20 Hz	Aud. cortices	d = 0.5	d = 0.7	d = 0.9	d = 0.6
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Subcortical	d = −1.2	d = -1.1	d = −0.1	d = -1.1
		$p < 0.001$	$p < 0.001$	n.s.	$p < 0.001$
	Main effect	d = −0.6	d = -0.6	d = 0.1	d = -0.7
		$p < 0.001$	$p < 0.001$	n.s.	$p < 0.001$
40 Hz	Aud. cortices	d = −0.8	d = 0.2 n.s.	d = 0.1	d = 2.0
		$p < 0.001$		n.s.	$p < 0.001$
	Subcortical	d = −1.6	d = 0.7	d = 1.0	d = 1.5
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Main effect	d = −1.3	d = 0.5	d = 0.8	d = 1.7
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
80 Hz	Aud. cortices	d = −1.5	d = -0.1 n.s.	d = −1.0	d = 0.3 n.s.
		$p < 0.001$		$p < 0.001$	
	Subcortical	d = −4.0	d = 0.7	d = −1.9	d = 0.8
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
	Main effect	d = −3.2	d = 0.4	d = −1.6	d = 0.6
		$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

The *post-hoc* testing was performed per age cohort and modulation frequency for the neural generators in the auditory cortices and subcortical region. Cohen's d and p-value were reported for different age cohorts and different modulation frequencies. No significant differences were indicated with "n.s".

exception was for the cortical generators with larger response strengths for NH than participants with HI in response to the 20 Hz stimuli. The effect sizes suggest a large difference [d ≥ 0.8; (69, 70)] between HI and NH middle-aged listeners in response to the four different modulation frequencies. The results of *post-hoc* testing are summarized in Table 1.

For the older listeners, *post-hoc* testing revealed significantly smaller response strengths for listeners with HI compared to NH participants in the subcortical category of neural generators for all modulation frequencies, except for 20 Hz. Similarly, *post-hoc* testing revealed significantly smaller response strengths for listeners with HI compared to NH participants for neural generators in the auditory cortex in response to 4 and 20 Hz acoustic stimuli. The effect sizes demonstrate a large difference (d ≥ 0.8) between HI and NH older listeners for 4 Hz and a medium difference (d ≥ 0.5) for other frequencies.

Briefly, the response strength of the listeners with HI showed two different patterns of the changes in the middle-aged and older participants for most modulation frequencies. With the middle-aged participants, the response strength of listeners with HI was larger than those of NH listeners. In contrast, significantly smaller response strengths were observed in the listeners with HI compared to the NH ones for most modulation frequencies for the older participants.

Effect of hearing impairment on the phase coherence of the neural generators

Phase coherence reflects the changes in phase-locking of the responses regardless of the strength of the responses. Figure 4 illustrates the mean phase coherence for the cortical and subcortical neural generators (for anatomical locations, see Figure 2) for young, middle-aged, and older listeners for each of the four different modulation frequencies. A significant main effect of hearing impairment was observed for the middle-aged and older participants for most of the modulation frequencies. Detailed results are summarized in Table 1. Again, two different patterns of the changes were observed in the middle-aged with

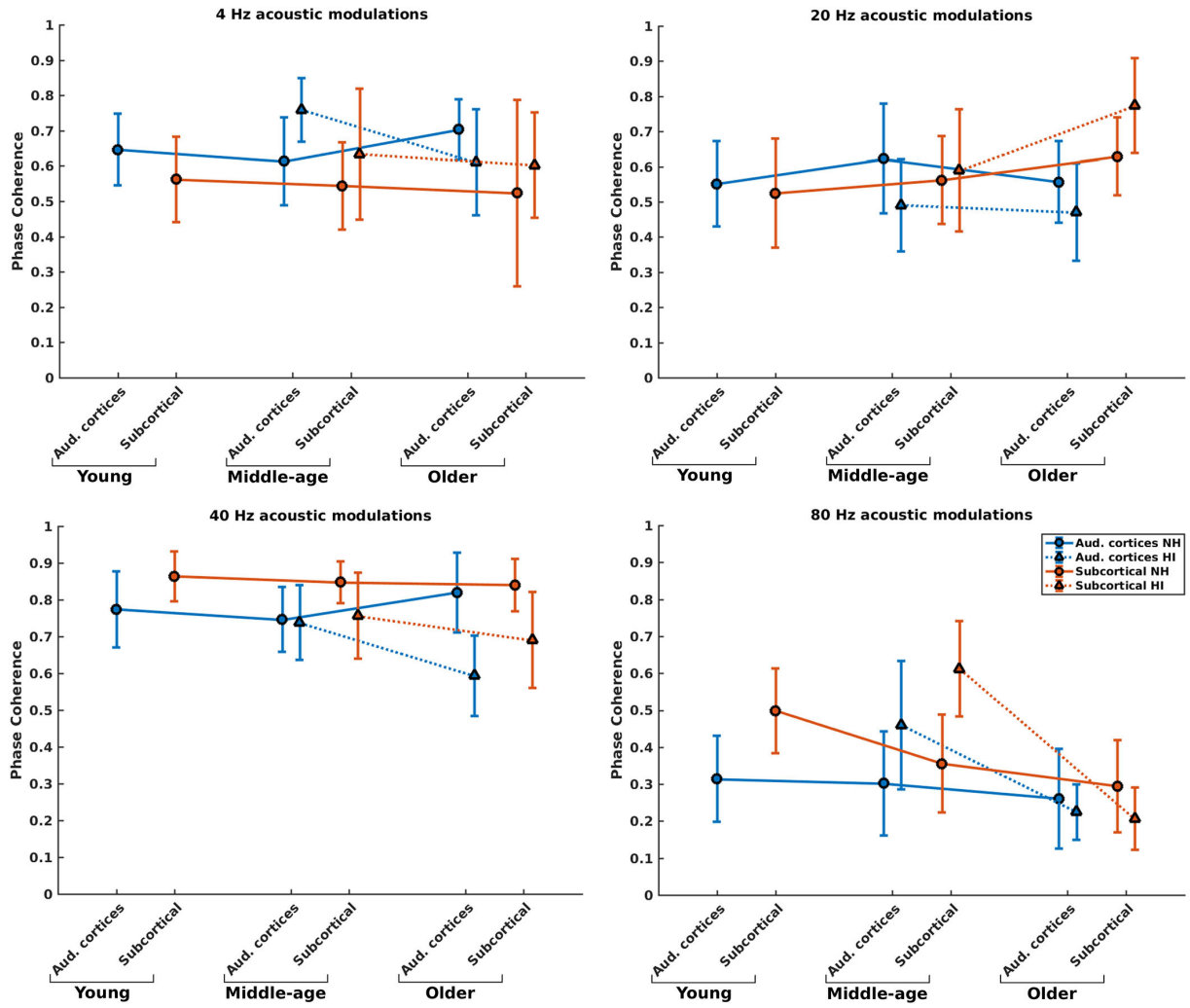


FIGURE 4
Phase coherence of the neural generators in auditory cortices and subcortical area in NH and HI participants regardless of the side of stimulation across age and modulation frequency. The circle and triangle symbols indicate the pooled means, and error bars represent the pooled standard deviations (69).

HI and older participants with HI. In most of the middle-aged participants' comparisons, HI listeners' phase-locking was larger than those of NH listeners. In contrast, a significantly smaller phase-locking was observed for the older HI participants than for the older NH ones.

Post-hoc testing in middle-aged participants showed a significantly larger phase coherence for listeners with HI than NH listeners in the cortical and subcortical neural generators for 4 and 80 Hz amplitude-modulated stimuli. The effect sizes of mean differences (Cohen's d) in these comparisons were medium or large [$d \geq 0.5$; (69, 70)]. However, there was less phase coherence in listeners with HI than NH listeners for cortical neural generators at 20 Hz and subcortical neural generators at 40 Hz stimulation conditions.

For the older listeners, *post-hoc* testing revealed significantly less phase coherence for listeners with HI compared to NH participants in the auditory cortices for all modulation frequencies, except for 80 Hz. In these modulation frequencies, Cohen's d suggests a medium or large effect size ($d \geq 0.5$) of mean differences (69, 70). A similar effect was observed for the subcortical neural generators in response to 40 and 80 Hz acoustic stimuli. The effect sizes were large [$d \geq 0.8$; (69, 70)].

Hemispheric lateralization and hearing impairment

To investigate potential changes in hemispheric asymmetry of envelope processing in listeners with HI and NH ones, we determined the LIs for the 4, 20, and 40 Hz modulation frequencies based on the ASSR amplitudes of the left and

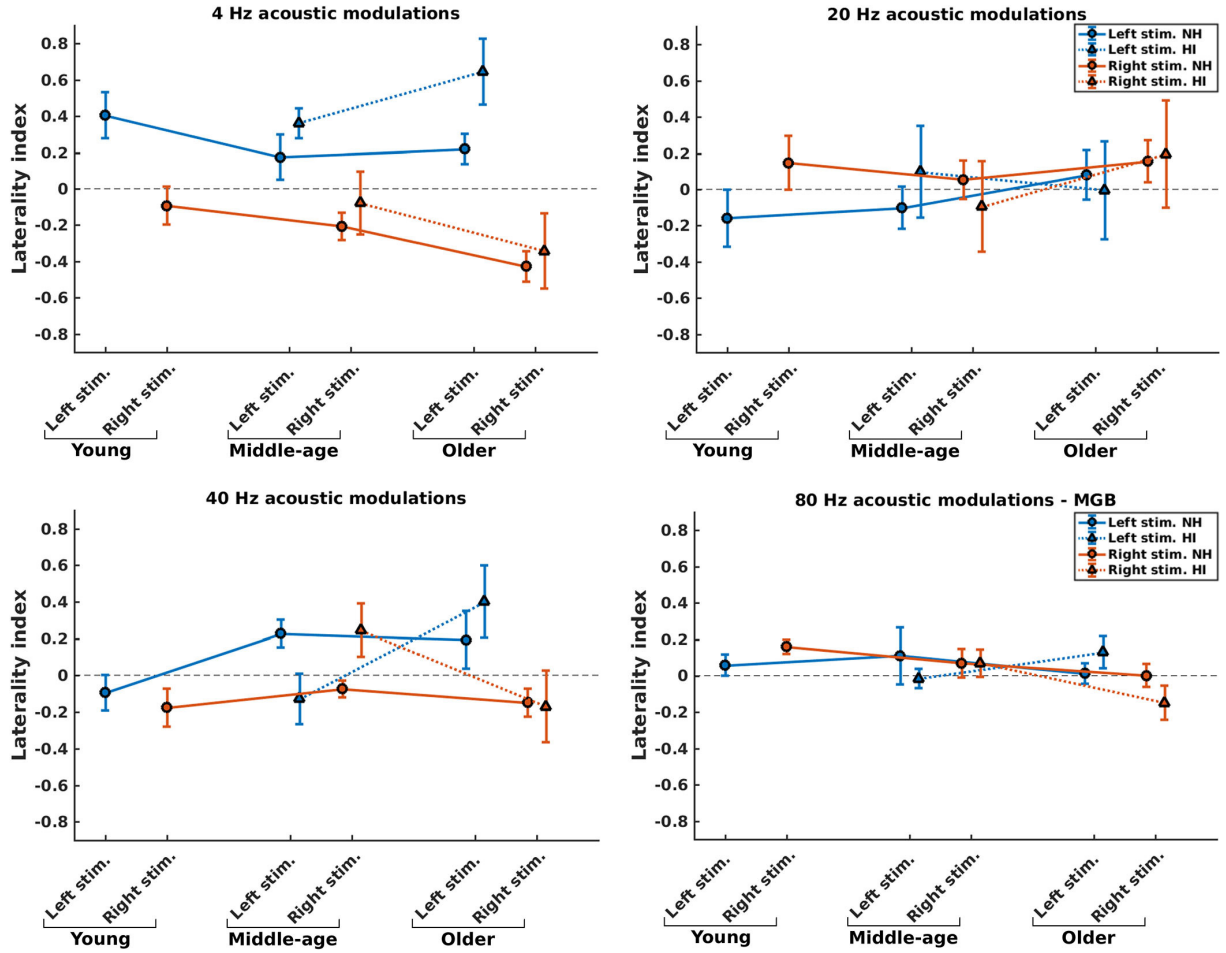


FIGURE 5
Hemispheric lateralization for normal-hearing (NH) and hearing-impaired (HI) listeners (indicated by solid lines and dotted lines, respectively) in different stimulation conditions (indicated by different colors) and different age groups. For 4, 20, and 40 Hz stimuli, the laterality indexes (LIs) were calculated based on the auditory cortex (AC), while for 80 Hz stimuli the LIs were calculated based on the medial geniculate body (MGB). The error bars illustrate the estimated standard deviations using the jackknife method (65).

right auditory cortices. For 80 Hz modulation frequency, we calculated the LI based on the ASSR amplitudes of the MGB, given the importance of subcortical activities (36). Figure 5 illustrates the LIs of the AC for 4, 20, and 40 Hz ASSRs in three age groups and two sides of stimulation and the LIs of the MGB for 80 Hz ASSRs. The groups with significant hemispheric asymmetry to the left or right hemisphere were determined using a one-sample *t*-test (the results are summarized in Supplementary Table 1).

The effect of hearing impairment on hemispheric asymmetry was investigated for middle-aged and older listeners. In most stimulation conditions, the hemispheric asymmetry in the listeners with HI was significantly more toward the right hemisphere than the hemispheric asymmetry of the NH ones. More specifically, with middle-aged participants, the LIs of listeners with HI were significantly more positive (toward the right hemisphere) than those of the NH ones for the 4 Hz (both sides of stimulation), 20 Hz (left side of stimulation), and 40 Hz (right side of stimulation) modulation frequencies. However, for 80 Hz AM stimuli, the hemispheric asymmetry was less or similar for the listeners with HI than for the NH ones for the left and right sides of stimulation, respectively. In these comparisons, Cohen's d suggests a large effect size ($d \geq 0.8$) of mean differences (69, 70). The results of statistical tests are summarized in Table 2.

For the older participants, the LIs of listeners with HI were similar or significantly more positive (toward the right hemisphere) than those of the NH ones for the 4, 20, and 40 Hz modulation frequencies for both the left and the right sides of stimulation. A similar effect was observed for the 80 Hz modulations presented to the left ear, while for the right side of stimulation, the LI of the listener with HI is more negative (toward the left hemisphere) than that of the NH group. In these comparisons, the effect sizes were large [$d \geq 0.8$; (69, 70)].

TABLE 2 The results of the statistical comparison between the laterality index of normal-hearing (NH) and hearing-impaired (HI) listeners in different stimulation conditions.

Stimulation condition		Middle-aged NH, HI	Older NH, HI
4 Hz	Left ear	d = −1.7	d = -3.2
		$p < 0.001$	$p < 0.001$
	Right ear	d = −1.1	d = -0.5 n.s.
		$p < 0.01$	
20 Hz	Left ear	d = −1.1	d = 0.4 n.s.
		$p < 0.01$	
	Right ear	d = 0.8	d = -0.1 n.s.
		n.s.	
40 Hz	Left ear	d = 3.3	d = -1.2
		$p < 0.001$	$p < 0.05$
	Right ear	d = −3.1	d = 0.1 n.s.
		$p < 0.001$	
80 Hz	Left ear	d = 1.0	d = -1.5
		$p < 0.05$	$p < 0.001$
	Right ear	d = −0.1	d = 1.9
		n.s.	$p < 0.001$

Discussion

Effect of age-related hearing impairment on the dynamics of neural generators

Our results indicated meaningful changes in the neural dynamics of middle-aged and older listeners with HI compared to those of middle-aged and older NH listeners. The effect of hearing impairment on the dynamics of the cortical and subcortical neural generators was investigated in persons with no indication of mild cognitive impairment to avoid the confounding factors of age and cognitive ability as much as possible. The acoustic modulations were presented at equal loudness levels to the participants with HI to correct for stimulus audibility. The cortical and subcortical neural generators' activity was reconstructed using the MNI approach. It should be noted that the selected parameters in the MNI approach, such as the number of layers of the head model, the conductivity of brain tissues, and the regularization parameters, may influence the results of the source reconstruction. Since the same methods and parameters were used for the different age cohorts with and without hearing impairment, the comparisons and the conclusions drawn from them remain reasonable.

Two different patterns of alterations were observed in the middle-aged participants with HI and older participants with HI. For middle-aged participants, we mainly found enhanced response strength and higher phase-locking in the HI group than NH, while for the older ones, we found decreased response strength and less phase-locking in the listeners with HI. The findings of middle-aged people agree with the literature (6, 21, 22). However, our results for the older participants are novel and different from sensor-level analysis on the same data as here (11). These findings for middle-aged and older participants with HI are elaborated on below.

Our observation of enhanced response strength in HI middle-aged listeners' auditory cortex follows Millman et al. (22) and Fuglsang et al. (21). Millman and colleagues investigated the neural synchronizations in response to 2 Hz acoustic modulated noise between HI and NH similarly aged persons (∼60 years old). Fuglsang et al. (21) reported magnified cortical responses in participants with HI compared to NH participants for tone sequences modulated at slow rates (4 Hz) during a passive listening task. They had also corrected for the audibility of auditory stimuli for the participants with HI, and the age range of participants was similar (∼65 years old).

The enhanced neural responses in the subcortical generators of middle-aged adults with HI are in line with animal studies which have shown that peripheral hearing loss is associated with increased neural responses to amplitude-modulated stimuli in the auditory nerve fibers (6–8) and the midbrain (9). Similarly, human electrophysiological studies reported enhanced neural responses in the brainstem of adults around 60 years old with HI relative to NH ones in the same age range (10, 11).

Only a few studies report how age-related hearing loss affects temporal envelope processing in older people (70–80 years old). Using source analysis, we observed significantly less response strength for the older adults with HI than the NH ones. However, sensor-level analysis on the same data yielded no significant difference in response strengths between the older adults with HI and NH ones (11). Note that the response strengths in the sensor-level reflect a weighted average of the activity (due to the volume conduction). Therefore, this approach may not be as sensitive to small changes as brain source analysis which estimates the original neural activity of each generator.

The reduced neural synchronization (response strength and phase-locking) in the older adults with HI in the current study agrees with the observations of Hao et al. (71). They found reduced frequency-following responses (FFRs), under quiet and noise conditions, in the older adults with presbycusis (60–82 years old) compared to NH similarly aged persons. However, data regarding the effect of hearing impairment on FFRs are not very consistent [for review, see (72)]. For instance, Presacco et al. (73) did not find significant differences between the FFRs in the older adults with HI (average 71 years old) and those in the NH adults (average 65 years old). The discrepancies between the findings of different FFR studies could be due to the different age ranges involved.

In an experiment using continuous speech, Decruy et al. (74) found evidence of enhanced envelope tracking to the target talker in older adults with HI compared to NH listeners. In a similar experiment, Presacco et al. (73) found

no differences. These results are different from our findings in the older participants with HI. The first possible reason could be differences between experimental conditions. The envelope tracking in our experiment is unattended, while in the experiment of Decruy et al. (74), the participant should attend to the stimuli. In speech envelope tracking onset responses play an important role, while it is not applicable for ASSRs. The second reason for different results refers to the source-level analysis in our study and reconstructing the activity of neural generators along the auditory pathway, while Decruy et al. (74) and Presacco et al. (73) used sensor-level analysis which considers all cortical activities.

For the relatively low frequencies (below 50 Hz), there is an age-related enhancement in the neural responses of NH older adults compared to those of young and middle-aged adults (19). Considering the age-related enhancement in the NH older adults and the enhancement effect in the middle-aged adults with HI (the current study), we expected to find an aggravated effect of hearing impairment in the older participants with HI. However, our results for the older adults with HI showed reduced responses compared to NH participants in the same age cohort. This novel finding suggests that the reduced effect of age-related hearing loss and age-related degradation in the older cohort (70–80 years) may be greater than a compensatory enhancement effect in the representation of envelope processing in this age cohort.

Potential mechanisms underlying the changes in temporal envelope processing

Homeostatic compensatory mechanisms can explain the enhanced response strength and phase-locking in the middle-aged adults with HI. It is known that diminished cochlear output in adults with HI activates various mechanisms which induce central gain to increase neural excitability (75–77). However, the potential compensatory mechanisms could be considered maladaptive, because the response strength and phase-locking in the middle-aged adults with HI were even higher than those of NH middle-aged listeners.

For example, the hearing-impaired auditory nerve fibers at the subcortical level show steeper loudness growth than NH ones (7, 78) and enhanced onset responses (79). Spontaneous activity is enhanced in the inferior colliculus (80) and the auditory cortex of older compared to young animals (75, 81, 82). Along the auditory pathway (from the brainstem up to the cortex), the influx of inhibitory neurotransmitters into excitatory neurons decreases, while it is preserved for inhibitory neurons (83–85).

The reduced response strength in the older adults with HI can be explained by the normal age-related changes in this age cohort. In a previous study on the adults with normal audiometric thresholds, we observed enhanced neural responses to envelope modulations for NH older persons compared to young and middle-aged NH individuals (19). This age-related enhancement can be attributed to the loss of functional inhibition in older adults as a compensatory mechanism (19, 86, 87). These mechanisms are used in normal-hearing older persons. On top of it, hearing impairment impacts neural processing in the older adults with HI. Consequently, the reduced response strength is detected for hearing impairment at an older age despite correcting for audibility.

Both middle-aged with HI and older adults with HI have similar patterns of hearing loss, with no significant differences in pure-tone average (PTA) across all audiometric thresholds (0.25–8 kHz) (88). However, age-related structural changes, such as cerebral atrophy and demyelination, increase with age (89, 90). The animal study of Wang et al. (91) showed that, in addition to known cochlear synaptopathy, the central synapses of spiral ganglion neurons are also pathologically changed during aging, which suggests a central synaptopathy. This central synaptopathy plays a significant role in weakened auditory input and altered central auditory processing during age-related hearing loss (91). The above-mentioned could also explain the different results for middle-aged and older adults.

Hemispheric asymmetry

Generally, our results suggest that hearing impairment is associated with altered hemispheric asymmetry in auditory temporal processing. In most cases, this alteration occurs through shifting toward the right hemisphere. This observation follows previous studies suggesting altered hemispheric asymmetry of event-related potentials in older adults with HI (32, 92).

To the best of our knowledge, this study is one of the first to investigate the association between hearing impairment and hemispheric asymmetry in temporal envelope processing using source analysis. In line with the HAROLD model (31), it was previously documented that hemispheric asymmetry for temporal envelope processing is reduced (more symmetric) for the NH older adults compared to those of the younger ones (29, 30). Using source analysis, Farahani et al. (19) reported that hemispheric asymmetry is reduced in NH older adults compared to NH younger ones in response to the 20 and 80 Hz amplitude-modulated stimuli. Although NH older is thus expected to be associated with less asymmetrical neural processing, our older participants with HI exhibit asymmetrical processing patterns. The LI in the middle-aged and older participants with HI exhibits a hemispheric asymmetry more toward the right hemisphere than the hemispheric asymmetry of the NH ones. This novel observation may be explained by the reduced integrity of white matter tracts related to presbycusis (33). The corpus callosum is a large bundle of white matter tracts that play a key role in interhemispheric interactions (93). As such, white

matter deficits in people with severe age-related hearing loss can impact the hemispheric asymmetry in temporal envelope processing. However, further research is needed to clarify the relationship between the changes in the white matter and the altered hemispheric asymmetry in older adults with HI.

The role of source-level analysis

In electrophysiological measurements, the recorded data at each sensor are a weighted average of the activity of several neural generators due to the volume conduction of the brain tissue. However, brain source analysis allows us to estimate the original activity of each neural generator. Such an analysis increases our understanding of the potential alterations at different levels of the auditory pathway across age and with or without hearing impairment.

Furthermore, brain source analysis enables us to detect relatively small changes in the activity of a neural generator which may not be detectable in the sensor-level analysis. For example, values of Cohen's d (ASSR amplitude, Table 1) suggest that the differences in the responses between listeners with HI and NH are larger than those between HI and NH older adults. In middle-aged adults, the results of sensor-level analyses (i.e., enhanced response strengths in listeners with HI, 10) were in line with the results of source-level analysis (i.e., the current study). However, in older adults, where the differences are smaller, the sensor-level analysis yielded no significant difference in response strengths between the older adults with HI and NH ones (11), while brain source analysis using MNI on the same data revealed significant changes for the neural generators.

Conclusion

The present study investigated the effect of age-related hearing loss on the dynamics of the neural generators involved in the temporal envelope processing for middle-aged and older adults. The activity of the cortical and subcortical neural generators of ASSRs was reconstructed for participants with HI and NH ones using the MNI approach. This approach allows for a detailed analysis of the neural generators' activity along the auditory pathway (25). Our results showed that age-related hearing loss, with correction for audibility, is accompanied by changes in response strength and phase-locking of the neural generators of the ASSRs. However, the patterns of the changes in the middle-aged participants are different from those of older ones. With the middle-aged participants, the response strength and phase coherence of listeners with HI were larger than those of NH listeners. In contrast, for the older participants, a significantly smaller response strength and phase coherence were observed in the listeners with HI compared to the NH ones for most modulation frequencies. This is an essential finding to develop rehabilitation strategies for hearing-impaired persons across the aging life span.

With our novel approach, we observed that middle-aged and older participants with HI exhibit a hemispheric asymmetry more toward the right hemisphere than the hemispheric asymmetry of the NH ones. This observation can be explained by the brain structural changes associated with presbycusis in the middle-aged and older adults.

Data availability statement

The datasets presented in this article are not readily available because of ethical and privacy restrictions. Requests to access the datasets should be directed to Astrid van Wieringen, astrid.vanwieringen@kuleuven.be.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethical Committee of the University Hospitals and University of Leuven. The patients/participants provided their written informed consent to participate in this study.

Author contributions

EF, JW, and AW designed the study, contributed to the interpretation of the results, and critically revised the manuscript. EF analyzed data, performed statistical analyses, and wrote the manuscript draft. JW and AW verified the analytical methods and supported data analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Research Council, KU Leuven, through projects C14/19/110, C14/17/046, and by the Research Foundation Flanders through FWO-projects G066213 and G0A9115. This work was also partly funded by Flanders Innovation & Entrepreneurship through the VLAIO research grant HBC.20192373.

Acknowledgments

Our special thanks go to Dr. Tine Goossens for accumulating the ASSR recording data used in this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2022.905017/full#supplementary-material

References

1. Eisenberg LS, Dirks DD, Bell TS. *Speech Recognition in Amplitude-Modulated Noise of Listeners With Normal and Listeners With Impaired Hearing.* (1995). Available online at: http://pubs.asha.org/ss/rights_and_permissions.aspx

2. George ELJ, Festen JM, Houtgast T. Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am.* (2006) 120:2295–311. doi: 10.1121/1.2266530

3. Summers V, Molis MR. *Summers and Molis: Speech Recognition in Fluctuating Maskers 245 Speech Recognition in Fluctuating and Continuous Maskers: Effects of Hearing Loss and Presentation Level.* (2004). Available online at: http://pubs.asha.org/ss/rights_and_permissions.aspx

4. Goossens T, Vercammen C, Wouters J, van Wieringen A. Masked speech perception across the adult lifespan: impact of age and hearing impairment. *Hear Res.* (2017) 344:109–24. doi: 10.1016/j.heares.2016.11.004

5. Noordhoek IM, Houtgast T, Festen JM. Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region. *J Acoust Soc Am.* (2001) 109:1197–212. doi: 10.1121/1.1349429

6. Henry MJ, Herrmann B, Obleser J. Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc Natl Acad Sci U S A.* (2014) 111:14935–40. doi: 10.1073/pnas.1408741111

7. Kale S, Heinz MG. Envelope coding in auditory nerve fibers following noise-induced hearing loss. *J Assoc Res Otolaryngol.* (2010) 11:657–73. doi: 10.1007/s10162-010-0223-6

8. Kale S, Heinz MG. Temporal modulation transfer functions measured from auditory-nerve responses following sensorineural hearingloss. *Hear Res.* (2012) 286:64–75. doi: 10.1016/j.heares.2012.02.004

9. Zhong Z, Henry KS, Heinz MG. Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas. *Hear Res.* (2014) 309:55–62. doi: 10.1016/j.heares.2013.11.006

10. Anderson S, Parbery-Clark A, White-Schwoch T, Drehobl S, Kraus N. Effects of hearing loss on the subcortical representation of speech cues. *J Acoust Soc Am.* (2013) 133:3030–8. doi: 10.1121/1.4799804

11. Goossens T, Vercammen C, Wouters J, van Wieringen A. The association between hearing impairment and neural envelope encoding at different ages. *Neurobiol Aging.* (2019) 74:202–12. doi: 10.1016/j.neurobiolaging.2018.10.008

12. Stone MA, Füllgrabe C, Mackinnon RC, Moore BCJ. The importance for speech intelligibility of random fluctuations in "steady" background noise. *J Acoust Soc Am.* (2011) 130:2874–81. doi: 10.1121/1.3641371

13. Peelle JE, Davis MH. Neural oscillations carry speech rhythm through to comprehension. *Front Psychol.* (2012) 3:320. doi: 10.3389/fpsyg.2012.00320

14. Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* (1995) 270:303–4. doi: 10.1126/science.270.5234.303

15. Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci.* (1992) 336:367–73. doi: 10.1098/rstb.1992.0070

16. Cogan GB, Poeppel D. A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *J Neu-rophysiol.* (2011) 106:554–63. doi: 10.1152/jn.00075.2011

17. Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron.* (2007) 54:1001–10. doi: 10.1016/j.neuron.2007.06.004

18. Picton T. Hearing in time: evoked potential studies of temporal processing. *Ear Hear.* (2013) 34:385–401. doi: 10.1097/AUD.0b013e31827ada02

19. Farahani ED, Wouters J, van Wieringen A. Neural generators underlying temporal envelope processing show altered responses and hemispheric asymmetry across age. *Front Aging Neurosci.* (2020) 12:596551. doi: 10.3389/fnagi.2020.596551

20. Ananthakrishnan S, Krishnan A, Bartlett E. Human frequency following response: neural representation of envelope and temporal fine structure in listeners with normal hearing and sensorineural hearing loss. *Ear Hear.* (2016)37:*e*91–103. doi: 10.1097/AUD.0000000000000247

21. Fuglsang SA, Märcher-Rørsted J, Dau T, Hjortkjær J. Effects of sensorineural hearing loss on cortical synchronization to competing speech during selective attention. *J Neurosci.* (2020) 40:2562–72. doi: 10.1523/JNEUROSCI.1936-19.2020

22. Millman RE, Mattys SL, Gouws AD, Prendergast G. Magnified neural envelope coding predicts deficits in speech perception in noise. *J Neurosci.* (2017) 37:7727–36. doi: 10.1523/JNEUROSCI.2722-16.2017

23. Grech R, Cassar T, Muscat J, Camilleri KP, Fabri SG, Zervakis M, et al. Review on solving the inverse problem in EEG source analysis. *J Neuroeng Rehabil.* (2008) 5:25. doi: 10.1186/1743-0003-5-25

24. Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hämäläinen MS. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage.* (2006) 31:160–71. doi: 10.1016/j.neuroimage.2005.11.054

25. Farahani ED, Wouters J, van Wieringen A. Brain mapping of auditory steady-state responses: a broad view of cortical and subcortical sources. *Hum Brain Mapp.* (2021) 42:780–96. doi: 10.1002/hbm.25262

26. Luke R, de Vos A, Wouters J. Source analysis of auditory steady-state responses in acoustic and electric hearing. *Neuroimage.* (2017) 147:568–76. doi: 10.1016/j.neuroimage.2016.11.023

27. Popescu M, Popescu EA, Chan T, Blunt SD, Lewine JD. Spatio-temporal reconstruction of bilateral auditory steady-state responses using MEG beamformers. *IEEE Transac Biomed Eng.* (2008) 55:1092–102. doi: 10.1109/TBME.2007.906504

28. Popov T, Oostenveld R, Schoffelen JM. FieldTrip made easy: an analysis protocol for group analysis of the auditory steady state brain response in time, frequency, and space. *Front Neurosci.* (2018) 12:711. doi: 10.3389/fnins.2018.00711

29. Goossens T, Vercammen C, Wouters J, van Wieringen A. Aging affects neural synchronization to speech-related acoustic modulations. *Front Aging Neurosci.* (2016) 8:133. doi: 10.3389/fnagi.2016.00133

30. Bellis TJ, Nicol T, Kraus N. Aging affects hemispheric asymmetry in the neural representation of speech sounds. *J Neurosci.* (2000) 20:791–7. doi: 10.1523/JNEUROSCI.20-02-00791.2000

31. Cabeza R. Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychol Aging.* (2002) 17:85–100. doi: 10.1037/0882-7974.17.1.85

32. Greenwald RR, Jerger J. Aging affects hemispheric asymmetry on a competing speech task. *J Am Acad Audiol.* (2001) 12:167–73. doi: 10.1055/s-0042-1745594

33. Mudar RA, Husain FT. Neural alterations in acquired age-related hearing loss. *Front Psychol.* (2016) 7:828. doi: 10.3389/fpsyg.2016.00828

34. Chambers AR, Resnik J, Yuan Y, Whitton JP, Edge AS, Liberman MC, et al. Central gain restores auditory processing following near-complete cochlear denervation. *Neuron.* (2016) 89:867–79. doi: 10.1016/j.neuron.2015.12.041

35. Herrmann B, Parthasarathy A, Bartlett EL. Ageing affects dual encoding of periodicity and envelope shape in rat inferior colliculus neurons. *Eur J Neurosci.* (2017) 45:299–311. doi: 10.1111/ejn.13463

36. Herdman AT, Lins O, Roon P, van Stapells DR, Scherg M, Picton TW. Intracerebral sources of human auditory steady-state responses. *Brain Topogr.* (2002) 15:69-—86. doi: 10.1023/a:1021470822922

37. Ross B. A novel type of auditory responses: temporal dynamics of 40-Hz steady-state responses induced by changes in sound localization. *J Neurophysiol.* (2008) 100:1265–77. doi: 10.1152/jn.00048.2008

38. Margolis RH, Saly GL. Distribution of hearing loss characteristics in a clinical population. *Ear Hear.* (2008) 29:524–32. doi: 10.1097/AUD.0b013e3181731e2e

39. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x

40. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* (1971) 9:97–113. doi: 10.1016/0028-3932(71)90067-4

41. John MS, Picton TW. Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements C. *Hear Res.* (2000) 141:57–79. doi: 10.1016/s0378-5955(99)00209-9

42. van Eeckhoutte M, Wouters J, Francart T. Auditory steady-state responses as neural correlates of loudness growth. *Hear Res.* (2016) 342:58–68. doi: 10.1016/j.heares.2016.09.009

43. Emara AAY, Kolkaila EA. Prediction of loudness growth in subjects with sensorineural hearing loss using auditory steady state response. *J Int Adv Otol.* (2010) 6:371. Available online at: https://www.advancedotology.org/content/files/sayilar/77/buyuk/IAOOct2010p371-379.pdf

44. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci.* (2011) 2011:156869. doi: 10.1155/2011/156869

45. Farahani ED, Wouters J, van Wieringen A. Contributions of non-primary cortical sources to auditory temporal processing. *Neuroimage.* (2019) 191:303–14. doi: 10.1016/j.neuroimage.2019.02.037

46. Gramfort A, Papadopoulo T, Olivi E, Clerc M. *OpenMEEG: Opensource Software for Quasistatic Bioelectromagnetics.* (2010). Available online at: http://www.biomedical-engineering-online.com/content/9/1/45

47. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage.* (2011) 54:313–27. doi: 10.1016/j.neuroimage.2010.07.033

48. Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci.* (2011) 2011:879716. doi: 10.1155/2011/879716

49. Tadel F, Bock E, Niso G, Mosher JC, Cousineau M, Pantazis D, et al. MEG/EEG group analysis with brainstorm. *Front Neurosci.* (2019) 13:76. doi: 10.3389/fnins.2019.00076

50. Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, et al. Neurotechnique mapping : combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron.* (2000) 26:55–67. doi: 10.1016/S0896-6273(00)81138-1

51. Hauk O, Wakeman DG, Henson R. Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *Neuroimage.* (2011) 54:1966–74. doi: 10.1016/j.neuroimage.2010.09.053

52. Bradley A, Yao J, Dewald J, Richter CP. Evaluation of electroencephalography source localization algorithms with multiple cortical sources. *PLoS ONE.* (2016) 11:e0147266. doi: 10.1371/journal.pone.0147266

53. Hincapié AS, Kujala J, Mattout J, Daligault S, Delpuech C, Mery D, et al. MEG connectivity and power detections with minimum norm estimates require different regularization parameters. *Comput Intell Neurosci.* (2016) 2016:3979547. doi: 10.1155/2016/3979547

54. Ghumare EG, Schrooten M, Vandenberghe R, Dupont P. A Time-varying connectivity analysis from distributed EEG sources: a simulation study. *Brain Topogr.* (2018) 31:721–37. doi: 10.1007/s10548-018-0621-3

55. Dobie RA, Wilson MJ. A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion product otoacoustic emissions, or other sinusoids. *J Acoust Soc Am.* (1996) 100:2236–46. doi: 10.1121/1.417933

56. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B.* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

57. Langers DRM, van Dijk P, Backes WH. Lateralization, connectivity and plasticity in the human central auditory system. *Neuroimage.* (2005) 28:490–9. doi: 10.1016/j.neuroimage.2005.06.024

58. Steinmann I, Gutschalk A. Potential fMRI correlates of 40-Hz phase locking in primary auditory cortex, thalamus and midbrain. *Neuroimage.* (2011) 54:495–504. doi: 10.1016/j.neuroimage.2010.07.064

59. Overath T, Zhang Y, Sanes DH, Poeppel D. Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. *J Neurophysiol.* (2012) 107:2042–56. doi: 10.1152/jn.00308.2011

60. Coffey EBJ, Herholz SC, Chepesiuk AMP, Baillet S, Zatorre RJ. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun.* (2016) 7:11070. doi: 10.1038/ncomms11070

61. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage.* (2006) 31:968–80. doi: 10.1016/j.neuroimage.2006.01.021

62. Picton TW, Dimitrijevic A, Sasha John M, Van Roon P. The use of phase in the detection of auditory steady-state responses. *Clin Neurophysiol.* (2001) 112:1698–711. doi: 10.1016/S1388-2457(01)00608-3

63. Koerner TK, Zhang Y. Effects of background noise on inter-trial phase coherence and auditory N1-P2 responses to speech stimuli. *Hear Res.* (2015) 328:113–9. doi: 10.1016/j.heares.2015.08.002

64. Rueda-Delgado LM, Solesio-Jofre E, Mantini D, Dupont P, Daffertshofer A, Swinnen SP. Coordinative task difficulty and behavioural errors are associated with increased long-range beta band synchronization. *Neuroimage.* (2017) 146:883–93. doi: 10.1016/j.neuroimage.2016.10.030

65. Efron B, Stein C. The jackknife estimate of variance. *Ann Stat.* (1981) 9:586–96. doi: 10.1214/aos/1176345462

66. Cohen BH. Calculating a factorial ANOVA from means and standard deviations. *Understand Stat Stat Issues Psychol Educ Soc Sci.* (2002) 1:191–203. doi: 10.1207/S15328031US0103_04

67. Nagy P. *n-way ANOVA From Summary Statistics. MATLAB Cent. File Exch.* (2013). Available online at: https://www.mathworks.com/matlabcentral/fileexchange/ 41036-n-way-anova-from-summary-statistics (accessed September 11, 2018).

68. Sullivan GM, Feinn R. Using effect size—or why the p value is not enough. *J Grad Med Educ.* (2012) 4:279–82. doi: 10.4300/JGME-D-12-00156.1

69. Cohen J. *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Routledge (1988). doi: 10.4324/9780203

70. Sawilowsky SS. Very large and huge effect sizes. *J Modern Appl Stat Method.* (2009) 8:597–9. doi: 10.22237/jmasm/1257035100

71. Hao W, Wang Q, Li L, Qiao Y, Gao Z, Ni D, et al. Effects of phase-locking deficits on speech recognition in older adults with presbycusis. *Front Aging Neurosci.* (2018) 10:397. doi: 10.3389/fnagi.2018.00397

72. Anderson S, Karawani H. Objective evidence of temporal processing deficits in older adults. *Hear Res.* (2020) 397:108053. doi: 10.1016/j.heares.2020.108053

73. Presacco A, Simon JZ, Anderson S. Speech-in-noise representation in the aging midbrain and cortex: effects of hearing loss. *PLoS ONE.* (2019) 14:e0213899. doi: 10.1371/journal.pone.0213899

74. Decruy L, Vanthornhout J, Francart T. Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hear Res.* (2020) 393:107961. doi: 10.1016/j.heares.2020.107961

75. Herrmann B, Butler BE. Hearing loss and brain plasticity: the hyperactivity phenomenon. *Brain Struct Func.* (2021) 226:2019–39. doi: 10.1007/s00429-021-02313-9

76. Kujawa SG, Liberman MC. Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hear Res.* (2015) 330:191–9. doi: 10.1016/j.heares.2015.02.009

77. Salvi R, Sun W, Ding D, Chen G, di Lobarinas E, Wang J, et al. Inner hair cell loss disrupts hearing and cochlear function leading to sensory deprivation and enhanced central auditory gain. *Front Neurosci.* (2017) 10:621. doi: 10.3389/fnins.2016.00621

78. Heinz MG, Young ED. Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss. *J Neurophysiol.* (2004) 91:784–95. doi: 10.1152/jn.00776.2003

79. Crumling MA, Saunders JC. Tonotopic distribution of short-term adaptation properties in the cochlear nerve of normal and acoustically overexposed chicks. *J Assoc Res Otolaryngol.* (2007) 8:54–68. doi: 10.1007/s10162-006-0061-8

80. Parthasarathy A, Herrmann B, Bartlett EL. Aging alters envelope representations of speech-like sounds in the inferior colliculus. *Neurobiol Aging.* (2019) 73:30–40. doi: 10.1016/j.neurobiolaging.2018.08.023

81. Hughes LF, Turner JG, Parrish JL, Caspary DM. Processing of broadband stimuli across A1 layers in young and aged rats. *Hear Res.* (2010) 264:79–85. doi: 10.1016/j.heares.2009.09.005

82. Overton JA, Recanzone GH. Effects of aging on the response of single neurons to amplitude-modulated noise in primary auditory cortex of rhesus macaque. *J Neurophysiol.* (2016) 115:2911–23. doi: 10.1152/jn.01098.2015

83. Sanes DH, Kotak VC. Developmental plasticity of auditory cortical inhibitory synapses. *Hear Res.* (2011) 279:140–8. doi: 10.1016/j.heares.2011.03.015

84. Sarro EC, Kotak VC, Sanes DH, Aoki C. Hearing loss alters the subcellular distribution of presynaptic GAD and postsynaptic GABAA receptors in the auditory cortex. *Cerebral Cortex.* (2008) 18:2855–67. doi: 10.1093/cercor/bhn044

85. Vale C, Sanes DH. The effect of bilateral deafness on excitatory and inhibitory synaptic strength in the inferior colliculus. *Eur J Neurosci.* (2002) 16:2394–404. doi: 10.1046/j.1460-9568.2002.02302.x

86. Caspary DM, Ling L, Turner JG, Hughes LF. Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J Experiment Biol.* (2008) 211:1781–91. doi: 10.1242/jeb.013581

87. Chen JL, Ros T, Gruzelier JH. Dynamic changes of ICA-derived EEG functional connectivity in the resting state. *Hum Brain Mapp.* (2013) 34:852–68. doi: 10.1002/hbm.21475

88. Goossens T, Vercammen C, Wouters J, van Wieringen A. Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. *Hear Res.* (2018) 370:189–200. doi: 10.1016/j.heares.2018.07.012

89. Giroud N, Hirsiger S, Muri R, Kegel A, Dillier N, Meyer M. Neuroanatomical and resting state EEG power correlates of central hearing loss in older adults. *Brain Struct Func.* (2018) 223:145–63. doi: 10.1007/s00429-017-1477-0

90. Giroud N, Keller M, Hirsiger S, Dellwo V, Meyer M. Bridging the brain structure—brain function gap in prosodic speech processing in older adults. *Neurobiol Aging.* (2019) 80:116–26. doi: 10.1016/j.neurobiolaging.2019.04.017

91. Wang M, Zhang C, Lin S, Wang Y, Seicol BJ, Ariss RW, et al. Biased auditory nerve central synaptopathy is associated with age-related hearing loss. *J Physiol.* (2021) 599:1833–54. doi: 10.1113/JP281014

92. Jerger J, Estes R. Asymmetry in event-related potentials to simulated auditory motion in children, young adults, and seniors. *J Am Acad Audiol.* (2002) 13:1–13. doi: 10.1055/s-0040-1715943

93. Hoptman MJ, Davidson RJ. How and why do the two cerebral hemispheres interact? *Psychol Bull.* (1994) 116:195. doi: 10.1037/0033-2909.116.2.195

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 23 August 2022
DOI 10.3389/fneur.2022.960012

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Davide Chicco,
University of Toronto, Canada
Katarzyna Tarnowska,
University of North Florida, United States

*CORRESPONDENCE
Mareike Buhl
mareike.buhl@uni-oldenburg.de
Andrea Hildebrandt
andrea.hildebrandt@uni-oldenburg.de

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 02 June 2022
ACCEPTED 20 July 2022
PUBLISHED 23 August 2022

CITATION
Buhl M, Akin G, Saak S, Eysholdt U, Radeloff A, Kollmeier B and Hildebrandt A (2022) Expert validation of prediction models for a clinical decision-support system in audiology. *Front. Neurol.* 13:960012. doi: 10.3389/fneur.2022.960012

Expert validation of prediction models for a clinical decision-support system in audiology

Mareike Buhl[1,2]*, Gülce Akin[2,3], Samira Saak[1,2], Ulrich Eysholdt[1,2,4], Andreas Radeloff[2,4], Birger Kollmeier[1,2,5,6] and Andrea Hildebrandt[2,3]*

[1]Department of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, [2]Cluster of Excellence Hearing4all, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, [3]Department of Psychological Methods and Statistics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, [4]Universitätsklinik für Hals-Nasen-Ohren-Heilkunde, Evangelisches Krankenhaus Oldenburg, Oldenburg, Germany, [5]Hörzentrum Oldenburg gGmbH, Oldenburg, Germany, [6]Hearing Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology (IDMT), Oldenburg, Germany

For supporting clinical decision-making in audiology, Common Audiological Functional Parameters (CAFPAs) were suggested as an interpretable intermediate representation of audiological information taken from various diagnostic sources within a clinical decision-support system (CDSS). Ten different CAFPAs were proposed to represent specific functional aspects of the human auditory system, namely hearing threshold, supra-threshold deficits, binaural hearing, neural processing, cognitive abilities, and a socio-economic component. CAFPAs were established as a viable basis for deriving audiological findings and treatment recommendations, and it has been demonstrated that model-predicted CAFPAs, with machine learning models trained on expert-labeled patient cases, are sufficiently accurate to be included in a CDSS, but it requires further validation by experts. The present study aimed to validate model-predicted CAFPAs based on previously unlabeled cases from the same data set. Here, we ask to which extent domain experts agree with the model-predicted CAFPAs and whether potential disagreement can be understood in terms of patient characteristics. To these aims, an expert survey was designed and applied to two highly-experienced audiology specialists. They were asked to evaluate model-predicted CAFPAs and estimate audiological findings of the given audiological information about the patients that they were presented with simultaneously. The results revealed strong relative agreement between the two experts and importantly between experts and the prediction for all CAFPAs, except for the neural processing and binaural hearing-related ones. It turned out, however, that experts tend to score CAFPAs in a larger value range, but, on average, across patients with smaller scores as compared with the machine learning models. For the hearing threshold-associated CAFPA in frequencies smaller than 0.75 kHz and the cognitive CAFPA, not only the relative agreement but also the absolute agreement between machine and experts was very high. For those CAFPAs

with an average difference between the model- and expert-estimated values, patient characteristics were predictive of the disagreement. The findings are discussed in terms of how they can help toward further improvement of model-predicted CAFPAs to be incorporated in a CDSS for audiology.

KEYWORDS

precision audiology, CDSS, expert validation, audiological diagnostics, expert knowledge, machine learning, CAFPAs

Introduction

Audiological diagnostics mostly relies on test batteries of audiological measures conducted on a patient in need. Experts in audiology characterize patients' hearing impairment by combining the knowledge derived from those audiological measures and additional information from anamnesis as well as their subjective impression of the respective patient. However, experts' experience differs depending on the number of previously treated patients and the range of seen cases (1). On the other hand, large amounts of diverse patient data are available in clinical databases which originate from different audiological tests. Thus, theoretically, the knowledge saved in different databases could be made available to any audiologist with different levels of expertise. This is one long-term goal of the current research.

Toward precision audiology, the clinical decision-support system (CDSS) provides the potential to improve the objectivity of audiological diagnostics by supporting experts with information about probabilities for different audiological findings or treatment recommendations, such as the usage of hearing devices (2). Thereby, less experienced professionals could be supported by a CDSS with an expanded basis of diagnostic knowledge. However, more experienced experts could benefit from the statistical knowledge fed into a CDSS, which exploits a large amount of data and derives knowledge about base rates and association patterns between features that are relevant for audiological recommendations (2, 3).

Currently, CDSSs are not widely adopted in audiology. This is due to a couple of challenges to be solved, such as the integration of different data sources for the same audiological finding (4), the integration of CDSS into the clinical decision-making process of experts (5), and the accomplishment of interpretability of algorithms implemented into a CDSS by clinicians (3). To overcome the latter challenge, it has been recommended to develop CDSS in collaboration with domain experts in the respective medical field (6–8). Expert knowledge can be incorporated into the developmental process in different regards: First, when planning a CDSS, concepts and definitions need to be discussed with domain experts (2). Second, highly-experienced experts can be asked to provide insights into their decision-making process or can be asked to gain insights into the decision-making process of a trained algorithm to be implemented in a CDSS (3). Furthermore, domain experts are needed to provide labels, i.e., to estimate audiological findings, if those are not yet available in a certain database (unlabeled data) [e.g., (9, 10)]. Finally, whenever algorithms were trained on an existing database (3, 11), domain experts can be asked to validate machine-predicted labels (10, 12, 13), and the concordance between experts' and algorithmic decisions can be statistically evaluated (9).

In audiology, some CDSS approaches exist for different decision types of the field. For example, a CDSS has been designed for tinnitus diagnosis and therapy (14) and another one for diagnosing idiopathic sudden hearing loss (15), and for the selection of a suitable hearing aid device type (16). However, these approaches do not rely on test batteries containing a combination of audiological measurements to comprehensively characterize patients. For such a purpose, Sanchez-Lopez et al. (17, 18) performed a classification of hearing-impaired patients based on published research data. Their auditory profiles classify patients along the dimensions of audibility- and non-audibility-related distortions. Importantly, their approach combines data-driven knowledge with audiological model assumptions (17).

Aiming to further ameliorate clinical applicability, Buhl et al. (19–22) and Saak et al. (23) rendered a series of development steps toward a CDSS for audiology, which strongly relies on expert knowledge and is targeted toward future interpretability and integration across different data sources. The CDSS should operate on diverse clinical databases, and it aims at covering the complete audiological decision-making process, including the classification of audiological findings for given patients, as well as suggesting appropriate treatment recommendations (summarized as diagnostic cases). In the proposed CDSS, Common Audiological Functional Parameters (CAFPAs; 19) were employed as an interpretable intermediate layer between audiological tests and diagnostic cases (cf. Figure 1B). CAFPAs were thus introduced as abstract parameters that aim to cover all relevant functional aspects of the human auditory system, while not depending on the exact choice of audiological measures applied to a patient (19). Figure 1A provides an overview of the defined CAFPAs which represent an abstract and common data format based on which different audiological test batteries can be combined and compared, given that a link from

a respective measurement to CAFPAs has been established. Buhl et al. (19) introduced the choice of 10 CAFPAs and established the first link to audiological measures and diagnostic cases by means of an expert survey in the inverse direction of the audiological diagnostic process. Thus, 11 audiological experts estimated CAFPAs and distributions of audiological measurement outcomes for given diagnostic cases. This study provided a proof of concept and demonstrated the feasibility of the CAFPA approach.

Aiming to establish a link to individual patients which can be used as training data for machine learning approaches, by means of a second expert survey conducted with 12 experts, Buhl et al. (20) collected CAFPA labels and diagnostic cases for the given measurement outcomes of an existing audiological database. The respective database of individuals with mild-to-moderate hearing impairment contained patients' results on the audiogram, one speech test, and loudness scaling. The audiological measures were visually summarized on result sheets for every patient. The patient data was sorted into categories corresponding to expert-estimated diagnostic cases (labels), and probability density functions were derived for each category and each measurement parameter as well as CAFPA. Thereby, plausible distributions that can be used as training data for classifying diagnostic cases were obtained.

Furthermore, Buhl et al. (21) investigated if CAFPAs provide similar information as included in the audiological measurements and, consequently, if the classification in a CDSS can be performed based on the CAFPAs as intermediate representation instead of directly based on the measurements. For this purpose, classification was performed based on measurements and CAFPAs, employing the training distributions from Buhl et al. (20), including cross-validation. These analyses revealed that, in most cases, approximately the same classification performance in terms of sensitivity and specificity was achieved by CAFPAs as with direct measurements. This means that they contain all the relevant information that is important for classification.

In the above-summarized studies, the relationships between audiological measurements and CAFPAs were established based on expert knowledge only. Thus, the link was not quantified by prediction models and therefore the association pattern could not be used as envisaged in the use case of a CDSS, where CAFPAs for individual, new patients need to be automatically predicted. Aiming to establish an automatic prediction of CAFPAs, Saak et al. (23) statistically derived CAFPAs based on the CAFPA expert labels (collected for 240 out of 595 patients included in the database) and the corresponding outcomes of audiological measures from Buhl et al. (20). This was done by means of regularized regression models (with lasso and elastic net penalties) and random forests. The trained prediction models were shown to have an adequate to good performance, with coefficients of determination (R^2) between 0.6 and 0.7 for the CAFPAs related to the hearing threshold. However, the neural CAFPA CN showed insufficient predictive performance (0.17). As compared with the expert labels, the statistical models tended to predict fewer extreme values for CAFPAs (23). Saak et al. (23) also analyzed the importance of different audiological measures (features) for the prediction and demonstrated that the models indicated audiologically plausible relationships between the measurement outcomes and the CAFPAs. Finally, Saak et al. (23) applied the trained models to predict CAFPAs for the unlabeled part of the database and provided the first consistency check of the model-derived CAFPAs by means of an unsupervised learning approach. More specifically, cluster analysis identified five plausible groups of individuals which were in line with the audiological findings. However, no comparison with "true" labels for audiological findings was possible as expert-estimated diagnostic cases (assumed as ground truth) were not available for the unlabeled patients.

Aiming for further validation of statistically derived CAFPA values, to connect all components, and to finally build a CDSS operable for individual patients (based on labeled data), Buhl (22) applied the classification approach from Buhl et al. (21) to technically evaluate the predictions in the use case of a CDSS (Figure 1B, lower left part). The classification was performed on expert-estimated CAFPAs and model-predicted CAFPAs. It has then been investigated which CAFPAs were relevant for high classification performance in different diagnostic decisions. Furthermore, the interpretability of the system was assessed. It was shown that predicted CAFPAs lead to a similar classification of patients into the different diagnostic cases [prediction accuracy of 0.64–0.78 (depending on the investigated audiological parameter) for optimal weighting of CAFPAs]. The predicted CAFPAs can in general already be used in the classification, but some misclassifications occur that can both be related to the fact that less extreme CAFPAs are predicted by the regression models (23), and to the properties of the data set. However, for a definitive validation of the statistically derived CAFPAs, especially for unlabeled patients, their evaluation by independent experts remains indispensable.

For the purpose of investigating if the current CAFPA prediction can plausibly be applied to unlabeled patients (and consequently to new individual patients in the use case of a CDSS) and to further investigate the properties of the prediction models, the present study aims at an expert validation of the statistically derived CAFPAs [blue and green (dashed) arrows in Figure 1B, right part]. Two highly-experienced audiological experts were asked to assess model-predicted CAFPAs given the measurement outcomes of individual patients and to update the values if they considered a given model-derived CAFPA to be inappropriate. The deviations between model-predicted and expert-validated CAFPAs are statistically analyzed to investigate how disagreements between the model and experts might depend on audiological measurements and to understand how the CAFPA prediction could further be

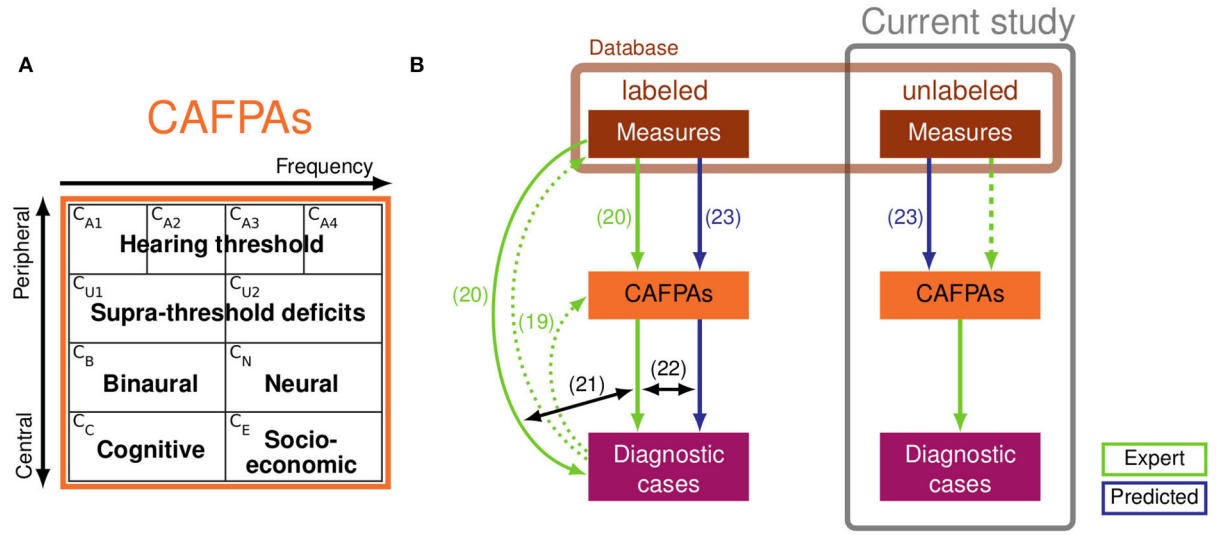


FIGURE 1
(A) Definition of Common Audiological Functional Parameters (CAFPAs). From left to right, the functional aspects CA1-CA4 and CU1-CU2 are frequency-dependent, and from top to bottom, the functional aspects range from peripheral to central. The CAFPAs are defined on a continuum ranging on the interval [0 1], with 0 representing "normal" and 1 representing "maximally impaired". Panel **(A)** of the figure was taken from Buhl (22). **(B)** Schematic representation of the clinical decision-support system (CDSS) by Buhl (22) (left part, based on labeled data) and relationships to the current study (right, based on unlabeled data). Labeled and unlabeled measurement data originate from the same database (light brown box). Light green arrows depict expert knowledge and blue arrows depict statistical predictions of CAFPAs. Numbered arrows represent contributions of previous studies: collection of expert knowledge in the opposite direction of audiological diagnostics (19); collection of expert knowledge based on individual patients from the currently used database (20); comparison of classification based on audiological measures vs. expert-estimated CAFPAs (21); comparison of classification based on expert-estimated CAFPAs vs. model-predicted CAFPAs. The prediction models were developed by Saak et al. (23) based on the expert-estimated CAFPAs from Buhl et al. (20). The prediction models were derived based on labeled patients (left) and applied to unlabeled patients (right). The experts' task in the current study was to validate the model-predicted CAFPAs (dashed light green arrow) and to estimate audiological findings for unlabeled patients.

improved. In addition, experts were asked to also estimate audiological findings based on the given measurement data (for the purpose of collecting corresponding labels for diagnostic cases, cf. Figure 1B, lower right part) and to fill out a short questionnaire asking about how they approached the CAFPA evaluation task.

Specifically, the study aimed to provide an answer to the following research questions (RQs):

1. What is the magnitude of relative and absolute agreement of experts with model-predicted CAFPAs? Whereas the relative agreement indicates whether experts and statistical models provide CAFPAs leading to equivalent rank orders of the evaluated patients, the absolute agreement indicates average deviations from the opinion of experts and models across all patients. Both are relevant criteria to understand the overlap between automatic and expertise-based audiological decision-making based on CAFPAs.
2. If a disagreement between model-predicted and expert-validated CAFPAs exists, does it depend on certain characteristics of the patients' test data?
3. Are the estimated audiological findings consistent with expert labels from previous studies collected from patients in the same database?
4. Is the applied expert validation approach a reliable check of the model-predicted CAFPAs?

Materials and methods

Data set and audiological experts

For the present study, patients' data displayed to the experts along with model-predicted CAFPAs [as estimated by Saak et al. (23)] were provided by the Hörzentrum Oldenburg gGmbH. The dataset contained $N = 595$ cases for which data were available on medical history, speech recognition in noise performance [Goettingen sentence test, GOESA (24)], two audiological measurements [audiogram and adaptive categorical loudness scaling (25)], and performance on two cognitive tests [German vocabulary test, WST (26); and DemTect (27)]. Patients varied with respect to their degree of hearing loss. A detailed description of the database can be found in Gieseler et al. (28). For $n = 240$ patients, expert labels for CAFPAs and audiological findings were collected by Buhl et al. (20).

The model-predicted CAFPAs for unlabeled patients were taken from Saak et al. (23), where three statistical learning models (lasso regression, elastic net, and random forests) were

trained based on 80% of the labeled patients of Buhl et al. (20) and evaluated based on the remaining 20%. The prediction for the 355 existing unlabeled patients was performed using these trained models. Thus, for each statistical learning algorithm, the predictions were obtained by averaging the predicted CAFPAs across 20 models derived from 20 different missing imputed data sets. The code running the prediction models was published along with Saak et al. (23), and it has been applied without any changes. All models performed well, but they were slightly different in their performance accuracy. To account for variation in model performance for the CAFPAs to be evaluated by the experts enrolled in the present study, 50% of the evaluated cases were displayed with estimated CAFPAs based on the best performing model for the respective CAFPA. For the second half of the cases, CAFPAs were taken from the respective worst-performing models.

Two highly-experienced experts (authors AR and UE) evaluated the model-predicted CAFPAs. Both have substantial scientific and clinical experience of more than 20 years (with more than 7,500 seen patients), including all degrees of hearing loss and treatment options. The experts are familiar with the measurements presented in the expert validation survey as well as with measurements performed in clinical practice and their combined interpretation with additional information about patients.

Due to their elaborated experience, two experts were estimated to be sufficient for the purpose of this study. In addition, the experts involved here did not participate in the previous surveys (19, 20) and thereby their expert knowledge was not yet depicted in the current prediction models. This allows for an independent view on the predicted CAFPAs. Moreover, the statistical analysis of differences between the model-predicted and expert-validated CAFPAs (cf. Section Statistical analyses) is better interpretable if the comparison between statistical and expertise-based prediction is performed by individual experts.

Expert survey design

The original survey design from Buhl et al. (20) was adopted and implemented as an electronic survey on PsychoPy 3 Builder (29). Same as in Buhl et al. (20), the information sheet of a given patient was presented to the expert on the left side of

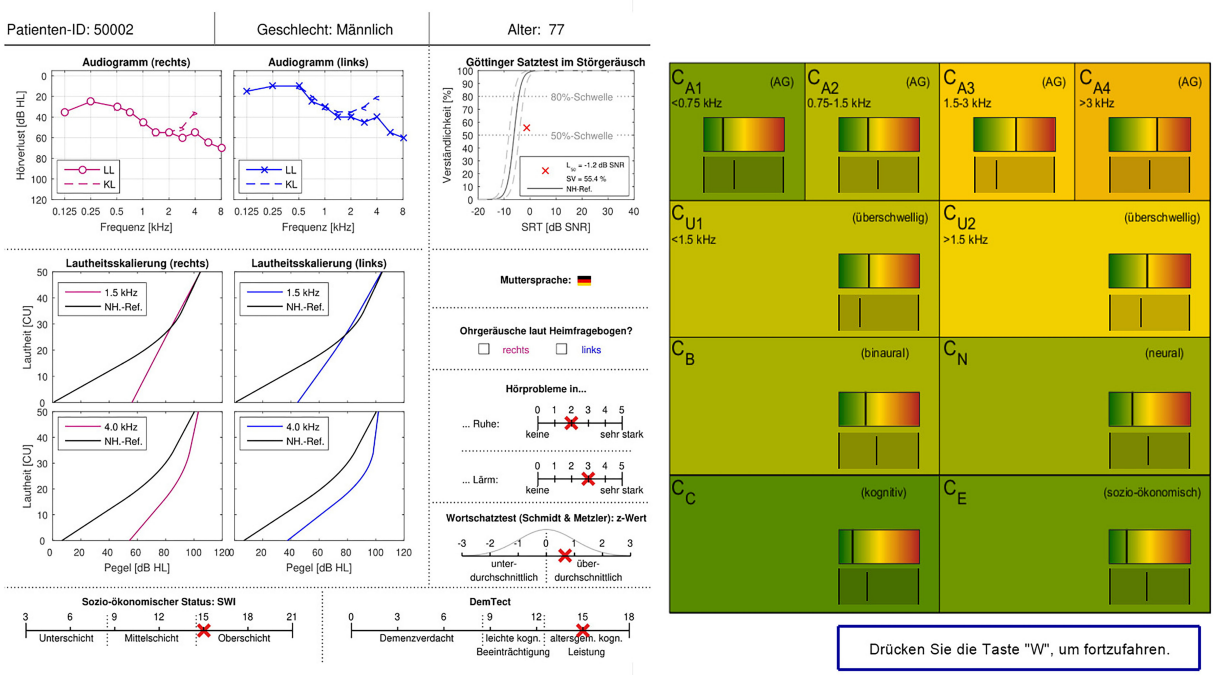


FIGURE 2
Patient data and CAFPA evaluation sheet as implemented in the electronic version of the expert survey. Patient cases were displayed one at a time. The survey sheet is shown in German as the survey was conducted in Germany. For the main terms, a translation is given in the following. Upper row: Patient ID, gender, and age. Measurements: Audiogram (right and left), LL: air conduction, KL: bone conduction, and hearing loss plotted over frequency. Goettingen sentence test (GOESA) in noise, intelligibility plotted over SRT. Loudness scaling (Adaptive CAtegorical LOudness Scaling (ACALOS); right and left), loudness plotted over level, and black line: normal-hearing reference. Native language. Tinnitus according to the home questionnaire (right and left). Hearing problems in quiet and in noise (scale from none to very much). Verbal intelligence test: z-score (negative scores: below average, positive scores: above average). Socio-economic status: lower class, middle class, and upper class. DemTect: suspicion of dementia, mild cognitive impairment, and age-specific normal cognitive abilities. CAFPAs: the meaning of the different parameters is given in Figure 1A.

the screen (see Figure 2), one patient at a time. On the right side of the screen, statistically predicted CAFPAs for the given patient were presented on the range highlighted by the traffic-light color. A visual analog scale of the same range was displayed below. Experts were requested to use this scale and indicate their estimate for all 10 CAFPAs using the respective slider. They were instructed that their slider setting could be perfectly overlapping with the bar indicating the model estimate, or it could deviate from it. The experts were clearly informed about the meaning of the displayed CAFPAs. They thus knew that these were estimates originating from trained statistical algorithms by Saak et al. (23). After placing the slider for all CAFPAs, the experts were able to proceed to the next page by pressing the button displayed at the lower right corner of the screen. On the next page, the same patient's data were displayed again, but on the right side of the screen, audiological findings were now listed, asking the experts to select those that they considered appropriate (multiple answers were allowed). Audiological findings were as follows: 1. normal hearing; 2. cochlear hearing loss (with the options high-frequency, middle-frequency, low-frequency, or broadband hearing loss); 3. conductive hearing loss; 4. central hearing loss. After indicating the appropriate audiological finding(s), experts could proceed with evaluating the next patient. There were separate blocks of 15 patients each, such that experts could interrupt their evaluation for shorter or longer breaks. It was possible to restart the survey on another day and continue with the block of patients who were not yet evaluated before. Experts were not informed about the repeated patients. These were just displayed randomly to them in between new patient cases. Expert 1 evaluated CAFPAs predicted for 150 cases which were randomly selected out of the 355 existing unlabeled patient cases. The cases were chosen to equally correspond to the five clusters of Saak et al. (23) to represent different hearing loss degrees as uniformly as possible. Half of them were predicted with the best and worst performing models, respectively. For evaluating the within-expert agreement, 15 of these cases were presented two times to Expert 1. Expert 2 evaluated 15 patient cases repeatedly, 12 out of those were also evaluated by Expert 1. Expert 2 only received patient cases associated with the CAFPAs predicted by the models with the best performance accuracy.

After each session of 15 cases, a form was displayed, and experts were asked to indicate their confidence in deciding on the CAFPAs' values and the suggested audiological findings. Furthermore, at the end of the survey, they were requested to reveal their expert validation approach and to indicate which measurement information they used while updating each CAFPA. More specifically, we asked whether experts have evaluated the displayed measurements or the statistically estimated CAFPAs first and whether they considered the predicted CAFPAs at all. Furthermore, for each measurement, a list of all CAFPAs was displayed to the experts one by one, and they were asked to mark whether a certain CAFPA was relevant for a given measurement. If none of the CAFPAs was considered to be related to a specific measurement, experts were asked to choose the reason from the options, "The measurement is not known to me," "The measurement is not important for the characterization of patients," or "Not possible to decode or represent in CAFPAs." In addition, the expert's approach to the expert validation task was assessed by a multiple-choice question where different potential approaches or components of those were suggested (Supplementary Tables A1, A2 for details).

Statistical analyses

All analyses were conducted with the R Software for Statistical Computing (30). To estimate the stability of the CAFPA ratings within and relative agreement across experts, as well as the relative agreement between the model-predicted and expert-validated CAFPA, intraclass correlation coefficients (ICCs) were computed along with their 95% confidence intervals (CIs). The ICC is a widely used tool for measuring inter-rater agreement. It indicates a correlation within the same class of data (here repeated measurements of CAFPAs by different sources: Statistical model, Expert 1, and Expert 2). Whereas the correlation coefficient refers to different variables, the ICC is a correlation of the same variable measured in different conditions. The psych package (31) has been used for this purpose by applying a two-way mixed-effects model [ICC3k (32)]. The relative agreement between experts, as well as between statistical models and experts, indicates whether the raters were ranking the patient cases in terms of CAFPAs in an approximately equivalent order. If the patients' rank orders were approximately overlapping between raters, the ICC would take on a value close to 1. Within-expert stability and cross-expert agreement were taken as necessary preconditions (reliability) for estimating the relative overlap between experts' ratings vs. those of the statistical models.

Not only rank order agreement but also absolute agreement was relevant to understand the overlap between model-predicted and expert-validated CAFPAs. To estimate absolute agreement, a series of linear mixed effect regression (LMER) models were fitted by means of the package lme4 (33), separately for each CAFPA as an outcome variable. The condition model-predicted vs. expert-validated was dummy coded (0 = statistical model). Random intercepts were included when regressing a CAFPA onto the within-patient condition factor to estimate the absolute difference between CAFPA ratings of Expert 1 vs. the statistical models. Given the dummy coded within-patient factor, a negative β-weight (fixed effect) will indicate higher CAFPA values provided by the statistical models on average across patients as compared with the expert. In analogy, a positive β-weight indicates the expert to rate a certain CAFPA higher than the model. These analyses were only based on data from Expert 1, because Expert 2 evaluated only a few patients,

TABLE 1 Agreement between experts and stability of experts' ratings.

	E1–E2 (agreement; $N = 15$; rated first time by both experts)		E1–E1 (stability; $N = 15$; rated 2 times)		E2–E2 (stability; $N = 15$; rated 12 times)	
CAFPAs	**ICC [CI]**	***p*-Value**	**ICC [CI]**	***p*-Value**	**ICC [CI]**	***p*-Value**
CA1	0.90 [0.72; 0.97]	0.00	**0.49** [−0.53; 0.83]	0.11	0.99 [0.99; 1.00]	0.00
CA2	0.96 [0.87; 0.98]	0.00	0.97 [0.92; 0.99]	0.00	0.99 [0.98; 1.00]	0.00
CA3	0.95 [0.86; 0.98]	0.00	0.99 [0.97; 1.00]	0.00	0.99 [0.98; 1.00]	0.00
CA4	0.92 [0.75; 0.97]	0.00	0.84 [0.53; 0.95]	0.00	0.98 [0.96; 0.99]	0.00
CU1	**0.52** [−0.43; 0.84]	0.09	0.89 [0.68; 0.96]	0.00	0.96 [0.92; 0.98]	0.00
CU2	0.94 [0.81; 0.98]	0.00	0.90 [0.71; 0.97]	0.00	0.98 [0.96; 0.99]	0.00
CB	singular	0.00	0.85 [0.54; 0.95]	0.00	0.92 [0.84; 0.97]	0.00
CN	**0.00** [−1.98; 0.66]	0.00	0.82 [0.47; 0.94]	0.00	0.96 [0.91; 0.98]	0.00
CC	**0.71** [0.15; 0.90]	0.01	0.96 [0.88; 0.99]	0.00	0.94 [0.88; 0.98]	0.00
CE	0.86 [0.58; 0.95]	0.00	0.97 [0.91; 0.99]	0.00	0.96 [0.92; 0.98]	0.00

CA1–CA4, hearing threshold-related CAFPAs; CU1–CU2, Suprathreshold-deficits related CAFPAs; CB, binaural hearing; CN, neural processing; CC, cognitive components of hearing; CE, socio-economic status; E1, Expert 1 who rated 15 patient cases two times; E2, Expert 2 who rated 15 patient cases 12 times; ICC, intra-class correlation; CI, confidence interval.
Bold numbers indicate estimated agreements with a lower than acceptable effect size.

but repeatedly multiple times. Per design, the data from Expert 2 were collected for reliability estimates with many repetitions.

Last, we aim to test whether the measured audiological data of the patients can explain potentially observed differences between the model-predicted and expert-validated CAFPAs. Thus, patients' audiological measures were included as additional predictors in the above described within-patient factor models, estimated separately for each CAFPA. Cross-level interactions between the within-patient condition variable and measurements tested whether the difference between the expert and the statistical model depended on the audiological measurements.

After performing the described statistical analyses, a post-survey interview with the experts was conducted. In a semi-structured discussion with all coauthors (from which two acted as experts), all results and links among the results were discussed, while especially focusing on the experts' perspective.

Results

Stability of experts' ratings and agreement between experts

Prior to assessing the agreement between statistical CAFPA predictions vs. experts' evaluations, the reliability of experts' ratings needs to be quantified. Table 1 provides a comprehensive summary of these reliability analyses for the 10 CAFPAs (displayed as columns). Within-expert agreements were very high as indicated by the ICC values close to 1. The ICCs expressing very high stability within Expert 2, who rated the CAFPAs many times repeatedly, are all above 0.90, with a very narrow CI. Thus, learning effects during the first round of ratings were adjusted by multiple repetitions in this case. The ICCs indicating stability within Expert 1 are somewhat lower, but satisfactory (all above 0.80), except for the CA1. However, CA1 was the CAFPA to be rated first, and the 15 patients used for stability estimates were presented as the first cases to the expert and repeated later. Thus, the low ICC of this first CAFPA can be explained by the fact that the expert had to familiarize himself with the task at the beginning of the survey. This was probably the case for the second expert as well; however, by analyzing "12 repetitions in that case," the agreements were adjusted, and one run of ratings will not have such a substantial effect on the agreement estimates across 12 columns of 15 patients' ratings.

Experts 1 and 2 were in high agreement with respect to all but three CAFPAs (refer to the first column of Table 1). The outlier CAFPAs were CU1, CB, and CN. In the case of CB and CN, the two experts did not agree with each other at all, such that the model returned a hint toward singularity. By exploring the distribution of the CB estimates within Expert 1 and Expert 2, it became obvious that the first expert evaluated all 15 patient cases used for reliability estimates with an approximately zero CB value and a very narrow value range slightly above zero in the case of CN. This was not the case for Expert 2 who used a somewhat broader but also restricted value range for these two CAFPAs. A post-survey interview with both experts provided further insights into the experts' reasoning on these patient cases with respect to CB and CN. These qualitative reports are outlined below in the discussion section and used for interpreting the quantitative findings summarized in Table 1. Overall, we can conclude that, for most of the CAFPAs, the experts' evaluations were reliable in terms of stability within experts and agreement of two different experts with different experience backgrounds.

Relative agreement between CAFPAs predicted by statistical models vs. experts (RQ 1)

Table 2 provides a comprehensive summary of the ICC estimates indicating an agreement between the statistically predicted CAFPAs and the two experts based on 15 cases rated by all. The second column of the table indicates an agreement of CAFPA predictions between the statistical model and Expert 1 on the basis of 150 patients. These relative agreements between the models and Expert 1 are also displayed as scatterplots in Figure 3, separately for each CAFPA. The table and the scatterplots clearly reveal high agreement rates of experts with the statistically predicted CAFPAs, except for CB and CN. We can thus conclude that 8 out of 10 CAFPAs are valid and can be readily used in a CDSS for audiological decision-making. Reasons for the low validity of the statistically predicted CB and CN, as well as potential measures for improving the prediction of these two CAFPAs in the future, are discussed below.

TABLE 2 Relative agreement between statistically predicted CAFPAs and experts' opinion.

	M-E1-E2		M-E1	
CAFPAs	**ICC [CI]**	***p*-Value**	**ICC [CI]**	***p*-Value**
CA1	0.94 [0.85; 0.98]	0.00	0.94 [0.92; 0.96]	0.00
CA2	0.98 [0.94; 0.99]	0.00	0.96 [0.94; 0.97]	0.00
CA3	0.97 [0.93; 0.99]	0.00	0.96 [0.95; 0.97]	0.00
CA4	0.94 [0.87; 0.98]	0.00	0.94 [0.91; 0.95]	0.00
CU1	**0.73** [0.36; 0.90]	0.00	0.86 [0.80; 0.90]	0.00
CU2	0.94 [0.86; 0.98]	0.00	0.90 [0.86; 0.93]	0.00
CB	**0.63** [0.13; 0.87]	0.01	**0.56** [0.39; 0.68]	0.00
CN	**0.39** [−0.43; 0.78]	0.13	**0.43** [0.21; 0.59]	0.00
CC	0.88 [0.72; 0.96]	0.00	**0.75** [0.65; 0.82]	0.00
CE	0.91 [0.79; 0.97]	0.00	0.82 [0.75; 0.87]	0.00

CA1–CA4, hearing threshold-related CAFPAs; CU1–CU2, Suprathreshold-deficits related CAFPAs; CB, binaural hearing; CN, neural processing; CC, cognitive components of hearing; CE, socio-economic status; M, model = statistical model-predicted CAFPA, refer to Saak et al. (23); E1, Expert 1 who rated 15 patient cases two times and in total 150 different patients (used in second column M-E1); E2, Expert 2 who rated 15 patient cases 12 times; ICC, intra-class correlation; CI: confidence interval.
Bold numbers indicate estimated agreements with a lower than acceptable effect size.

Absolute agreement between CAFPAs predicted by statistical models vs. experts (RQ 1)

We next investigated the absolute agreement between CAFPAs predicted by statistical models vs. experts. Despite proximal rank order equivalence of patients between experts and statistical decisions on the CAFPAs, the question remains whether, on average, across patients, experts, and the models agree. Table 3 provides a numeric summary of the results (see above for explanations of the modeling approach). As indicated by the first column of the table (β-weights), all but two differences were negative. This means that the CAFPAs CA1–CA4, CU1–CU2, CB, and CN were on average corrected across patients to lower values by Expert 1 as compared with the predictions of statistical models. On a scale between 0 and 100 (rescaled CAFPAs to range between 0 to 100, instead of 0 to 1), these negative differences ranged between 2.09 and 17.79 scale point units. Thus, most of the average differences between the expert's vs. the statistical models' CAFPA estimates were very

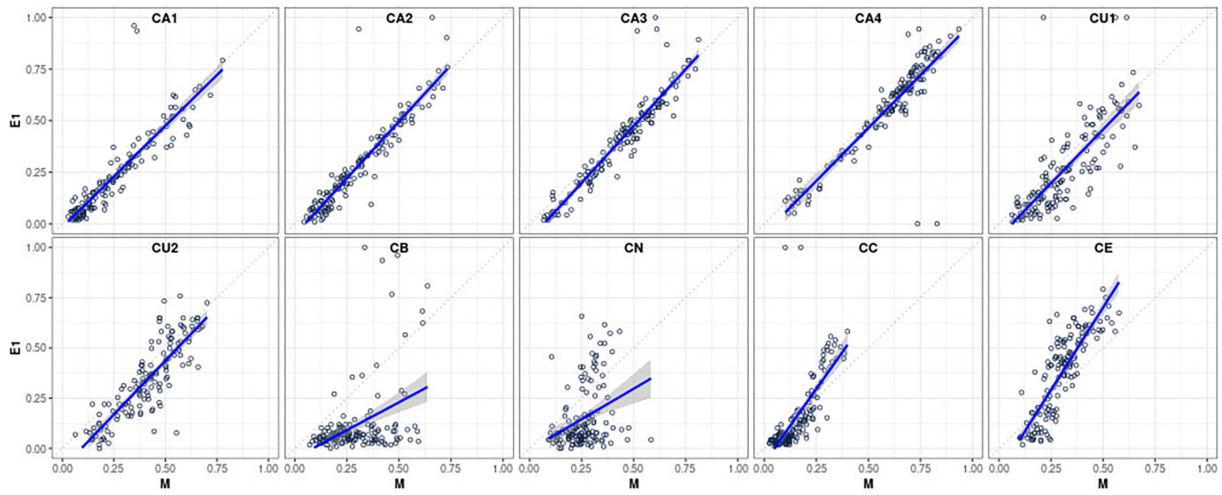


FIGURE 3
Scatterplots visualizing the relative agreement between statistically predicted CAFPAs and the expert's opinion (for Expert 1, $N = 150$ patients; corresponding to the second column of Table 2).

10.3389/fneur.2022.960012

TABLE 3 Main effect of evaluator in the linear mixed effects regression (LMER) models with evaluators (M and E1) nested within patients.

CAFPAs	β (SE)	CI	*p*-Value
CA1	−2.09 (0.70)	−3.48; −0.71	0.00
CA2	−2.33 (0.64)	−3.60; −1.06	0.00
CA3	−3.20 (0.64)	−4.46; −1.94	0.00
CA4	−3.37 (0.85)	−5.05; −1.68	0.00
CU1	−5.14 (1.07)	−7.24; −3.04	0.00
CU2	−6.95 (0.83)	−8.60; −5.31	0.00
CB	−17.79 (1.43)	−20.61; −14.98	0.00
CN	−10.71 (1.24)	−13.61; −8.27	0.00
CC	0.27 (1.04)	−1.79; 2.33	0.79
CE	7.21 (1.05)	5.15; 9.27	0.00

CA1–CA4, hearing threshold-related CAFPAs; CU1–CU2, Suprathreshold-deficits related CAFPAs; CB, binaural hearing; CN, neural processing; CC, cognitive components of hearing; CE, socio-economic status.
Evaluator was dummy coded with 0 = machine learning model, 1 = expert (1). $N_{patients}$ = 150. β: regression weight (fixed effect) of CAFPAs depending on the within-patient factor (machine learning model vs. expert); it indicates the difference between experts' ratings across patients on average as compared with the statistical model; SE, standard error of the regression weight estimate; CI, confidence interval.

small but significant. Larger deviations only occurred for CB and CN, for which statistical predictions turned out to be currently still insufficiently valid in terms of relative agreements as well. The cognitive processing and socio-economic CAFPAs (CC and CE) were rated on average across patients slightly higher by the expert as compared with the statistical models. However, the difference was not significant for CC.

On the dependency of the disagreement between statistical models and the expert from patients' characteristics (RQ 2)

Given that expert and statistical predictions slightly but significantly differed on average, we explored whether patient characteristics (their audiological measurements) explain these differences. The modeling approach has been outlined above and the results are summarized in Table 4. For better readability, only significant effects are provided in the table. However, note that all listed interactions were estimated as explained above and in the note of the table.

The difference for CA4 does not depend on any patient characteristics, and for none of the CAFPAs, the difference between the expert and the model was associated with the age of the patients. In the post-survey interview (see also discussion below), experts also confirmed not to have considered the age when concluding about any of the CAFPAs. The difference between the statistical model and expert evaluation of the socio-economic CAFPA depended on the biological sex of the patients, which is plausible, given sex differences in status evaluations in society in general. Patient differences in pure tone average (PTA) explained the difference between the expert and the model on CA1-CA3. PTA also explained differences in the neural processing CAFPA; however, in general, the results of this CAFPA need to be interpreted with caution. The speech recognition in noise performance (see above GOESA) was relevant for the observed differences on CU1–CU2, CB, and CN. These results were also discussed with the experts in the post-survey interview and were in line with the experts' reports with respect to which measurements they considered when intending to correct the displayed model's estimated value for a given CAFPA. Finally, Adaptive Categorical Loudness Scaling (ACALOS) further contributed to accounting for the difference between the expert and the statistical model.

Questionnaire about experts' approach and relationships between measurements and CAFPAs (RQ 4)

The general questionnaire part of the survey provided additional subjective information to be linked with the analysis outcomes. The answers (by Expert 1) about the expert validation approach revealed that the expert considered patient characteristics as a complete picture. In addition, specific links between measurements and CAFPAs were considered from both directions, that is, thinking about which measurement information was important for a certain CAFPA, as well as to which CAFPAs a certain measurement contributed. The exact choice and formulation of answers are provided in the Supplementary Table A1.

Related to that, the questions about associations between CAFPAs and a respective measurement provided more detailed information about the links indicated by the expert. The CAFPAs CA1–CA4 were clearly related to the audiogram; the cognitive CAFPA CC to the verbal intelligence test (WST) and to DemTect; and the socio-economic CAFPA CE to the SWI. In contrast, CU1–CU2 and CN were related to a combination of audiogram, ACALOS, GOESA, native language, and verbal intelligence test. The binaural CAFPA was not linked to any measurement, meaning that the expert found no information about this aspect in the patient characteristics. These links are plausible and comparable to the results of the statistical analyses as described above, as well as to the variable importance analysis by (23).

CAFPA distributions for given audiological findings (RQ 3)

Finally, we investigated the differences between model-predicted and expert-validated CAFPAs sorted to audiological findings as estimated by the experts, for

TABLE 4 β-weights (of the cross-level interaction) indicating whether the difference between the expert and statistical model depends on the patients' audiological measures.

	Δ_{CA1}		Δ_{CA2}		Δ_{CA3}		Δ_{CA4}		Δ_{CU1}		Δ_{CU2}		Δ_{CB}		Δ_{CN}		Δ_{CC}		Δ_{CE}	
Predictors	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**	**β**	***p*-Value**
Age																				
Sex			**3.24**	0.00					**6.15**	0.02									**−4.31**	0.02
PTA	**0.10**	0.01	**0.17**	0.00	**0.17**	0.00									**−0.42**	0.00				
SES																			**−2.70**	0.00
GOESA									**2.59**	0.00	**1.75**	0.00	**−2.03**	0.00	**4.18**	0.00				
WST																				
DemTect															**1.05**	0.03	**−2.10**	0.00		
$Tinnitus_{right}$			**−4.73**	0.00																
$Tinnitus_{left}$																				
$ACALOS_{1.5L2.5}$	**−0.10**	0.00											**−0.21**	0.02						
$ACALOS_{1.5L50}$			**0.11**	0.00					**0.22**	0.04					**−0.46**	0.00				
$ACALOS_{4L2.5}$																				

Note that only significant results have been listed and an empty cell in the table indicates a null effect. Shaded rows or columns indicate that no significant results were obtained at all for the respective predictor or CAFPA.
CA1–CA4, hearing threshold-related CAFPAs; CU1–CU2, Suprathreshold-deficits related CAFPAs; CB, binaural hearing; CN, neural processing; CC, cognitive components of hearing; CE, socio-economic status.
Δ indicates the difference between the expert and the statistical models. p-values indicate the probability of observing the respective prediction of the difference, or more extreme ones, assuming the null hypothesis of no difference is true. The coefficient estimates originate from 10 different models, one model for each CAFPA. All predictors listed in the table were simultaneously included in the model, along with their interaction with the within-patient condition variable (model = 0; expert = 1). Thus, β-weights indicate cross-level interaction effects (within-patient condition variable and between-patient predictors as listed in the first column of the table).

10.3389/fneur.2022.960012

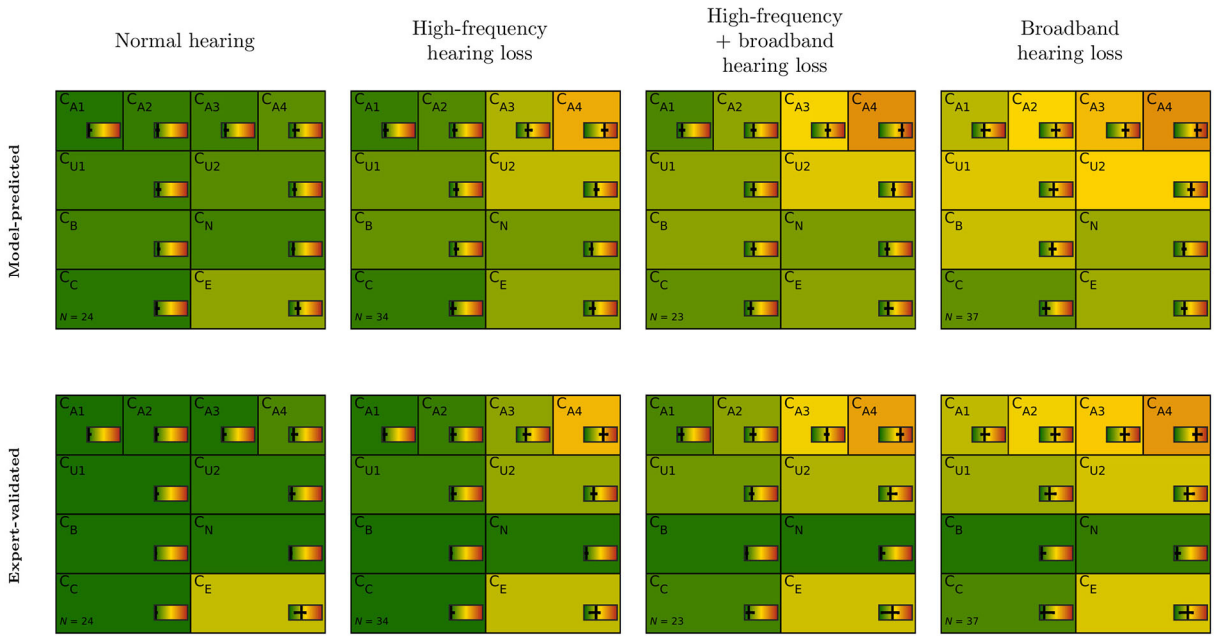


FIGURE 4
CAFPA patterns for the four most frequent audiological findings (columns) as indicated by Expert 1. Model-predicted (first row) and expert-validated CAFPAs (second row). The background color represents the median of the respective CAFPA for all patients associated to the respective audiological finding. The horizontal color bar includes the interquartile range in addition to the median.

the purpose of performing a plausibility check in the applied context toward a CDSS. From the 150 patient cases evaluated by Expert 1, the combinations of four audiological findings were mainly chosen: normal hearing, high-frequency hearing loss, broadband hearing loss, and the combination of high-frequency and broadband hearing loss. Other findings were chosen very rarely (less than six).

Figure 4 depicts model-predicted and expert-validated CAFPAs for different audiological findings. Usually, only small differences are visible by comparing the median (background color) of model-predicted and expert-validated CAFPAs. Thus, the differences as described above comprise a small influence of CAFPAs as compared to the possible range and vary only a little across audiological findings. Interquartile ranges of CAFPAs within audiological findings are partly larger for expert-validated CAFPAs, showing that the expert found slightly more variability across patient cases than was covered by the prediction models. For CB (binaural) and CN (neural), the correction toward zero as described above influenced all audiological findings in the same way, resulting in median values close to zero and a very small interquartile range. A more detailed view on interquartile ranges along with distributions of the different CAFPAs is displayed in Supplementary Figure A3.

Discussion

The present study aimed at an expert validation of model-predicted CAFPAs to be used as an intermediate layer in a CDSS for audiology. For this purpose, we performed an expert survey with two highly-experienced audiological experts and statistically analyzed differences between model-predicted and expert-validated CAFPAs, as well as associations of the observed differences with audiological measurements and patient characteristics.

Expert validation of model-predicted CAFPAs

The main finding was that experts agreed on most model-predicted CAFPA values, except for the binaural CAFPA CB, and the neural CAFPA CN (RQ 1). For these, in a considerable number of patients, large corrections were proposed by experts. This finding was consistently revealed by different statistical analyses, i.e., the assessment of relative and absolute agreement between experts and prediction models, the questionnaire inquiring about the experts' validation approach and their understanding of the relationships between audiological measurements and the

different CAFPAs, and the evaluation of CAFPAs aligned to expert-estimated audiological findings.

For all CAFPAs, except for CB and CN, experts proposed only small corrections on the model-predicted CAFPAs given the measurement data of a sample of patients. Therefore, we conclude that the model-based prediction of these CAFPAs is already well applicable to unlabeled patients. Slight potential for improvement can however be inferred based on the results obtained. The relative agreement between prediction models and both experts was high, except for the supra-threshold CAFPA CU1. The same was applied to the cognitive CAFPA CC when assessing the agreement between model-predicted CAFPAs and Expert 1. Consequently, the agreement among the two experts was rather narrow, but still acceptable for CU1 and CC.

Interestingly, the main evaluator effect (absolute agreement) assessed between prediction models and Expert 1 was significant for all CAFPAs, but not CC. That is, the cognitive CAFPA was on average across patients not corrected by the expert. This could be due to the fact that the range of available patient data is restricted especially in the case of CC where low CAFPA values represent typical functioning. According to the variable importance analyses by Saak et al. (23) and the experts' reports, the CC CAFPA was mainly estimated and concluded on the basis of the DemTect scores, which is a screening test for cognitive impairment. DemTect scores in the present sample, however, are rather in the typically functioning range.

Linear mixed effects regression models revealed that the small, but statistically significant evaluator effects, reflecting differences between the model-predicted and expert-validated CAFPAs, on all remaining seven CAFPAs followed mostly plausible associations with audiological measurements (RQ 2). For instance, analyses indicated that patients' GOESA scores were significantly associated with four CAFPAs, namely CU1, CU2, CB, and CN. This relationship is especially plausible for the supra-threshold CAFPAs, CU1, and CU2, as well as the neural CAFPA CN (see below). However, theoretically one would expect that the binaural CAFPA would not be associated with GOESA, which was measured in the S0N0 condition (speech and noise from the frontal direction), i.e., binaural processing should not be characterized by the given speech test outcome. Furthermore, these empirical relationships were in line with the experts' responses in the questionnaire where they were asked to indicate expected links between audiological measurements and the different CAFPAs. This procedure is similar to the variable importance analysis of Saak et al. (23), which illustrated the links between audiological measurements (features) and the CAFPAs by means of statistical associations learned from the labeled part of the dataset.

In contrast, for the binaural CAFPA CB and the neural CAFPA CN, the relative agreement between experts and the prediction model was limited. The absolute agreement analyses showed the largest differences between model-predicted and expert-validated CAFPAs, for these among all other CAFPAs as well (RQ 1). These findings can be interpreted in the light of all analyses conducted in the present study. The difference between the model-predicted vs. expert-validated binaural CAFPA CB was associated with patients' scores on GOESA and ACALOS, while the expert indicated in the questionnaire that none of the provided measurements allows for conclusions about this CAFPA. In a post-survey interview with both experts, the questionnaire statement was confirmed one more time. That is, according to both experts, the available measurements displayed in the expert survey and used for statistical predictions of CAFPAs do not provide sufficient information about binaural processing (RQ 2). This assessment is consistent with the literature (34–39). Both experts agreed in the joint interview that information from a localization task, as well as speech intelligibility measured in a spatial condition, would be needed for CB evaluation, whereas the displayed condition for GOESA was S0N0. However, Expert 1 also reported being able to gain an impression of the binaural hearing abilities of patients from the available data. A potential decision strategy would be as follows: One would adapt the CAFPA CB toward zero (green, normal) if no binaural problem was expected in the light of all other measurements provided. Therefore, in the case of CB, the absolute agreement and relationships with the audiological measurements need careful interpretation in line with these reports of the expert. Nevertheless, the revealed associations by the statistical analyses may also indicate experts' implicit assumptions about the measurements which are not explicated in their decision-making process.

The evaluator effects for the neural CAFPA CN were associated with several measurements, namely the audiogram (PTA), GOESA, DemTect, and ACALOS. Out of these, GOESA was most strongly associated with CN updates by the expert. These associations are mainly consistent with the questionnaire reports. However, in the post-survey interview, Expert 1 emphasized again his decision-making strategy and commented on the importance of these measurements for the assessment of the neural CAFPA CN. According to both experts, generally in clinical practice, the challenge persists with evaluating neural aspects of hearing loss. These can be characterized by certain measurements such as brainstem-evoked response audiometry or electrocochleography (31), but there is no common and established selection of measurement approaches, and the availability of such measures largely varies across patient cases. Therefore, experts' diagnostic decision-making process contains several steps. They reported to first consider the audiogram and a speech test in combination, and only if inconsistencies pop up, additional measurements, such as brainstem-evoked response audiometry or electrocochleography would be potentially suggested. This diagnostic rationale explains the approach explicated by Expert 1 on how he approached the validation task: CN for patients with consistent results among the audiogram and GOESA has been corrected toward zero. Thereby, the expert

validation of CN relies on the partially explicated diagnostic rationale only, given that no additional information on neural sources of hearing loss was available in the studied patient database. These aspects need improvement toward a reliable CDSS algorithm in the domain of CN and also CB.

The audiological findings as estimated by the experts provided further opportunities to assess how decisive differences between model-predicted and expert-validated CAFPAs were for the final diagnostic outcome (RQ 3). The CAFPA patterns of patients sorted into distinct classes according to the experts' labels for audiological findings were consistent with those which were statistically derived by Saak et al. (23) when clustering unlabeled cases based on model-predicted CAFPAs. The most frequently occurring diagnostic findings (normal hearing, high-frequency hearing loss, broadband hearing loss, and a combination of high-frequency and broadband hearing loss) are approximately equally distributed. This is a consistency check, given that the patients for the current survey were chosen to equally represent the clusters of Saak et al. (23). By comparing the CAFPA distributions (median) of model-predicted and expert-validated CAFPAs, we found in general no noticeable changes in the CAFPA patterns for all CAFPAs except for CB and CN. That is, the above-discussed approach of the experts (correcting these CAFPAs toward zero if no inconsistencies in the data were present) had a similar impact on all audiological findings. This is plausible given that the employed categories of audiological findings [as introduced in Ref. (20)] mainly relate to audibility, and most of the patients did not show extreme findings with regard to binaural hearing or neural aspects of hearing loss. This is in general a property of the database which contains mainly mild-to-moderate hearing impairment collected in a pre-clinical context for the purpose of hearing aid fitting.

In summary, the performed expert validation and corresponding statistical analyses revealed that the CAFPA prediction models as trained by Saak et al. (23) are applicable to unlabeled patient cases. For all CAFPAs except for CB and CN, the expert-validated CAFPAs as well as the audiological findings collected in this study can be additionally used for further training of the prediction models.

For CB and CN, the current prediction models need improvement by considering additional measurements. In these cases, with the measurement data at hand, experts indicated the respective CAFPAs to be normal if no inconsistencies were observed in the data. They both concluded that additional information was necessary to evaluate CB and CN. It is thus plausible that the expert's diagnostic decision-making approach for these two CAFPAs is not reflected by the models that learn from the multivariate association pattern of the audiological test battery taken as input and are by design not able to apply If-Then rules in a similar way as experts do. However, the current predictions are still useful as a starting point or the first best guess for CB and CN. Future models need to be trained on additional information for these two CAFPAs on a potentially more comprehensive clinical sample.

On the importance of experts' qualitative reports on their decision-making approach to improving statistical predictions

The present study clearly demonstrated the importance of combining expert knowledge and statistical learning in the design of a CDSS for audiology. The expert validation and corresponding statistical analyses to investigate agreement between model-predicted and expert-derived CAFPAs provided important insights into the current properties and the necessary future improvement of the CDSS proposed by Buhl (22) and the prediction models of CAFPAs (23). Furthermore, the collected qualitative data on the experts' decision-making process are highly valuable to complement statistical conclusions.

Questionnaire reports revealed that the experts were confident in evaluating model-predicted CAFPAs and combining these statistical proposals with their views on the respective audiological findings (RQ 4). First, this conclusion is supported by plausible expert-validated CAFPAs, which are consistent with the indicated links between measurements and CAFPAs by experts in the questionnaire. Second, the questionnaire also assessed the experts' approach to the task. These data confirmed that Expert 1 was comfortable with the task of making diagnostic decisions on the basis of proposed solutions achieved by statistical predictions. The concept of CAFPAs was also valued by the expert. In summary, the expert concluded a case based on an overall impression of the patient in terms of measurements as well as CAFPAs and additionally reflected upon the respective links between these two information sources. As a limitation, it should be however mentioned that only two audiological experts were involved in this study, and future studies will need to validate a designed CDSS on additional experts with different levels of experience. The two experts involved in this study are highly experienced and provided valuable insights and opinions in a post-survey interview. Their suggestions are consistent with literature, e.g., regarding their reported limitations, such as insufficient available measurements for CB and CN hitherto considered for deriving these CAFPAs. Future studies with more experts with varying levels of experience could assess how the approach to correcting CAFPAs and associations between measurements and CAFPAs implied by the experts' opinion depend on experts' experience. Also, it could be investigated which level of experience is required to perform the expert validation task accurately. It will be crucial that only experts are included who are sufficiently familiar with the typical audiological diagnostic process and are well acquainted with the CAFPA concept.

Their knowledge may be structured differently depending on the experience. Potentially, experts have more implicit links between different aspects of the audiological diagnostic process given higher levels of experience.

The current expert validation was highly informative for the successful implementation of CAFPAs for designing a CDSS for audiology (RQ 4): (1) The model-predicted CAFPAs were validated here by experts, (2) the expert-validation data were statistically analyzed, and (3) qualitative questionnaire and post-survey interview reports of the experts provided a consistency check and additional insights on the experts' decision-making process (9, 10, 13), as discussed above. Thereby, experts' opinions collected here assure the use of CAFPAs in the context of CDSS (2). It should be mentioned that the present expert survey was closely related to the expert survey procedure of Buhl et al. (20). This ensures comparability of the obtained experts' labels and diagnostic conclusions. However, there was a crucial difference. The present study employed an expert validation of model-predicted CAFPAs for previously unlabeled cases instead of simple labeling of CAFPAs. This has the advantage to provide information on how experts accept diagnostic conclusions suggested by a data-driven diagnostic approach.

In summary, the present study contributed to linking expert knowledge and machine learning toward the development of a CDSS for audiology. This link needs to be interpretable. Interpretability was assured in several regards in the current CDSS (22) as well as in the analysis applied in this study. First, the CAFPAs themselves act as an interpretable intermediate layer of a CDSS (19). Second, the variable importance assessments in Saak et al. (23) provided a basis for interpretability of the statistical learning models and allowed insights into the underlying measurements for the different CAFPAs. Third, in the present study, by means of linear mixed effect models, we investigated how differences between model-predicted and expert-validated CAFPAs depend on audiological measurements of the patients. Thereby, we could learn about the experts' implicit approach and interpretation of the CAFPA concept. Although the current version of the CDSS based on CAFPAs was built upon only one audiological database, the proposed methodological approach is generalizable to further data of a similar structure.

Toward future application in the clinical decision-support system and outlook

The outcomes of the present study provide insights into how the CDSS of Buhl (22) could be further improved toward applicability for new patients. For all CAFPAs except for the binaural CAFPA CB and the neural CAFPA CN, the prediction models of Saak et al. (23) can be improved by including the expert-validated CAFPAs as additional labels in the training process and thereby taking the proposed corrections of the two experts involved in this study into account. In the future, this could be done even more efficiently, for example, by using a procedure as described by Baur et al. (13). There, an iterative data annotation approach has been suggested. First, a machine learning algorithm is trained based on a number of available labeled data points, and then, expert labeling is included iteratively by presenting experts with those respective data points that show the most uncertain labels.

For CB and CN, the prediction models of Saak et al. (23) are not yet accurate enough in their current version for use in a CDSS. The automatic prediction of the binaural CAFPA should be included in the future as soon as a database with appropriate audiological measurements is available. The neural CAFPA will require even more research to be included in the decision-support system. This is because the diagnostic process for neural aspects of hearing loss is not well-defined by domain experts, not even with respect to the choice of necessary measurements for a straightforward diagnostic. More specifically, including CN, further discussions with clinicians from different sites are needed to learn more about which measurements are employed for which patients in the clinical practice. Second, appropriate datasets need to be accessed that contain consistent measurement outcomes across patients. This step may include existing datasets, but it may also be necessary to collect structured data for a new group of patients. Third, if data are available, expert labels for CAFPAs can be collected, and/or CAFPAs can be predicted, and a subsequent expert validation be performed (see below for a discussion about expert validation for including additional databases).

The integration of additional databases including more balanced and more severe patient cases is required not only to back up the CDSS with a larger number of patients but also to cover the whole range of potential audiological findings and treatment recommendations. Therefore, the CAFPAs provide great potential, as they are defined as a measurement-independent representation of audiological knowledge. The applied expert-validation approach can be used in the future to validate CAFPAs that were predicted on the basis of different audiological measurements and variable amounts of information available for different patients. This is relevant because clinical practice is characterized by heterogeneity in data availability for different patient cases. In this respect, the expert validation approach could be included in two ways in a hybrid ML-based CDSS combining machine learning and expert knowledge. On the one hand, as explained above, expert validation can be used to derive corrected CAFPAs for additional measurement information in a to-be-connected database. Thereby, it could also be beneficial if the specialization of a respective expert corresponds to the new measurements contained in a dataset. On the other hand, the expert validation could be used on the basis of single patients during the operation of the CDSS in clinical practice, i.e., if the uncertainty of

the predicted CAFPAs (or classified audiological finding or treatment recommendation) exceeds a certain threshold, the system would ask for an expert validation of CAFPAs for the respective patient [related to the approach of Ref. (13)]. In this case, either the current physician could be asked to expert-validate the CAFPAs, or the CDSS would not continue for the current patient, but the patient's data and CAFPAs would be stored to later perform (offline) expert validation on such stored cases.

In contrast to knowledge- or rule-based CDSS (40), expert knowledge would not explicitly be modeled to be incorporated in an ML-based CDSS. Instead, expert knowledge is implicitly incorporated into the CDSS, as it is included in the data (labels for CAFPAs or diagnostic cases) and the relationships between different layers of the CDSS (audiological measures, CAFPAs, and diagnostic cases) are derived from data (supervised ML). With expert validation as performed in this study, the data (CAFPAs) underlying these relationships can be optimized to best fit to experts' implicit understanding of the relationships.

Overall, the present study demonstrated not only the need, but also the potential to incorporate diverse information on expert knowledge in the development (and application) of a CDSS.

Conclusion

The present study provided important insights into the advantages, limitations, and potential improvement of the current prediction of CAFPAs.

The performed expert validation and corresponding statistical analyses revealed that the current CAFPA prediction models are applicable to unlabeled patient cases. For all CAFPAs except for the binaural CAFPA CB and neural CAFPA CN, the experts' agreement with the model-predicted CAFPAs was high, and only small corrections were performed, which were associated with plausible underlying audiological measures by the linear mixed effect models. Therefore, the expert-validated CAFPAs can be employed as additional labels for further training of the respective CAFPAs 'prediction models.

In contrast, large corrections were performed for the CAFPAs CB and CN. The expert's approach of correcting these CAFPAs toward zero if the overall impression of the patient was normal was revealed by the post-interview, along with the fact that appropriate measurement information was missing in the database. The current predictions are useful as a starting point or the first best guess for CB and CN, but future models need to be trained on additional information for these two CAFPAs on a potentially more comprehensive clinical sample.

Audiological findings were found to be consistent with previous expert labels on the same data set. Due to the definition of these categories mainly in threshold-related terms, the large corrections for CB and CN similarly affected all audiological findings.

In summary, the present study contributed to linking expert knowledge and machine learning toward the development of a CDSS for audiology. By means of linear mixed effect models, we investigated how differences between model-predicted and expert-validated CAFPAs depend on audiological measurements of the patients. Thereby, we could learn about the experts' implicit approach and interpretation of the CAFPA concept. Although the current version of the CDSS based on CAFPAs was built upon only one audiological database, the proposed methodological approach is generalizable to further data of a similar structure.

In the future, the expert validation approach could also be used to establish relationships with additional measurements included in different databases. If a prediction is performed on parts of a database, experts could be asked to validate and correct the predicted CAFPAs based on a larger choice of measurements presented within the expert validation survey.

Data availability statement

The data analyzed in this study was obtained from Hörzentrum Oldenburg gGmbH, the following licenses/restrictions apply: According to the Data Usage Agreement of the authors, the datasets analyzed in this study can only be shared upon motivated request. Requests to access these datasets should be directed to MB, mareike.buhl@uni-oldenburg.de and AH, andrea.hildebrandt@uni-oldenburg.de. The analysis scripts can be found at Zenodo, https://zenodo.org/, https://doi.org/10.5281/zenodo.6817974.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements'.

Author contributions

AH, MB, and GA contributed to the conception and design of the study. MB organized the database. GA implemented

and conducted the expert survey. AR and UE participated as experts. GA, SS, MB, and AH contributed to the analysis of the results. MB and AH wrote the first draft of the manuscript. All authors discussed the results in the post-interview, contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 – Project ID 390895286.

Acknowledgments

We thank Hörzentrum Oldenburg gGmbH for the provision of the patient data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2022.960012/full#supplementary-material

References

1. Lamond D, Farnell S. The treatment of pressure sores: a comparison of novice and expert nurses' knowledge, information use and decision accuracy. *J Adv Nurs.* (1998) 27:280–6. doi: 10.1046/j.1365-2648.1998.00532.x

2. Shortliffe EH, Cimino JJ. *Biomedical Informatics: Computer Applications in Health care and Biomedicine*. London: Springer (2014). doi: 10.1007/978-1-4471-4474-8

3. Belle V, Papantonis I. Principles and practice of explainable machine learning. *Front Big Data*. (2021) 4:688969. doi: 10.3389/fdata.2021.688969

4. Bietenbeck A, Streichert T. Preparing laboratories for interconnected health care. *Diagnostics.* (2021) 11:1487. doi: 10.3390/diagnostics11081487

5. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. *Decis Support Syst.* (2013) 54:953–61. doi: 10.1016/j.dss.2012.09.018

6. Spreckelsen C, Spitzer K. *Wissensbasen und Expertensysteme in der Medizin: KI-Ansätze Zwischen Klinischer Entscheidungsunterstützung und Medizinischem Wissensmanagement*. Wiesbaden: Vieweg + Teubner GWV Fachverlage GmbH (2008). doi: 10.1007/978-3-8348-9294-2

7. Sandryhaila A, Moura JM. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Process Mag.* (2014) 31:80–90. doi: 10.1109/MSP.2014.2329213

8. Medlock S, Wyatt JC, Patel VL, Shortliffe EH, Abu-Hanna A. Modeling information flows in clinical decision support: key insights for enhancing system effectiveness. *J Am Med Inform Assoc.* (2016) 23:1001–6. doi: 10.1093/jamia/ocv177

9. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. In: *Proceedings of the AAAI conference on artificial intelligence* (Honolulu), Vol. 33, No. 01. (2019), p. 590–7. doi: 10.1609/aaai.v33i01.3301590

10. Liu M, Jiang L, Liu J, Wang X, Zhu J, Liu S. Improving learning-from-crowds through expert validation. In: *IJCAI* (Melbourne), (2017), p. 2329–36. doi: 10.24963/ijcai.2017/324

11. Walter Z, Lopez SM. Physician acceptance of information technologies: role of perceived threat to professional autonomy. *Decis Support Syst.* (2008) 46:206–15. doi: 10.1016/j.dss.2008.06.004

12. Bruun M, Frederiksen KS, Rhodius-Meester HF, Baroni M, Gjerum L, Koikkalainen J, et al. Impact of a clinical decision support tool on prediction of progression in early-stage dementia: a prospective validation study. *Alzheimers Res Ther.* (2019) 11:1–11. doi: 10.1186/s13195-019-0482-3

13. Baur T, Heimerl A, Lingenfelser F, Wagner J, Valstar MF, Schuller B, et al. eXplainable cooperative machine learning with NOVA. *KI-Künstliche Intelligenz.* (2020) 34:143–64. doi: 10.1007/s13218-020-00632-3

14. Tarnowska KA, Dispoto BC, Conragan J. Explainable AI-based clinical decision support system for hearing disorders. In: *Proceedings of the AMIA Annual Symposium, San Diego, CA, USA, 30 October—3 November 2021* (San Diego, CA), (2021), p. 595.

15. Liao W-H, Cheng Y-F, Chen Y-C, Lai Y-H, Lai F, Chu Y-C. Physician decision support system for idiopathic sudden sensorineural hearing loss patients. *J Chin Med Assoc.* (2021) 84:101–7. doi: 10.1097/JCMA.0000000000000450

16. Naveed Anwar M, Philip Oakes M. Decision support system for the selection of an ITE or a BTE hearing aid. *Int J Comput Appl.* (2013) 76:37–42. doi: 10.5120/13318-0936

17. Sanchez-Lopez R, Bianchi F, Fereczkowski M, Santurette S, Dau T. Data-driven approach for auditory profiling and characterization of individual hearing loss. *Trends Hear.* (2018) 22:233121651880740. doi: 10.1177/2331216518807400

18. Sanchez-Lopez R, Fereczkowski M, Neher T, Santurette S, Dau T. Robust data-driven auditory profiling towards precision audiology. *Trends Hear.* (2020) 24:233121652097353. doi: 10.1177/2331216520973539

19. Buhl M, Warzybok A, Schädler MR, Lenarz T, Majdani O, Kollmeier B. Common Audiological Functional Parameters (CAFPAs): statistical and compact representation of rehabilitative audiological classification based on expert knowledge. *Int J Audiol.* (2019) 58:231–45. doi: 10.1080/14992027.2018.1554912

20. Buhl M, Warzybok A, Schion of rehabilitative audiological classification based octional Parameters (CAFPAs) for single patient cases: deriving statistical models from an expert-labelled data set. *Int J Audiol.* (2020) 59:534diold data se1080/14992027.2020.1728401

21. Buhl M, Warzybok A, Sch s) for single patient cases: deriving statistical models from an enowledge. rameters (CAFPAs)acterization of individual hearing loss. eness. Audiological Functional Parameters (CAFPAs). *Int J Audiol.* (2021) 60:16udioll Functi1080/14992027.2020.1817581

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

23. Saak SK, 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.* (2020) 2:596433. doi: 10.3389/fdgth.2020.596433

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

25. Brand T, Hohmann V. An adaptive procedure for categorical loudness scaling. *J Acoust Soc Am.* (2002) 112:1597–604. doi: 10.1121/1.1502902

26. Schmidt KH, Metzler P. *WST-Wortschatztest.* Göttingen: Beltz Test (1992).

27. Kalbe E, Kessler J, Calabrese P, Smith R, Passmore AP, Brand MA, et al. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry.* (2004) 19:136–43. doi: 10.1002/gps.1042

28. Gieseler A, Tahden MA, Thiel CM, Wagener KC, Meis M, Colonius H. Auditory and non-auditory contributions for unaided speech recognition in noise as a function of hearing aid use. *Front Psychol.* (2017) 8:219. doi: 10.3389/fpsyg.2017.00219

29. Peirce JW, Gray JR, Simpson S, MacAskill MR, Höchenberger R, Sogo H, et al. PsychoPy2: experiments in behavior made easy. *Behav Res Methods.* (2019) 51:195–203. doi: 10.3758/s13428-018-01193-y

30. RStudio Team. *Rstudio: Integrated Development Environment for R* [Computer software manual]. Boston, MA (2020). Available online at: http://www.rstudio.com/ (accessed April 4, 2022).

31. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, IL: Northwestern University (2022). R package version 2.2.5. Available online at: https://CRAN.R-project.org/package=psych (accessed April 4, 2022).

32. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012

33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* (2015) 67:1–48. doi: 10.18637/jss.v067.i01

34. van Esch TE, Kollmeier B, Vormann M, Lyzenga J, Houtgast T, Hällgren M, et al. Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *Int J Audiol.* (2013) 52:305–21. doi: 10.3109/14992027.2012.759665

35. Beutelmann R, Brand T, Kollmeier B. Revision, extension, and evaluation of a binaural speech intelligibility model. *J Acoust Soc Am.* (2010) 127:2479–97. doi: 10.1121/1.3295575

36. Bronkhorst AW. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acta Acust United Acust.* (2000) 86:117–28.

37. Ching TY, Van Wanrooy E, Dillon H, Carter L. Spatial release from masking in normal-hearing children and children who use hearing aids. *J Acoust Soc Am.* (2011) 129:368–75. doi: 10.1121/1.3523295

38. Noble W, Byrne D, Ter-Horst K. Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners. *J Acoust Soc Am.* (1997) 102:2343–52. doi: 10.1121/1.419618

39. Lenarz T, Boenninghaus HG. *Hals-Nasen-Ohren-Heilkunde.* Berlin, Heidelberg: Springer-Verlag (2012). doi: 10.1007/978-3-642-21131-7

40. Ali SI, Jung SW, Bilal HSM, Lee S-H, Hussain J, Afzal M, et al. Clinical decision support system based on hybrid knowledge modeling: a case study of chronic kidney disease-mineral and bone disorder treatment. *Int J Environ Res Public Health.* (2022) 19:226. doi: 10.3390/ijerph19010226

frontiers | Frontiers in Neurology

TYPE Study Protocol
PUBLISHED 26 August 2022
DOI 10.3389/fneur.2022.933940

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Pedro A. Moreno-Sanchez,
Tampere University, Finland
Mareike Buhl,
Carl von Ossietzky Universität Oldenburg, Germany

*CORRESPONDENCE
Eleftheria Iliadou
iliadoue@med.uoa.gr
Qiqi Su
qiqi.su@city.ac.uk

†These authors have contributed equally to this work and share first authorship
‡These authors have contributed equally to this work and share senior authorship

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 01 May 2022
ACCEPTED 08 August 2022
PUBLISHED 26 August 2022

CITATION
Iliadou E, Su Q, Kikidis D, Bibas T and Kloukinas C (2022) Profiling hearing aid users through big data explainable artificial intelligence techniques. *Front. Neurol.* 13:933940. doi: 10.3389/fneur.2022.933940

Profiling hearing aid users through big data explainable artificial intelligence techniques

Eleftheria Iliadou[1*†], Qiqi Su[2*†], Dimitrios Kikidis[1], Thanos Bibas[1‡] and Christos Kloukinas[2‡]

[1]1st Department of Otorhinolaryngology-Head and Neck Surgery, National and Kapodistrian University of Athens Medical School, Athens, Greece, [2]Department of Computer Science, University of London, London, United Kingdom

Debilitating hearing loss (HL) affects ∼6% of the human population. Only 20% of the people in need of a hearing assistive device will eventually seek and acquire one. The number of people that are satisfied with their Hearing Aids (HAids) and continue using them in the long term is even lower. Understanding the personal, behavioral, environmental, or other factors that correlate with the optimal HAid fitting and with users' experience of HAids is a significant step in improving patient satisfaction and quality of life, while reducing societal and financial burden. In SMART BEAR we are addressing this need by making use of the capacity of modern HAids to provide dynamic logging of their operation and by combining this information with a big amount of information about the medical, environmental, and social context of each HAid user. We are studying hearing rehabilitation through a 12-month continuous monitoring of HL patients, collecting data, such as participants' demographics, audiometric and medical data, their cognitive and mental status, their habits, and preferences, through a set of medical devices and wearables, as well as through face-to-face and remote clinical assessments and fitting/fine-tuning sessions. Descriptive, AI-based analysis and assessment of the relationships between heterogeneous data and HL-related parameters will help clinical researchers to better understand the overall health profiles of HL patients, and to identify patterns or relations that may be proven essential for future clinical trials. In addition, the future state and behavioral (e.g., HAids Satisfiability and HAids usage) of the patients will be predicted with time-dependent machine learning models to assist the clinical researchers to decide on the nature of the interventions. Explainable Artificial Intelligence (XAI) techniques will be leveraged to better understand the factors that play a significant role in the success of a hearing rehabilitation program, constructing patient profiles. This paper is a conceptual one aiming to describe the upcoming data collection process and proposed framework for providing a comprehensive profile for patients with HL in the context of EU-funded SMART BEAR project. Such patient profiles can be invaluable in HL treatment as they can help to identify the characteristics making patients more prone to drop out and stop using their HAids, using their HAids sufficiently long during the day, and being more satisfied by their HAids experience. They can also help decrease the number of needed remote sessions with their Audiologist for counseling, and/or HAids

10.3389/fneur.2022.933940

fine tuning, or the number of manual changes of HAids program (as indication of poor sound quality and bad adaptation of HAids configuration to patients' real needs and daily challenges), leading to reduced healthcare cost.

KEYWORDS

explainable AI (XAI), Deep Learning, big data, hearing loss, Hearing Aids, prognosis prediction, Long Short-Term Memory (LSTM), attention mechanism

Introduction

Hearing Loss (HL) is a public health problem that affects one out of three people over the age of 65, while debilitating HL is estimated to affect 6% of the population (466 million people) according to World Health Organization (WHO) statistics[1]. As per the same statistics, its annual management cost is estimated at more than 555 billion Euros (1) for the European countries and at 750 billion Dollars globally. HL should not be considered as an isolated health problem. Apart from the associated financial cost, HL severely affects communication and is associated with various comorbidities. Multiple studies have suggested that hearing impairment is associated with psychological and physical illness, such as cognitive disorders and dementia. An increase in the hearing threshold of 25 decibels (dB) corresponds to a loss of 7 cognitive years (2), and is associated with increased anxiety and depression (3), and even higher mortality rate (4). On the other hand, adults with hearing impairment tend to isolate themselves by limiting their participation in social events (5), thereby reducing their quality of life significantly (6).

Although the only available and validated management solution that currently exists for HL is the fitting and use of hearing assistive devices, only one in five people in need of a Hearing Aid (HAid) will eventually seek, acquire, and continue to use one efficiently (7, 8). A "HAid experience" refers to the process of living with a HAid and involves all the real-life challenges, coping strategies, and facilitations that the uses of HAid may evoke. Improvements in the HAid experience can lead to minimization of drop-out risk and enhancement of the overall quality of life (9).

The key factors in improving the HAid experience include, but are not limited to, proper fitting, affordability and accessibility of the follow-up services, and their combination with thorough and evidence-based personalized counseling and training on how to use the selected HAid (10). Since everyday patient needs and HL degree are not static and might change over time, there are still many factors that audiologists find challenging to address, including selecting optimal HAid configurations or best counseling approach according to individual patient profile and lifestyle (7, 11–13). Dynamic monitoring and collecting information about a patient's hearing and cognitive capacity, as well as their ability to control settings in real time in order to cope in different sound environments, could be very helpful toward this direction (14, 15). The development and validation of prediction models using the collected information and making accurate prognoses of how each patient's HAid experience will unfold are of major priority.

The use of Artificial Intelligence (AI) models in prognosis studies has gained traction increasingly in recent years due to its ability to handle large amounts of messy data (16), to learn from different types of data (17), and to facilitate clinical management of patients (18). Researchers have incorporated AI models in prognosis in clinical cancer research, such as breast cancer with Support Vector Machine (SVM) (19), colorectal cancer with Long Short-Term Memory (LSTM) (20), and glioblastoma with Prognosis Enhanced Neural Network (PENN) (21). As well as the prognosis for adult congenital heart disease with Convolutional Neural Network (CNN)-LSTM (22), rate of kidney disease with an ensemble of Logistic Regression, Decision Tree, Random Forest (RF), and K-Nearest Neighbor (KNN) (23), and COVID-19 with a segmentation network (24).

The effectiveness of AI models in HL prognosis has also been investigated by many researchers. Sensorineural Hearing Loss (SNHL) is the most common form of permanent HL resulting from the damage to the auditory nerve and/or the hair cells in the inner ear. Abdollahi et al. (25) constructed eight Machine Learning (ML) models to predict SNHL after chemoradiotherapy, including Decision Stump, Hoeffding, C4.5, Bayesian Network, Naïve, Adaptive Boosting (AdaBoost), Bootstrap Aggregating, Classification *via* Regression, and Logistic Regression (LR). The average predictive power of all models was found to be more than 70% in terms of accuracy, precision, and Area Under Curve (AUC). Idiopathic Sensorineural Hearing Loss (ISSHL) is characterized by an acute dysfunction of the inner ear. Zhao et al. (26) developed several ML models for ISSHL prediction, including SVM, Multilayer Perceptron (MLP), RF, and AdaBoost. A similarly high level of accuracy is also reported and varies between 78.6 and 80.1%. Bing et al. (27) evaluated several Deep Learning (DL)

1 https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

and ML models to predict the dichotomised hearing outcome of ISSHL in order to identify the best predictive model for clinical application. Six input feature collections derived from 149 potential predictors have been used with Deep Belief Network, LR, SVM, and MLP. Best predictive performance was achieved by Deep Belief Network when evaluated with accuracy, precision, recall, F-score, Receiver Operating Characteristic Curve (ROC), and AUC, achieving 77.58% of accuracy and 0.84 of AUC. Ototoxic-induced HL, more specifically, the ototoxic effects in participants who were exposed to cigarette smoke and/or pesticides were evaluated by Artificial Neural Network, KNN, and SVM (28). While all models showed a good performance during training, KNN achieved the highest training accuracy with about 90% in two of the five datasets.

Attention-based DL models have also gained popularity in the medical domain recently. Bahdanau et al. (29) proposed the first attention mechanism, also known as the Soft Attention, for a Neural Machine Translation task using LSTM. The advantage of using attention mechanisms with LSTM is that it prevents the LSTM from forgetting certain input features when analyzing long-term dependencies and from putting too much weight on certain input features. Despite the lack of research using attention-based LSTM for HL patients specifically, a similar approach has been adapted for other comorbidities. Park et al. (30) used a Frequency-aware Attention-based LSTM (FA-Attn-LSTM) to investigate medical features that can be considered as critical for predicting the risk of cardiovascular disease. Wall et al. (31) proposed a framework for audio classification, specifically for chronic and non-chronic lung disease and COVID-19 diagnosis, with attention-based bidirectional LSTM (A-BiLSTM).

AI, particularly DL models, in general are appreciated for their ability to achieve high prediction accuracy. However, for sensitive domains, such as health care, accuracy is not the only determining factor (32). The inherent limitation of many AI systems is their black box nature, which means that humans are unable to easily understand the inner workings of these systems or how they arrive at their conclusions. Thus, automated decision-making systems that employ AI models are not widely accepted (32) due to a lack of trust from the end users. The integration of AI models into medical domains also faces criticisms where the models may fail to adhere to high standards of accountability, reliability, and transparency for medical decisions (33). It also complicates the issue of accountability in the event of a wrong decision (34).

Explainable AI (XAI) aims to overcome these limitations by explaining the learned decisions of AI models, thus giving end-users the ability to trust the models (35) and understanding why the models made certain decisions (32). Different XAI methods have been proposed over the years, particularly in the fields of computer vision and natural language processing. Yet very few studies have explored the potential applications of XAI methods to the medical field (34), especially in prognosis studies. A number of researchers have adapted Local Interpretable Model-agnostic Explanation (LIME) (36) to explain a CNN-based diagnostic model, including chronic wound classification (37), gastral image classification (38), and Alzheimer's diagnosis (39). Gu et al. (40) proposed an auxiliary decision system for breast cancer diagnosis and prediction with Extreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) (41). Chakraborty et al. (17) developed a similar framework that was inspired by Gu et al. (40) using XGBoost and SHAP for prognosis in breast cancer patients. In the HL domain, Lenatti et al. (42) applied SHAP to explain the classification results of RF in predicting whether or not a patient has HL. In particular, SHAP is used to investigate the local predictions for each of the two output classes in four scenarios: true positive, true negative, false positive, and false negative. They have found that Age is the most important feature that impacts the classifier. In particular, values of age equal to 74 contribute positively to the model correctly predicting participants with HL (true positive), whereas values of age equal to 25 contribute negatively to the model correctly predicting participants without HL (true negative).

To the best of our knowledge, this is the first conceptual paper on a framework that leverages AI and XAI for prognosis for HL benefit and usage. ML techniques have been implemented previously in studies focusing on the prognosis of SNHL, ISSHL, and HL induced by ototoxic drugs and other substances (25–28), and modeling has also been attempted with synthetic data in more progressive types of HL, such as age-related or noise-induced HL (43). Nevertheless, we are unaware of any such attempts with real multi-source big data to date.

In the EU-funded SMART BEAR project[2], we are developing and validating a prognosis framework to address this scientific gap for HL patients. AI and XAI techniques will help identify and explain particular trends and factors in the large amount of heterogeneous data collected that correlate with the success or failure of hearing rehabilitation. In particular, the proposed framework composes the predictive power of LSTM with Attention Mechanism with the explanatory abilities of SHAP, and it will be used to answer several questions to provide a comprehensive profiling of HL patients.

The purpose of this article is to describe the planned data collection process, as well as the upcoming analyses to identify and explain particular trends and factors that correlate with the success or failure of hearing rehabilitation: drop-out of HAids usage, more hours of HAids usage and higher benefit from it, and less frequent need for manual adjustments or fine tuning of the HAids. As this is a conceptual paper, data collection is expected to begin in autumn 2022, followed by the experiments of the proposed methods.

2 https://www.smart-bear.eu/

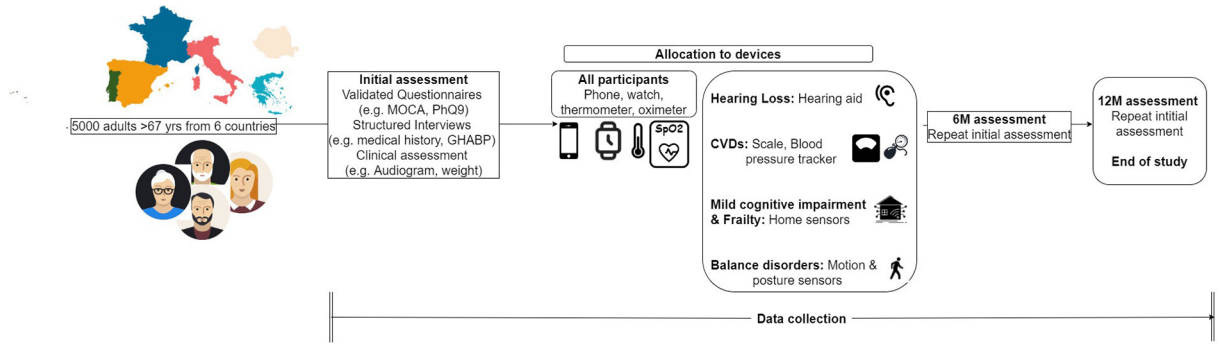


FIGURE 1
Participants' flow of action.

Materials and methods

Participants

Five thousand elderly participants from six different EU countries will be included in the study. In particular, these six countries are divided into five study groups and 1,000 participants are recruited from each, namely France, Greece, Italy, Romania, and Portugal-Spain. A smaller-scale pilot study with 100 participants is already underway in the island of Madeira. The large-scale project is scheduled to begin in autumn 2022 and run for 24 months. Subjects will be included in the study based on the following eligibility criteria:

1. Age and birth gender: males and females, 67–80 years old.
2. Medical history: at least 2 of the following conditions: cardiovascular diseases (CVDs: hypertension, coronary disease, heart failure), hearing loss, balance disorders, mild depression, mild cognitive impairment, frailty.
3. Cognitive function according to MoCA score: participants with 26–30/30 (no cognitive impairment), and 18–26/30 (mild cognitive impairment) will be included (44). Score lower than 18/30 corresponds to mild dementia which is not addressed in SMART BEAR so those participants scoring < 18/30 will be excluded.
4. Excellent to Moderate level of mobility, which corresponds to be able to perform simple tasks such as walking and jumping independently, with or without the help of a mechanical equipment, for example, a cane.
5. Ability to read.
6. Ability to use the basic functions of a smartphone (answer, call, check a notification, open an application).

Participants who meet the aforementioned criteria but present a severe or life-threatening condition, such as severe depression or high risk of heart failure, will be excluded from the study. All participants willing to provide their informed consent and voluntarily participate in the study will undergo an initial clinical assessment as shown in Figure 1. According to the results of this screening assessment, a specific set of devices and clinical procedures will be allocated to each participant. These devices are being obtained through joint procurement for all six countries and will be the same in terms of type, model, and configuration for all participants.

Participants with hearing loss

We intend to recruit one thousand people with HL to a degree that requires amplification. Participants with a moderate to severe unilateral or bilateral HL, as indicated by their pure tone audiogram, are considered eligible for HAid fitting if their HL negatively impacts their communication ability, cannot be treated surgically, or can be treated but the surgery is contra-indicated for the particular participant. Participants will only be excluded from Fitting if they do not wish to be fitted with a HAid, or if they have profound HL (Pure tone average 0.5–4 kHz > 80 dB), and have not received any benefit from recent previous HAid fitting and use.

Audiological assessment

The same audiometric assessment (Figure 2) will be conducted on all participants with suspected or diagnosed HL by experienced personnel who have undergone additional internal training on every procedure of the clinical protocol by the clinical coordination team of the SMART BEAR. Joint procurement will ensure that the equipment (including HAids) and relevant software will be the same for all countries. Following the audiometric assessment, all participants will be fitted with HAids according to the same fitting protocol. The exact fitting protocol will be defined once the specific model and manufacturer of the HAids is selected during the international procurement procedure as discussed above.

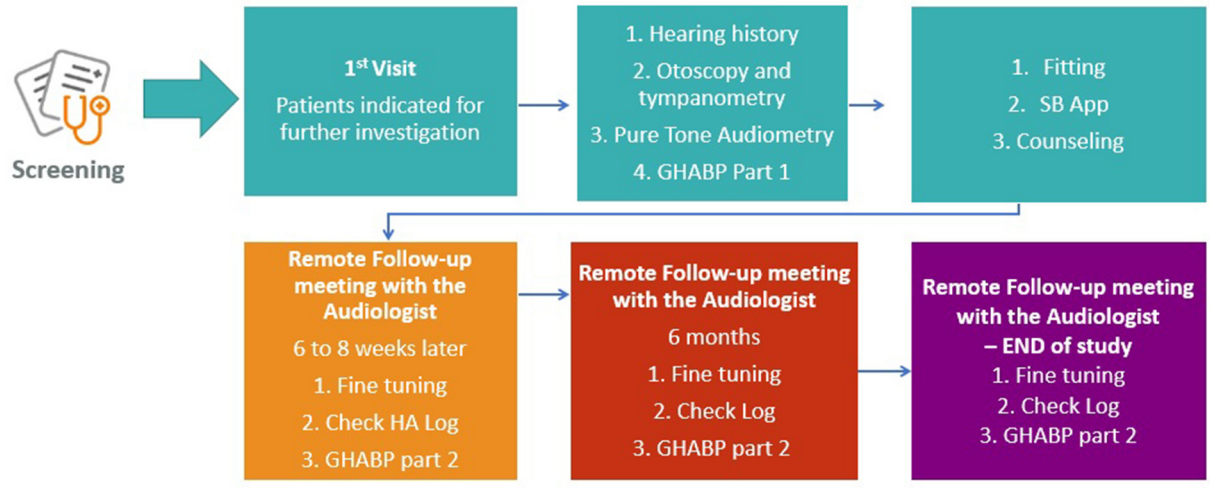


FIGURE 2
Audiological assessment flow of action.

HAids configuration will then be fine-tuned in accordance with the participant's experience level, listening preferences, and language preferences. There will be a predefined HAids program for all participants, other programs may be added based on the judgment of the audiologists and the needs of the participants. Pure tone audiometry will follow the British Society of Audiology[3] guidelines.

In accordance with the SMART BEAR fitting protocol, participants will be monitored for 12 months after they have been fitted with either one or two HAids (same manufacturer, same model). As shown in Figure 2, participants will also have continuous access to remote and face-to-face fine-tuning services provided by the SMART BEAR audiologists. Through the SMART BEAR clinician's dashboard, the audiologists will have access to participants' data and HAids log throughout this period.

Source of data

SMART BEAR is a large-scale multi-centric clinical study that aims to integrate state-of-the-art technology into everyday life of senior citizens with specific comorbidities, composing off-the-shelf and user-friendly devices onto an innovative platform. There are three subsystems in the SMART BEAR architecture as shown in Figure 3, namely the mobile phone application, the SMART BEAR HomeHub, and the SMART BEAR Cloud (SB@Cloud). Data are collected (i) during participants' clinical assessments *via* the clinician dashboard (e.g., anamnestic history, physiological and audiometric measurements), (ii) from all linked portable devices *via* the mobile phone application (e.g., HAid program, heart rate, and steps measurement), and (iii) through the mobile phone application itself (e.g., through questionnaires about their mood, diet, medication adherence and sleep quality). The HomeHub accumulates data from different home-based device sensors, such as weight scales and movement sensors. Finally, SB@Cloud securely stores and analyses the collected data through model and data-driven big data analytics during a 12-month period for each participant.

A total of 24 variable and covariates are collected through SMART BEAR HAids, including timestamp of the measurement, environmental noise, and manual program adjustments. Supplementary Table 1 provides a detailed description of each variable and covariate. Several other covariates are also being considered and are shown in Supplementary Table 2. The additional 241 covariates are collected in order to monitor the participants' other comorbidities based on their demographics, biological, environmental, and behavioral characteristics. There is a need to consider the impact of these additional covariates on the outcomes since they have been previously shown to affect to HL and HAid experiences, such as age, occupation, education, family history, mood disorders, cognitive function, diet, glucose levels and medication (3, 45–47). They are also currently being investigated for their correlation to hearing, as in the case of cardiovascular diseases, poorer mobility, frailty, and balance disorders (46, 48, 49). Furthermore, the medical and audiological assessment will also be supplemented by

3 https://www.thebsa.org.uk/wp-content/uploads/2018/11/OD104-32-Recommended-Procedure-Pure-Tone-Audiometry-August-2018-FINAL.pdf

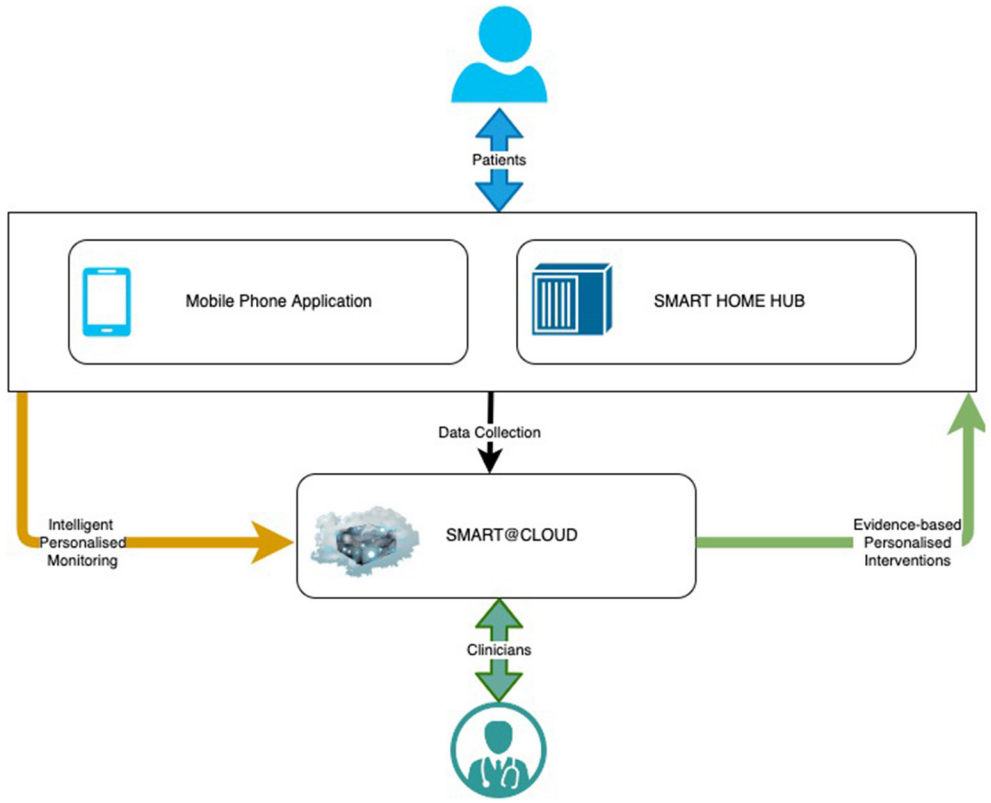


FIGURE 3
The SMART BEAR architecture.

additional sensor data as listed in Supplementary Table 3, such as blood pressure measured by the blood pressure tracker and physical activity measured by the smart watch. These variables are collected as a part of SMART BEAR's commitment to collect a wide range of data which will be explored as a part of data-driven analysis.

Sample size

SMART BEAR is aiming at collecting and analyzing big data—integrating information from many thousands of participants and different data sources. In Big Data, common sample size calculations cannot apply (50). Big data studies need to consider the marginal costs vs. the marginal value of possible sample sizes and include as many participants as possible (51). In SMART BEAR, the maximum number of participants that can be recruited based on available resources and time is 5,000. In accordance with the requirements of the study, this number is considered sufficient for ensuring the impact analysis obtained at the end of the project to be significant. In the case of HL, 200 participants with HL will be recruited from each of the five study groups, creating a sample of 1,000 participants with HL. These participants will then be fitted with either one or two HAids depending on whether one or both ears require amplification. Therefore, the total number of HAids to be used in the planned data collection is estimated between 1,000 and 2,000. The SMART BEAR platform is designed to facilitate the collection of data from a maximum number of 2,000 HAids, in case all participants suffer from bilateral HL. Data collected from up to 2,000 HAids are also considered to be sufficient based on previous experience (50).

Analysis methods

The questions that will be addressed with the proposed framework are based on future events. The prediction model will be used, for example, to predict future HAid usage or future drop-out rate. As a result, the model is fundamentally constructed with participants' historical medical history, HAid usage and habit, as well as the outcomes of medical and audiological assessments. As such, the collected SMART BEAR

data are sequential in nature and can be viewed as time series data.

The proposed framework uses an attention-based LSTM (attn-LSTM) as the prediction model and then applies SHAP to interpret the model predictions. More specifically, SHAP is employed to identify those characteristics that influence the model predictions. To enable continuous learning and provision of personalized solutions, the pipeline for the proposed framework is to pre-process the data, hyper-tune the model, train/test the model with the optimal set of hyper-parameters selected from hyper-tuning, and then apply the XAI method. The performance of the prediction models is evaluated using different set of evaluation metrics for classification and regression problems.

Pre-processing the data

The temporal element of the collected data is determined by the Time variable, which records the date and time of the collected variables every 60 s when the SMART BEAR HAids are active in use. In SMART BEAR, clinicians also have the option of choosing how the data are aggregated for different analysis. Due to this, the data frequency is transformed first in order to allow hourly, daily, weekly, monthly, or yearly predictions, depending on the choice of clinician.

Transforming the distribution of the features allows the ML and DL algorithms to converge faster and minimize the weight of any variable with extreme values. *Standardization* and *normalization* are two pre-processing techniques that are particularly important for training an LSTM algorithm, since standardization on the data centers the noise from trend reverse signals and prevents activation functions to saturate (52), whereas normalization prevents the weights of the model being skewed (53).

Ordinal variables will be transformed with ordinal encoding and nominal variables will be transformed with one-hot encoding in order to convert these variables into either binary or multiple values with a numerical form. If the expected outcome variable is categorical then these will be treated label encoding.

Another important pre-processing step is to handle missing data. Several studies regarding data completeness in medical data were reviewed by Chan et al. (54) and found that the percentage of missing values of a variable, such as clinical status, laboratory results, and clinical actions or procedures, can reach as high as 98%. There is a possibility that this phenomenon might also be observed with data collected through SMART BEAR HAids due to connectivity issue and lack of participant adherence. As a result, simply deleting rows with missing values is not feasible for treating missing data, and imputation and model-based approaches should be used instead. There are several types of both imputation and model-based methods. For imputation methods, there are mean, median, zero, linear interpolation, forward, and backward, whereas for model-based methods, there are linear regression, KNN, and Multiple-value Imputation. A generic method was suggested by Salgado et al. (55) for the purpose of evaluating the performance of various methods for handling missing data. To start with, use a sample of the dataset that contains no missing data as ground truth, and then introduce the proportions of missing data at random in increments of say 5%. In the next step, compute the sum of squared errors (SSE) between the ground truth and the reconstructed data, for each method and for each proportion of missing data. Repeat these steps for each method and calculate the average SSE. Lastly, select the method that performed best at the level of missing data in the given dataset.

In addition, there is the question of how to deal with outliers—"samples that are exceptionally far from the mainstream data" (56). Even with a thorough understanding of the data, outliers can still be difficult to detect (56); however, statistical methods can assist in the identification of them. As standard deviation method is more suited for data with a normal distribution, therefore, it is used after the data have been standardized and normalized. Given the mean and standard deviation of the dataset, z-score can be computed for every ξ_i, which is the number of standard deviations away from the mean, as a way to identify outliers (57). Data points can be declared as outliers if their z-score standard deviation is greater than a predefined threshold. The threshold used in this analysis is three, as it is common practice to identify outliers in data with Gaussian or Gaussian-like distributions.

Lastly, it is important to determine whether there is multicollinearity among the variables. Multicollinearity refers to when there is a lack of orthogonality among two or more variables, and it often creates problems in a regression model (58) because the model results tend to fluctuate significantly when changes are made to independent variables that are highly correlated. In terms of hearing data, multicollinearity is often met among several variables. A typical example is the pure tone thresholds across different frequencies. Pure tone thresholds are measured in frequency bands with each representing a cochlear region, and the neighboring frequencies tend to be highly correlated (59). Moreover, pure tone audiogram also shows a high correlation among the sensitivity of the two ears for each participant when symmetric hearing is present (59). A common method of checking whether the data are multicollinear is to use the Variance Inflation Method (VIF) for each independent variable. In general, a VIF value of 10 indicates weak multicollinearity, and a variable with a higher value is typically considered to have a high correlation with another independent variable (58). A simple way to eliminate high multicollinearity variables is to remove them. However, this may not be feasible in practice. As a result, alternative methods, such as transforming the variables or performing Principal Component Analysis, should be considered instead, depending on the data and the expected outcome. Finally, data will be split into training, validation, and testing sets.

In this conceptual paper, the pre-processing steps discussed here are generic. While these techniques should be considered regardless of the questions to be answered, specific pre-processing methods, such as handling missing data and multicollinearity variables, will only become apparent following the data collection.

Hyper-tuning the model

The model is validated on the validation set during hyper-tuning in order to determine the set of optimal hyper-parameters. The hyper-tuning is performed using the Keras Tuner[4] library to determine the set of optimal hyper-parameters for model trained with TensorFlow[5]. There are many hyper-parameters that need to be determined when training an LSTM model. For this analysis, the number of hidden states in each layer, choice of activation function, learning rate, dropout rate, and batch size are hyper-tuned.

It is imperative to adjust the number of hidden units according to the complexity of the data and select an activation function that is capable of learning the complex relationship in the data. Learning rate is also important because if it is too fast, the model converges too quickly, while if it is too slow, it reaches some local minima. Dropout is a regularization technique while training a DL model, aiming at improving generalization and reducing overfitting. Last but not least, the batch size is the number of samples of training data that will be propagated through the model and should be adjusted accordingly as it impacts the stability of the learning process. Furthermore, the model will also be trained with early stopping in order to prevent overfitting. Early stopping is implemented through a callback function, which monitors the progress of the training, and if no improvements are made during the course of training, the training is terminated early.

Proposed model architecture

The proposed prediction model, attn-LSTM, will be trained on the training set with the set of optimal hyper-parameters from hyper-tuning, and the results are reported by predicting the unseen testing set. Table 1 shows the proposed model architecture of attn-LSTM and hyper-parameters setting for each layer. It should note that the choice of learning rate and batch size is hyper-tuned for the entire model and not for each individual layer.

LSTM (60) is a refined variant of the Recurrent Neural Network that is designed with a feedback architecture such that the current time step prediction is influenced by the network activation from the previous time steps as inputs. LSTM is one of the widely used DL technique for analyzing time series data and is capable of learning long-term time series data as well as short-term time series data (61). The hidden layer inside an LSTM network contains recurrently connected special units called memory cells and their corresponding gate units: input gate, forget gate, and output gate (60) as shown in Figure 4. The input gate is responsible for preventing the memory stored in a memory cell from perturbations by irrelevant inputs. Similarly, the output gate is there so other units are protected from perturbations by currently irrelevant stored memory. To optimize the performance of the LSTM, information that is no longer required by the LSTM is removed in the mechanism of the forget gate.

TABLE 1 Proposed model architecture.

Layer no.	Layer description	Hyper-parameters setting
1	Input layer	N/A
2	LSTM layer	Hidden units are hyper-tuned between 32 and 512. Activation function is hyper-tuned between Sigmoid and Tanh.
3	Self-attention layer	N/A
4	Dropout layer	Dropout rate is hyper-tuned between 0.001 and 0.1.
5	Flatten layer	N/A
6	Output (dense) layer	Regression problems: hidden unit is 1, and activation function is hyper-tuned between ReLu, Sigmoid, and None. Binary classification problem: hidden unit is 2, and activation function is Softmax and Sigmoid.

At each timestep t, the cell takes an input vector, x_t, and produces an output vector, h_t, which also refers to the hidden state of the LSTM. Firstly, the cell needs to determine whether the information from the previous timestep, $t-1$, should be kept or not with the forget gate, f_t. The forget gate takes the input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t*-1}, and produces an output between 0 and 1 where 0 represents "completely forget this information" and 1 represents "completely keep this information". The forget gate, f_t, is calculated as follows:

$$f_t = \sigma(w_x\, x_t + w_h\, h_{t-1} + b)\,,$$

where σ is the sigmoid function, w_x, w_h are the weighting factor, and b is the bias vector. More specifically, the sigmoid function is calculated as:

$$\sigma\,(\mathrm{x}) = \frac{1}{1+\mathrm{e}^{-\mathrm{x}}}.$$

4 https://keras.io/keras_tuner/

5 https://www.tensorflow.org/

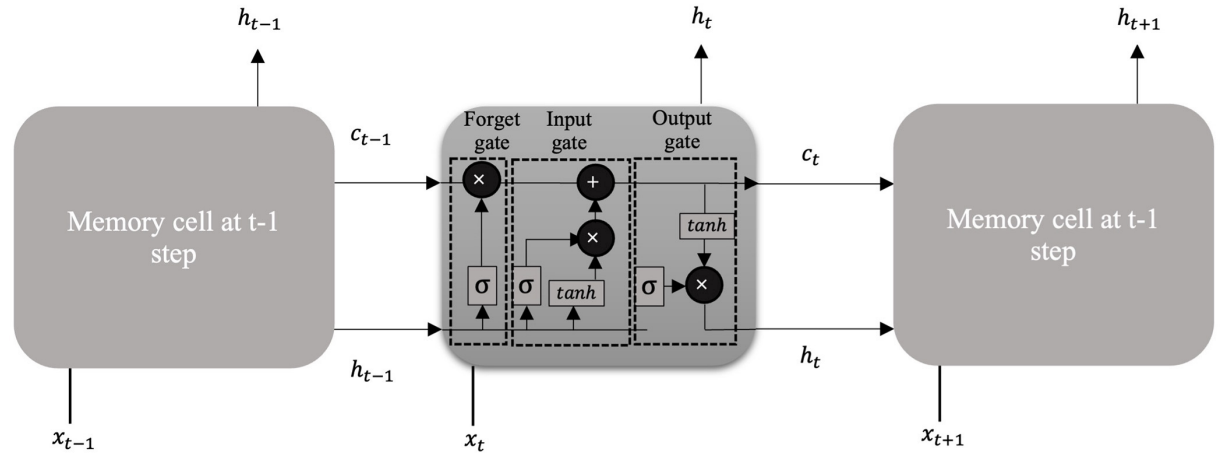


FIGURE 4
An illustration of the LSTM network.

The next step is to quantify the importance of the new information with the input gate, i_t:

$$i_t = \sigma(w_x\, x_t + w_h\, h_{t-1} + b),$$

which is also a function of input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t-1}. Then, a new vector named s_t is created which decides if the new information should be stored in the cell state or not. This is done by applying a hyperbolic tangent function, tanh, to the input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t-1}. It is calculated as:

$$s_t = \tanh(w_x\, x_t + w_h\, h_{t-1} + b)\,,$$

and the value of new information is transformed to a value between−1 and 1, where−1 means the new information is subtracted from the cell state and 1 means the new information is added to the cell state. The current cell state, c_t, is finally updated by taking the previous cell state, c_{t-1}, the forget gate, f_t, the input gate, i_t, and s_t into consideration by:

$$c_t = f_t \odot c_{t-1} + i_t \odot s_t\,,$$

where $\odot$ is the element-wise product. Then, the output gate, o_t, determines what information from the cell state is going to be the output. The output gate is also a function of input vector at current timestep, x_t, and the hidden state from the previous timestep, h_{t-1}, and outputs a value between 0 and 1. It is calculated as follows:

$$o_t = \sigma(w_x\, x_t + w_h\, h_{t-1} + b)\,.$$

Finally, the hidden state, h_t, at timestep t is updated with the current cell state, c_t, and the output gate, o_t, by:

$$h_t = \tanh(c_t) \odot o_t\,.$$

The use of attention-based LSTM was initially designed for natural language processing tasks and has been extended to other areas such as computer vision and time series prediction. The attention mechanism is also inspired by the human biological system, such that humans do not process large amounts of data all at once, but instead selectively focus on certain distinct parts of information (62). Moreover, integrating an attention mechanism into an LSTM model architecture may also enhance the interpretability of the model (63), since the attention mechanism can be used to demonstrate which features are important for predicting a particular outcome. The specific attention mechanism adopted in this framework is the Self-attention similar to the one proposed by Vaswani et al. (64), where the mechanism is relating different positions of a single sequence in order to gain a representation of the sequence.

Vaswani et al. (64) introduced a generalized definition for attention functions in which the inputs of the function consist of three vectors: queries (q), keys (k), and values (v). In practice, the attention function is computed on a set of queries simultaneously and packed into the matrix Q, and similarly the keys and values are packed into the matrix K and V, respectively. The concepts of Q, K, and V were first introduced in the context of NLP, specifically with Encoder-Decoder models. Taking the task of machine translation as an example, the query is derived from the Decoder layers reading the current translated text, whereas the key and value are derived from the Encoder layers reading the original sentence.

However, Self-attention is a special case of the attention mechanism where all of the queries, keys, and values come from the same place, such that $Q = K = V$ (64). The mechanism queries only the inputs to obtain the self-attention, and from the self-attention a new representation of the inputs

can be constructed. In this framework, the inputs of the attention function are the sequence of hidden state vectors for all timesteps produced by LSTM, $H = (h_1, h_2, \ldots, h_n)$, therefore, $H = Q = K = V$.

The next step is to calculate a compatibility score for each hidden state vector in the LSTM. More specifically, it involves scoring the compatibility of each hidden state vector in H against the hidden state vector for which the self-attention is calculated. The specific compatibility score used in this framework is similar to the proposed by Vaswani et al. (64) and calculated as follows[6]:

$$Compatibility\ score = \frac{HH^\top}{\sqrt{d_H}},$$

where d_H is the dimension of the sequence of hidden state vectors and it is a dot-product-based compatibility score. For example, the compatibility score of the first hidden state vector, h_1, is calculated by scoring each hidden state vector, $h_2, \ldots, h_n$, against h_1, with $h_1 \cdot h_1^\top / \sqrt{d_H}$, $h_1 \cdot h_2^\top / \sqrt{d_H}$, …, $h_1 \cdot h_n^\top / \sqrt{d_H}$. The other commonly used compatibility score is the additive-based one, where the compatibility score is computed using a single hidden layer feed-forward network. Dot-product-based compatibility scores can be space-efficient and much faster in practice when compared to additive-based compatibility scores (64).

Each compatibility score for each hidden state vector is then sent through to the Softmax function in order to normalize the scores so that all scores are positive and sum to 1. Finally, the output of the self-attention function is calculated as a weighted sum of the hidden state vectors and the compatibility score. The matrix of the output is calculated as follows[7]:

$$Attention\,(H) = softmax\left(\frac{HH^\top}{\sqrt{d_H}}\right)H.$$

Evaluating the model performance

The results of the trained attn-LSTM are reported by predicting the unseen testing set and evaluated using different sets of metrics for classification and regression problems. For classification problems, the evaluation metrics are accuracy, precision, recall, F1 score, and AUC. Accuracy, precision, and recall can be derived from a confusion matrix, and F1 score is the harmonic mean of precision and recall. Each of the metric is calculated as follows:

$$\begin{aligned} Accuracy &= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \\ Precision &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \\ Recall &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \\ \mathrm{F1\ score} &= 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \end{aligned}$$

Finally, AUC measures the area under the ROC curve, which is a graphical representation of how well the model performed and shows the relationship between True Positive Rate and False Positive Rate.

For regression problems, four standard error estimators are used, namely Symmetric Mean Absolute Percentage Error (sMAPE), Mean Absolute Scaled Error (MASE), Mean Absolute Percentage Error (MAPE), and Weighted Average Percentage Error (WAPE). The error estimators are calculated as follows:

$$\begin{aligned} sMAPE &= \frac{200}{N}\sum_{t=1}^{N}\frac{|y_i - \tilde{y_i}|}{|y_i| + |\tilde{y_i}|}, \\ MASE &= \frac{1}{N}\sum_{t=1}^{N}\frac{|y_i - \tilde{y_i}|}{\frac{1}{t+N-1}\sum_{j=2}^{t+N}|y_j - y_{j-1}|}, \\ MAPE &= \frac{1}{N}\sum_{t=1}^{N}\frac{|y_i - \tilde{y_i}|}{y_i}, \\ WAPE &= \frac{\sum_{i=1}^{N}|y_i - \tilde{y_i}|}{\sum_{i=1}^{N}|y_i|}, \end{aligned}$$

where y_i is the true value, $\tilde{y_i}$ is the predicted value, and N is the number of data points.

Since sMAPE, MASE, and MAPE are percentage-based error estimators, they are scaled-independent so that they can also be used for comparing prediction performance across different datasets. In addition, all error estimators are symmetric, which means that both positive and negative prediction errors are penalized equally. However, MAPE has the disadvantage that the errors tend to blow-up when the variable values are low, causing the results to be misleading. Thus, WAPE is also applied here since the errors are weighted by the total values.

Explaining the model

SHAP (41), more specifically, Kernel SHAP, is a local, *post-hoc*, and model-agnostic XAI method that can be used for both classification and regression problems. *Post-hoc* interpretation means that the interpretability is created after the model has been constructed (32) and aims to provide an explanation for the black-box models (65). Another method is ante-hoc, in which the decision-making process or the basis of a technique of a model can be understood by humans without additional information (65). Some of the ante-hoc methods

6 The original notation for the generalized compatibility score in Vaswani et al. (64) is $\frac{QK^\top}{\sqrt{d_k}}$.

7 The original notation for the generalized output of the attention function in Vaswani et al. (64) is $Attention\,(Q,\ K,\ V) = softmax\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$.

include LR, Decision Tree, and KNN. Both ante-hoc and *post-hoc* methods can be further divided into two approaches, *Model (Global) Explanation* and *Instance (Local) Explanation*. The Local Explanation approach explains only the model prediction for the single data instance, whereas the Global Explanation approach explains the inner workings of the entire model trained on a dataset. Model-agnostic is a subcategory of *post-hoc* methods, such that it can be applied to a variety of models, whereas model-specific can only be applied to one specific type of model.

SHAP uses the Shapley value from Game Theory to assign importance to each feature. In effect, the feature contributions (Shapley values) are calculated by the marginal contribution of the feature over every feature so that how the model behaves in its absence is analyzed, and then the prediction of the model can be written as the sum of bias and single feature contributions (41). According to Lundberg et al. (79), SHAP belongs to the family of *Additive Feature Attribution Methods*, meaning that the Shapley values are applied to binarised features, where a value of 0 corresponds to an unknown feature value, and a value of 1 corresponds to a feature being observed. The explanation model can be written mathematically as:

$$g\left(z^{'}\right) = \phi_0 + \sum_{i=1}^{M} \phi_i z_i^{'},$$

where g is the explanation model of the prediction model, $z^{'} \in \{0,1\}^M$ where $z'$is the binarised feature and M is the number of binarised input features, ϕ_0 is the model output without binarised inputs, and $\phi_i \in R$ are the Shapley values (41). When compared with the other state-of-the-art explanation approach, LIME (36), SHAP satisfies three crucial properties that LIME does not: Local Accuracy, Missingness, and Consistency (41). Local accuracy requires consistency between the outputs of the explanation model and the prediction model. Missingness requires features missing in the original input to have no impact on the output. Lastly, consistency ensures that the impact of a feature does not decrease as it increases or remains the same.

Local accuracy is particularly important for providing explanations, as it ensures that the explanation model is less susceptible to adversarial attacks (66). Adversarial attacks refer to when the outputs of a classifier can be manipulated by a small perturbation of an input to conceal the biases of a system. In the study of Slack et al. (67), the authors attempted to fool both LIME and SHAP in order to determine if the feature contributions can be manipulated through the use of biased classifiers. It was found that the SHAP is less vulnerable to adversarial attacks than LIME due its local accuracy property. It is for these reasons that SHAP was chosen over LIME in our framework.

SHAP is a local XAI method that has been used to explain local predictions in many studies. For instance, Lenatti et al. (42) investigated the contribution of specific feature values to an individual prediction based on SHAP values. It is nevertheless also possible to obtain a global SHAP explanation by calculating the mean absolute SHAP values for each feature across the datasets allowing the global importance of each feature and the relative impact of all features over the entire dataset to be determined.

The results of SHAP will therefore be presented in the form of a visualization, in particular, the summary plots[8] will be used where it combines the feature importance with feature effects. The x-axis of the plots represents the SHAP value, or the impact on the model prediction, of each feature, the y-axis lists all the features and ordered according to their importance, and the color depicts the value of the feature from low to high.

In addition to the summary plots proposed to be used here, SHAP values can be analyzed in a variety of ways, including a dependence plot to demonstrate the global interaction effects between features. SHAP values may also be useful for assessing the contribution of features to an incorrect prediction, as demonstrated in the work of Lenatti et al. (42).

Expected outcome and predictors

The objectives of the SMART BEAR project in relations to HL are to answer several questions using the collected SMART BEAR data and the proposed predictive framework that leverages XAI techniques in order to develop a comprehensive profiling of patients with HL. Table 2 summarizes the expected outcome and its associated predictors (characteristics) for each question, and how this framework is applied to each question is discussed below.

As mentioned previously, this is a conceptual paper meaning that the precise details of the pre-processing techniques, optimal hyper-parameters for each question, and the prediction and explanation results will only be available once the study is commenced in autumn 2022.

Q1—Identification of those characteristics that make patients more prone to drop-out and stop using their HAids

The optimal drop-out rate should be less than the general population with HL (7), therefore, the expected outcome for Q1 is to be <45–50% for aged populations. Clinicians have the option of choosing how the data are aggregated in order to determine what the drop-out rate will be in the future in days, weeks, months, or years. In cases where a weekly analysis is required, for example, the average of HL chronicity, degree of HL, and manual adjustments of volume/program, and the sum of time of HAids usage are calculated for each week to

8 https://shap-lrjball.readthedocs.io/en/latest/generated/shap.summary_plot.html

TABLE 2 A description of the predictive models, their expected outcome, and associated predictors.

Prediction models (PM)	Predictors	Outcome variables	Expected outcome	Value type
Q1	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, manual adjustments of volume/program, overall HAids satisfaction, time, time of hearing aids usage	Dropout	<45–50%	Y/N
Q2	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, time	Time of HAid usage	Adults should use their HAids >10 h a day.	Minutes/day
Q3	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, number of visits, manual adjustments of volume/program, time	GHABP score	Described in detail below.	(Integer)
Q4	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, overall HAids satisfaction, manual adjustments of volume/program, time, time of hearing aids usage	Number of face-to-face sessions	<4 visits to the Audiologist's in the first 6 months.	(Integer)
Q5	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, overall HAids satisfaction, manual adjustments of volume/program, time, time of hearing aids usage	Number of remote sessions	<4 visits to the Audiologist's in the first 6 months.	(Integer)
Q6	Age, biological gender, hearing loss type, hearing loss chronicity, degree of hearing loss, noise exposure, overall HAids satisfaction, time, time of hearing aids usage	Number of manual changes per day	<3 per day.	(Integer)

convert the data frequency. Apart from handling missing data, outliers, and multicollinearity among the variables, continuous variables such as age, degree of HL, and time of HAids usage are standardized and normalized, nominal variables such as gender are one-hot encoded, and ordinal variables such as HL chronicity, HL type, and manual adjustment of volume/program are ordinal encoded. In addition, the outcome variable is also treated with label encoding, with 1 representing Yes and 0 representing No, for making a binary classification.

Attn-LSTM is then employed to predict whether or not a participant will stop using their HAids in the future and the identification of characteristics that have an impact on this prediction is carried out through SHAP. Finally, the predicted future number of drop-out participants is compared to the general population with HL in order to compute the drop-out rate.

Q2—Identification of those characteristics that make patients more prone to use their HAids sufficiently long during the day

It is recommended that adults should use their HAids for more than 10 hours a day (76). Due to this, data are aggregated to have a daily frequency by default. This is done by taking the average of HL chronicity, degree of HL, manual adjustments of volume/program, and overall HAids satisfaction for each day, and the sum of time of HAids usage for each day in minutes. It should note that, although the data are transformed to have a daily frequency by default, clinicians will still have the option to choose to analyse monthly HAid usage, for example, if required. Similarly to Q1, continuous variables are standardized and normalized, while nominal and ordinal variables are one-hot and ordinal encoded, respectively. Missing data, outliers, and multicollinearity will also be treated with appropriate pre-processing techniques.

As a regression problem, attn-LSTM is used to predict participants' future HAids usage. SHAP is then used to interpret the model prediction to identify which characteristics influence participants to use their HAids more often.

Q3—Identification of those factors augmenting the benefit of patients from using their HAid

The Glasgow Hearing-Aid Benefit Profile (GHABP)[9] is a questionnaire that was designed to assess the operational management for HAid benefit, both at the systematic and clinical levels (15). The questionnaire will assess 4 situations with 6 questions, which are scored with 1 being the best score and 5 being the worst score. Whitmer et al. (77) recruited 1,574 participants and were asked to rate their hearing disability, handicap, HAid use, HAid benefit, HAid satisfaction, and residual (aided) disability with the GHABP questionnaire. The participants were divided into none, unilateral, and bilateral aided users and assessed in the four situations: quiet conversations, TV listening, noisy conversations, and group

9 https://www.hey.nhs.uk/wp/wp-content/uploads/2020/09/HEY1167-2020-GHABP.pdf

conversations. Their findings regarding the normative GHABP score for HAid benefit will be used as the expected outcome for Q3.

Q3 is also a regression problem as the future GHABP score is predicted with attn-LSTM, and the reasons for this prediction are provided by SHAP. When clinicians require a monthly analysis, for example, the average of the GHABP score, HL chronicity, degree of HL, number of visits, and manual adjustments of volume/program, and the sum of time of HAids usage are calculated for each month to convert the data frequency. For Q3, pre-processing steps are similar to those used for previous questions, where continuous variables such as age, degree of HL, and time of HAids usage are standardized and normalized, nominal variable such as gender are one-hot encoded, and ordinal variables such as GHABP score, HL chronicity, HL type, number of visits, and manual adjustments of volume/program are ordinal encoded.

Q4—Identification of those factors decreasing the number of needed face-to-face sessions with their audiologist for counseling and/or HAid fine tuning, as an indicator of better self-management and optimal initial HAid configuration

The number of face-to-face with the audiologists is suggested to be <4 times in the first 6 months (78). Following this, the data are transformed to have a monthly frequency by default, with the options of analyzing the data at other frequencies still available. Therefore, the average of HL chronicity, degree of HL, number of visits, overall HAids satisfaction, and manual adjustments of volume/program, and the sum of time of HAids usage are calculated for each month. Nominal variables such as gender are one-hot encoded, ordinal variables such as overall HAids satisfaction, HL chronicity, HL type, number of visits, and manual adjustments of volume/program are ordinal encoded, and continuous variables such as age, degree of HL, and time of HAids usage are standardized and normalized.

As a regression problem, the future number of face-to-face sessions is predicted using attn-LSTM, and the characteristics affecting the prediction are investigated with SHAP.

Q5—Identification of those factors decreasing the number of needed remote sessions with their audiologist for counseling and/or HAid fine tuning, as an indicator of better self-management and optimal initial HAid configuration

Similar with Q4, the suggested number of remote sessions with the audiologists is also to be <4 times in the first 6 months (Tecca, 2018). Therefore, the default frequency is also set to be monthly, and attn-LSTM is used to predict the number of remote sessions with the audiologists in future months. SHAP is then used to identify the characteristics that influence participants to request fewer sessions with their audiologist. The pre-processing steps are also in line with Q4.

Q6—Identification of those factors decreasing the number of manual changes of HAid program, as indication of poor sound quality and bad adaptation of hearing aid configuration to patients' real needs and daily challenges

Although there is no precise definition for the optimal number of manual adjustments of the HAids, clinical experience has shown that fewer than three manual changes per day is considered as acceptable. By default, data are transformed to have a daily frequency in order to predict future daily manual adjustments with attn-LSTM, with SHAP providing information on the characteristics that impact the prediction.

It is also possible for clinicians to select a different data frequency for this analysis if required. The average of HL chronicity, degree of HL, number of visits, overall HAids satisfaction, and manual adjustments of volume and program, and the sum of time of HAids usage are calculated for each day to convert the data frequency. Pre-processing steps also consists of handling missing data, outliers, multicollinearity. As well as transforming continuous variables with standardization and normalization, ordinal variables with ordinal encoding, and nominal variables with one-hot encoding.

As a final point, SHAP values are analyzed with the same principle for all questions. The y-axis on the SHAP summary plot would indicate the most important feature on average for attn-LSTM to predict a certain outcome. The x-axis, along with the color, would show the impact of each feature value on the model prediction. For example, the SHAP values for Q1 may indicate that perhaps Age is the most important feature on average for participants to stop using their HAids. More specifically, younger participants might be less likely to drop out, whereas perhaps participants with a lower HAids usage might be more likely to stop using their HAids. As for Q3, SHAP result might show that perhaps HL type influences future GHABP score the most on average, where participants with a mixed type of HL might be more likely to benefit from their HAids.

Results—Discussion

This paper is a conceptual paper that synthesizes previous work on prediction models in healthcare and audiology (20, 27, 30, 31), and further describes the design and methods of the Big Data research project SMART BEAR with which we

are aiming to fill the identified knowledge gaps. To the best of our knowledge, SMART BEAR represents the first research initiative in hearing research aimed at integrating such large and heterogeneous datasets and analyzing them using AI and XAI methods.

According to Mellor et al. (12), many factors beyond the pure tone audiogram should be monitored and dynamically adapted in order to achieve optimal hearing rehabilitation. Prognostic prediction models using audiometric and other lifestyle or medical data may be helpful toward achieving this goal. Education level (68), cognitive performance (69), and performance on speech recognition tests (70) have previously been suggested as potential prognostic factors. Following this, a wide range of data is collected in SMART BEAR as shown in Supplementary materials 2, 3, such as demographics, audiometric data, cognitive status, mental status, habits, and biological gender. Taking advantage of the ability of modern HAids to record their dynamic operation will also enable a relatively low-cost collection of data, such as hours of HAid use, from a large population, while clinical assessment will provide insight into the clinical context of the collected data. Furthermore, instead of assessing patients in a laboratory environment, SMART BEAR is collecting data both at the office and in real life through clinical assessments and smart sensors.

The created and continuously updated data can then be viewed as sequences with temporal elements and contain high-dimensional clinical variables (63). Therefore, collected SMART BEAR data will be analyzed through time-dependent multivariate prediction models that are capable of handling both classification and regression problems while ensuring a high level of accuracy. The XAI method will then be applied in order to explain the model to clinicians so that they will be able to better understand how the model arrives at the predicted results. In this study, attention-based LSTM is proposed to be the prediction model and then using SHAP to interpret the model. The proposed framework introduced in this conceptual paper can also be applied to other comorbidities within the SMART BEAR project.

The findings of this analysis will have implications in clinical practice, health policies and research.

Clinical and research implications

With proper analysis and interpretation of SMART BEAR results, the most accurate patient profile to date can be created for HL patients, allowing it to serve as a valid proxy for anticipated behavior even before the initial HAid fitting session. According to the analysis of synthetic hearing data conducted within the context of the H2020 project EVOTION[10], higher levels of physical activity are associated with longer daily HAid use (43). Therefore, SMART BEAR results also aim to provide a better understanding how physical activity, such as walking, affects HAid experience in order to incorporate physical activity promotion into hearing rehabilitation for different populations. Furthermore, different factors relating to hearing rehabilitation might be identified with different participants. This is shown in the data-driven analysis with the subjective data of 572 HAid users conducted by Sanchez-Lopez et al. (71), where participants with different HL degree preferred different types of hearing rehabilitation. Other factors may include presence of particular comorbidities or different living situations, therefore, the combinations and interactions between the factors will also be examined in SMART BEAR.

The patient profiling proposed by SMART BEAR may be able to assist manufacturers and clinicians in making optimal choices in terms of HAid model and configuration options, or, in future stages, it could create automatic fine-tuning of HAids (12). In this context, after the end of the study, SMART BEAR is considering providing access, upon request, to the de-identified dataset for future exploration. Participants will be fully informed and will provide their consent so access to their de-identified data can be granted in the future for specific scientific purposes. Open Access will be provided for the following SMART BEAR datasets: anonymised data from demographics, questionnaires, interviews, anonymised sensor raw data, video of the protocols for annotation, and anonymised data from basic clinical information for annotation. It is envisaged that this policy will facilitate the use of SMART BEAR's gained knowledge by a range of different stakeholders.

Limitations

All participants in SMART BEAR will be fitted with the same HAid model, following the same fitting protocol, with the use of the same algorithm. Although the fine-tuning and the program selection of the HAids will be based on the needs and preferences of each participant, the fitting of the HAids may not be optimal for every participant when only one universal fitting protocol is used. However, this choice was made since the comparison of programs or algorithms is not in the scope of SMART BEAR, as well as in order to avoid unnecessary heterogeneity or lower quality of the data as a result of systematic errors. This limitation will be taken into account in the interpretation of our results. Moreover, SMART BEAR participants will only be between the ages of 67 and 80, which means that its results cannot be generalized to a population younger than that. Data like hours of usage and changes in programs will be subject to connectivity loss, which is a significant barrier in similar projects (50). The impact of loss of follow-up patients, such as the unavailability of information regarding continuation of usage, is also expected to be low, provided that this percentage will remain in the predicted range (below 20%). Close follow-ups and dedicated helpdesks

10 https://h2020evotion.eu/

will help minimize these risks, while imputation and model-based approaches will facilitate dealing with missing data, as explained above. Another limitation will be the variation in the population between six different countries with socioeconomic and cultural diversities; however, comparison between study groups is expected to produce useful results. Finally, speech audiometry in quiet or in noise is not part of the SMART BEAR data collection. This is due to the fact that there do not currently exist any universally validated materials that could be used across all six countries and thus in all languages. Speech audiometry, while recognized as having clinical value in fitting choices, does not fall under the scope of SMART BEAR. As an alternative approach to assess HAid benefit, we are aiming to collect other parameters, including real-life data, such as hours of usage and manual changes of programs, as well as interview data, such as the GHABP questionnaire.

It is noteworthy that unlike the evaluation metrics used in this paper to evaluate a prediction model, there are currently no widely accepted objective metrics for evaluating XAI methods. Though the proposed XAI method will be validated by clinicians and medical experts in SMART BEAR, this will only provide a subjective assessment of the XAI method. To this end, existing evaluation metrics for XAI metrics, such as Rosenfield's set (72), should be tested in the future with the collected data in order to obtain both objective and subjective validation. Although SHAP is one of the best known XAI methods, it is often criticized for long computation time and Shapley values do not work if features are correlated (73). As a result, the proposed framework may be unable to deliver what clinicians require in cases where the characteristics to be identified are correlated. Therefore, alternative methods of XAI should be considered in the future. Among them is Attention Mechanism-based XAI methods, such as the one proposed by Choi et al. (74) and Schockaert et al. (75). An attention mechanism-based XAI method can provide an explanation for Recurrent Neural Network or its variants by assigning corresponding values to the importance of the different sub-sequence of the input sequence according to the model and may be more suitable for the proposed prediction model.

Conclusion

SMART BEAR is, to the best of our knowledge, the first big data study whose goal is to integrate heterogeneous and contextualized HAid, medical, societal, and environmental data in order to develop and validate a prognosis framework using AI and XAI methods. The outcomes of the project are expected to benefit multiple stakeholders in the field of Audiology, such as HAid users, manufacturers, clinicians, researchers, and health policy makers, as well as to influence current practice and future research. These outcomes could also improve confidence in integrating AI models in the medical field, particularly with encouraging AI to be used in the medical decision-making process by utilizing XAI methods to enhance its interpretability, transparency, and accountability.

Ethics statement

This is a conceptual paper describing the rationale and design of the large scale H2020 project SMART BEAR. The SMART BEAR protocols have obtained, or are in the process to obtain, ethical approval in all six countries. All participants will have to provide their voluntary consent after oral and written information about the details of the project. General Data Protection Regulation (EU) 2016/679 (GDPR) principles will be implemented in all stages of data collection, storage and sharing.

Author contributions

EI, QS, CK, and TB: conceptualization, methodology: writing—original draft, and writing—review and editing. DK: writing—original draft. CK and TB: supervision. All authors contributed to the article and approved the submitted version.

Funding

SMART BEAR was funded by the European Commission (Grant Agreement No.: 857172/H2020-SC1-FA-DTS-2018-2).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2022.933940/full#supplementary-material

References

1. Shield B. Evaluation of the social and economic costs of hearing impairment. *Hear-It AISBL*. (2006). p. 1–202. Available online at: https://www.hear-it.org/sites/default/files/multimedia/documents/Hear_It_Report_October_2006.pdf

2. Lin FR, Ferrucci L, Metter EJ, An Y, Zonderman AB, Resnick SM. Hearing loss and cognition in the Baltimore Longitudinal Study of Aging. *Neuropsychology*. (2011) 25:763–70. doi: 10.1037/a0024238

3. Dawes P, Emsley R, Cruickshanks KJ, Moore DR, Fortnum H, Edmondson-Jones M, et al. Hearing loss and cognition: the role of hearing aids, social isolation and depression. *PLoS ONE*. (2015) 10:e0119616. doi: 10.1371/journal.pone.0119616

4. Li L, Simonsick EM, Ferrucci L, Lin FR. Hearing loss and gait speed among older adults in the United States. *Gait Posture*. (2013) 38:25–9. doi: 10.1016/j.gaitpost.2012.10.006

5. Saunders JE, Rankin Z, Noonan KY. Otolaryngology and the global burden of disease. *Otolaryngol Clin North Am*. (2018) 51:515–34. doi: 10.1016/j.otc.2018.01.016

6. Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. (2016) 388:1545–602. doi: 10.1016/S0140-6736(16)31678-6

7. McCormack A, Fortnum H. Why do people fitted with hearing aids not wear them? *Int J Audiol*. (2013) 52:360–8. doi: 10.3109/14992027.2013.769066

8. Saunders GH, Dillard LK, Zobay O, Cannon JB, Naylor G. Electronic health records as a platform for audiological research: data validity, patient characteristics, and hearing-aid use persistence among 731,213 U.S. veterans. *Ear Hear*. (2021) 42:927–40. doi: 10.1097/AUD.0000000000000980

9. Newman CW, Weinstein BE, Jacobson GP, Hug GA. The hearing handicap inventory for adults. *Ear Hear*. (1990) 11:430–3. doi: 10.1097/00003446-199012000-00004

10. Gatehouse S, Naylor G, Elberling C. Linear and nonlinear hearing aid fittings – 2. Patterns of candidature. *Int. J. Audiol*. (2006) 45:153–71. doi: 10.1080/14992020500429484

11. Dillon H. *Hearing Aids*, 2nd ed. New York, NY: Thieme Medical Publishers (2012).

12. Mellor J, Stone MA, Keane J. Application of data mining to a large hearing-aid manufacturer's dataset to identify possible benefits for clinicians, manufacturers, and users. *Trends Hearing*. (2018) 22:233121651877363. doi: 10.1177/2331216518773632

13. Timmer BHB, Hickson L, Launer S. Adults with mild hearing impairment: are we meeting the challenge? *Int J Audiol*. (2015) 54:786–95. doi: 10.3109/14992027.2015.1046504

14. Ferguson MA, Henshaw H. Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Front. Psychol*. (2015) 6:556. doi: 10.3389/fpsyg.2015.00556

15. Gatehouse S. Glasgow hearing aid benefit profile: derivation and validation of a client-centered outcome measure for hearing aid services. *J Am Acad Audiol*. (1999) 10:24. doi: 10.1055/s-0042-1748460

16. Wang L, Wang H, Song Y, Wang Q. MCPL-Based FT-LSTM: medical representation learning-based clinical prediction model for time series events. *IEEE Access*. (2019) 7:70253–64. doi: 10.1109/ACCESS.2019.2919683

17. Chakraborty D, Ivan C, Amero P, Khan M, Rodriguez-Aguayo C, Başagaoglu H, et al. Explainable artificial intelligence reveals novel insight into tumor microenvironment conditions linked with better prognosis in patients with breast cancer. *Cancers*. (2021) 13:3450. doi: 10.3390/cancers13143450

18. Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett*. (2020) 471:61–71. doi: 10.1016/j.canlet.2019.12.007

19. Ferroni P, Zanzotto F, Riondino S, Scarpato N, Guadagni F, Roselli M. Breast cancer prognosis using a machine learning approach. *Cancers*. (2019) 11:328. doi: 10.3390/cancers11030328

20. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep*. (2018) 8:3395. doi: 10.1038/s41598-018-21758-3

21. Vasudevan P, Murugesan T. Cancer subtype discovery using prognosis-enhanced neural network classifier in multigenomic data. *Technol Cancer Res Treat*. (2018) 17:1533033818790509. doi: 10.1177/1533033818790509

22. Diller GP, Kempny A, Babu-Narayan SV, Henrichs M, Brida M, Uebing A, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. *Eur. Heart J*. (2019) 40, 1069–1077. doi: 10.1093/eurheartj/ehy915

23. Javed Mehedi Shamrat FM, Ghosh P, Sadek MH, Kazi MdA, Shultana S. Implementation of machine learning algorithms to detect the prognosis rate of kidney disease. In: *2020 IEEE International Conference for Innovation in Technology (INOCON)*. (2020). p. 1–7.

24. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell*. (2020) 181:1423–33.e11. doi: 10.1016/j.cell.2020.04.045

25. Abdollahi H, Mostafaei S, Cheraghi S, Shiri I, Rabi Mahdavi S, Kazemnejad A. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: a machine learning and multi-variable modelling study. *Physica Medica*. (2018) 45:192–7. doi: 10.1016/j.ejmp.2017.10.008

26. Zhao Y, Li J, Zhang M, Lu Y, Xie H, Tian Y, et al. Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: a pilot study. *Ear Hear*. (2019) 40:690–9. doi: 10.1097/AUD.0000000000000649

27. Bing D, Ying J, Miao J, Lan L, Wang D, Zhao L, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. *Clin. Otolaryngol*. (2018) 43:868–74. doi: 10.1111/coa.13068

28. Tomiazzi JS, Pereira DR, Judai MA, Antunes PA, Favareto APA. Performance of machine-learning algorithms to pattern recognition and classification of hearing impairment in Brazilian farmers exposed to pesticide and/or cigarette smoke. *Environ Sci Pollut Res*. (2019) 26:6481–91. doi: 10.1007/s11356-018-04106-w

29. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *ArXiv*:1409.0473. (2014). doi: 10.48550/arXiv.1409.047363

30. Park HD, Han Y, Choi JH. Frequency-aware attention based LSTM networks for cardiovascular disease. In: *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. (2018). p. 1503–5.

31. Wall C, Zhang L, Yu Y, Mistry K. Deep recurrent neural networks with attention mechanisms for respiratory anomaly classification. In: *2021 International Joint Conference on Neural Networks (IJCNN)*. (2021). p. 1–8.

32. Burkart N, Huber MF. A survey on the explainability of supervised machine learning. *J Artif Int Res*. (2021) 70:245–317. doi: 10.1613/jair.1.12228

33. Anderson C. Ready for prime time?: AI influencing precision medicine but may not match the hype. *Clin OMICs*. (2018) 5:44–6. doi: 10.1089/clinomi.05.03.26

34. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. In: *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32 (2021). p. 4793–813.

35. Schlegel U, Arnout H, El-Assady M, Oelke D, Keim DA. Towards a rigorous evaluation of XAI methods on time series. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. (2019) 4197–201. doi: 10.1109/ICCVW.2019.00516

36. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?": explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016). 1135-1144.

37. Sarp S, Kuzlu M, Wilson E, Cali U, Guler O. The enlightening role of explainable artificial intelligence in chronic wound classification. *Electronics*. (2021) 10:1406. doi: 10.3390/electronics10121406

38. Malhi A, Kampik T, Pannu H, Madhikermi M, Framling K. Explaining machine learning-based classifications of in-vivo gastral images. In: *2019 Digital Image Computing: Techniques and Applications (DICTA)*. (2019). p. 1–7.

39. Das D, Ito J, Kadowaki T, Tsuda K. An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ*. (2019) 7:e6543. doi: 10.7717/peerj.6543

40. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast cancer recurrence prediction. *Artif Intell Med*. (2020) 107:101858. doi: 10.1016/j.artmed.2020.101858

41. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. (2017). p. 4768–4777.

42. Lenatti M, Moreno-Sánchez PA, Polo EM, Mollura M, Barbieri R, Paglialonga A. Evaluation of machine learning algorithms and explainability techniques to detect hearing loss from a speech-in-noise screening test. *Am J Audiol.* (2022) 1–19. doi: 10.1044/2022_AJA-21-00194 [Epub ahead of print].

43. Saunders GH, Christensen JH, Gutenberg J, Pontoppidan NH, Smith A, Spanoudakis G, et al. Application of big data to support evidence-based public health policy decision-making for hearing. *Ear Hear.* (2020) 41:1057–63. doi: 10.1097/AUD.0000000000000850

44. Nasreddine ZS, Phillips NA, BÃ©dirian V, Charbonneau S, Whitehead V, Collin I, et al. The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* (2005) 53:695–9. doi: 10.1111/j.1532-5415.2005.53221.x

45. Carpenter MG, Campos JL. The effects of hearing loss on balance: a critical review. *Ear Hear.* (2020) 41 (Suppl. 1):107S—19S. doi: 10.1097/AUD.0000000000000929

46. Oishi N, Shinden S, Kanzaki S, Saito H, Inoue Y, Ogawa K. Influence of depressive symptoms, state anxiety, and pure-tone thresholds on the tinnitus handicap inventory in Japan. *Int J Audiol.* (2011) 50:491–5. doi: 10.3109/14992027.2011.560904

47. Samocha-Bonet D, Wu B, Ryugo DK. Diabetes mellitus and hearing loss: a review. *Ageing Res Rev.* (2021) 71:101423. doi: 10.1016/j.arr.2021.101423

48. Manson J, Alessio H, Cristell M, Hutchinson KM. Does cardiovascular health mediate hearing ability? *Med Sci Sports Exerc.* (1994) 26:866–71. doi: 10.1249/00005768-199407000-00009

49. Simões JFCPM, Vlaminck S, Seiça RMF, Acke F, Miguéis ACE. Cardiovascular risk and sudden sensorineural hearing loss: a systematic review and meta-analysis. (2022) *Laryngoscope*. doi: 10.1002/lary.30141 [Epub ahead of print].

50. Dritsakis G, Kikidis D, Koloutsou N, Murdin L, Bibas A, Ploumidou K, et al. Clinical validation of a public health policy-making platform for hearing loss (EVOTION): protocol for a big data study. *BMJ Open.* (2018) 8:e020978. doi: 10.1136/bmjopen-2017-020978

51. Nayak B. Understanding the relevance of sample size calculation. *Indian J Ophthalmol.* (2010) 58:469. doi: 10.4103/0301-4738.71673

52. Sethia A, Raut P. Application of LSTM, GRU and ICA for stock price prediction. In: *Information and Communication Technology for Intelligent Systems*. Singapore: Springer (2019). p. 479–487.

53. Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock price prediction using LSTM, RNN and CNN-sliding window model. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).* (2017). p. 1643–7.

54. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev.* (2010) 67:503–27. doi: 10.1177/1077558709359007

55. Salgado CM, Azevedo C, Proença H, Vieira SM. Missing data. In: *Secondary Analysis of Electronic Health Records*, MIT Critical Data, editor (New York, NY: Springer International Publishing) (2016). p. 143–62.

56. Kuhn M, Johnson K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Boca Raton, FL: CRC Press (2019).

57. Ilyas IF, Chu X. *Data Cleaning*. New York, NY: ACM (2019).

58. Alin A. Multicollinearity. *Wiley Interdiscip Rev Comput Stat.* (2010) 2:370–4. doi: 10.1002/wics.84

59. Coren S. Summarizing pure-tone hearing thresholds: the equipollence of components of the audiogram. *Bull Psychon Soc.* (1989) 27:42–4. doi: 10.3758/BF03329892

60. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

61. Preeti BR, Singh RP. Financial and non-stationary time series forecasting using LSTM recurrent neural network for short and long horizon. In: *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT).* (2019). p. 1–7.

62. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing.* (2021) 452:48–62. doi: 10.1016/j.neucom.2021.03.091

63. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. *Adv Neural Inf Process Syst.* (2016) 9:29. doi: 10.48550/arXiv.1608.05745

64. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in Neural Information Processing systems 30 (NIPS 2017)*. Long Beach, CA: Curran Associates (2017).

65. Zhang Y, Weng Y, Lund J. Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics.* (2022) 12:237. doi: 10.3390/diagnostics12020237

66. Janizek JD, Sturmfels P, Lee S-I. Explaining explanations: axiomatic feature interactions for deep networks. *J Mach Learn Res.* (2021) 22:1–54.

67. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H. Fooling LIME and SHAP. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* (2020). p. 180–6.

68. Fuentes-López E, Fuente A, Valdivia G, Luna-Monsalve M. Does educational level predict hearing aid self-efficacy in experienced older adult hearing aid users from Latin America? Validation process of the Spanish version of the MARS-HA questionnaire. *PLoS ONE.* (2019) 14:e0226085. doi: 10.1371/journal.pone.0226085

69. Meister H, Rählmann S, Walger M, Margolf-Hackl S, Kießling J. Hearing aid fitting in older persons with hearing impairment: the influence of cognitive function, age, and hearing loss on hearing aid benefit. *Clin Interv Aging.* (2015) 10:435. doi: 10.2147/CIA.S77096

70. Davidson A, Marrone N, Wong B, Musiek F. Predicting hearing aid satisfaction in adults: a systematic review of speech-in-noise tests and other behavioral measures. *Ear Hear.* (2021) 42:1485–98. doi: 10.1097/AUD.0000000000001051

71. Sanchez-Lopez R, Dau T, Whitmer WM. Audiometric profiles and patterns of benefit: a data-driven analysis of subjective hearing difficulties and handicaps. *Int J Audiol.* (2022) 61:301–10. doi: 10.1080/14992027.2021.1905890

72. Rosenfeld A. Better metrics for evaluating explainable artificial intelligence. In: *20th International Foundation for Autonomous Agents and Multiagent Systems (AAMAS '21).* (2021), 45–50.

73. Molnar, C. (2020). *Interpretable Machine Learning*. Available online at: https://www.lulu.com/ (accessed July 02, 2022).

74. Choi KS, Choi SH, Jeong B. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro Oncol.* (2019) 21:1197–209. doi: 10.1093/neuonc/noz095

75. Schockaert C, Leperlier R, Moawad A. Attention mechanism for multivariate time series recurrent model interpretability applied to the ironmaking industry. *arXiv[Preprint].arXiv:*2007.12617 (2020).

76. Laplante-Lévesque A, Nielsen C, Jensen LD, Naylor G. Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport. *J Am Acad Audiol.* (2014) 25:187–98. doi: 10.3766/jaaa.25.2.7

77. Whitmer WM, Howell P, Akeroyd MA. Proposed norms for the glasgow hearing-aid benefit profile (Ghabp) questionnaire. *Int J Audiol.* (2014) 53:345–51. doi: 10.3109/14992027.2013.876110

78. Tecca JE. Are post-fitting follow-up visits not hearing aid best practices? *Hear. Rev.* (2018) 25:12–22.

79. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, CA: Curran Associates (2017). p. 4766–75.

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fneur.2022.943816

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Institute of Electronics, Information Engineering and Telecommunications (CNR), Italy

REVIEWED BY
Andrej Kral,
Hannover Medical School, Germany
Waldo Nogueira,
Hannover Medical School, Germany

*CORRESPONDENCE
Stefan Weder
stefan.weder@insel.ch

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 14 May 2022
ACCEPTED 25 July 2022
PUBLISHED 29 August 2022

CITATION
Schuerch K, Wimmer W, Dalbert A, Rummel C, Caversaccio M, Mantokoudis G and Weder S (2022) Objectification of intracochlear electrocochleography using machine learning. *Front. Neurol.* 13:943816. doi: 10.3389/fneur.2022.943816

Objectification of intracochlear electrocochleography using machine learning

Klaus Schuerch[1,2], Wilhelm Wimmer[1,2], Adrian Dalbert[3], Christian Rummel[4], Marco Caversaccio[1,2], Georgios Mantokoudis[1] and Stefan Weder[1]*

[1]Department of ENT, Head and Neck Surgery, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland, [2]Hearing Research Laboratory, ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland, [3]Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital Zurich, University of Zurich, Zurich, Switzerland, [4]Support Center for Advanced Neuroimaging (SCAN), University Institute for Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

Introduction: Electrocochleography (ECochG) measures inner ear potentials in response to acoustic stimulation. In patients with cochlear implant (CI), the technique is increasingly used to monitor residual inner ear function. So far, when analyzing ECochG potentials, the visual assessment has been the gold standard. However, visual assessment requires a high level of experience to interpret the signals. Furthermore, expert-dependent assessment leads to inconsistency and a lack of reproducibility. The aim of this study was to automate and objectify the analysis of cochlear microphonic (CM) signals in ECochG recordings.

Methods: Prospective cohort study including 41 implanted ears with residual hearing. We measured ECochG potentials at four different electrodes and only at stable electrode positions (after full insertion or postoperatively). When stimulating acoustically, depending on the individual residual hearing, we used three different intensity levels of pure tones (i.e., supra-, near-, and sub-threshold stimulation; 250–2,000 Hz). Our aim was to obtain ECochG potentials with differing SNRs. To objectify the detection of CM signals, we compared three different methods: correlation analysis, Hotelling's T^2 test, and deep learning. We benchmarked these methods against the visual analysis of three ECochG experts.

Results: For the visual analysis of ECochG recordings, the Fleiss' kappa value demonstrated a substantial to almost perfect agreement among the three examiners. We used the labels as ground truth to train our objectification methods. Thereby, the deep learning algorithm performed best (area under curve = 0.97, accuracy = 0.92), closely followed by Hotelling's T^2 test. The correlation method slightly underperformed due to its susceptibility to noise interference.

Conclusions: Objectification of ECochG signals is possible with the presented methods. Deep learning and Hotelling's T^2 methods achieved excellent discrimination performance. Objective automatic analysis of CM

signals enables standardized, fast, accurate, and examiner-independent evaluation of ECochG measurements.

KEYWORDS

ECochG, signal processing, deep learning, Hotelling's T^2, correlation analysis, residual hearing, electroacoustic stimulation, cochlear implant

1. Introduction

Electrocochleography (ECochG) measures electrical potentials generated by the inner ear in response to acoustic stimulation. In patients with cochlear implant (CI), using the implanted electrode, these potentials can be picked up directly from the inner ear. The technique is increasingly used to monitor the inner ear function during and after implantation. Research groups were able to correlate changes in the ECochG signal with traumatic events during implantation (1–6).

In order to assess ECochG potentials (either intra or postoperatively), the analysis is most commonly performed by visual inspection, which is currently the gold standard. Therefore, the interpretation is heavily relying on the expertise of the examiner. This entails several problems: i) a high level of experience is needed to interpret the signals correctly. Thus, inexperienced clinicians and researchers are unable to exploit the technique; ii) the examiner determines whether or not an ECochG response is present, which may result in a lack of reproducibility; iii) longitudinal comparisons are hampered as the assessment is not absolutely identical. iv) research groups use different types of analysis, which makes the comparability of clinical findings and study results difficult or impossible (4, 7–12); v) due to the inconsistent assessment, patients with a poor signal-to-noise ratio (SNR) are often not reported. However, in order to draw correct conclusions, all measurements should be reported (13, 14); and vi) the analysis of ECochG signals is complex, which makes immediate judgment difficult. This is, of course, a prerequisite when an instant assessment is required (e.g., in the operating theater).

ECochG itself is an umbrella term for different electrophysiological signal components of the inner ear (i.e., the cochlear microphonic, CM, the auditory neurophonic, ANN, the compound action potential, CAP, the summating potential, SP). These signal components can be highlighted by measurements with different acoustic polarities (condensation, CON and rarefaction, RAR). The difference potential (DIF) is calculated by subtracting the CON and RAR polarities. The DIF response mainly represents the CM signal (15). In addition, the sum highlights the summating potential (SUM), which mainly represents the ANN (16). However, CM and ANN potentials cannot be isolated, especially at high stimulation levels and low frequencies (17). In intra and postoperative recordings, most commonly the CM/DIF signal is used as it is the largest and most robust signal component (18). For this reason, in this article, we will limit the analysis to the CM/DIF signal. Even though the CM/DIF signal is the strongest potential, there are some things to keep in mind. The amplitude of the signal is in the microvolt range and varies greatly between individuals. While certain patients show large amplitudes, in others, the potentials are very small, resulting in a poor SNR. Furthermore, the morphology and latency of the CM/DIF signal might vary significantly depending on the remaining intact hair cells (19–21). These factors (i.e., poor SNR, different wave morphology) must be taken into account when analyzing ECochG potentials.

For the reasons given above, an automated and objective evaluation would be highly desirable. This would standardize and significantly simplify the analysis of the signals and make it independent of the examiner. For ECochG signals, an approach using Fast Fourier Transform (FFT) has been proposed (18, 22–24). However, this method is not always applicable, especially for short signals, since they do not have a stationary period and adjacent frequencies cannot be accurately distinguished. For other electrophysiological signals, objectified analyses have become established in clinical practice. For example, for auditory brainstem responses (ABR), correlation analysis is used (25, 26). In the evaluation of cortical auditory evoked potentials (CAEP), Hotelling's T^2 test has yielded a sensitivity at least comparable to that of visual inspection (27–29). In other medical disciplines (i.e., identification of cardiac arrhythmias in electrocardiograms, ECGs), deep learning (DL) strategies could be successfully implemented (30–33).

The aim of this study was to automate and objectify the analysis of CM/DIF signals in ECochG recordings. The employed method should i) be comparable to visual analysis, (ii) allow the interpretation of intra- and postoperative ECochG signals by clinicians and researchers who do not have much experience in the field, (iii) allow immediate feedback, (iv) should be replicable by other clinical and research centers, (v) allow reproducible comparison of longitudinal data (since the same analysis is performed).

2. Materials and methods

This prospective cohort study was conducted in accordance with the Declaration of Helsinki and was approved by the

local institutional review board (KEK-BE 2016-00887 and 2019-01578). All participants gave written informed consent before participation.

2.1. ECochG data

We performed ECochG measurements in 36 subjects (n = 41 ears). All subjects used a Med-El implant (MED-EL, Austria). Pure tone audiograms were performed in a certified acoustic chamber with a clinical audiometer (Interacoustics, Denmark). Hearing thresholds were collected either immediately preoperatively or, in the case of postoperative measurements, on the same day as the ECochG measurement. We obtained pure tone air conduction hearing thresholds in dB hearing level (HL) at 125, 250, 500, 750, 1,000, 1,500, 2,000, and 4,000 Hz using either headphones or plug-in earphones. Pure tone averages (PTAs) were calculated as the mean hearing threshold at 125, 250, 500, and 1,000 Hz. PTAs and patient demographics are shown in Table 1.

We recorded ECochG potentials using the Maestro Software (version 8.03 AS and 9.03 AS, MED-EL, Austria). The system setup was identical to our previous study (10). We measured ECochG potentials at electrodes 1, 4, 7, and 10 (with electrode 1 at the tip) and only at a stable electrode position (i.e., either intraoperatively after completed electrode insertion or in a postoperative setting). When stimulating, depending on the individual hearing threshold, we used three different intensity levels: supra-threshold level (5 dB below discomfort level), near-threshold level (10 dB above hearing threshold), and sub-threshold level (10 dB below hearing threshold). Thereby, the acoustic amplitude level was restricted as shown in Table 2. Our aim was that not all stimulations would elicit an ECochG response and that, depending on the stimulation level, the SNR was different. As an acoustic stimulus, we used pure tones with settings shown in Table 2. ECochG potentials were recorded with two polarities (i.e., CON, and RAR). For each ECochG response, we recorded 100 epochs per polarity. The two polarities were subtracted to form the CM/DIF signal.

2.2. Preprocessing of ECochG signals

As preprocessing, we used the following steps: i) if present, removal of stitching artifacts, ii) application of a Gaussian weighted averaging method to increase the SNR and exclude uncorrelated epochs from further analysis, and iii) a 2nd order, forward-backward filtered Butterworth bandpass filter (cutoff frequencies 10 Hz / 5 kHz for visual analysis, and 100 Hz / 5 kHz for objective evaluation methods). To increase the SNR in our ECochG recordings, we calculated the Gaussian weighted epochs $S_{GE(i)}$ as described by Davila et al. (34) and Kumaragamage et al. (35). We used the following equation:

$$S_{GE(i)} = \sum_{l=-2}^{2} \left(e^{-[0.5(\frac{l}{\sigma\cdot(5-1)/2})^2]} \cdot S_{E(i+l)}\right)$$

whereas, l is the index number, starting from −2 to 2 that accounts for five epochs S_E averaged under the Gaussian window, and i is the index number of the epochs in S_E. The SD of the Gaussian window σ was set to 0.4. Each Gaussian weighted epoch $S_{GE(i)}$ was then correlated with the mean of all epochs S_{approx}. $S_{GE(i)}$ with a correlation less than −0.2 were excluded to form the final ECochG response S. If more than 10% of epochs had to be removed, only the 10 worst correlated were discarded. Finally, we calculated the SNR using the +/- averaging method (36).

2.3. Visual analysis

ECochG data were visually analyzed by three examiners with extensive experience in the field. The goal was to have a labeled data set that was used i) to train and test the objective algorithms, and ii) to obtain a benchmark for evaluating the accuracy, specificity, and sensitivity of the objective detection methods. Using Labelbox (37), the data were presented to the examiners as a subplot with six individual graphs representing i) the DIF response, ii) the SUM response, iii) the CON and RAR responses, and iv-vi) their individual FFT traces (an example is shown in the Supplementary material). Each examiner had to assess 4133 ECochGs with the question if a CM/DIF response was present or not (dichotomous question). Thereby, we used a blinded design in which the investigators did not discuss the assessment to avoid bias in the individual assessment. Signals classified as CM/DIF response by two examiners (and noise by one examiner) were presented a second time to all three examiners (to minimize volatility errors). Only ECochG signals that were finally considered valid responses by all three investigators were classified as responses. These were used as ground truth for the objective classification. We used Fleiss' kappa to compare the raters. Fleiss' kappa is a measure of agreement between multiple raters in classifying items (38).

2.4. Objective detection methods

We included the following objective detection methods: i) Hotelling's T^2 test, ii) correlation analysis, and iii) a DL convolutional neural network (CNN). To train and evaluate our objective analysis, we benchmarked these methods against the visual analysis of the three experts.

The dataset was divided into two parts: 70% for training and 30% for testing purposes. We used the training subset

10.3389/fneur.2022.943816

TABLE 1 Demographic of included subjects.

Subject ID	Gender	Age (years)	Side	Etiology	Electrode	ToM (month)	PTA (dB HL)
io 1	M	49	L	Meningitis	Flex 28	io	52.5
io 2	M	69	L	Progressive HL	Flex 28	io	58.8
io 4	F	45	L	Progressive HL	Flex 28	io	93.8
io 5	F	60	L	Progressive HL	Flex 24	io	66.3
io 6	M	51	R	Progressive HL	Flex 28	io	60.0
io 7	M	75	R	Progressive HL	Flex 28	io	52.5
io 8	F	77	L	Progressive HL	Flex 28	io	75.0
io 9	M	36	R	Congential genetic	Flex 26	io	48.8
io 10	M	71	R	Progressive HL	Flex 28	io	71.3
io 11	F	70	L	Progressive HL	Flex 28	io	50.0
io 12	F	27	R	Congential genetic	Flex 28	io	62.5
io 13	M	66	R	Meniere's disease	Flex 28	io	72.5
io 14	F	53	L	Progressive HL	Flex 28	io	78.8
io 15	M	59	R	Progressive HL	Flex 28	io	48.8
io 16	F	78	L	Progressive HL	Flex 28	io	86.3
io 17	F	28	R	Progressive HL	Flex 26	io	33.8
io 18	M	86	L	Progressive HL	Flex 26	io	91.3
io 19	M	21	R	Progressive HL	Flex 28	io	78.8
io 20	F	61	R	Sudden HL	Flex 28	io	81.3
io 23	M	59	L	Progressive HL	Flex 28	io	77.5
io 24	F	37	L	Sudden HL	Flex 26	io	83.8
po 0	F	60	R	Progressive HL	Flex 28	10	68.8
po 1	M	73	R	Progressive HL	Flex 28	17	110.0
po 2	M	75	L	Progressive HL	Flex 24	46	66.3
po 3	M	80	L	Congential genetic	Flex 28	9	85.0
po 4	F	27	R	Congential genetic	Flex 28	20	101.3
po 5	F	66	R	Progressive HL	Flex 28	28	92.5
po 6	F	73	R	Meniere's disease	Flex 28	78	90.0
po 7	M	82	L	Progressive HL	Flex 28	75	113.8
po 8	F	25	R	Congential genetic	Flex 28	57	85.0
po 9	F	43	R	Progressive HL	Flex 28	22	83.8
po 10	F	60	R	Progressive HL	Flex 24	13	97.5
po 11	F	73	L	Progressive HL	Flex 28	70	100.0
po 12	M	50	R	Meningitis	Flex 28	11	81.3
po 13	F	68	L	Progressive HL	Flex 28	22	93.8
po 14	F	52	R	Congential genetic	Flex 24	174	95.0
po 15	M	50	L	Meningitis	Flex 28	6	75.0
po 16	M	66	R	Meniere's disease	Flex 28	7	106.3
po 17	M	56	R	Sudden HL	Flex 28	11	91.3
po 18	M	75	R	Progressive HL	Flex 28	70	96.3
po 19	F	63	R	Progressive HL	Flex 24	131	91.3
Mean		58.4				43.9	79.2

PTA, pure tone average; HL, hearing loss; ToM, time of measurement in months after implantation; io, intraoperative; po, postoperative.

to train and validate the models. For training, both features (ECochG signals) and labels (ground truth determined by the examiners) were provided. The test set was used to evaluate the performance of the model. Here, only features were provided. The predictions of the model were then compared to the labels.

2.4.1. Hotelling's T^2 test

Based on Hotelling's T^2 method described by Golding et al. and Chesnaye et al. for objective detection of CAEP signals, we adapted the method to ECochG signals (27, 29). The Hotelling's T^2 test for one sample is a multivariate extension of the Student's *t*-test (39, 40). With Hotelling's T^2 test, we can test the null

TABLE 2 Settings for acoustic stimulation and maximum possible acoustic stimulation level (maximum amplitude).

Frequency (Hz)	Stimulus duration (ms)	Recording delay (ms)	Measurement window (ms)	Maximum amplitude (dB HL)
250	12	1	19.1	109
500	8	1	9.6	115
750	6.67	1	9.6	123
1,000	5	1	8.0	122
1,500	4	1	8.0	122
2,000	3	1	6.5	122

hypothesis (H0) whether Q features are statistically different from Q hypothesized values.

In our case, the ECochG recordings were the features and the hypothesized values were noise. The ECochG recordings were divided into Q windows along the time axis called 'time-voltage-means' (TVMs). The mean value was taken from each Q-window, resulting in the following N × Q voltage matrix V:

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1Q} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{NQ} \end{bmatrix}$$

Where N was the number of epochs and v_{ij} the j^{th} voltage means from the i^{th} epoch. The corresponding hypothetical values (noise) were an array of size 1 × Q filled with zeros. The noise was zero because the expected mean value of an ECochG signal should be zero due to the bandpass filtering. The number of used TVMs resulted in a down sampling, illustrated in Figure 1.

We performed the calculations using a python (v 3.9.7) script and the *hotellings* function from the spm1d module (v 0.4) (41, 42). As significance level α, we used 0.01 to tune the number of voltage means Q for each acoustic stimulus frequency individually. The optimal number of TVMs for the Hotelling T^2 test was calculated based on the maximum accuracy. For this purpose, the number of TVMs was successively increased in steps of five from 5 to 195 and the Hotelling's T^2 test was calculated on the training set.

2.4.2. Correlation analysis

Our correlation algorithm is based on the method of Wang et al. which explores the correlation of ABR signals (26). The correlation procedure relies on the repeatability of the similarity of two waveforms. The degree of similarity can be quantified by calculating the Pearson correlation coefficient. A positive correlation close to one reflects the presence of a response, while a zero correlation shows the absence of response (25).

In our calculations, we treated the two polarities (CON/RAR) separately and finally averaged the correlation coefficients. The two polarities were separate, treated as they evolve inversely (which is caused by condensation and rarefaction phased acoustic stimuli). The procedure is shown in Figure 2. Finally, we fitted a logistic regression model based on the correlation coefficients.

2.4.3. Deep learning

Our DL classification approach was based on the method used to automatically identify cardiac arrhythmia in ECG signals. Several DL approaches to cardiac arrhythmia detection have been proposed in the literature (30–33). Among them, time frequency scalograms using continuous wavelet transform (CWT) and AlexNet showed convincing results (32, 33). AlexNet is a large convolutional neural network (CNN) containing about 6,50,000 neurons and 60 million parameters. It consists of five convolutional layers, and three fully connected layers and is optimized for image classification (43).

Time frequency scalogram images for the classifier were generated from our dataset using CWT and the Python module PyWavelets (44). In this process, a Morlet wavelet shrinks and expands to map the signals into a time-frequency scalogram. We chose the Morlet wavelet because it offers a good compromise between spatial and frequency resolution (33, 45). We normalized the scalograms and compressed them to a dimension of 224 × 224 × 3 for width, height, and depth (red, green, blue). ECochG DIF traces and their wavelet transformation are shown in Figure 3.

We used PyTorch (v 1.11.0) and the pre-trained (on the ImageNet database) AlexNet loaded from torchvision (v 0.6.0) to take advantage of the already good classification properties (46, 47). We substituted the last two classifiers of the AlexNet for binary classification output. The rest of the network was left exactly as it was during initialization. Stochastic gradient descent with momentum was used to train the model. The mini-batch size was 8 and the maximum epoch was 25 with the learning rate being 1e-4, and a momentum of 0.9. We used 10-fold cross-validation to detect overfitting. We then trained the model with the full training set to increase model performance.

2.5. Statistical analysis

We used accuracy, sensitivity, and specificity to evaluate our algorithms. The algorithms were compared using the area under the receiver operating characteristic (ROC) curve, also known as the area under the curve (AUC). We used a one-sided DeLong test with a confidence level of 0.95 using the *roc.test* function of the pROC package (v 1.18.0) with R (v 4.1.2) (48, 49).

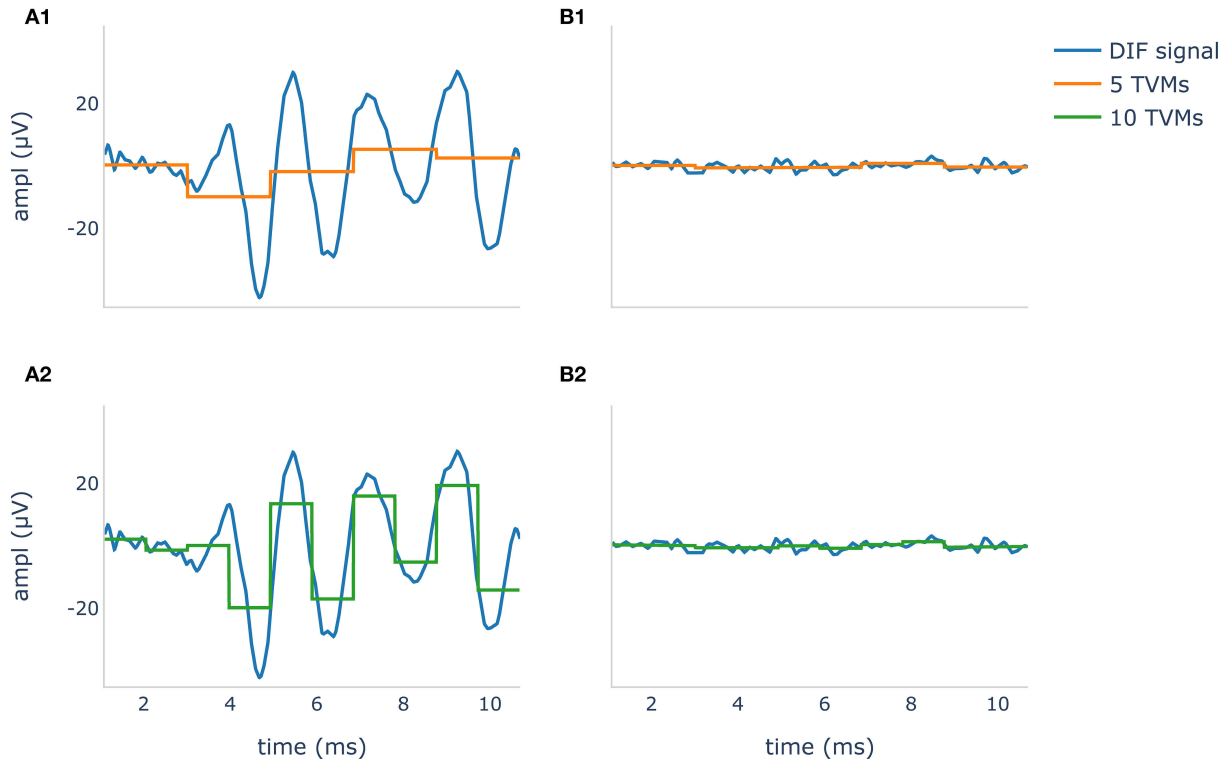


FIGURE 1
Difference potential (DIF) curves in blue show a recognizable CM/DIF signal **(A1,A2)** and noise with no visible CM/DIF component (B1, B2) in response to a 500 Hz stimulus. The orange curve in **(A1,B1)** shows 5 time-voltage-means (TVMs), and the green curve in **(A2,B2)** shows 10 TVMs used to calculate Hotelling's T^2 test. It is evident that in this example, an increase of the TMVs leads to better mapping of the CM/DIF signal with higher accuracy.

3. Results

3.1. ECochG recordings and preprocessing

Gaussian weighted averaging significantly increased the mean SNR from 2.50 dB (standard deviation, SD, 2.39) to 4.18 dB (SD 1.86) as demonstrated by the one-tailed paired-samples *t*-test ($p < 0.001$). In total, 4133 DIF signals were labeled visually by the three experts. Labeling took between 13.5 and 15 h (on average, 12 s per signal). In contrast, objective analysis using the algorithms took less than 25 ms per signal (the duration was determined on a notebook XPS 13 9360 (Dell, Round Rock, TX, USA) and does not include the training time of the algorithms, which was substantially longer).

3.2. Visual analysis

The Fleiss' kappa value of the agreement for the examiners and all stimulation frequencies are shown in Table 3. Results demonstrated a substantial to almost perfect agreement among the examiners (50). Particularly, for the mid-frequencies (500 Hz – 1 kHz), the examiners were very much in agreement. This agreement was a little lower for the lowest (i.e., 250 Hz) and the two highest frequencies (i.e., 1,500 and 2,000 Hz), but still substantial. However, between the three examiners, there was a systematic discrepancy in the visual assessment. The false-positive rates (FPRs) for examiners 1, 2, and 3 were 0.110, 0.068, and 0.032, respectively. That is examiner 1 still considered signals with a lot of noise as valid responses, whereas examiner 3 only accepted clearer neurophysiological traces.

Table 4 shows an overview of the stimulation frequencies, the stimulation levels, the SNR, and the number of signals where the experts identified a CM/DIF response. For frequencies of 500 Hz and above, when stimulated at supra-threshold level, a clear CM/DIF component was found in 53.3%.

For all frequencies, the supra-threshold stimulation showed the largest amplitudes ($p < 0.001$, one-tailed paired-samples *t*-test), the biggest SNR ($p < 0.001$) as well as the most visible signals. Near-threshold stimulation showed larger

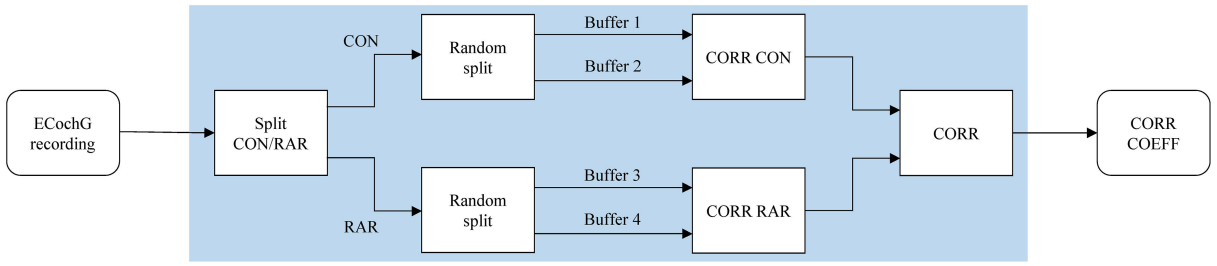


FIGURE 2
The correlation analysis handles CON and RAR recordings separately and proceeds as follows: (i) the ECochG recordings are divided into CON and RAR; (ii) CON and RAR are each divided into two randomly arranged buffers of the same size (Buffers 1–4, 50 epochs each); (iii) the Pearson correlation coefficients for CORR CON and CORR RAR are calculated from buffer 1 and 2 and buffer 3 and 4, respectively; (iv) CORR is calculated from the mean of CORR CON and CORR RAR. Since CORR depends on the subdivision of buffers, steps ii–iv (shaded area) are repeated 100 times and averaged to get the final correlation coefficient CORR COEFF. CON, condensation; RAR, rarefaction; CORR, correlation; COEFF, coefficient.

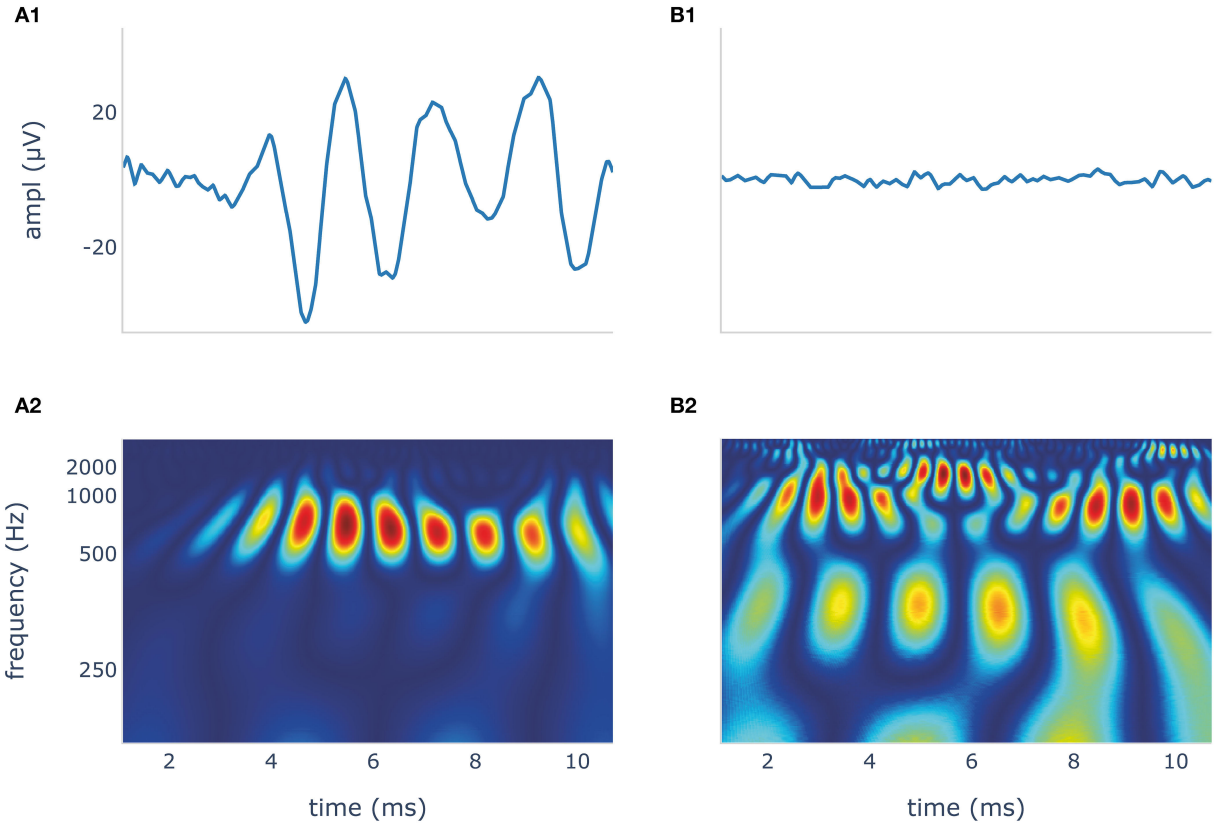


FIGURE 3
The blue DIF curves **(A1,B1)** show a recognizable CM/DIF signal **(A1)** and noise with no visible CM/DIF component **(B1)**, respectively, in response to a 500 Hz stimulus. Their corresponding time frequency scalograms generated using continuous wavelet transformation (CWT) are shown in **(A2,B2)**. These scalograms were then used to train and test the deep learning algorithm. DIF, difference; CM, cochlear microphonic.

amplitudes ($p < 0.001$), and bigger SNR ($p < 0.001$) than sub-threshold stimulation. However, this was not the case for 250 Hz stimulation amplitudes ($p = 0.104$). Regarding visual analysis, near-threshold levels showed significantly more visible CM/DIF responses than sub-threshold levels, except at 250 Hz. At this frequency, we identified the same number of responses for near-threshold and sub-threshold levels.

3.3. Comparison of objectification methods

All objectification methods presented in Table 5 showed good performance in detecting CM/DIF responses (51). The ROC curves of the objectification methods for all mixed frequencies are shown in Figure 4. The DL method performed best (AUC = 0.97, accuracy = 92%), followed closely by Hotelling's T^2 test (AUC = 0.96, accuracy = 91%). Statistically, this difference was not significant ($p = 0.14$). In contrast, the correlation analysis method underperformed as a classifier (AUC = 0.85; accuracy = 83%). This difference was statistically significant (DL $p < 0.001$; Hotelling's T^2 test $p < 0.001$). Table 5 shows the performance of the algorithms for all frequencies.

TABLE 3 Fleiss' kappa among all three examiners.

Frequency (Hz)	Fleiss' kappa	Interpretation
250	0.748	Substantial agreement
500	0.860	Almost perfect agreement
750	0.868	Almost perfect agreement
1,000	0.858	Almost perfect agreement
1,500	0.799	Substantial agreement
2,000	0.740	Substantial agreement
Mean	**0.815**	Almost perfect agreement

Interpretation according to Landis and Koch et al. (50).

4. Discussion

This study demonstrates that it is possible to objectively and automatically determine whether a CM/DIF response is present or not. All three algorithms investigated showed very good to excellent discrimination performance. Especially Hotelling's T^2 test and the DL method revealed excellent results (mean accuracy was 91 and 92% with an AUC of 0.96 and 0.97, respectively).

4.1. Preprocessing

ECochG traces are usually displayed as averaged signals (both, intra,- and postoperatively). During signal recordings, noisy epochs can affect the signal quality and reduce SNR (34). In addition, there are large inter-individual differences. While some patients show very prominent potentials, in others the signal amplitude is small (1, 3, 10, 12, 52). If ECochG is to be used routinely in the operating room and postoperative setting, however, all patients (including those with small signals) must be analyzed. In our cohort, the previously described Gaussian weighted averaging method (34, 35) showed a substantial increase in SNR of ECochG signals of all frequencies. Our calculations improved the mean SNR by 1.68 dB. Kumarange et al. were able to improve the SNR by 3.5 dB. However, they used extracochlear ECochG recordings, whereas we measured from inside the cochlea.

TABLE 4 Overview of the stimulation frequencies, the individual intensities, and the SNR.

Frequency (Hz)	Threshold	*n*	Ampl (dB)	Ampl STD	SNR (dB)	SNR STD	*n* visible	% n visible
250	Supra	226	27.33	2.71	2.68	1.34	49	21.7
	Near	222	26.19	2.32	2.32	0.40	10	9.0
	Sub	135	26.50	2.21	2.28	0.27	6	4.4
500	Supra	301	27.61	5.32	4.41	5.55	144	47.8
	Near	283	25.00	2.48	2.62	0.66	43	15.2
	Sub	161	24.50	2.60	2.37	0.39	2	1.2
750	Supra	225	28.00	5.92	4.20	5.30	114	50.1
	Near	272	25.30	3.45	2.80	2.70	63	23.2
	Sub	190	24.92	2.84	2.41	1.17	12	6.3
1,000	Supra	212	29.62	7.88	5.64	6.96	120	56.6
	Near	333	25.24	3.86	2.95	2.83	123	36.9
	Sub	200	24.09	2.60	2.38	0.78	10	5.0
1,500	Supra	193	27.44	6.31	3.98	5.18	102	52.8
	Near	301	24.38	3.11	2.40	1.20	67	22.2
	Sub	187	23.84	3.43	2.30	1.03	8	4.3
2,000	Supra	176	27.35	7.05	4.24	5.39	110	62.5
	Near	270	24.09	3.23	2.40	0.95	81	30.0
	Sub	227	23.23	3.23	2.42	1.02	35	15.4

Frequency, pure tone frequency in Hz; Ampl, peak-to-peak amplitude (dB re 1 μV): SD, standard deviation; SNR, signal-to-noise ratio; n, number of entries. n visible: signals where all examiners indicated a visible cochlear microphonic component.

TABLE 5 Performance of objectification methods.

Frequency	Correlation analysis					Hotelling's T^2 test						Deep learning				
(Hz)	Acc	Sens	Spec	CI	AUC	Acc	Sens	Spec	CI	AUC	Q	Acc	Sens	Spec	CI	AUC
250	0.92	0.19	1	0.04	0.64	0.90	0.83	0.91	0.04	0.96	90	0.96	0.88	0.98	0.03	0.98
500	0.85	0.50	0.99	0.05	0.91	0.93	0.95	0.92	0.03	0.97	80	0.92	0.94	0.91	0.04	0.97
750	0.84	0.50	0.98	0.05	0.88	0.93	0.91	0.93	0.04	0.98	100	0.94	0.91	0.95	0.03	0.97
1,000	0.81	0.58	0.94	0.05	0.84	0.86	0.95	0.82	0.05	0.97	85	0.91	0.88	0.92	0.04	0.97
1500	0.82	0.42	0.97	0.05	0.82	0.95	0.95	0.94	0.03	0.99	105	0.93	0.95	0.92	0.04	0.99
2,000	0.77	0.44	0.95	0.06	0.81	0.89	0.78	0.96	0.05	0.91	100	0.84	0.71	0.92	0.06	0.92
all	0.83	0.52	0.95	0.02	0.85	0.91	0.91	0.91	0.02	0.96		0.92	0.88	0.94	0.02	0.97

Q is the number of TVMs used in Hotelling's T^2 test. The optimal number of TVMs depends on the frequency and is given in the Q column. Stim, stimulus frequency (Hz); Acc, accuracy; Sens, sensitivity; Spec, specificity; CI, 95% confidence interval; AUC, area under the receiver operator characteristic curve; Q, number of used TVMs for Hotelling's T^2 test.

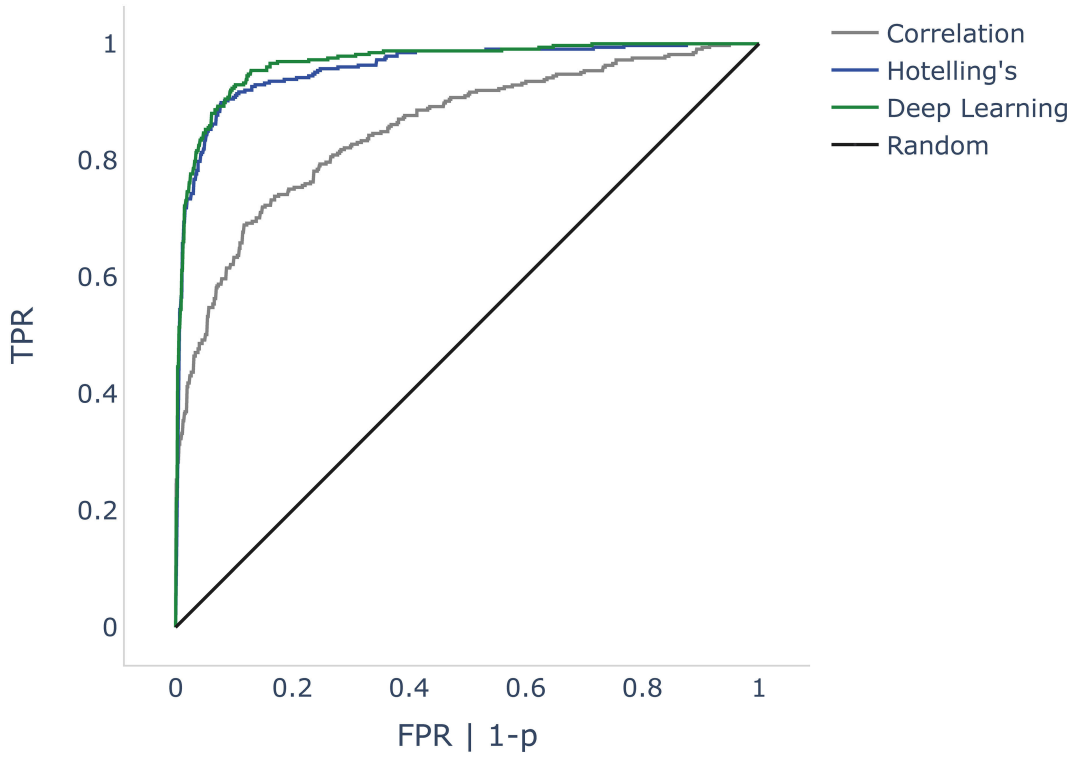


FIGURE 4
ROC curves comparing correlation analysis, Hotelling's T^2 test, and deep learning (DL) methods. The false positive rate (FPR) is the dependent variable (x-axis) in the DL and correlation algorithms. As Hotelling's T^2 test does not specify probabilities, we used *p*-values instead. Since p is inversely proportional to probabilities, we mapped 1-p. The black line shows a random classifier. ROC, receiver operator characteristic; p, *p*-value.

4.2. Visual analysis

In our study, the visual evaluation of the data was carried out by three independent examiners who have many years of experience in this field. Per recording, it took them 12 s on average to judge if a signal was present or not. In contrast, with the described computer algorithms, the evaluation was available after a few milliseconds. This time span may not sound like much. But it is crucial, especially in the intraoperative real-time setting, where immediate decisions must be made to prevent possible inner ear injury.

Regarding the visual analysis, the agreement of the three examiners was very good, especially in the frequency range between 500 and 1,000 Hz. Disagreements occurred mainly in borderline cases with low SNRs (another reason why the

SNR needs to be improved, if possible). The agreement among the experts was still substantial, but lower 250 Hz and for the two highest frequencies (i.e., 1,500 and 2,000 Hz). At 250 Hz, among all measured intracochlear ECochG, the SNR was the lowest (also refer to Table 4) (8). For the two highest frequencies, in some cases, it was difficult to distinguish between natural signal fluctuations and reproducible CM/DIF signal components.

It is important to note that a low SNR can affect the waveform morphology. In our data, e.g., CON and RAR responses did not evolve in opposite directions to each other, or there was a change in the usual morphology (e.g., the characteristic frequency of the CM/DIF signal was too low, or the ECochG traces had an irregular shape). This resulted in one examiner detecting a CM/DIF response while the other detected only noise. In our analysis, we found that the overall agreement was high, but one expert was rather cautious and another more tolerant in his assessment. This issue can be addressed by using an automated, quantitative and objective evaluation method, as suggested by our study. This allows for a uniform evaluation of the signals, which simplifies the comparison between individuals and different implantation centers or even makes it possible in the first place.

The analysis of the three stimulation levels showed that supra-threshold stimulation most frequently elicited a visually present CM/DIF signal. In addition, the SNR (except at 250 Hz) was substantially higher compared to the near- and sub-threshold levels. With supra-threshold stimulation, in our cohort, for the frequencies 500 Hz and above, a clear CM/DIF response was detected in 53.3% of cases. This implies that in a significant proportion of cases, no clear response could be detected. Additionally, this is despite the fact that most of the measurements took place in a postoperative setting and patients had a measurable residual hearing on the day of examination. However, it should be noted that the PTA of our study population shows a large variance and was, in some individuals, above 90 dB (compare Table 1). Consequently, the stimulus intensity was not always equally above the hearing threshold. In addition, recordings were measured from 4 different electrodes. For many subjects, ECochG responses were not visible at all electrodes. In literature, the situation regarding the prevalence of CM/DIF responses when stimulating above the hearing threshold is controversial. While some authors have found a close correlation between hearing threshold and CM/DIF signal threshold (11), other scientists have not found a clear relationship (1, 2, 8, 9, 12, 22, 23). Based on our data (refer to Table 4), we must assume that this correlation is both level- and frequency-dependent. For near-threshold and sub-threshold simulations, we detected significantly fewer visually detectable ECochG signals. Interestingly, the sub-threshold stimulation also showed CM/DIF responses in some cases (9, 23). Especially at 2,000 Hz, this finding was more pronounced.

4.3. Comparison of the objectification methods

In our study, DL with CNN AlexNet on time-frequency scalogram plots using CWT showed the best discrimination performance. The advantage of this method is that the morphology of the electrophysiological signal is taken into account. Similar to visual inspection, our algorithm was able to identify the CM/DIF response in the time-frequency scalograms shown in Figure 3. Another advantage of DL is its independence from preprocessing steps of ECochG signals (e.g., filtering). We trained our network with both, filtered and unfiltered data and could observe an almost identical accuracy of 90%.

Hotelling's T^2 test showed the highest sensitivity of our tested algorithms. This high sensitivity is also known from other research (27–29). However, in order to achieve good results with the Hotelling T^2 method, the signal must be free of artifacts and baseline wander. Both signal phenomena occur in ECochG measurements and must be addressed by using preprocessing steps. Furthermore, an optimal length of the TVMs must be defined. This is a trade-off; if the TVMs are too long, they contain the natural fluctuation of the ECochG signal (e.g., peaks and valleys). This results in TVMs with zero amplitude (similar to noise). If the TVMs are too short, the robustness and thus the test sensitivity decreases (overfitting) (29, 39).

Finally, the correlation analysis gave good objectivity to our data, although it did not reach the performance of the other two methods. It should be noted that signal artifacts can also have a high correlation and thus reduce the accuracy of this method. Such artifacts arise, e.g., from stitching or other unwanted effects (25). To overcome this, one could try to eliminate artifacts with more elaborate techniques or correlate only segments that are not affected.

In summary, the DL algorithm and Hotelling's T^2 test are very well suited for the objective assessment of ECochG signals; we achieved a high accuracy with both approaches. By using one of these methods, we can evaluate CM/DIF signals independently of the expertise of the examiner. In this article, we focused on the methodology itself with the question of whether a CM/DIF response was present or not. In the next step, further calculations could be included. For example, the evolution of amplitude or latency during electrode insertion. Furthermore, the advantages of the methodology are the immediate result as well as the reproducibility, which allows the comparison i) between individuals, ii) between different implant centers as well as iii) of longitudinal data. Finally, an automated ECochG assessment tool would pave the way for future standardized and widespread use in the clinical setting.

4.4. Limitations

Our data set was limited to 4133 ECochG recordings. Additional signals would further improve the methodology, increase the generalization of our models and reduce overfitting. Moreover, the data were visually reviewed by three experts. If more experts were incorporated into the algorithm, this may also refine the evaluation. Systemic noise can hamper the use of objective algorithms. In particular, the correlation analysis and Hotelling's T^2 test were found to be vulnerable. The DL method on the other hand was less dependent on data preprocessing and less sensitive to noise interference.

We have applied our methodology only when the electrode position was stable. In the next step, the objectification methods must also be tested during insertion, i.e., when the electrode is in motion. Furthermore, in the current study, we restricted ourselves to the CM/DIF signal. However, the methodology could also be used for the other signal components (i.e., ANN/SUM, CAP, SP). The combination of different data features is also advisable (4, 53) and must be evaluated in a future study.

5. Conclusion

Objectification of ECochG signals is possible with the methods presented in this paper. Our DL algorithm and Hotelling's T^2 test achieved a high accuracy to detect CM/DIF responses that had previously been identified by three ECochG experts. Objective automatic analysis of CM/DIF signals enables standardized, fast, accurate, and examiner-independent evaluation of ECochG measurements.

Data availability statement

The raw data supporting the conclusions of this article will be provided by the authors upon request.

Ethics statement

The studies involving human participants were reviewed and approved by the Cantonal Ethics Committee of Bern (BASEC ID 2019-01578). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

KS performed the measurement, wrote the software and article, and labeled and analyzed the data. WW analyzed the data and provided interpretive analysis and critical revision. AD labeled the data. CR, MC, and GM provided interpretive analysis and critical revision. SW designed the experiment, analyzed the data, labeled the data, and provided interpretive analysis and critical revision. All authors contributed to the article and approved the submitted version.

Funding

This study was partly funded by the Department of Otorhinolaryngology, Head and Neck Surgery at the Inselspital Bern, the Clinical trials unit (CTU) research grant, and the MED-EL company. GM was supported by the Swiss National Science Foundation #320030_173081. The authors declare that this study received funding from MED-EL Germany. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Acknowledgments

The authors would like to thank Marek Polak and his team from MED-EL, Austria, for their support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2022.943816/full#supplementary-material

References

1. Campbell L, Kaicer A, Sly D, Iseli C, Wei B, Briggs R, et al. Intraoperative real-Time cochlear response telemetry predicts hearing preservation in cochlear implantation. *Otol Neurotol.* (2016) 37:332–8. doi: 10.1097/MAO.0000000000000972

2. Dalbert A, Pfiffner F, Hoesli M, Koka K, Veraguth D, Roosli C, et al. Assessment of cochlear function during cochlear implantation by extra- and intracochlear electrocochleography. *Front Neurosci.* (2018) 12:18. doi: 10.3389/fnins.2018.00018

3. Weder S, Bester C, Collins A, Shaul C, Briggs RJ, O'Leary S. Toward a better understanding of electrocochleography: analysis of real-time recordings. *Ear Hear.* (2020) 41:1560–7. doi: 10.1097/AUD.0000000000000871

4. Weder S, Bester C, Collins A, Shaul C, Briggs RJ, O'Leary S. Real Time monitoring during cochlear implantation: increasing the accuracy of predicting residual hearing outcomes. *Otol Neurotol.* (2021) 42:E1030–6. doi: 10.1097/MAO.0000000000003177

5. Bester C, Weder S, Collins A, Dragovic A, Brody K, Hampson A, et al. Cochlear microphonic latency predicts outer hair cell function in animal models and clinical populations. *Hear Res.* (2020) 398:108094. doi: 10.1016/j.heares.2020.108094

6. Bester C, Collins A, Razmovski T, Weder S, Briggs RJ, Wei B, et al. Electrocochleography triggered intervention successfully preserves residual hearing during cochlear implantation: results of a randomised clinical trial. *Hear Res.* (2021) 20:108353. doi: 10.1016/j.heares.2021.108353

7. Sijgers L, Pfiffner F, Grosse J, Dillier N, Koka K, Röösli C, et al. Simultaneous intra- and extracochlear electrocochleography during cochlear implantation to enhance response interpretation. *Trends Hear.* (2021) 25: 2331216521990594. doi: 10.1177/2331216521990594

8. Haumann S, Imsiecke M, Bauernfeind G, Büchner A, Helmstaedter V, Lenarz T, et al. Monitoring of the inner ear function during and after cochlear implant insertion using electrocochleography. *Trends Hear.* (2019) 23:2331216519833567. doi: 10.1177/2331216519833567

9. Dalbert A, Pfiffner F, Röösli C, Thoele K, Sim JH, Gerig R, et al. Extra- and intracochlear electrocochleography in cochlear implant recipients. *Audiol Neurootol.* (2015) 20:339–48. doi: 10.1159/000438742

10. Schuerch K, Waser M, Mantokoudis G, Anschuetz L, Wimmer W, Caversaccio M, et al. Performing intracochlear electrocochleography during cochlear implantation. *J Vis Exp).* (2022) 8:e63153. doi: 10.3791/63153

11. Koka K, Saoji AA, Litvak LM. Electrocochleography in cochlear implant recipients with residual hearing: comparison with audiometric thresholds. *Ear Hear.* (2017) 38:e161–7. doi: 10.1097/AUD.0000000000000385

12. Campbell L, Kaicer A, Briggs R, O'Leary S. Cochlear response telemetry: intracochlear electrocochleography via cochlear implant neural response telemetry pilot study results. *Otol Neurotol.* (2015) 36:399–405. doi: 10.1097/MAO.0000000000000678

13. Yin LX, Barnes JH, Saoji AA, Carlson ML. Clinical utility of intraoperative electrocochleography (ECochG) during cochlear implantation: a systematic review and quantitative analysis. *Otol Neurotol.* (2021) 42:363–71. doi: 10.1097/MAO.0000000000002996

14. Schuerch K, Waser M, Mantokoudis G, Anschuetz L, Caversaccio M, Wimmer W, et al. Increasing the reliability of real-time electrocochleography during cochlear implantation: a standardized guideline. *Eur Arch Otorhinolaryngol.* (2022) 1:1–11. doi: 10.1007/s00405-021-07204-7

15. Dallos P, Cheatham MA, Ferraro J. Cochlear mechanics, nonlinearities, and cochlear potentials. *J Acoust Soc Am.* (2005) 55:597. doi: 10.1121/1.1914570

16. Snyder RL, Schreiner CE. The auditory neurophonic: basic properties. *Hear Res.* (1984) 15:261–80. doi: 10.1016/0378-5955(84)90033-9

17. Forgues M, Koehn HA, Dunnon AK, Pulver SH, Buchman CA, Adunka OF, et al. Distinguishing hair cell from neural potentials recorded at the round window. *J Neurophysiol.* (2014) 111:580–93. doi: 10.1152/jn.00446.2013

18. Fitzpatrick DC, Campbell AT, Choudhury B, Dillon MP, Forgues M, Buchman CA, et al. Round window electrocochleography just before cochlear implantation: relationship to word recognition outcomes in adults. *Otol Neurotol.* (2014) 35:64–71. doi: 10.1097/MAO.0000000000000219

19. Kim JS, Tejani VD, Abbas PJ, Brown CJ. Postoperative electrocochleography from hybrid cochlear implant users: an alternative analysis procedure. *Hear Res.* (2018) 370:304–15. doi: 10.1016/j.heares.2018.10.016

20. Polak M, Lorens A, Walkowiak A, Furmanek M, Skarzynski PH, Skarzynski H. *In vivo* basilar membrane time delays in humans. *Brain Sci.* (2022) 12:400. doi: 10.3390/brainsci12030400

21. Lorens A, Walkowiak A, Polak M, Kowalczuk A, Furmanek M, Skarzynski H, et al. Cochlear microphonics in hearing preservation cochlear implantees. *J Int Adv Otol.* (2019) 15:345. doi: 10.5152/iao.2019.6334

22. Imsiecke M, Büchner A, Lenarz T, Nogueira W. Psychoacoustic and electrophysiological electric-acoustic interaction effects in cochlear implant users with ipsilateral residual hearing. *Hear Res.* (2020) 386:107873. doi: 10.1016/j.heares.2019.107873

23. Krüger B, Büchner A, Lenarz T, Nogueira W. Amplitude growth of intracochlear electrocochleography in cochlear implant users with residual hearing. *J Acoust Soc Am.* (2020) 147:1147–62. doi: 10.1121/10.0000744

24. Krüger B, Büchner A, Lenarz T, Nogueira W. Electric-acoustic interaction measurements in cochlear-implant users with ipsilateral residual hearing using electrocochleography. *J Acoust Soc Am.* (2020) 147:350. doi: 10.1121/10.0000577

25. Arnold SA. Objective versus visual detection of the auditory brain stem response. *Ear Hear.* (1985) 6:144–50. doi: 10.1097/00003446-198505000-00004

26. Wang H, Li B, Lu Y, Han K, Sheng H, Zhou J, et al. Real-time threshold determination of auditory brainstem responses by cross-correlation analysis. *iScience.* (2021) 24:103285. doi: 10.1016/j.isci.2021.103285

27. Golding M, Dillon H, Seymour J, Carter L. The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. doi: 10.3109/14992020903140928

28. Dun BV, Carter L, Dillon H. Sensitivity of cortical auditory evoked potential detection for hearing-impaired infants in response to short speech sounds. *Audiol Res.* (2012) 2:e13. doi: 10.4081/audiores.2012.e13

29. Chesnaye MA, Bell SL, Harte JM, Simonsen LB, Visram AS, Stone MA, et al. Efficient detection of cortical auditory evoked potentials in adults using bootstrapped methods. *Ear Hear.* (2020) 42:574–83. doi: 10.1097/AUD.0000000000000959

30. Sodmann P, Vollmer M, Nath N, Kaderali L. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms. *Physiol Meas.* (2018) 39:104005. doi: 10.1088/1361-6579/aae304

31. Xiong Z, Nash MP, Cheng E, Fedorov VV, Stiles MK, Zhao J. ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network. *Physiol Meas.* (2018) 39:094006. doi: 10.1088/1361-6579/aad9ed

32. Mashrur FR, Roy AD, Saha DK. Automatic identification of arrhythmia from ECG using AlexNet convolutional neural network. In: *2019 4th International Conference on Electrical Information and Communication Technology, EICT.* Khulna (2019). doi: 10.1109/EICT48899.2019.9068806

33. Aqil M, Jbari A. *Continuous Wavelet Analysis and Extraction of ECG Features.* Springer Nature (2021).

34. Davila CE, Mobin MS. Weighted averaging of evoked potentials. *IEEE Trans Biomed Eng.* (1992) 39:338–45. doi: 10.1109/10.126606

35. Kumaragamage CL, Lithgow BJ, Moussavi ZK. Investigation of a new weighted averaging method to improve SNR of electrocochleography recordings. *IEEE Trans Biomed Eng.* (2016) 63:340–7. doi: 10.1109/TBME.2015.2457412

36. Drongelen WV. *Signal Processing For Neuroscientists.* Elsevier (2018). doi: 10.1016/C2010-0-65662-8

37. Labelbox. Labelbox (2022). Available online at: https://labelbox.com

38. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* (1971) 76:378–82. doi: 10.1037/h0031619

39. Hotelling H. The generalization of student's ratio. In: Kotz S, Johnson NL, editors. *Breakthroughs in Statistics. Springer Series in Statistics.* New York, NY: Springer (1992). p. 54–65.

40. Rencher AC. Methods of multivariate analysis. *Methods Mult Anal.* (2002) 2:0471271357. doi: 10.1002/0471271357

41. Rossum GV, Drake FL. *Python 3 Reference Manual.* Scotts Valley, CA: CreateSpace (2009). doi: 10.5555/1593511

42. Pataky T. *spm1d.* (2021). Available online at: https://spm1d.org/.

43. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM.* (2017) 60:84–90. doi: 10.1145/3065386

44. Lee GR, Gommers R, Waselewski F, Wohlfahrt K, O'Leary A. PyWavelets: a python package for wavelet analysis. *J Open Source Softw.* (2019) 4:1237. doi: 10.21105/joss.01237

45. Mallat S. A wavelet tour of signal processing (2009). doi: 10.1016/978-0-12-374370-1.X0001-8

46. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Vancouver, BC: Curran Associates, Inc. (2019). doi: 10.5555/3454287.3455008

47. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE (2009). p. 248–55.

48. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team (2018).

49. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. (2011) 12:77. doi: 10.1186/1471-2105-12-77

50. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. (1977) 33:174. doi: 10.2307/2529310

51. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: John Wiley & Sons (2013). p. 1–510.

52. Dalbert A, Sijgers L, Grosse J, Veraguth D, Roosli C, Huber A, et al. Simultaneous intra- and extracochlear electrocochleography during electrode insertion. *Ear Hear*. (2021) 42:414–24. doi: 10.1097/AUD.0000000000000935

53. Fontenot TE, Giardina CK, Fitzpatrick DC. A model-based approach for separating the cochlear microphonic from the auditory nerve neurophonic in the ongoing response using electrocochleography. *Front Neurosci*. (2017) 11:592. doi: 10.3389/fnins.2017.00592

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fneur.2022.959582

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Raul Sanchez-Lopez,
University of Nottingham, United Kingdom
Jeppe Høy Christensen,
Eriksholm Research Centre, Denmark

*CORRESPONDENCE
Samira Saak
samira.saak@uni-oldenburg.de

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 01 June 2022
ACCEPTED 11 August 2022
PUBLISHED 15 September 2022

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

A flexible data-driven audiological patient stratification method for deriving auditory profiles

Samira Saak[1,2]*, David Huelsmeier[1,2], Birger Kollmeier[1,2,3,4] and Mareike Buhl[1,2]

[1]Medical Physics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, [2]Cluster of Excellence Hearing4all, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, [3]Hörzentrum Oldenburg gGmbH, Oldenburg, Germany, [4]Hearing Speech and Audio Technology, Fraunhofer Institute of Digital Media Technology (IDMT), Oldenburg, Germany

For characterizing the complexity of hearing deficits, it is important to consider different aspects of auditory functioning in addition to the audiogram. For this purpose, extensive test batteries have been developed aiming to cover all relevant aspects as defined by experts or model assumptions. However, as the assessment time of physicians is limited, such test batteries are often not used in clinical practice. Instead, fewer measures are used, which vary across clinics. This study aimed at proposing a flexible data-driven approach for characterizing distinct patient groups (patient stratification into auditory profiles) based on one prototypical database ($N = 595$) containing audiogram data, loudness scaling, speech tests, and anamnesis questions. To further maintain the applicability of the auditory profiles in clinical routine, we built random forest classification models based on a reduced set of audiological measures which are often available in clinics. Different parameterizations regarding binarization strategy, cross-validation procedure, and evaluation metric were compared to determine the optimum classification model. Our data-driven approach, involving model-based clustering, resulted in a set of 13 patient groups, which serve as auditory profiles. The 13 auditory profiles separate patients within certain ranges across audiological measures and are audiologically plausible. Both a normal hearing profile and profiles with varying extents of hearing impairments are defined. Further, a random forest classification model with a combination of a one-vs.-all and one-vs.-one binarization strategy, 10-fold cross-validation, and the kappa evaluation metric was determined as the optimal model. With the selected model, patients can be classified into 12 of the 13 auditory profiles with adequate precision (*mean across profiles* = 0.9) and sensitivity (*mean across profiles* = 0.84). The proposed approach, consequently, allows generating of audiologically plausible and interpretable, data-driven clinical auditory profiles, providing an efficient way of characterizing hearing deficits, while maintaining clinical applicability. The method should by design be applicable to all audiological data sets from clinics or research, and in addition be flexible to summarize

information across databases by means of profiles, as well as to expand the approach toward aided measurements, fitting parameters, and further information from databases.

KEYWORDS

auditory profiles, precision audiology, data mining, machine learning, patient stratification, audiology

Introduction

It has become increasingly evident that characterizing hearing deficits by the audiogram alone is not enough. In addition to a loss of sensitivity, other factors, such as suprathreshold distortions, determine how well individuals can understand speech in daily life and communicate efficiently (1–5). However, it is yet an open issue which measures should be applied to achieve "precision audiology," i.e., to characterize the individual patient as completely and exactly as necessary without losing too much time on comparatively irrelevant measurements. Hence, a number of approaches were described in the literature that differ in their general purpose, their amount of measurements included, and their evaluation method to characterize the most relevant measures.

For instance, van Esch et al. (6) proposed a test battery ("*auditory profile*") for standardized audiological testing comprising eight domains (pure-tone audiometry, loudness perception, spectral and temporal resolution, speech perception in quiet and in noise, spatial hearing, cognitive abilities, listening effort, and self-reported disability and handicap) aiming to describe all major aspects of hearing impairment without introducing redundancy among measures. Similarly, the BEAR test battery was proposed for research purposes to characterize different dimensions of hearing and was evaluated with patients with symmetric sensorineural hearing loss (7). In spite of the benefit of the proposed test batteries, widespread adoption in clinical practice is currently lacking. The complete BEAR test battery, for instance, takes ∼2.5 h to complete (7), even though a shorter version for clinical purposes was also proposed in (8). Nevertheless, in clinical practice, time is short and the assessment of patients on such a multitude of tests may not be feasible.

To tackle time constraints, Gieseler et al. (9) aimed at determining clinically relevant predictors for unaided speech recognition from a large test battery, thus, reducing the amount of required tests. They showed that pure-tone audiometry, age, verbal intelligence, self-report measures of hearing loss (e.g., familial hearing loss), loudness scaling at 4 kHz, and an overall physical health score were most important in predicting unaided speech recognition, with the pure-tone audiometry serving as the best predictor. Their model, however, left 38% of the variance in predicting unaided speech recognition unexplained, indicating that further measures may be related to unaided speech recognition. At the same time, their analyses were tailored toward explaining unaided speech recognition performance as an outcome measure. Predictors for aided speech recognition performance, in contrast, or other outcome measures, may vary. In Lopez-Poveda et al. (10), for instance, temporal processing deficits as measured by the frequency-modulation detection threshold (FMDT) were shown to be most relevant in predicting aided speech recognition performance. When including only predictors available in clinical situations, however, the unaided speech recognition threshold (SRT) in quiet was determined to be the best predictor. This demonstrates the discrepancy between research and clinical applications and highlights the importance to analyze insights from both clinical and research datasets in combination. It further shows that the relevance of predictors depends on the outcome measures, as different predictors were determined most relevant for unaided and aided speech recognition.

To improve patient characterization in the field of audiology, patient data, therefore, need to be summarized efficiently and flexibly. By summarizing patient data flexibly, the generated knowledge could be used in a variety of settings (e.g., in clinics, for mobile assessments, and decision-support systems in general), and for a variety of outcome measures (e.g., diagnostic outcomes or unaided and aided speech recognition performance). This, however, poses several challenges. First, patients need to be characterized across different dimensions of hearing loss. Second, to gain insights from a diverse patient population, data aggregation across databases is required, which, however, is hindered by the heterogeneity in the applied measures across clinical and research databases in the field of audiology (11). Lastly, for the general applicability of the stored information, it needs to be accessible *via* measures also applied in clinical settings, such that physicians can be supported.

To tackle these challenges, different approaches toward patient stratification exist that involve identifying subgroups in patient populations based on measurement data from single measures or from interrelations of measures. An example of a data-driven stratification based on single measures is the Bisgaard standard audiograms by (12). There, a set of 10 standard audiogram patterns occurring in clinical practice

were defined. This has subsequently resulted in a variety of studies investigating outcome measures such as aided SRTs in relation to the 10 audiograms [(9, 13–15), to name a few], aiming toward precision audiology, thus, demonstrating the promising nature of finding sub-classes in the field of audiology. In contrast, an expert-based approach, based on single measures, was proposed by Dubno et al. (16) that linked four audiometric phenotypes to knowledge about possible etiologies from animal models of presbyacusis *via* expert decisions. Schematic boundaries for the five phenotypes "older-normal," "pre-metabolic," "metabolic," "sensory," and "metabolic+sensory" are provided which allow for inferences of etiologies, given patient presentations of presbyacusis.

In contrast to patient stratification based on single measures, Sanchez-Lopez et al. (17) introduced a data-driven profiling method based on multiple measures using a combination of unsupervised and supervised machine learning. Based on the hypothesis that two distortion types for the characterization of hearing loss exist, four distinct profiles were generated by means of principal component analysis and archetypal analysis. Thereby, the most important variables for the characterization of each distortion dimension were estimated and employed to identify the most extreme data combinations (archetypes). All patients of two existing research data sets (containing a certain battery of tests) were labeled with the most similar archetype. In a second step, decision trees were built to allow for the classification of new patients into the four auditory profiles. The obtained profiles are interpretable as they were defined based on the hypothesis of two distortion components and the variables used for classification are known. The meaning of the two distortions, however, was different depending on the available measures in the respective data set.

Sanchez-Lopez et al. (18) improved the profiling method to be more robust (e.g., due to bootstrapping, a more flexible number of allowed variables, and estimating the association of a patient to a profile based on probability) and applied it to the BEAR test battery (7), which was designed for the purpose of including all relevant measures according to the literature and previous work. As a result, a plausible interpretation of the two distortion dimensions was obtained, namely being associated with speech intelligibility and loudness perception, respectively (18). However, by tailoring their analyses toward four extreme distinct profiles and by using archetypal analysis, a priori hypotheses were included in the derivation of the profiles. Consequently, further distinctions between patient groups may be lost.

A further example of summarizing audiological data efficiently is provided by Buhl et al. (11, 19). The Common Audiological Functional Parameters (CAFPAs) were derived by experts and aim at representing audiological functions in an abstract and measurement-independent way. The CAFPAs further act as an interpretable intermediate layer in a clinical decision-support system. Prediction models allow for a data-driven prediction of CAFPAs (20) and a subsequent classification into audiological findings (21). However, to relate new measures from further data sets to the CAFPAs, experts are currently required for labeling purposes, which consequently does not allow for the automatic integration of new data sets containing additional measures.

The aforementioned methods all contribute toward enhancing patient characterization but are either restricted to single measures or include prior assumptions regarding the distinction of patient groups or audiological functions. Consequently, not all existent differences between patient groups may be detected. In this study, we aim at (1) providing a method for a fully data-driven stratification of patients into subgroups based on audiological measures, namely *auditory profiles*. This patient stratification approach is not restricted in terms of prior assumptions, the number of patient groups, and contained measures. In that way, all differences between patient groups can be summarized independently of outcome measures. The auditory profiles aim to describe patient groups with similar measurement ranges across audiological measures and are defined based on the contained patient patterns, instead of prior assumptions. In future, profiles could, hence, be combined, added, or removed, depending on the provided insights gained from applying the profiling approach to further data sets, as well as based on the relevance of profile distinctions in clinical routine. The applicability of defined profiles to different settings (e.g., clinical settings) can, however, only be obtained if the knowledge from within the profiles, in the form of plausible ranges for the contained measures, can be linked to patients, given their results on widely used measures (e.g., pure-tone and speech audiometry). We, therefore, further aim at (2) maintaining clinical applicability by building classification models using random forests, based on measures available in clinical routine. This allows for classifying new patients into the auditory profiles. In clinics, it could support physicians to associate a new patient to a profile and in that way exploit statistical knowledge available for the respective profile.

The current study, thus, aims at answering the following two research questions:

RQ1: Does our proposed profiling approach result in a meaningful and distinct grouping (auditory profiles) of patients with respect to important hearing loss factors contained in the employed data set?

RQ2: Which classification model can provide high precision and sensitivity in classifying patients into the auditory profiles using only a subset of the contained audiological measures?

Materials and methods

Data set

To define the first set of auditory profiles, we analyzed an existing data set that was provided by Hörzentrum

Oldenburg gGmbH and is described in detail in Gieseler et al. (9). In contrast to Gieseler et al. (9), we did not exclude any patients with, e.g., an air-bone gap >10 dB HL but aimed for a diverse patient sample. Our patient sample, consequently, consisted of all patients that completed the full test battery, resulting in 595 patients (mean age = 67.6, SD = 11.9, female = 44%) with normal to impaired hearing. For each patient, information with respect to a broad range of measures, including audiogram data, loudness scaling, speech tests, cognitive measures, and anamnesis questions is contained.

The contained measures either are, or can easily be integrated into clinical routine. The audiogram and the Göttingen sentence test (GOESA) (22) are commonly used for the assessment of individuals' hearing status. The former assesses an individual's thresholds across frequencies; the latter assesses the speech recognition threshold (SRT), here, in noise for the collocated condition (S0N0). Both the audiogram and the GOESA are used in hearing aid fitting, for gain adjustments, and as an outcome measure, respectively. From the contained measures, we used several features to generate the auditory profiles (see Table 1 for an overview of the features). For the audiogram, the pure-tone average (PTA, threshold averaged across 0.5, 1, 2, and 4 kHz) for air-, and bone conduction was used for the more severely affected ear. Asymmetric hearing loss was accounted for *via* the inclusion of an asymmetry score (absolute difference between PTA of left and right ear). Additionally, the air–bone gap (ABG), the PTA of the Uncomfortable Loudness Level (UCL), and the Bisgaard standard audiograms (12) were derived from the audiogram. The Bisgaard standard audiograms were included to allow for a separation of different audiogram patterns (e.g., moderately and steeply sloping audiograms), while reducing the dimensionality of the audiogram. A further speech test [digit triplet test (DTT)] (23) was included to add information to the auditory profiles from a measure mainly used for screening purposes. The adaptive categorical loudness scaling (ACALOS) (24) provides relevant information with respect to an individual's loudness perception and recruitment, and has also shown its effectiveness in hearing aid fitting (25). To characterize both the lower and upper part of the loudness curves, both L15, L35, and the difference between L15 and L35 were selected as features. As a relation between cognition and hearing exists (26), the age-normed sum score from a screening test for dementia (Demtect) (27) and the raw score from a measure of verbal intelligence [Vocabulary test (WST)] (28) were also included. Further, information regarding the socio-economic status (sum score of education, income, and occupation) (29), the presence of tinnitus [none (1), unilateral (2), bilateral (3)], and the age of the patients were available.

TABLE 1 Overview of audiological domains and features used for the generation of the profiles.

Domain	Number of features	Features
Audiogram	6	**AC PTA**, BC PTA, **Asymmetry (left/right ear)**, ABG, UCL PTA, **Bisgaard standard audiograms**
Loudness Scaling	6	**ACALOS (L15,L35, L15-L35) for 1.5 & 4 kHz**
Speech tests	3	**GOESA (SRT, slope)**, DTT (SRT)
Cognitive measures	2	DemTect score, WST score
Anamnesis	3	Tinnitus, Socio-economic status, **age**

Features used for the classification into the profiles are shown in **bold**.

Generating auditory profiles using model-based clustering

To generate auditory profiles that are capable of separating patients with respect to ranges of audiological tests, we applied clustering, as it has shown promising for purposes of patient stratification. For the current analyses, the clustering pipeline consists of two steps, namely robust learning and profile generation (see Figure 1 for visualization).

Robust learning

Bootstrapping and imputation of missing data

As bootstrapping techniques have shown to improve the robustness of clustering solutions (30, 31), we first subsampled the data set 1,000 times containing 95% of the original data set. We chose subsampling over resampling with replacement, in order to avoid duplicate samples being seen as a "mini"-cluster, hence, artificially increasing the number of clusters. As missing values existed in the original data set, each of the 1,000 subsamples also contained missing values and needed to be imputed. Missing values pose a common problem in clinical data sets, and a loss of patient information, e.g., complete-case analysis, is often undesirable, thus, requiring an adequate technique to solve it.

Consequently, for audiogram data, prior to extracting pure-tone averages and Bisgaard standard audiograms, missing thresholds were interpolated if the thresholds prior to and after missing values were available. For the remainder of missings (on average 1.5% with a maximum of 2.5%), multivariate imputations with chained equations (MICE) (32) was applied. MICE results in multiple completed data sets that account for the uncertainty that stems from imputing missings. With MICE, the

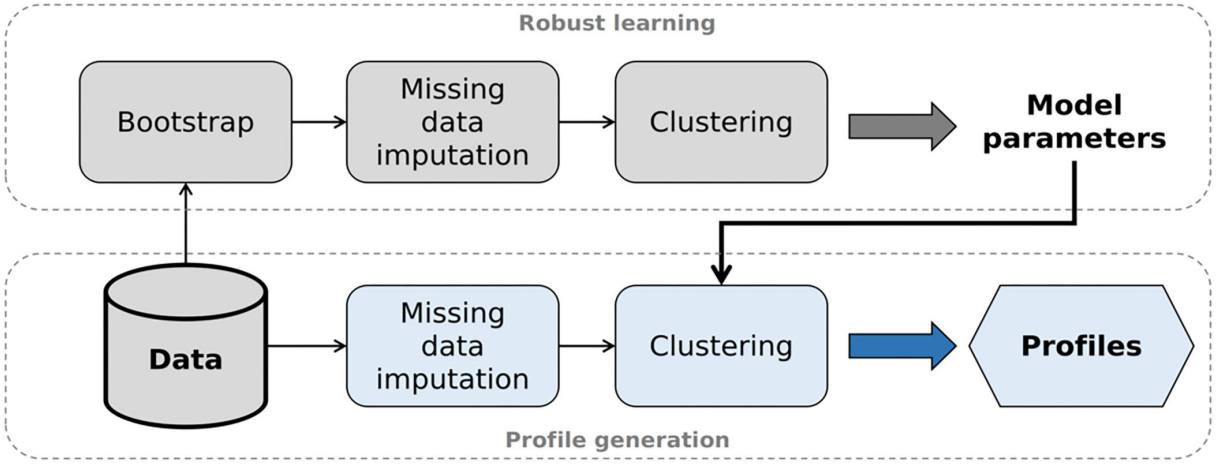


FIGURE 1
Analysis pipeline to generate auditory profiles. After selecting the optimal model parameters (robust learning, upper part), model-based clustering is applied to the original data set (profile generation, lower part).

analyses of interest are subsequently performed on all completed data sets and the results are combined (32). For the present analyses, we generated 20 completed data sets. Accordingly, clustering was performed on each of the 1,000 × 20 data sets.

Model-based clustering

Before clustering, we transformed the features of Bisgaard standard audiograms and tinnitus and treated them as continuous for clustering purposes. Bisgaard standard audiograms were ordered with respect to increasing PTA; tinnitus with respect to its absence, unilateral, or bilateral presence. All features (see Table 1) were then scaled using min–max scaling, resulting in values between 0 and 1. As the number of features ($N = 20$) can be considered small, we refrained from further dimensionality reduction and instead aimed at maintaining a balance of the number of features stemming from the different measures. Depending on the clustering goal, dimension reduction with, e.g., principal component analysis can prove problematic as the reduction of dimensionality could also lead to the removal of information that would have proved to be discriminatory for the clustering goal (33).

On the scaled feature set, we applied model-based clustering. Model-based clustering was especially suitable for our purposes of uncovering patient groups existent in the data set, as it assumes that the data stem from a mixture of subgroups. The mixture of subgroups is further assumed to be generated by an underlying model which model-based clustering aims to recover (34, 35). For this purpose, the number of clusters k and a parameterization of the covariance matrices with respect to their shape, size, and orientation [see (36) for possible covariance parameterizations] need to be specified beforehand. Subsequently, each cluster's mean vector mu_k and covariance matrix Σk is learned and a likelihood estimate for the given clustering solution is computed.

In contrast to simpler clustering techniques such as k-means clustering, model-based clustering is able to detect more complex shapes in the data (37). It is, therefore, more suitable for our purposes of detecting all plausible differences in the data. At the same time, the parameterization of the covariance matrices can constrain the complexity of the clustering solution by enforcing stronger restrictions and reducing the number of parameters that need to be estimated (38). To select the most suitable model, all candidate parameterizations (k and covariance matrix parameterization) are computed and the model with the highest likelihood of explaining the underlying data structure is selected using the Bayesian Information Criterion (BIC) (39). More complex clustering structures (i.e., less covariance matrix restrictions) may suffice in explaining the dataset with fewer clusters but require the estimation of a much larger number of parameters and are, thus, not always feasible with smaller datasets. Less complex clustering structures, in contrast, could explain the same underlying data structure by increasing the number of clusters (38). This also holds for increasing the number of features used for clustering. Increasing the number of features increases the number of parameters to be estimated (i.e., the complexity), which, however, can be reduced by restraining the covariance matrices. This may, in turn, increase the number of estimated clusters required to explain the data. To avoid increasing the number of clusters beyond clusters that enhance the explanation of the data structure, however, the BIC penalizes for the complexity of the covariance parameterization and number of clusters k, and

thus, results in a trade-off between model complexity and over-parameterization (34).

Here, for each of the 1,000 × 20 data sets, we computed all potential parameterizations for 2–30 clusters and then derived the optimal model for each data set using the BIC, which resulted in 1,000 × 20 candidate models. The dimensionality of the candidate models was then reduced across the 20 completed data sets of each of the 1,000 subsamples. The most frequently occurring model parameterization was selected as a candidate model, resulting in a reduced set of 1,000 candidate models. We then defined the overall optimal model *via* its frequency across the 1,000 candidate models, which resulted in an estimate for the model parameters (i.e., the number of profiles and the model's covariance parameterization).

Profile generation

In the profile generation step, we generated the auditory profiles using the original data set without prior subsampling. First, we imputed missings using multivariate imputation with chained equations (MICE) in the same manner as described in Section Bootstrapping and imputation of missing data. Thus, 20 completed data sets were generated with differing estimates for missings. Second, we applied model-based clustering using the estimated optimal model structure from the robust learning step for each completed data set, which resulted in 20 candidate clustering solutions. From these 20 candidate clustering solutions, we aimed to select the solution showing the highest overlap with the remaining solutions regarding patient allocation into the clusters. The rationale behind this is that, since model parameters are kept constant, differences between clustering solutions stem from differences in the imputed values. The solution showing the highest overlap can then be assumed to be least influenced by imputed values, as patient allocations into the clusters were agreed upon by most solutions.

Building classification models to classify patients into auditory profiles

Features and labels

To allow for the usage of the auditory profiles for different purposes (e.g., clinical applications), it is necessary to classify patients into the profiles based on a subset of measures widely available. Therefore, we built classification models using the profiles as labels and a reduced set of measures as features. From the aforementioned features used for clustering (see Table 1), only the features from ACALOS, GOESA, and the air-conduction audiogram (PTA, Asym PTA, Bisgaard) were used next to the age of the patients (12 features), to simulate the case that these measures were conducted for a to-be-classified patient.

Model training

For model training, we split the reduced data set, containing the above-mentioned 12 features, into a training (75% of patients) and test data set (25% of patients). The training data set was used for training the model, which included cross-validation (CV), model tuning, and the selection of the best model tuning parameters containing different binarization strategies, CV procedures, and evaluation metrics defining the prediction error, and are described in more detail in the following. The best model is defined as the model minimizing prediction error. We then evaluated the training data set's best model on the test data set to estimate its predictive performance on patient cases not used for model training, which indicates how the classification model would generalize on unseen patient cases.

To build the classification models on the training data set, we used random forests (40), as it has shown competitive classification performance, while remaining interpretable. It is also less prone to overfitting and handles relatively small sample sizes well (41, 42). Random forests are an extension of simple decision trees. Multiple decision trees are built, each segmenting the predictor space into several smaller regions, based on derived decision rules. Predictions are consequently derived from the ensemble of trees. For classification purposes, the label predicted most frequently among trees is selected. In other words, it has the highest estimated probability among candidate labels. To avoid building correlated trees, the tuning parameter *mtry* defines the number of features considered at each split. At each split, the specified number of features is then randomly sampled from the feature set, thus, enforcing different tree structures, which in turn reduce the variance of the predictions (41). For the current analyses, we tuned *mtry* using cross-validation.

To provide optimal prediction models for each of the profiles, we applied different binarization techniques. Binarization strategies to tackle multi-class problems have proved beneficial in enhancing predictive performance. They involve building base learners for binary classification tasks which are subsequently aggregated to provide a prediction (43, 44).

Consequently, we compared multi-class classification to three different binarization strategies. First, we built predictive models for each auditory profile separately (k models), with the one-vs.-all (OVA) technique, allowing the model to learn the specific differences of a profile, as compared to all remaining ones. Thus, for each profile, we built a classification model that decides whether a patient belongs to a given profile, or not. If more than one of the k OVA models predicted that a patient belonged to its profile, the profile with the highest probability among candidate profiles is selected, as defined by the frequency of its prediction in the random forest. Second, we used a one-vs.-one (OVO) technique to build predictive models for all $k(k\text{-}1)/2$ profile combinations. Thus, differences between each pair of profiles were learned. To provide a prediction, voting aggregation was applied, which means that the most frequently

predicted profile was selected. Lastly, we used a combination of OVA and OVO (OVAOVO). Here, again, we used OVA to predict profile classes. However, for uncertain cases, if more than one profile was predicted, instead of selecting the profile with the higher probability, we used OVO to decide upon the final profile prediction.

Across profiles, a class imbalance exists, either due to differing profile sizes or due to the applied binarization strategy. Classifiers trained on imbalanced data sets tend to favor the majority class over the minority class in order to reduce the prediction error, which leads to undesirable results if the minority class is of interest (e.g., in an OVA or OVO model). Consequently, we upsampled all profiles to contain at least the number of patients of the largest profile p in terms of sample size ($maxN_p$). Upsampled patients were selected randomly from each profile and across features Gaussian noise was added to the observations (+/- 1 *SD*). Upsampling with Gaussian noise was shown to be especially suitable for clinical data sets (45). As a result, no class imbalance was present for multi-class and OVO. For OVA, the class imbalance was still present due to the OVA design. As upsampling would require upsampling for several magnitudes of the original profile size, and downsampling would discard too much valuable information, a different technique was applied. In addition to upsampling to $maxN_p$, we used a weighted random forest model using cost-sensitive learning. Thus, weights were introduced, which more severely punished for the misclassification of the minority class over the majority class (46). The issue of the tendency toward majority predictions was, therefore, addressed also for the OVA binarization strategy.

Further, we compared two different CV schemes for optimal model tuning, namely, leave-one-out CV (LOOCV) and 10-fold CV repeated 10 times (RepCV). LOOCV is a special case of CV, in which the validation set consists of only one observation; RepCV splits the training set randomly into 10-folds, which is then repeated 10 times. LOOCV provides advantages for small data sets, as models are trained on larger sample size as compared to RepCV. However, in return, predictions may have high variance, as the variation in training sets is small. RepCV, in contrast, has lower variance due to differing training sets, but may be biased due to smaller sample size (41).

Lastly, we compared different evaluation metrics which optimize classifiers to different aspects of predictive performance. The main measures to evaluate the performance of a classifier are accuracy, sensitivity, specificity, and precision. Accuracy defines the ratio between correctly classified instances and the total sample size. Sensitivity (also called recall) and specificity are evaluation metrics for binary classification problems, but can be easily extended toward multi-class classification problems by employing an OVA binarization of the classification problem. This, however, again introduces an imbalance in the data regarding the evaluation. Sensitivity refers to correctly classifying all classes of interest as positive, whereas specificity refers to the ability to correctly classify all remaining classes as negative. The precision of a classifier, in contrast, determines the preciseness of a classifier. That means precision is high if no other class was misclassified as the class of interest (47). The four evaluation metrics we compared in the current study, namely, Cohen's kappa, balanced accuracy, F1-score, and the Area under the precision–recall curve (AUPRC) differently weight aspects of accuracy, sensitivity, specificity, and precision. Cohen's kappa is inherently capable of evaluating multi-class problems, by comparing the accuracy to the baseline accuracy obtained by chance (48). Balanced accuracy weights sensitivity with specificity, and is consequently less able to handle multi-class problems, since specificity increases with imbalanced data sets. The F1-score addresses this issue by calculating the harmonic mean between sensitivity and precision, instead of sensitivity and specificity. Likewise, the AUPRC has shown to be especially suitable for imbalanced data (49). To determine the optimal classifier, it is important to select an adequate evaluation metric, suitable for the class distribution in the data set. Since we have different class distributions across our four classification strategies (multi-class, OVA, OVO, OVAOVO), we compared different evaluation metrics.

Model selection and evaluation

To select the optimal classification model, we evaluated the four different classification strategies (multi-class, OVA, OVO, OVAOVO) on the training data set with respect to the different metrics (Kappa, balanced accuracy, F1-score, and AUPRC) and cross-validation procedures (repCV, LOOCV). To compare the performance of the models that were optimized with the different evaluation metrics, after training, a general *post-hoc* performance measure is needed. Here, we chose the F1-score as it summarizes both sensitivity and precision, and can adequately describe the performance of a classifier in case of imbalance. Accordingly, we determined the model leading to the highest F1-score by averaging the F1-scores across profiles and then selected it as the best performing classification model. Lastly, to evaluate the predictive performance of the selected classification model and its generalizability to new data, we evaluated the model on the test data set. Here, instead of the F1-score, we used both sensitivity and precision to provide a more thorough assessment of the classifiers' performance for the distinct auditory profiles.

Results

Generation of profiles

Estimation of profile number and covariance parameters

To generate auditory profiles which characterize a diverse range of patient patterns across measures, the number of separable patient groups and the covariance parameter were determined. Figure 2 depicts the distribution of estimated

cluster numbers across the 1,000 bootstrapped samples. Across bootstrapped samples, 11–19 profiles were estimated as an optimal model with a maximum of 13 clusters. Further, the covariance parameterization "VEI" was selected across all 1,000 subsamples. VEI (variable volume, equal shape, coordinate axes orientation) is a rather parsimonious model as it restricts both the shape and axis alignment of the clusters and requires a diagonal cluster distribution. The sizes of the clusters, however, may vary. Hence, 13 clusters with the covariance parameter "VEI" are estimated to represent the data structure best.

Subsequently, the above-defined parameterization ($k = 13$, "VEI") was used to generate profiles on all 20 completed data sets of the original data set. The completed data set showing the highest overlap with the remaining completed data sets regarding patient allocation into the profiles (*max_similarity* = 0.794) was selected to base the auditory profiles on. Mean classification similarity across all 20 completed data sets was 0.75 ($SD = 0.032$).

Profile ranges across audiological measures

Figure 3 shows the profile ranges of the generated auditory profiles and Table 2 contains the number of patients contained in each profile. The profiles cover a large range across audiological measures and show profile-based differences in patient presentation of the contained measures. All profiles can be distinguished from each other based on at least one audiological feature. The speech test results (Figure 3, blue box) regarding GOESA and the DTT are generally comparable. The profiles cover different extents of impairments, ranging from normal hearing (profile 1) to strong difficulties in understanding speech in noise (profile 13), as indicated by the increasing SRT. Likewise, the slope of the GOESA decreases with increasing SRT.

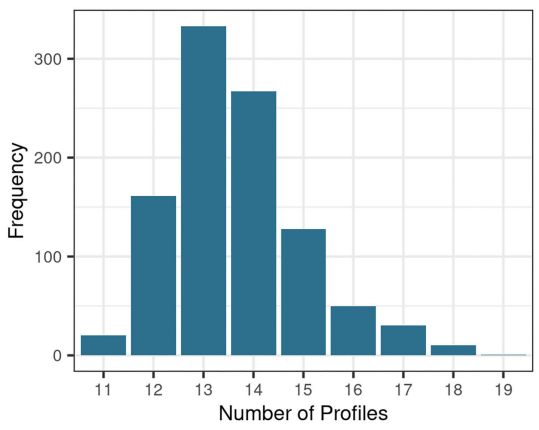


FIGURE 2
Distribution of optimal profile numbers across bootstrapped samples.

Within the SRT range of −5 to 0 dB SNR, most of the profiles are contained. Here, the different profiles show similarities regarding SRT ranges, and the difference between the profiles can be found *via* other measures. Audiogram results (Figure 3, green box) indicate the existence of normal hearing (profile 1), moderately (profiles 2, 3, 6, 7, 8, 9, 11, 13), and rather steeply sloping (profiles 4, 5, 10, 12) patterns. Generally, we observe a trend of increasing thresholds on the audiogram together with increasing SRTs. There are, however, also exceptions. Profile 11 displays the highest thresholds across frequencies and profiles, but does not show the strongest impairment on the GOESA. Instead, it includes patients with an air–bone gap and asymmetric hearing loss, as indicated by the asymmetry score. Profiles can also be distinguished based on the ACALOS (Figure 3, loudness scaling—yellow box) and the UCL. With increasing SRTs, we can observe an increase in the UCL, as well as a decrease in the dynamic range, as shown by the difference between L35 and L15 for both 1.5 and 4 kHz. In spite of this, differences exist across profiles unrelated to the increasing SRT. Profiles 4 and 5, for instance, show overlapping ranges regarding the SRT, but differ with respect to the UCL. Across cognitive measures (Figure 3, cognitive measures—orange box), no clear distinctions across profiles were found. Likewise, ranges for the age of patients and the socio-economic status (Figure 3, anamnesis—gray box) overlap across profiles, with the exception of profile 1 containing younger patients.

To summarize, similarities exist to varying extents between profiles. Some profiles can be easily distinguished. For instance, profiles 1 and 2 can be easily distinguished from profiles 11, 12, and 13 across audiogram, GOESA, and loudness scaling data. In contrast, other profiles only differ on certain measures. Profiles 2 and 3, for instance, show overlapping ranges on both the audiogram and the GOESA, but different average loudness curves and distinct distributions regarding the UCL.

Classification into profiles

Model selection

To allow for a classification of new patients into the auditory profiles based on a reduced set of measures widely available in clinical practice, classification models were built using random forests. Different parameterizations (optimization metrics, binarization strategies, and CV procedures) were compared with the aim to provide the classification model best suited for the auditory profiles. The *mtry* parameter was inherently determined within each model.

Figure 4 displays the results of the comparative performance with respect to the binarization strategies, optimization metric, and cross-validation procedure on the training data set. Model performances with respect to the F1-scores were averaged across profiles to result in an overall F1-score. This allowed for a selection of the best model parameterization. Profile 7 was not

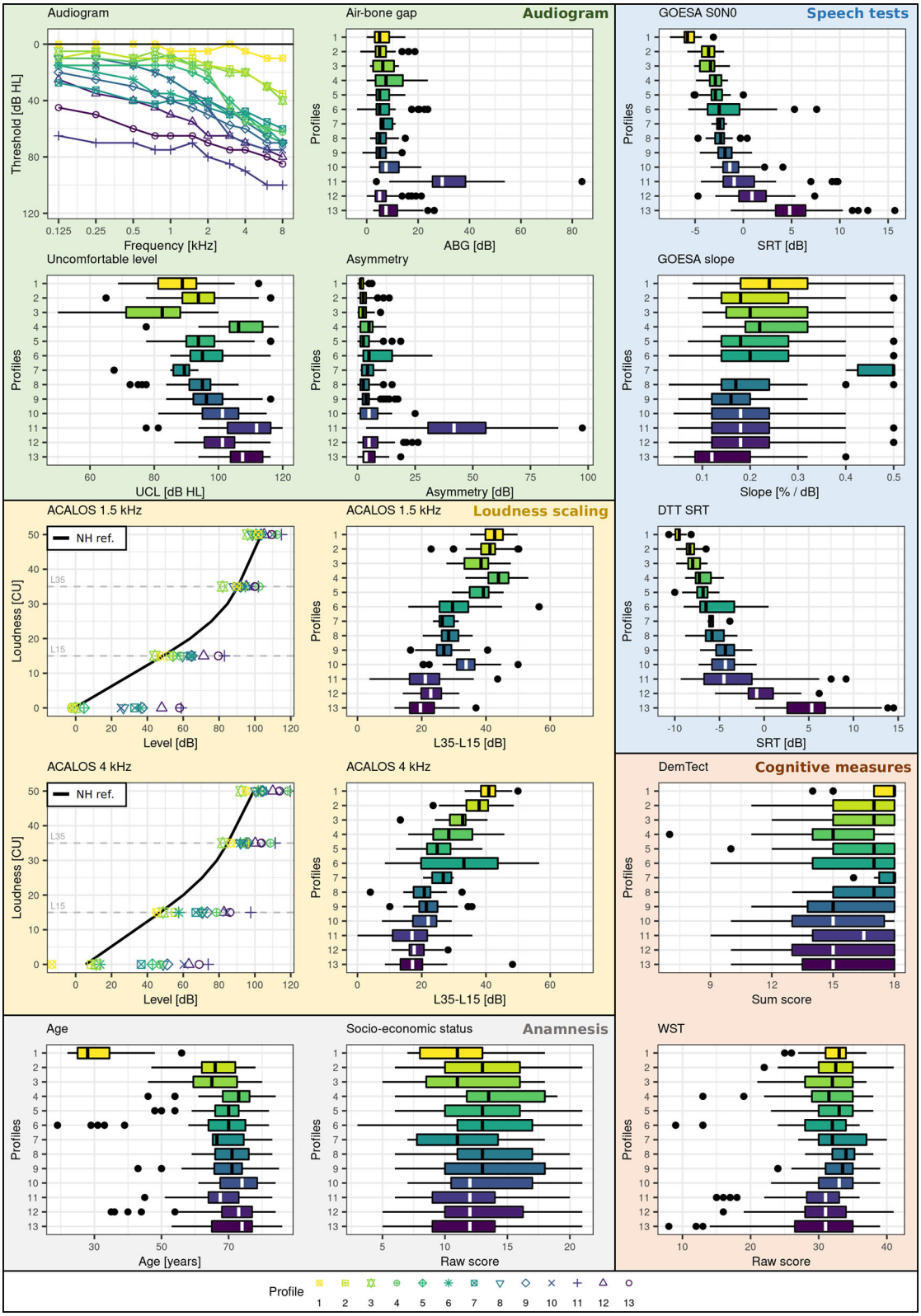


FIGURE 3
Profile ranges across measures. Plot backgrounds are colored according to underlying domains. Blue corresponds to the speech domain, green to the audiogram, yellow to the loudness domain, orange to the cognitive domain, and gray to the anamnesis. Profiles are color-coded (yellow to violet) and numbered (1-13) with respect to increasing SRT (impairment) on the GOESA.

selected for averaging, as the number of patients contained in the profile ($N = 6$) is not large enough to lead to reliable results and interpretations.

All models perform well in predicting profile classes, as indicated by the overall small and high range of mean F1-scores. The highest F1-score was obtained by the OVAOVO model using the kappa evaluation metric and repeated 10-fold CV. Consequently, the OVAOVO (kappa, repCV) model is selected as the classification model to allow for a prediction of patients into profiles. Across models, the kappa metric provided the best results, whereas optimal CV procedures differed across binarization strategies, with the exception of the OVAOVO model in which repCV provided the best results for all evaluation metrics.

Model evaluation

The previously selected optimal model (OVAOVO, repCV) was selected based on its performance on the training data set (75% of the patients). To investigate the generalizability of the classification model to new patients, its performance was subsequently evaluated on the test data set (25% of the patients). Figure 5 displays the performance results with respect to the sensitivity and precision across all profiles.

Generally, the classifier's performance is adequate regarding achieved sensitivity and precision on the test data set. Across profiles 1–6 and 8–13, average precision and sensitivity on the test data set are 0.9 and 0.84, respectively. Results for profile 7 were plotted for completeness, however, are unreliable due to the small sample size, since the test data set only consisted of two patients. Overall test performance is only slightly lower than training performance for most profiles, except for profiles 3, 6, and 7. For these profiles, the generalization of the learned classification approach toward unseen data is limited. Profile 3 and profile 6 show low levels of sensitivity, but high levels of precision. Thus, not all cases of the two profiles are detected, however, if the two profiles are predicted one can be highly certain that the patient does, indeed, belong to profile 3 or profile 6.

Discussion

The aim of this study was to propose a flexible and data-driven approach to patient stratification in the field of audiology that allows for a detailed investigation into the combination of hearing deficits across audiological measures. Our results demonstrate the feasibility and efficiency of our proposed profiling pipeline in characterizing hearing deficits in the form of patient groups, namely, auditory profiles. The proposed 13 auditory profiles separate patients with respect to ranges on audiological tests. Further, to ensure the applicability of the auditory profiles in clinical practice with only a basic set of audiological tests, classification models were built that allow for an adequate classification of the auditory profiles given such a reduced set of audiological measures.

Generation of profiles

The proposed profiles aim to represent the underlying patterns of the current data set best. Hence, the profiles describe the patterns across measures for the available patients and etiologies, rather than aiming to cover all generally existent patient groups with the current set of auditory profiles. Additionally, the number of profiles that can be generated is variable and dependent on the underlying data. This becomes evident when inspecting the distribution of optimal profile numbers in Figure 2. Across bootstrapped data sets different profile numbers were suggested. This may in part be due to the applied method. Different subsets of the bootstrapped data may miss extreme patient patterns, and thus, lead to a reduction or increase in suggested profile numbers. This, next to the added

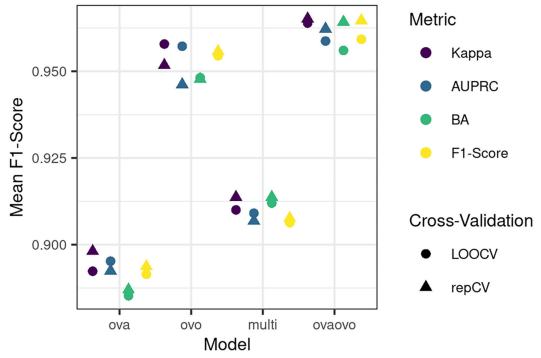


FIGURE 4
Performance of different models on the training data set. The mean F1-score was calculated as the mean of F1-scores across profiles 1–6 and 8–13. Metrics and cross-validation schemes can be distinguished by color and shape, respectively. BA refers to balanced accuracy. LOOCV refers to leave-one-out cross-validation; repCV to repeated 10-fold cross-validation.

TABLE 2 Number of patients contained in each auditory profile.

Profile	1	2	3	4	5	6	7	8	9	10	11	12	13
N	27	76	19	24	77	33	6	44	68	51	42	79	39

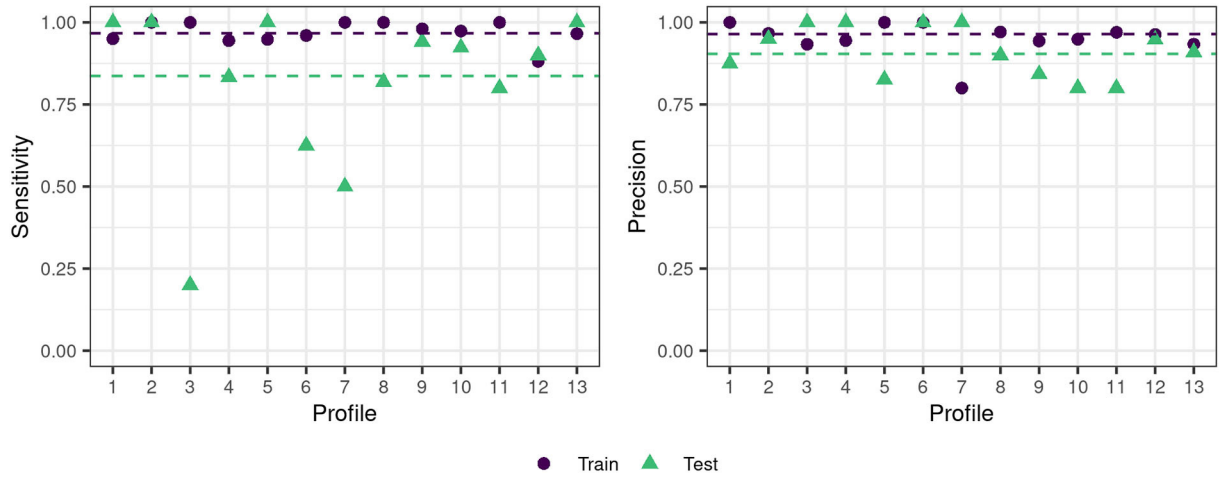


FIGURE 5
Train-test data set performance for the OVAOVO (kappa, repCV model) for both sensitivity and precision. The dashed lines indicate the mean across profiles 1–6 and 8–13 for the respective condition.

uncertainty that stems from imputing missings, may explain the variability in suggested profile numbers across bootstrapped samples. By using a bootstrapping approach, where the optimal number of profiles is defined as the most frequently proposed profile number, it can be assumed, however, that the effects of imputations and extreme patient patterns on the generated profile number were minimized.

The number of profiles may further be influenced by the employed model restrictions. Since the covariance parameterization "VEI" restricts both the shape and axis alignment and requires diagonal cluster distributions, a parsimonious model was selected as describing the underlying data structure best. The number of profiles, therefore, may be large in order to characterize the data structure best with the given restrictions (38). It would be of interest to apply the modeling approach to a larger dataset that allows for a less restrictive model in order to investigate if the resultant number of profiles would decrease. A more parsimonious model that leads to a larger number of profiles, however, is in line with our aim of detecting all plausible differences between patient groups.

Interpretation of profiles

The profiles, generally, cover a large range of different types and extents of hearing deficits and appear audiologically plausible. All profiles can be distinguished from each other by at least one audiological feature and can, thus, be considered as distinct patient groups regarding audiological measures (RQ1). The relevance of the distinction has to be evaluated with respect to the outcome measure of interest. Certain distinctions are, for instance, not necessarily relevant for diagnostic purposes. It can be assumed that profiles 4 and 5 would be categorized as bilateral sensorineural hearing loss (ICD code h90.3) (50) and could, thus, for purposes of coarse diagnostic classification be combined. Profile 5, however, shows a lower range of UCL levels, indicating that loudness would need to be compensated differently in a hearing aid for patients within profile 5 as compared to profile 4. The distinctions regarding loudness perception could influence the benefit that patients within the separate profiles may experience from hearing aids, if the same hearing aid parameters are applied to both groups. This highlights our motivation for flexible profiles that can be combined or separately considered given different outcome measures. The exact number of profiles may, therefore, change with the inclusion of further datasets and also depend on the targeted outcome measure. The proposed auditory profiles, however, enable a detailed investigation into differences that exist between patient groups.

Most of the profiles can be assumed to be caused by symmetrical sensorineural hearing loss. Profile 11, however, also contains an asymmetric conductive hearing loss, as indicated by the presence of both an asymmetry between the ears and an air–bone gap in the group (51). For the remainder of the profiles, however, we can interpret the profiles in the consideration of the four-factor model for sensorineural hearing loss by Kollmeier (52). The current profiles contain measures that allow for an estimation of the first two factors (attenuation and compression loss), but not binaural and central loss. The audiogram can provide an indirect indication for the attenuation loss, which is defined as the required amplification for each frequency to obtain an intermediate loudness perception (L25), whereas the ACALOS can indicate a compression loss *via* a reduced dynamic range (52). Overall, we can observe differences in both

the audiogram shapes and the dynamic ranges across profiles. Most importantly, similar audiogram shapes (e.g., profiles 2 and 3) do not necessarily lead to a similar compression loss and our profiles are able to detect these differences, which is in line with the assumption of the four-factor model, that the audiogram alone cannot explain all underlying characteristics of sensorineural hearing loss. We, therefore, conclude that the 13 auditory profiles provide meaningful information regarding two important factors of hearing deficits, i.e., attenuation and compression loss (RQ1), and that the profiling pipeline has the potential for the detection of patient group differences also for further datasets, if suitable measures are included.

In general, the interrelation across speech tests, loudness scaling, and audiogram data lead to a separation of patients into profiles. For instance, profiles 2 and 3 contain patients with both similar SRTs and audiogram thresholds. Profile 3, however, shows a reduced dynamic range with its uncomfortable loudness level (UCL) thresholds derived from the audiogram and the range between soft (L15) and loud (L35) sounds on the ACALOS reduced, which indicates recruitment. This, in turn, has implications for hearing aid fitting. It can be assumed that patients within the two profiles require different compression settings, in spite of similar audiograms (53, 54). In contrast, the main difference for profiles 8 and 9 lies within their thresholds on the audiogram, with profile 9 showing about 10 dB higher thresholds, while showing similar SRT and loudness curve ranges. The relevance of a distinction between these two profiles, for both diagnostics and hearing aid fitting, thus, needs to be further investigated. For other profiles, differences are more strongly pronounced and they can well be separated.

Certain profiles also align well with the proposed phenotypes by Dubno et al. (16). Profiles 6, 7, and 9 are consistent with the metabolic phenotype, and profiles 2 and 3 appear to be in between the pre-metabolic and metabolic phenotype with respect to the ranges on the audiogram. Profile 4 can be described in terms of the sensory phenotype and profiles 5 and 10 as the metabolic + sensory phenotype. However, the auditory profiles also contain different patterns, with either more severe presentations as described by the phenotypes (profiles 11 and 13), or different slopes in the lower frequency range of the audiogram (profiles 8 and 12). Further, instead of an older normal hearing profile to match the older normal hearing phenotype, only a young normal hearing profile is included. Regardless, certain probable etiologies can be inferred for the respective profiles, exemplifying how alternate stratification approaches could be connected to the auditory profiles proposed in this study. Since more than one profile can be matched to sensory and metabolic phenotypes, however, it can, again, be assumed that further contributors regarding individual presentations of hearing deficits exist, which are not assessed *via* the pure-tone audiogram.

No distinctions across profiles regarding the cognitive measures were found (WST, DemTect). Even though hearing deficits and cognitive impairments have been widely associated (55), the precise causal relationship remains unclear and some studies did not find significant relations (26). With the profiles, a slight trend toward increasing impairment on the DemTect with increasing SRT can be observed; however, the ranges across profiles overlap substantially. On the one hand, this may indicate, that none of the present profiles is significantly influenced by cognitive abilities and that the observed patterns of hearing deficits may occur for both cognitively impaired and non-impaired patients. This would require further investigations and the inclusion of patients with more severe cognitive impairments. On the other hand, the DemTect, as a screening instrument, may not be sensitive enough for detecting a further association between cognitive impairment and hearing deficits. For the auditory profiles, this indicates that cognitive differences are not well-represented, such that patients' cognitive abilities would need to be assessed *via* further cognitive measures that are currently not included in the database.

The currently available profiles naturally only provide a picture of the contained measures. It can be assumed that the inclusion of further measures will enhance the precision of patient characterization. Of the specified eight domains relevant for characterizing hearing deficits, defined by van Esch et al. (6), currently, four are contained in the defined profiles (pure-tone audiogram, loudness perception, speech perception in noise, and cognitive abilities). Spatial contributors, i.e., the intelligibility level difference (ILD) and binaural intelligibility level difference (BILD) measures, were—unfortunately—not included in the original database so no relation to the profiles given here can be provided. However, it can be assumed that they could provide an enhanced characterization of patients' hearing status, as well as prove valuable for hearing aid fitting. Similarly, measures describing the central factor of hearing loss could be incorporated if available in a data set, to comply with all four factors as suggested by Kollmeier (52). Consequently, future studies should work toward incorporating these measures into the profiles.

Classification into profiles

By building classification models to match patients into the auditory profiles using only features from the air-conduction audiogram, loudness scaling, and GOESA, we aimed for the applicability of the profiles in a variety of settings. First, in clinical routine, both the audiogram and a speech test, measuring the SRT, are the current standard in hearing aid fitting (56), and in Germany, the GOESA is included in the German guideline for hearing aid fitting (57). In addition, loudness scaling has proved promising for hearing aid adjustments (58). The three measures are, therefore, often available for hearing professionals and do not extend the testing time of patients and physicians. If fewer measures are available, e.g., only the audiogram and the GOESA, or a different set of measures,

the classification models would have to be retrained for this purpose. We believe, however, that loudness scaling provides valuable information for hearing aid fitting and should, thus, be included in the fitting process. Second, to use the profiles in further research and clinical data sets, it is important to include measures that are frequently measured and available. Thus, even though further measures may be contained in the data sets, it is necessary to provide classification models containing measures widely available across data sets.

The present results indicate the feasibility of classifying patients into most of the profiles. The OVAOVO model with the kappa loss function and 10-fold repeated CV reached the highest F1-score and was, therefore, selected as the optimal classification model for the analyzed dataset. With the model test set, sensitivity was >75% for all profiles but profiles 3, 6, and 7 (RQ2). For profile 7, this can be explained by the small sample size of the profile as only six patients were classified into the profile. Consequently, the training of a classifier for profile 7 does not lead to reliable results, and its generalizability is not assured. In spite of that, we included the results for profile 7 for completeness, since it may provide further separation from the remaining profiles for the multi-class classifier, by including counter-examples of patients. Profile 7, however, cannot yet reliably be used to classify new patients into it. Further information from databases is needed to investigate whether this profile represents rare cases or whether this profile was not represented enough in the present data set to provide a large enough sample size for classification purposes. Profile ranges for profile 6 are generally broader than for other profiles; therefore, misclassifications may have occurred more frequently, thus, reducing the sensitivity for profile 6.

The current classification model naturally only covers patient populations that were also contained in the analyzed dataset. Given the adequate classification performance of the classifier, it can be assumed that new patients with similar characteristics to the patients within the dataset would also be adequately predicted into the auditory profiles. At the same time, random forests allow for an estimation of the classification uncertainty when classifying patients into the profiles. This uncertainty estimation refers to how often a patient was predicted into a given profile across the decision trees of the random forest as compared to the remainder of the profiles. For certain predictions, there is a high amount of agreement of the random forest, whereas for uncertain predictions there is a lower amount of agreement of the random forest. New patients are, therefore, classified into a given profile with an estimate of uncertainty, which, in turn, could also indicate if none of the profiles adequately represents the given patient. This could then reveal a rare patient case or a patient belonging to an additional profile that has not yet been defined. Generally, patients would always be allocated to a profile based on all measures that are contained in the classification model (i.e., audiogram, ACALOS, age, GOESA) and no single feature would determine the classification. For instance, the analyzed dataset contains mainly elderly hearing impaired patients and younger normal hearing patients. Children and younger individuals may, however, also experience hearing deficits. A classification based solely on the feature age would lead to a misclassification into the normal hearing profile 1. The generated classification model, in contrast, would also consider information from the audiogram, ACALOS, and GOESA and in that way avoid misclassification into the normal hearing profile 1.

It can be argued that predictive performance would have been improved by including all measures in the classification models. However, we aimed at providing classification models that can be readily used with measures available across clinics in Germany, such that no additional testing is required and time constraints of physicians are met. Consequently, we decided on a reduced set of measures and aimed at predicting profiles with widely available measures. In future, it may be of interest to provide classification models for all combinations of measures, such that if, e.g., bone-conduction thresholds or more specific psychoacoustic tests are also available in clinical settings, they can be used to increase predictive performance with regard to, e.g., the "binaural" and "central noise" factor (52) involved in characterizing the individual hearing problem.

One limitation of the present classification is the number of patients contained in each profile. For further validation larger and more balanced data sets that also contain more severe patients are required, which can also be assumed to lead to improvements in the predictive performance. An increase in the size of the training set will support the training of the classifier, whereas an increase in the test set will improve the certainty of the predictions. Currently, test performance may have been artificially high for some profiles due to the small sample size in the test set. However, further reducing the training size would also not be desirable, as it would increase the bias of the classification models. Thus, further evaluations on additional data sets containing further patients are required.

Properties of the profiling approach and comparison to existing approaches

The current data-driven approach toward generating auditory profiles to characterize patient groups is not aimed at being contradictory with hitherto available profiling approaches but aims at providing a more detailed account of existing patient groups and offers several advantages.

First, its flexibility in the definition of profiles derived *via* purely data-driven clustering allows extending and refining the profiles, if in further data sets more extreme patient representations are contained. More specifically, it can be assumed that applying the profiling approach to additional data sets containing both similar and more extreme patient presentations will result in a set of auditory profiles that show overlap to herein proposed profiles, but also contain additional profiles. The new set of profiles could then be used to update

the current set of auditory profiles. As a result, the total number of auditory profiles is not fixed and instead remains flexible to include further profiles. Likewise, the presented profiling pipeline can be applied to additional data sets with varying measures. In case of differing measures across data sets, measures not used for clustering purposes could serve as descriptive features and allow for inference, if these features occur more frequently in certain profiles. The flexibility in terms of derived profiles and contained measures could, in future, aid in comparing patients across data sets. Appropriate means to combine profiles generated on different data sets, however, need to be defined. For this purpose, a profile similarity index based on, e.g., overlapping densities (59) could provide a cut-off score on when to combine or extend profiles.

Second, profiles are not tailored toward a certain outcome such as diagnostics or hearing aid fitting. This may, in part, explain the rather large number of generated profiles, since profiles may differ with respect to measurement ranges but not with respect to audiological findings, diagnoses, or treatment recommendations. By tailoring our analyses toward certain outcomes, we could have possibly reduced the number of generated profiles. Our aim, however, was to generate as many profiles as plausibly contained within the data set such that all differences between patient groups can be caught. More specifically, by using Bisgaard standard audiograms also as a feature for clustering, patients were already separated into 10 distinct audiogram ranges. Combining 10 separate audiogram ranges with different loudness curves and SRT ranges already leads to a larger amount of profiles, if these patterns across measures and patients (i.e., profiles) occur frequently and are well-distinguishable from other profiles. At the same time, the flexibility of the profiles by their definition directly on measurement ranges allows reducing the number of profiles if only certain outcomes are of interest. For instance, if, in future, profiles are connected to diagnostic information from further data sets, profiles leading to a distinction with respect to a diagnosis could be separated or merged. Similarly, if profiles are used for hearing aid fitting, only those profiles leading to separable groups with respect to aided parameters could be retained.

Third, all patients can be grouped into auditory profiles. In contrast, in Dubno et al. (16), around 80% of audiogram shapes were categorized as non-exemplar and could not be matched into one of the phenotypes, whereas in Sanchez-Lopez et al. (18), an "uncategorizable" category in addition to the four profiles exists.

A fourth advantage of the flexibility of our auditory profiles pertains to its ability to provide complementary knowledge compared to other profiling approaches, which allows analyzing the same data sets from different perspectives and potentially learning more about the inherent patterns. To exemplify, the profiling approach by Sanchez-Lopez et al. (18) is applicable to different audiological data sets as well and also comprises the two steps of profile generation and classification. Both approaches are data-driven; however, the approach by (18) is based on the hypothesis of two distortion types which limits the number of profiles to four. In contrast, our approach is purely data-driven, that is, the obtained number of profiles directly depends on the available combinations of measurement ranges in the respective data set, in order to detect all existing differences between patients. Each of our profiles (estimated by model-based clustering) characterizes the group of included patients in terms of underlying measurement data, while the profiles of (18) are characterized by one respective extreme prototypical patient (due to archetypal analysis) and all other patients classified into a respective profile show less extreme results on the variables identified by principal component analysis. The profiles of (18) are interpretable due to the hypothesis of two distortion types and the variables related to each distortion type; however, the obtained interpretation depends on the available measures in the dataset. That means that it needs to be ensured to employ an appropriate database, as was achieved in Sanchez-Lopez et al. (18) with the BEAR test battery (7), following the findings of (17) where the choice of data led to different, not completely plausible interpretations based on the two different analyzed datasets. In contrast, our profiling approach does not include explicit interpretability of every profile yet, but instead, interpretability needs to be added as an additional step. This can be done by relating the profiles to the literature as discussed above, or by including expert knowledge to label the different profiles. In addition, the type of interpretability required for different outcome measures considered in future analyses may be different, and can then be chosen appropriately.

For associating the profiles obtained by the two approaches, in a first step, the distributions of patient data grouped to profiles can be manually compared, for instance regarding audiogram and SRT ranges in Figure 6 of (18) and in our Figure 3. However, this comparison is limited as only a small subset of measures is common in the BEAR test battery and our dataset, as well as due to methodological differences as discussed above. Instead, it would be interesting to apply the two profiling approaches to the respective other datasets. As we have GOESA and ACALOS available to characterize speech intelligibility and loudness perception, it would be interesting if the profiling approach of (18) also estimates speech intelligibility and loudness as the two distortion dimensions based on our data. Vice versa, the application of our approach to the BEAR test battery would generate a certain number of profiles, which could be compared to the profiles obtained in this study (and thereby to a comparison and potential combination of datasets), as well as reveal measurement combinations leading to sub-classes of the four auditory profiles of (18).

Limitations of the profiling approach

Despite the advantages of our purely data-driven profiling approach, certain limitations persist. At the current stage,

the profiling approach can detect plausible patient subgroups in data sets. This property generalizes also to further data sets containing different sets of measures and a different patient population. A restriction in the application of the current profiling pipeline to additional databases is the current requirement for continuous or at least ordinal features. Relevant audiological measures may, however, also be categorical with no inherent ordering. Thus, to also incorporate these measures, the current pipeline would need to be adjusted to also allow for categorical features.

The ability to detect differences in patient groups also depends on the sample size, the contained measures, as well as the presence of distinctive patient groups within the data set. If sample sizes are small, a smaller number of patient groups may be detected in the data sets, which in turn, would be defined by broader ranges across measures. At the same time, this could result in an increase in profiles, each containing only a few patients. This, however, would indicate that the underlying data set is not suitable for the herein proposed profiling approach, as nearly no similarities between patients could be detected. In such a case, it would not be certain whether a profile corresponds to a patient group that could also be identified in larger datasets, or whether it corresponds to outliers in the analyzed data set. Likewise, if only a few measures are contained in new data sets, not all existent distinctions between patients may be detected. Instead, only distinctions regarding the included measures would be available. Combining profiles generated on further data sets with the current profiles may, thus, prove difficult. An estimate of profile "conciseness" could tackle this challenge. This estimate could refer to the average similarity of patients within a profile regarding relevant measures. The similarity between patients with broader profiles will be smaller than the similarity between profiles with smaller ranges across audiological measures. As a result, the conciseness estimate could indicate if the generated profiles on the new data set only result in a coarse grouping of patients. It could then be analyzed, whether the coarse grouping could be explained by a mixture of already available auditory profiles. This would, however, require an overlap between audiological measures across the profiles. If the profiling pipeline is applied to a data set with low overlap regarding measures, the generated profiles would have to be interpreted separately from the current set of profiles, until a relation between measures has been established. This could either occur *via* available knowledge or by analyzing a data set that contains an overlap between the measures of interest. Regardless, newly generated profiles on further data sets would first need to be analyzed in terms of general audiological plausibility.

At the same time, the relevance of the distinctions between patient groups, in general, and for clinical practice needs further evaluation. This could either comprise asking experts to rate the plausibility and clinical applicability of the distinctions between the profiles or incorporating expert knowledge from other approaches toward patient characterization. The Common Audiological Functional Parameters (CAFPAs) by Buhl et al. (21), for instance, provide an expert-based concept of describing patient characteristics; and in Saak et al. (20), regression models were built to predict CAFPAs based on features that are also available for the current auditory profiles. Hence, the predicted CAFPAs would be available as additional descriptive information for the profiles generated in this study, and a consistency check to previous CAFPA classification (60) could be obtained by analyzing the same data set from different perspectives (i.e., analysis tools). In that way, both approaches provide complementary insights, and both contribute to future combined analysis of different audiological databases. As a result, physicians' trust toward applications (e.g., clinical decision-support systems) using the auditory profiles could be enhanced, which has shown to be a relevant factor in the adoption of such systems in clinical routine (61). Additionally, it can be assumed that the inclusion of more severe patient cases, e.g., with indications for a cochlear implant, could enhance the current profiles toward more extreme profile representations. Currently, profiles can be mostly assigned to mild to moderate hearing loss. With the inclusion of further data sets, containing a higher prevalence of severe patient cases, this aspect could be addressed.

Application and outlook

The herein proposed profiling approach serves as a starting point for uncovering patient groups and patient presentations across audiological measures for the increasingly available amount of larger data sets. Consequently, the proposed profiling approach needs to be applied to additional data sets, which include more severe and diverse patient populations, as well as additional audiological measures to cover further important factors of hearing loss (e.g., binaural and central components). The set of auditory profiles would need to be updated after the inclusion of every further data set by either merging similar generated profiles or adding new profiles. In that way, it would conclude in a final set of auditory profiles, if generated profiles converge. This means that generated profiles on new datasets are already contained in the set of defined auditory profiles and no new information is added, thus, resulting in a final set of auditory profiles describing the audiological patient population.

If the generated auditory profiles describe the audiological patient population, they could be used in a variety of applications due to their flexibility. The profiles could efficiently summarize patient information for a clinical decision-support system. Likewise, they could also support mobile assessments of patients, in e.g., a "virtual hearing clinic." If patients are tested on the measures used for the classification models (or appropriate mobile implementations of those measures, ensuring that

measurements near the hearing threshold are feasible in realistic environments) they could be classified into a profile. In a clinical decision-support system, physicians could then be provided with statistical insights into patients' hearing statuses, whereas in virtual hearing clinic patients themselves could receive information regarding their hearing statuses. To also provide diagnostic decision-support as well as aided benefit predictions, however, data from additional data sets containing these measures need to be incorporated into the current profiles. A metric allowing for the combination or separation of profiles, if new profiles are generated on additional data sets, hence, needs to be defined.

After the final set of auditory profiles has been defined, it would also be of interest to define a minimum set of tests that allow for adequate classification of patients into the profiles across data sets. This could highlight the audiological measures that are most relevant across all profiles. Likewise, the profiles could contribute to the selection of the next to-be-performed measures for characterizing the patients. If classification models are available for all measurement combinations, measures leading to the best discriminatory performance across profiles could be selected next. This, in turn, could reduce the testing time of the patients, as well as support the derivation of test batteries covering all relevant aspects of hearing deficits, as in (5, 6), by highlighting the most important measures.

Conclusion

The proposed data-driven profiling approach resulted in 13 distinct and plausible auditory profiles and allows for efficiently characterizing patients based on the interrelations of audiological measures. All patients are characterized and patient groups with certain characteristics, such as asymmetry, are not excluded. Due to the profiles' flexibility by being defined on the contained patients' measurement ranges, profiles could be added or refined, given insights derived from applying the profiling approach to additional data sets. The profiles concur with other profiling approaches but are able to detect differences in patient groups regarding measurement ranges in more detail than hitherto available approaches.

New patients can be adequately classified into the auditory profiles for 12 of the 13 auditory profiles. For 10 profiles, both high precision and sensitivity were achieved (>0.75), and for two profiles, low to medium sensitivity and high precision were achieved, and for one profile no classification could be achieved due to the profiles' small sample size. Since the classification model was based on a reduced set of measures often available in clinical practice in Germany (GOESA, ACALOS, air-conduction audiogram, and age), clinicians could use the auditory profiles even without performing a complete audiological test battery, if a quick classification with less clinical detail is required. Likewise, all measures required for classifying patients into the auditory profiles are potentially available also on mobile devices, facilitating mobile assessments of the patient.

The proposed profiling approach depends on the underlying data set in terms of the number of profiles or the covered range of patients. Its properties such as flexibility, not being tailored toward a specific outcome, or ability to handle incomplete patient data, however, generalize to other data sets including additional measures. Appropriate means to combine profiles generated across data sets need to be defined.

Future research should extend the profiling toward integrating different data sets with more severe and diverse patient cases. In addition, binaural measures should be included, as well as aided data to investigate hearing device benefits with the profiles.

Data availability statement

The data analyzed in this study was obtained from Hörzentrum Oldenburg gGmbH, the following licenses/restrictions apply: According to the Data Usage Agreement of the authors, the datasets analyzed in this study can only be shared upon motivated request. Requests to access these datasets should be directed to MB, mareike.buhl@uni-oldenburg.de and SS, samira.saak@uni-oldenburg.de. The analyses scripts can be found here: Zenodo, https://zenodo.org/, https://doi.org/10.5281/zenodo.6604135.

Author contributions

SS conducted the data analysis which was continuously discussed with all authors and drafted the manuscript. All authors conceptualized, designed the study, and contributed to the editing of the manuscript.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germanys Excellence Strategy —EXC 2177/1 – Project ID 390895286.

Acknowledgments

We thank Hörzentrum Oldenburg gGmbH for the provision of the patient data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Musiek FE, Shinn J, Chermak GD, Bamiou DE. Perspectives on the pure-tone audiogram. *J Am Acad Audiol.* (2017) 28:655. doi: 10.3766/jaaa.16061

2. Houtgast T, Festen JM. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *Int J Audiol.* (2008) 47:287. doi: 10.1080/14992020802127109

3. Schoof T, Rosen S. The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Front Aging Neurosci.* (2014) 6:307. doi: 10.3389/fnagi.2014.00307

4. Humes LE. Factors underlying individual differences in speech-recognition threshold (SRT) in noise among older adults. *Front Aging Neurosci.* (2021) 13:702739. doi: 10.3389/fnagi.2021.702739

5. Van Esch TEM, Dreschler WA. Relations between the intelligibility of speech in noise and psychophysical measures of hearing measured in four languages using the auditory profile test battery. *Trends Hearing.* (2015) 19:2331216515618902. doi: 10.1177/2331216515618902

6. van Esch TEM, Kollmeier B, Vormann M, Lyzenga J, Houtgast T, Hitygren M, et al. Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *Int J Audiol.* (2013) 52:305. doi: 10.3109/14992027.2012.759665

7. Sanchez-Lopez R, Nielsen SG, El-Haj-Ali M, Bianchi F, Fereczkowski M, Cañete OM, et al. Auditory tests for characterizing hearing deficits in listeners with various hearing abilities: the bear test battery. *Front Neurosci.* (2021) 15:724007. doi: 10.3389/fnins.2021.724007

8. Sanchez Lopez R, Nielsen SG, Cañete O, Fereczkowski M, Wu M, Neher T et al. A clinical test battery for Better hEAring Rehabilitation (BEAR): Towards the prediction of individual auditory deficits and hearing-aid benefit. In: *Proceedings of the 23rd International Congress on Acoustics*. Deutsche Gesellschaft für Akustik e.V (2019). p. 3841–8. doi: 10.18154/RWTH-CONV-239177

9. Gieseler A, Tahden MAS, Thiel CM, Wagener KC, Meis M, Colonius H. Auditory and non-auditory contributions for unaided speech recognition in noise as a function of hearing aid use. *Front Psychol.* (2017) 8:219. doi: 10.3389/fpsyg.2017.00219

10. Lopez-Poveda EA, Johannesen PT, Pérez-González P, Blanco JL, Kalluri S, Edwards B. Predictors of hearing-aid outcomes. *Trends Hearing.* (2017) 21:2331216517730526. doi: 10.1177/2331216517730526

11. Buhl M, Warzybok A, Schädler MR, Lenarz T, Majdani O, Kollmeier B. Common audiological functional parameters (CAFPAs): statistical and compact representation of rehabilitative audiological classification based on expert knowledge. *Int J Audiol.* (2019) 58:231. doi: 10.1080/14992027.2018.1554912

12. Bisgard N, Vlaming MSMG, Dahlquist M. Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif.* (2010) 14:113. doi: 10.1177/1084713810379609

13. Dörfler C, Hocke T, Hast A, Hoppe U. Speech recognition with hearing aids for 10 standard audiograms. *HNO.* (2020) 68:933. doi: 10.1007/s00106-020-00843-y

14. Folkeard P, Bagatto M. Scollie S. Evaluation of hearing aid manufacturers' software-derived fittings to DSL v50 pediatric targets. *J Am Acad Audiol.* (2020) 31:354. doi: 10.3766/jaaa.19057

15. Kates JM, Arehart KH, Anderson MC, Muralimanohar RK, Harvey LO. Using objective metrics to measure hearing-aid performance. *Ear Hear.* (2018) 39:1165. doi: 10.1097/AUD.0000000000000574

16. Dubno JR, Eckert MA, Lee FS, Matthews LJ, Schmiedt RA. Classifying human audiometric phenotypes of age-related hearing loss from animal models. *JARO.* (2013) 14:687. doi: 10.1007/s10162-013-0396-x

17. Sanchez Lopez R, Bianchi F, Fereczkowski M, Santurette S, Dau T. Data-driven approach for auditory profiling and characterization of individual hearing loss. *Trends Hearing.* (2018) 22:2331216518807400. doi: 10.1177/2331216518807400

18. Sanchez-Lopez R, Fereczkowski M, Neher T, Santurette S, Dau T. Robust data-driven auditory profiling towards precision audiology. *Trends Hearing.* (2020) 24:2331216520973539. doi: 10.1177/2331216520973539

19. Buhl M, Warzybok A, Schädler MR, 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.* (2020) 59:534. doi: 10.1080/14992027.2020.1728401

20. Saak SK, 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.* (2020) 2:596433. doi: 10.3389/fdgth.2020.596433

21. Buhl M. Interpretable clinical decision support system for audiology based on predicted common audiological functional parameters (CAFPAs). *Diagnostics.* (2022) 12:463. doi: 10.3390/diagnostics12020463

22. Kollmeier B, Wesselkamp M. Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J Acoust Soc Am.* (1997) 102:2412.

23. Smits C, Kapteyn TS, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol.* (2004) 43:15. doi: 10.1080/14992020400050004

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

25. Oetting D, Hohmann V, Appell JE, Kollmeier B, Ewert SD. Restoring perceived loudness for listeners with hearing loss. *Ear Hear.* (2018) 39:664. doi: 10.1097/AUD.0000000000000521

26. Fulton SE, Lister JJ, Bush ALH, Edwards JD, Andel R. Mechanisms of the hearingudness for listeners with*Semin Hear.* (2015) 36:140. doi: 10.1055/s-0035-1555117

27. Kalbe E, Kessler J, Calabrese P, Smith R, Passmore AP, Brand M, et al. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry.* (2004) 19:136. doi: 10.1002/gps.1042

28. Schmidt KH, Metzler P. *WST-Wortschatz.* Göttingen: Beltz Test (1992).

29. Winkler J, Stolzenberg H. *Adjustierung des Sozialen-Schicht-Index für die Anwendung im Kinder- und Jugendgesundheitssurvey (KiGGS) (2009), Wismarer Diskussionspapiere, No. 07/2009, ISBN 978-3-939159-76-6, Hochschule Wismar.* Fakultät für Wirtschaftswissenschaften (2009).

30. Fang Y, Wang J. Selection of the number of clusters via the bootstrap method. *Comput Stat Data Anal.* (2012) 56:468. doi: 10.1016/j.csda.2011.09.003

31. von Luxburg U. Clustering stability: an overview. *FNT Machine Learn.* (2009) 2:235. doi: 10.1561/2200000008

32. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res.* (2011) 20:40. doi: 10.1002/mpr.329

33. Bouveyron C, Brunet-Saumard C. Model-based clustering of high-dimensional data: a review. *Comput Stat Data Anal.* (2014) 71:52. doi: 10.1016/j.csda.2012.12.008

34. Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc.* (2002) 97:611. doi: 10.1198/016214502760047131

35. Banerjee A, Shan H. Model-based clustering. In: Sammut C, Webb GI, editors *Encyclopedia of Machine Learning.* Boston, MA: Springer US (2010).

36. Fraley C, Raftery AE. Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *J Classification.* (2003) 20:263. doi: 10.1007/s00357-003-0015-3

37. Greve B, Pigeot I, Huybrechts I, Pala V, B clustering, Comparison of heuristic and model-based clustering methods for dietary pattern analysis. *Public Health Nutr.* (2016) 19:255. doi: 10.1017/S1368980014003243

38. Bouveyron C, Celeux G, Murphy TB, Raftery AE. *Model-Based Clustering and Classification for Data Science: With Applications in R.* Cambridge: Cambridge University Press (2019).

39. Schwarz G. Estimating the dimension of a model. *Ann Stat.* (1978) 6:461. doi: 10.1214/aos/1176344136

40. Breiman L. Random forests. *Mach Learn.* (2001) 45:5. doi: 10.1023/A:1010933404324

41. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer (2009).

42. Biau G, Scornet E. A random forest guided tour. *TEST*. (2015) 25:197–227. doi: 10.1007/s11749-016-0481-7

43. Galar M, Fern Fernández A, Barrenechea E, Bustince H, Herrera F. An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognit.* (2011) 44:1761–76. doi: 10.1016/j.patcog.2011.01.017

44. Adnan MN, Islam M. One-Vs-all binarization technique in the context of random forest. In: *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges (2015). p. 385–90.

45. Beinecke J, Heider D. Gaussian noise up-sampling is better suited than SMOTE and ADASYN for clinical decision making. *BioData Min.* (2021) 14:49. doi: 10.1186/s13040-021-00283-6

46. Thai-Nghe N, Gantner Z, Schmidt-Thieme L. Cost-sensitive learning methods for imbalanced data. In: *The 2010 International Joint Conference on Neural Networks*. Barcelona (2010). p. 1–8. doi: 10.1109/IJCNN.2010.5596486

47. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. (2022) 12:5979. doi: 10.1038/s41598-022-09954-8

48. Cohen JA. Coefficient of agreement for nominal scales. *Educ Psychol Meas.* (1960) 20:37. doi: 10.1177/001316446002000104

49. Sofaer HR, Hoeting JA, Jarnevich CS. The area under the precision-recall curve as a performance metric for rare binary events. *Meth Ecol Evol.* (2019) 10:565. doi: 10.1111/2041-210X.13140

50. World Health Organization. *ICD-10 : International Statistical Classification of Diseases and Related Health Problems : Tenth Revision*. Geneva: World Health Organization (2004).

51. Isaacson J, Vora NM. Differential diagnosis and treatment of hearing loss. *AFP*. (2003) 68:1125.

52. Kollmeier B. On the four factors involved in sensorineural hearing loss. *Psychophys Physiol Mod Hearing*. (1999) 211–8. doi: 10.1142/9789812818140_0036

53. Dreschler W, van Esch T, Larsby B, Hällgren M, Lutman M, Lyzenga J, et al. Charactering the individual ear by the "Auditory Profile". *J Acoust Soc Am*. (2008) 123:3714. doi: 10.1121/1.2935153

54. Launer S, Zakis JA, Moore BCJ. Hearing aid signal processing. In: Popelka GR, Moore BCJ, Fay RR, Popper AN, editors. *Hearing Aids*. Cham: Springer International Publishing (2016). p. 93–130.

55. Lin FR. Hearing loss and cognition among older adults in the United States. *J Gerontol Series A*. (2011) 66A:1131–6. doi: 10.1093/gerona/glr115

56. Hoppe U, Hesse G. Hearing aids: indications, technology, adaptation, and quality control. *GMS Curr Top Otorhinolaryngol Head Neck Surg*. (2017) 16:Doc08. doi: 10.3205/cto000147

57. Bundesausschuss G. *Richtlinie des Gemeinsamen Bundesausschusses nd quality control. ed SHilfsmitteln in der vertragsemeinsamen Bundesausschusses nd quality control. ed Stat*. Bundesanzeiger BAnz AT. Berlin (2021).

58. Kiessling J. Hearing aid fitting procedures - state-of-the-art and current issues. *Scand Audiol.* (2001) 30:57. doi: 10.1080/010503901300007074

59. Pastore M, Calcagnì A. Measuring distribution similarities between samples: a distribution-free overlapping index. *Front Psychol.* (2019) 10:1089. doi: 10.3389/fpsyg.2019.01089

60. Buhl M, Warzybok A, Schädler MR, Kollmeier B. Sensitivity and specificity of automatic audiological classification using expert-labelled audiological data and common audiological functional parameters. *Int J Audiol.* (2021) 60:16. doi: 10.1080/14992027.2020.1817581

61. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. *Decis Support Syst.* (2013) 54:953. doi: 10.1016/j.dss.2012.09.018

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fneur.2022.934650

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
National Research Council (CNR), Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Timothy Koscik,
The University of Iowa, United States
Carlo Cavaliere,
IRCCS SYNLAB SDN, Italy
Sara Narteni,
National Research Council (CNR), Italy

*CORRESPONDENCE
Rodrigo Moreno
rodmore@kth.se

SPECIALTY SECTION
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

RECEIVED 02 May 2022
ACCEPTED 19 September 2022
PUBLISHED 23 September 2022

CITATION
Siegbahn M, Engmér Berglin C and Moreno R (2022) Automatic segmentation of the core of the acoustic radiation in humans. *Front. Neurol.* 13:934650. doi: 10.3389/fneur.2022.934650

Automatic segmentation of the core of the acoustic radiation in humans

Malin Siegbahn[1,2], Cecilia Engmér Berglin[1,2] and Rodrigo Moreno[3]*

[1]Division of Ear, Nose and Throat Diseases, Department of Clinical Science, Intervention and Technology, Karolinska Institute, Stockholm, Sweden, [2]Medical Unit Ear, Nose, Throat and Hearing, Karolinska University Hospital, Stockholm, Sweden, [3]Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Stockholm, Sweden

Introduction: Acoustic radiation is one of the most important white matter fiber bundles of the human auditory system. However, segmenting the acoustic radiation is challenging due to its small size and proximity to several larger fiber bundles. TractSeg is a method that uses a neural network to segment some of the major fiber bundles in the brain. This study aims to train TractSeg to segment the core of acoustic radiation.

Methods: We propose a methodology to automatically extract the acoustic radiation from human connectome data, which is both of high quality and high resolution. The segmentation masks generated by TractSeg of nearby fiber bundles are used to steer the generation of valid streamlines through tractography. Only streamlines connecting the Heschl's gyrus and the medial geniculate nucleus were considered. These streamlines are then used to create masks of the core of the acoustic radiation that is used to train the neural network of TractSeg. The trained network is used to automatically segment the acoustic radiation from unseen images.

Results: The trained neural network successfully extracted anatomically plausible masks of the core of the acoustic radiation in human connectome data. We also applied the method to a dataset of 17 patients with unilateral congenital ear canal atresia and 17 age- and gender-paired controls acquired in a clinical setting. The method was able to extract 53/68 acoustic radiation in the dataset acquired with clinical settings. In 14/68 cases, the method generated fragments of the acoustic radiation and completely failed in a single case. The performance of the method on patients and controls was similar.

Discussion: In most cases, it is possible to segment the core of the acoustic radiations even in images acquired with clinical settings in a few seconds using a pre-trained neural network.

KEYWORDS

acoustic radiation, diffusion MRI, tractography, TractSeg, deep learning

1. Introduction

The acoustic radiation (AR) is a white matter fiber bundle that connects the Heschl's gyrus (HG) in the cortex with the medial geniculate nucleus (MGN) in the mid-brain (1, 2). The AR is one of the most important fiber bundles of the auditory system (3), and its analysis is relevant for understanding the mechanisms of acoustic stimuli processing and how they are affected by different diseases. For example, diseases such as tinnitus (4, 5), schwannoma (6), and putaminal hemorrhage (7, 8) have been associated with changes in the AR. Reliable methods for extracting the AR are crucial for performing such analyses.

Extracting the AR with tractography from diffusion MRI (dMRI) is challenging (9). First, the AR is a relatively short bundle of approximately 4–6 cm (2), making it especially sensitive to the low resolution of standard imaging acquisitions used in clinics. Second, the AR is very close to other bundles such as the cortico-spinal tract (CST), arcuate fasciculus (AF), the middle longitudinal fasciculus (MLF), the inferior fronto-occipital fasciculus (IFOF), and the optic radiation (OR) (10–12). We have also found that the AR is close to the inferior longitudinal fasciculus (ILF) in some cases. This closeness to other bundles can make it difficult for the tractography method to extract streamlines only related to the AR. Low-resolution dMRI might be unable to disentangle the crossing and kissing fiber bundles from the intersection regions along the AR. This has also been reported as a problem for segmenting neighboring fiber bundles (12). Moreover, MGN, HG, and AR have a large variability among subjects (2, 9, 11, 13).

The fiber bundle connecting the MGN with the HG can be considered the core of the AR. In their review, Maffei et al. (9) discussed that, in addition to the core of the AR, there is evidence from *ex vivo* studies on macaque monkeys that the AR might have extra layers of fibers that create a "belt" that can go beyond the HG and reach the superior temporal gyrus (STG) (14, 15). The core and this belt of the AR are thought to have different functions. The core of the AR might be involved in basic tone processing. In contrast, the belt might be involved in integrating auditory information with other sensory information. Since their purpose is different, neurological and auditory conditions can affect the core and the belt of the AR differently. Thus, having independent segmentation masks for the core and the belt is relevant for further analyses. In this article, we focus on generating segmentation masks of the core of AR.

Different atlases of AR have been proposed in the literature. For example, Bürgel et al. (2) used histology to create a high-resolution atlas of different fiber bundles of the white matter from ten donors, including the AR. More recently, Maffei et al. (16) created an atlas using dMRI acquisitions with ultra-high b-values (up to 10,000 s/mm^2) and high resolution (1.5 mm isotropic) from the MGH adult diffusion dataset of the human connectome project (HCP) (17, 18). However, as already mentioned, the use of atlases of AR is not ideal due to its reported anatomical variability (1, 2, 9, 16, 19).

Two automatic tools include the segmentation of the AR: XTRACT (20, 21) and TRACULA (22). XTRACT is a tool of the FMRIB Software Library (FSL) (23) that can segment 42 fiber bundles, including the AR. In order to segment the AR, XTRACT runs probabilistic tractography between the HG and the MGN and defines exclusion masks to remove anatomically implausible streamlines. In particular, it uses two coronal planes and an axial plane around the thalamus, a region covering the optic tract and the brainstem as exclusion masks. XTRACT also provides an atlas of the AR based on the HCP young adult dataset (24, 25) and the UK Biobank dataset (26). One potential issue of XTRACT is that its exclusion criteria might be too liberal with respect to knowledge from neuroanatomists (9, 10). Thus, there is a risk that segmentation masks might cover areas that should not be part of the AR.

TRACULA (27) is a tool of FreeSurfer (28) for fiber bundle segmentation. This method uses prior anatomical information of the fiber bundles to steer a Bayesian-based global tractography. The original method included 18 main fiber bundles and did not include the AR. Maffei et al. (22) extended the number of fiber bundles to 42, including the AR. For this, they manually segmented the 42 fiber bundles in 16 subjects of the MGH adult diffusion dataset of the HCP (17, 18).The new definitions were made available in the latest version of FreeSurfer (version 7.2, release date: July 2021).

Regarding the AR, Maffei et al. (22) used a subset of the segmentation masks used by Maffei et al. (16) to create their atlas of AR. One of the issues of TRACULA for segmenting the AR is that the manual dissections in the 16 subjects include too few streamlines. More specifically, the mean number of streamlines extracted per subject in the MGH dataset was 26 (ranging between 2 and 91) for the left side and 32 (ranging between 6 to 70) for the right side. As a comparison, TRACULA uses an average of 1,250 streamlines per subject (ranging between 333 and 2,726) for the left arcuate fascicle. This low number of streamlines used for the AR has the risk of making TRACULA less specific with respect to anatomical variations of the AR. An additional issue of TRACULA is that it uses global tractography, which makes it very time-consuming compared to other methods. Moreover, TRACULA requires the parcellation generated by FreeSurfer, which usually takes several hours.

Wasserthal et al. (29) proposed TractSeg, a method based on artificial intelligence (AI) that is able to segment 72 main fiber bundles from dMRI automatically. The advantages of this method are that it works with standard dMRI acquisitions, even with low b-values, is fast (takes a few seconds), does not require a previous registration of images, and, unlike atlases, the results are subject-specific. Due to the aforementioned difficulties in segmenting the AR, the original method did not include the AR. More recently, Wasserthal et al. (29) trained the original neural network using the masks generated by XTRACT (20,

21), including the AR. Thus, since version 2.2. of TractSeg, it is possible to obtain these segmentations with the option "*–tract_definition xtract*".

Both XTRACT and TRACULA allow the streamlines to go beyond the HG and reach the STG. This means that these methods are not designed to extract the core of the AR. Thus, the main goal of this paper is to assess the possibility of using TractSeg for the segmentation of the core of the AR in datasets acquired in clinical settings.

2. Methods

2.1. Datasets

We used two datasets in this study. The first one consists of dMRI data from 125 subjects of the HCP young adult dataset (24, 25). A total of 105 of these subjects are exactly the same used by Wasserthal et al. (29) and were used for training the TractSeg (29) models with masks generated using the segmentation methodology proposed in this paper, while the remaining 20 were used for independent testing. The dMRI data of HCP consists of 90 directions for each of the three *b*-values: 1,000, 2,000, and 3,000 s/mm^2, and the spatial resolution is 1.25 *mm* isotropic. These images were acquired in Siemens 3T scanners using a spin-echo EPI sequence with a multiband factor of 3, TR/TE is 5,520/89.5 ms, a flip angle of 78 degrees, and a refocusing flip angle of 160 degrees. The images were acquired using a head coil with 32 channels. More details on imaging parameters are available on the website of HCP[1]. The second dataset consists of dMRI data of 34 subjects acquired with the following parameters: isotropic resolution of 2.3 *mm* and 60 directions at b = 1,000 s/mm^2. The images were acquired at the MRI facility of Karolinska Institute at Karolinska University Hospital in Solna using a GE Discovery 3T MR750 scanner with a spin-echo EPI sequence with TR/TE of 7,000/80.9 ms and flip angle of 90 degrees. The images were acquired using a head coil with 8 channels. The cohort of this dataset consists of 17 patients with unilateral congenital ear canal atresia and 17 age- and gender-paired controls. The patients are adults with contralateral normal hearing, had no hearing aid or successful ear canal surgery before age 12, and have sufficient understanding of the Swedish language. Subjects with a history of severe psychiatric illness or neurological disease, any associated syndrome (Goldenhaar, CHARGE, etc.), or metallic artifacts were excluded from the cohort. In twelve of the patients, the right ear is affected. Eight of the patients are female and nine are male. The patients were all recruited in the Stockholm region. The ethical permit was granted by the Swedish ethical board (Dnr 2012/1661-31/3). The clinical dataset was pre-processed with the standard pre-processing pipeline of MRtrix3 (30) to remove artifacts and geometric distortions, which in turn uses methods from FSL (23).

1 https://www.humanconnectome.org/hcp-protocols

2.2. TractSeg

TractSeg is a method that trains deep neural networks for segmenting fiber bundles (29). Figure 1 shows the pipeline of TractSeg. The steps of TractSeg are the following. First, the dMRI data must be pre-processed to remove artifacts and geometric distortions. Notice that this step is not required for HCP data since this dataset is already pre-processed (25). The clinical dataset was pre-processed with the tools provided in MRtrix3 (30). Second, fiber orientation distribution functions (fODF) are estimated per voxel using constrained spherical deconvolution (CSD) (31). The maxima (also known as peaks) of the fODFs can be seen as estimations of the most likely orientation fiber bundles in every voxel. Thus, the next step is to extract the largest peaks of the fODFs per voxel. Every peak is a vector whose direction and magnitude encode the most likely orientation of a fiber bundle and its strength, respectively. This strength, among many factors, is related to the density of fibers at the specific orientation of the peak. TractSeg assumes that a maximum of three fiber bundles can traverse a voxel. Thus, only the three largest peaks are input to the neural network. Notice that the magnitude of only one peak is not negligible in regions traversed by a single fiber bundle and two for those with two crossing fiber bundles. We used the option "–super_resolution" from TractSeg, which upsamples the peaks to an isotropic resolution of 1.25 mm.

Expert neuroanatomists manually segmented 72 different fiber bundles in 105 HCP subjects. These segmentations were used in TractSeg to train U-Net-like neural networks (32). As shown in Figure 1, TractSeg uses 2D neural networks (one per axis) in two stages. The first stage is used to generate masks of the fiber bundles by only considering the 2D information contained in the training slices. The second stage is used to learn the best combination to generate the final segmentation of the 72 fiber bundles. Notice that TractSeg uses a so-called 2.5D approach, that is, segmenting 3D structures with multiple 2D neural networks. Although it is possible to use 3D U-Nets instead, the authors argue that a 2.5D approach is more efficient and less prone to overfitting (29), which is in agreement with studies dealing with other segmentation problems [e.g., (33)].

TractSeg can be seen as a powerful method that can be used out-of-the-box to segment 72 fiber bundles (29). One of the main advantages of TractSeg is that, although it was trained on high-quality data [HCP young adult dataset (24, 25)], the neural network is also able to segment these bundles in dMRI data of clinical quality without any need for training. This is because the 72 targeted fiber bundles are relatively big. It is interesting to assess whether or not TractSeg can achieve the same performance with smaller fiber bundles, specifically the AR

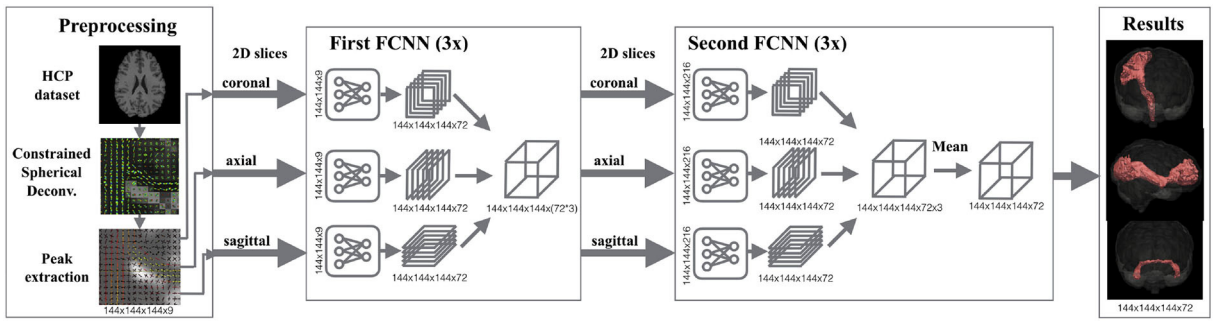


FIGURE 1
Segmentation pipeline of TractSeg. Left: The dMRI data is pre-processed for extracting the peaks of the fiber orientation distribution functions per voxel. These peaks are used as the input of the neural network. Middle: 2D U-Net-like fully convolutional neural networks (FCNNs) are trained to segment fiber bundles. Three networks are trained per axis (coronal, axial, sagittal) in two stages. While the goal in the first stage is to segment the fiber bundles using 2D information, the second stage aims at learning the best combination of the three intermediate results to generate the final segmentation. Right: Segmentation masks of 72 fiber bundles are generated. Figure reproduced from Wasserthal et al. (29), license CC BY 4.0.

in clinical data. Thus, we generated training data for the AR from the same 105 HCP subjects used in TractSeg as described in the following section.

Although TractSeg does not include the core of the AR, it can be trained for that purpose (29). The training procedure requires the segmentation of the new fiber bundles of interest, ideally using the same dataset of the original article. Following the same approach of TractSeg, we used five-fold cross-validation with 105 subjects: 63 training subjects, 21 validation subjects, and 21 test subjects per fold. An additional set of 20 subjects was used for independent testing. As mentioned, newer versions of TractSeg have the option of using segmentation masks from XTRACT, including the AR. However, these segmentations consider not only the core but also can contain fiber bundles reaching the STG.

By design, TractSeg is able to segment fiber bundles beyond the original 72. For this, it is crucial to use high-quality segmentation masks of the new bundles during training. The following subsection describes the proposed methodology for generating such segmentation masks for AR.

2.3. Generation of training data

Probabilistic tractography (iFOD2) with anatomically-constrained tractography (ACT) (34) from MRtrix3 (30) was used for creating streamlines connecting the left HG to the left MGN and the right HG to the right MGN targeting the left and right AR, respectively. Masks of the HG and MGN at both hemispheres extracted with FreeSurfer (28) are available in the HCP database and were used as independent seeds for tractography. Thus, two sets of streamlines were obtained per side: one for streamlines starting at the HG and ending at the MGN and the other reversing the roles of two masks. We used the command "*tckgen*" in MRTrix3 (30) with the default parameters of iFOD2. Moreover, we used the options from ACT "- *backtrack*", which tries to re-track partially truncated streamlines, and "- *crop_at_gmwmi*", which crops the streamlines once they cross the boundary between gray and white matter.

As mentioned, one of the challenges in obtaining the AR is that it is very close to other fiber bundles, as shown in Figure 2. Our approach to tackling this issue is to reject any streamline reaching segmentation masks of nearby fiber bundles. In particular, we used the masks of the CST, IFOF, and ILF created by Wasserthal et al. (29) for training TractSeg to reject implausible AR streamlines.

As shown in Figure 2, the AF, OR, and MLF are too close to the AR that even some voxels can contain streamlines of different bundles. Thus, masks of AF, IR, and MLF cannot be used to reject implausible AR streamlines. Instead, we removed the voxels from these masks that are closer than 4 cm from both the HG and the MGN and used them to reject implausible AR streamlines. With this procedure, streamlines are allowed to enter the voxels close to the MGN and HG, which are also covered by the AF, OR, and MLF segmentation masks.

An additional problem is that the HG and the superior temporal gyrus (STG) are very close to each other, as shown in Figure 3. Due to the closeness between the HG and the STG, some streamlines can leak to the latter, especially when the MGN is used as the origin of the streamlines. In order to avoid this from happening, we used the mask of the STG extracted with FreeSurfer, which is available in the HCP database, to reject streamlines not ending in the HG. This step is crucial to remove possible streamlines not belonging to the core of the AR.

Notice that the described restrictions for generating streamlines are stringent and make the generation of training data computationally expensive. Actually, around 150,000 generated streamlines were discarded per every single accepted one. Thus, as stopping criteria, we set a maximum of 1,000

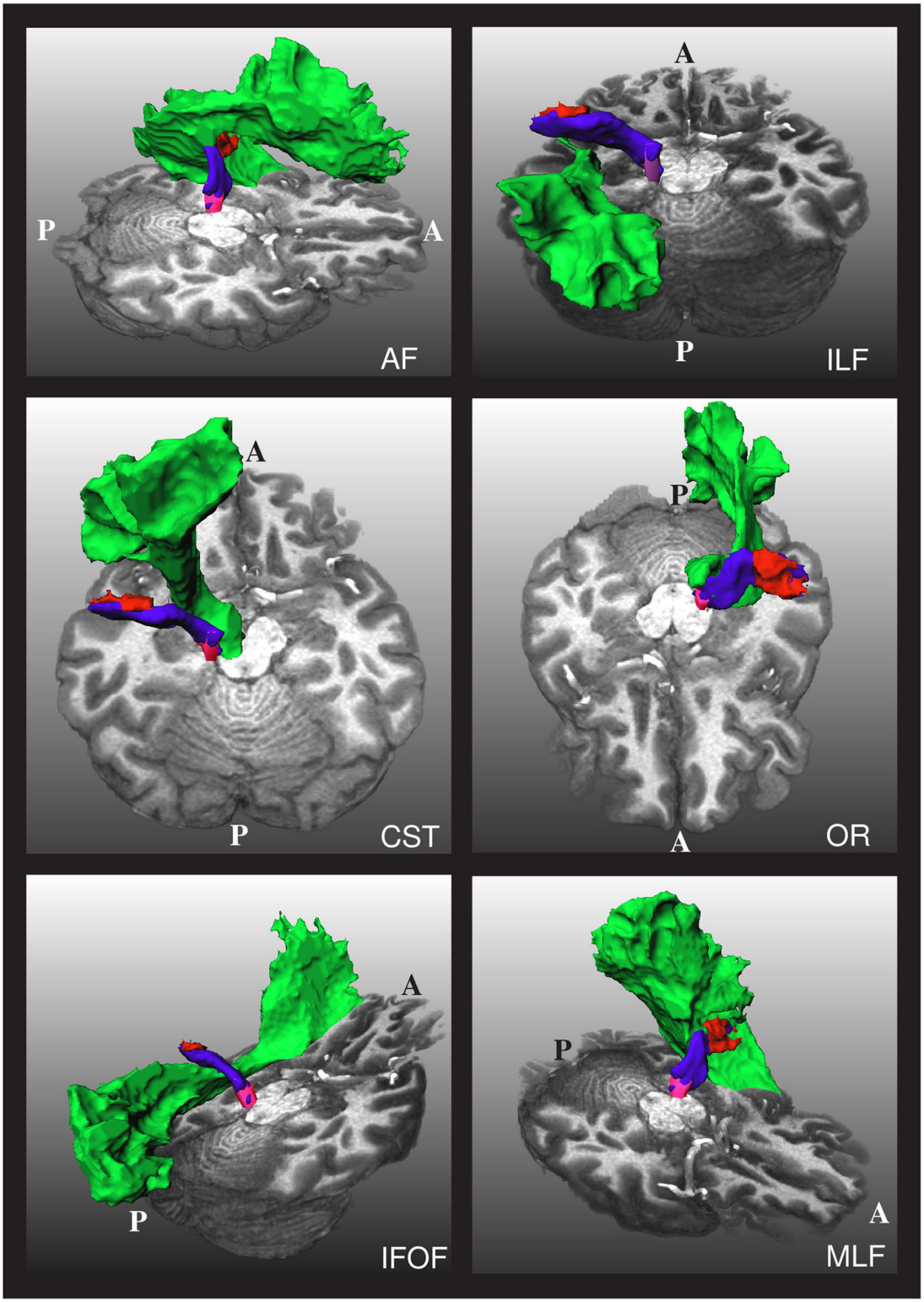


FIGURE 2
The relative position of the left acoustic radiation with six nearby fiber bundles for a subject of the human connectome project. The Heschl's gyrus, medial geniculate nucleus, and acoustic radiation of the left side of the brain are depicted in red, magenta, and blue, respectively. Each of the nearby fiber bundles is depicted in green, one per subfigure. A and P indicate the anterior and posterior sides of the brain, and T1w is used as a reference. The depicted acoustic radiation was computed using the methodology of Section 2.3.

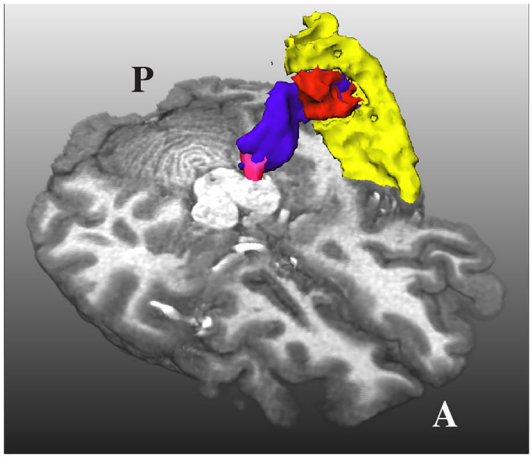


FIGURE 3
The acoustic radiation (in blue) from the medial geniculate nucleus (in magenta) and the Heschl's gyrus (in red) is also very close to the superior temporal gyrus (in yellow). A and P indicate the anterior and posterior sides of the brain, and T1w is used as a reference.

accepted streamlines, or 150 million generated streamlines in total per seed mask. The maximum length of each streamline was set to 60 mm. The two sets of streamlines per side were combined into a single tractogram. This procedure resulted in tractograms of at least 1,000 streamlines per side of the brain. Finally, a mask of the AR per side was created with the voxels traversed by at least ten streamlines. This procedure was successful in all HCP subjects.

It is important to emphasize that the original article of TractSeg (29) used whole-brain tractograms, each with 10 million streamlines with lengths between 40 and 250 mm. From these streamlines, only a few were part of the AR (fewer than 20 in all cases), which are not enough to generate reliable segmentation masks. The proposed procedure for generating streamlines of the core of the AR is expensive but effective for generating the masks that were used for training TractSeg.

3. Results

This section shows the results of the proposed methodology for segmenting the core of the AR applied to HCP data and the diffusion data acquired in a clinical setting on 17 patients with unilateral congenital ear canal atresia and 17 age- and gender-paired controls.

3.1. High-quality diffusion data

Figure 4 shows the curves of the F1 score during validation and testing on HCP data. The best performing network attained an F1 score of 0.73 during testing. The F1 score is equivalent to the Dice score for segmentation purposes.

We tested the trained network in 20 additional HCP subjects not used for training. As shown in Figures 5, 6 for one of these subjects, the segmentation results of the core of the AR at both sides are anatomically plausible. From the figure, it can be seen that there are differences between atlases. The segmentation generated from our methodology is more conservative than the atlases and XTRACT. For example, the generated segmentation masks always stop at the boundary between white matter and the HG, while, e.g., (2) usually overlaps with the HG and is more likely to reach the STG. Most of the generated masks of AR overlap with the two atlases and XTRACT.

As shown in Figure 6, the atlases and XTRACT tend to reach regions of the STG (see yellow arrows), sometimes in regions not adjacent to the HG. It can also be seen that the segmentation masks differ from each other, especially in the region close to the HG.

Using visual inspection, we found that the proposed methodology was able to extract anatomically plausible AR in all 20 subjects used for independent testing.

3.2. Diffusion data acquired in a clinical setting

We applied the trained network on dMRI data of 17 subjects with unilateral ear canal atresia and 17 controls. As mentioned, these images were acquired in a clinical setting ($b = 1{,}000 s/mm^2$, 60 directions, spatial resolution = 2.3 mm isotropic). This case is more challenging than the segmentation of the HCP data due to the low spatial and angular resolution and the relatively low b-value used in the acquisition. Table 1 shows the number of cores of the ARs that were completely reconstructed, were reconstructed in fragments, or where the method failed. As shown, the method was able to completely reconstruct the core of the AR in most cases (53/68 = 77.9%) with a similar performance between patients and controls (24 vs. 29). The method yielded fragmented cores of the ARs in 14 cases (20.5%) and more often in patients than in controls (9 vs. 5). The fragments were visually inspected. In most cases, the core of the AR was fragmented into two pieces, each of them closer to either the MGN or the HG. In a few cases, the core of the AR appeared as a blob in the middle between the MGN and the HG. In the 14 cases, the fragments were always located at the region where the AR is expected to be. The method only failed to reconstruct the left AR of a single patient. The trained network was also more consistent in yielding uncut segmentations on the left side (2 cases on the left vs. 12 on the right).

In the cases where TractSeg was not able to extract the complete core of the AR, it is possible to use the masks to guide

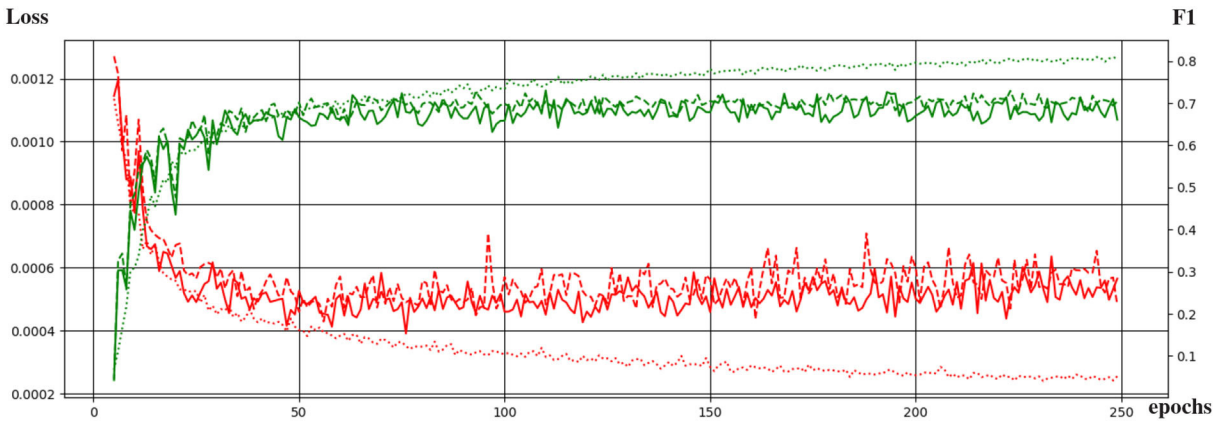


FIGURE 4
Evolution of the training of the neural network with the training epochs. The loss function and the F1 score are shown in red and green, respectively. Dotted, continuous, and dashed lines correspond to performance during training, validation, and testing.

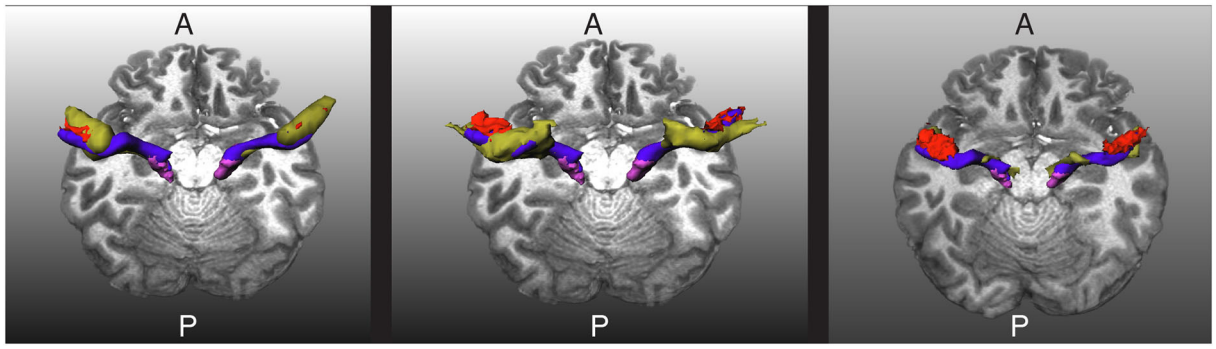


FIGURE 5
Visualization of the extracted acoustic radiation for one subject from the human connectome project in blue. The Heschl's gyrus and medial geniculate nucleus are depicted in red and magenta, respectively. Left: The atlas from Bürgel et al. (2) is shown as a reference in yellow. Middle: The atlas from Maffei et al. (16) is shown as a reference in yellow. Right: The segmentation obtained with XTRACT (20) is shown in yellow as a reference. A and P indicate the anterior and posterior sides of the brain, and T1w is used as a reference.

tractography. For this, not only the MGN and the HG are used as seed regions, but also the results of the segmentation with TractSeg. This makes it more likely for tractography to compute streamlines that comply with the strict restrictions described in Section 2.3. Figure 7 shows the results obtained for some of the subjects.

Figure 8 shows a visual comparison of the segmentation masks obtained with the proposed methodology, the atlases by Bürgel et al. (2) and Maffei et al. (16), and XTRACT for one subject from the clinical dataset where the methodology was able to extract the core of the AR. As shown, the atlases and XTRACT tend to reach more the STG. Except for the atlas by Bürgel et al. (2), the other methods have problems entering the cavity of the HG in this specific subject.

The extracted segmentation masks can be used for different group analyses. Among many other options, one can use the masks to restrict tractography and perform bundle analytics (35). To showcase this application, we used the implementation of TractSeg for bundle analytics. In brief, the method runs tractography, but unlike the procedure described in Section 2.3, the generated streamlines are only restricted to traversing the segmentation mask of the AR. Using the AR masks is much less restrictive than using the neighboring fiber bundle masks and, thus, is much less time-consuming (ca. 10–20 min. per subject). Then, the generated streamlines are used to sample the maps of fractional anisotropy (FA) or any other measurement along the path of the streamlines. This way, it is possible to assess differences between the groups along the trajectory of the AR. Figure 9 shows a bundle analysis of the FA applied to the AR for the clinical dataset. As shown, the FA starts at a very low value at the MGN, goes up in the middle, and down again to the end close to the Heschl's gyrus. It can be seen that the 95% CIs (shown with colored bands) are relatively large. In fact, these CI were 2–3 times larger than for the cortical

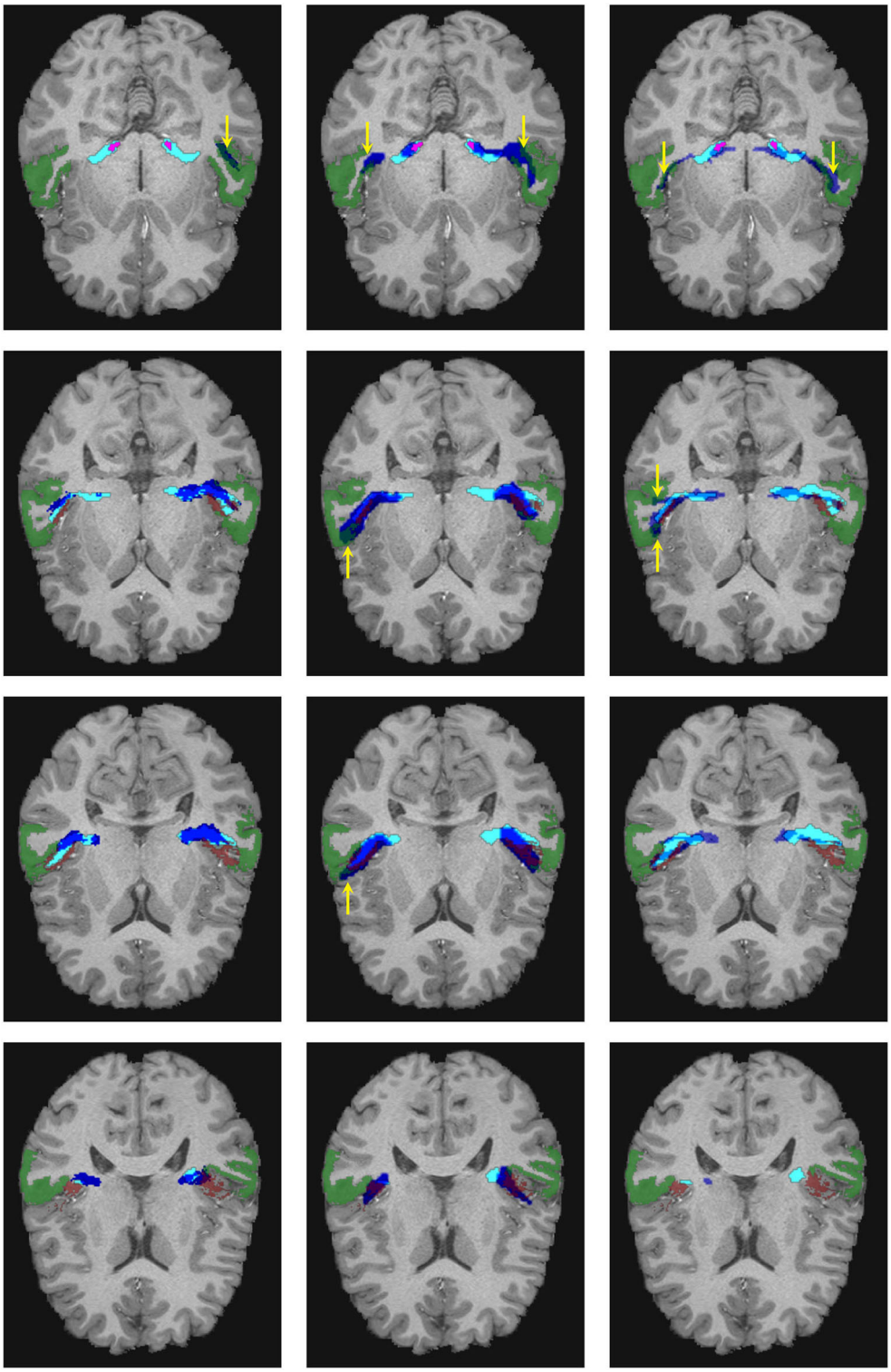


FIGURE 6
Visual comparison of the segmentation masks in one subject of the human connectome project. First column: Segmentation mask of the proposed methodology (in cyan) and the atlas by Maffei et al. (16) (in blue). Second column: Segmentation mask of the proposed methodology (in cyan) and the atlas by Bürgel et al. (2) (in blue). Third column: Segmentation mask of the proposed methodology (in cyan) vs. the result from XTRACT (in blue). Every row corresponds to a different axial slice. The superior temporal gyrus (STG), medial geniculate nucleus, and Heschl's gyrus are depicted in green, magenta, and brown, respectively. Yellow arrows indicate where the segmentation masks reach the STG.

TABLE 1 The number of subjects in which the proposed methodology was able to reconstruct the complete acoustic radiation (AR) (Uncut), split the AR into fragments (Fragm.), or completely failed (Fail) per side in the clinical dataset of unilateral ear canal atresia.

	Left AR			Right AR			ARs of both sides		
	Uncut	Fragm.	Fail	Uncut	Fragm.	Fail	Uncut	Fragm.	Fail
Patients R ($N = 12$)	10	1	1	8	4	0	29	5	1
Patients L ($N = 5$)	4	1	0	2	3	0	6	4	0
All Patients ($N = 17$)	14	2	1	10	7	0	24	9	1
Controls ($N = 17$)	17	0	0	12	5	0	29	5	0
All subjects ($N = 34$)	31	2	1	22	12	0	53	14	1

Patients R and Patients L indicate the side of the affected ear.

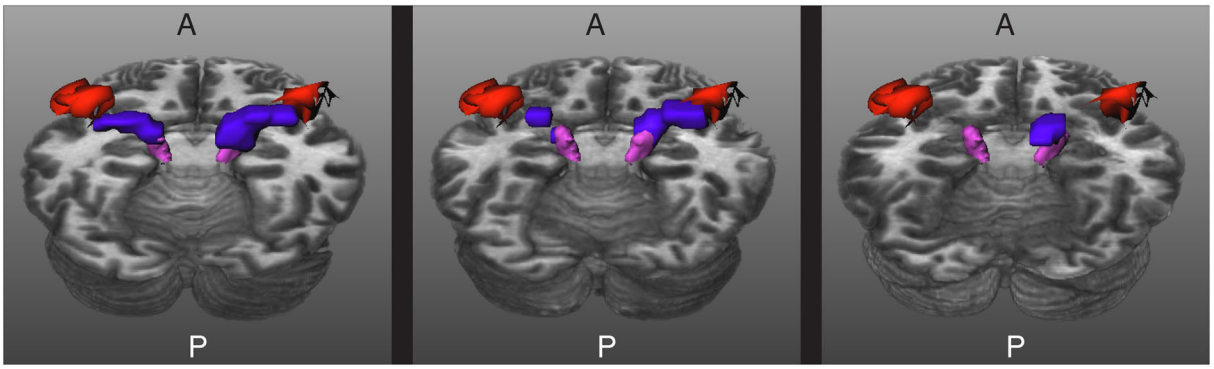


FIGURE 7
Results for three images acquired in a clinical setting. The core of the acoustic radiations (ARs) are depicted in blue, the Heschl's gyrus (HG) in red, and the medial geniculate nucleus (MGN) in magenta. Left: The core of the ARs are completely extracted. Middle: The core of the ARs are fragmented into two pieces. Right: The method gave a blob in between the MGN and the HG for the right side and was unable to segment the core of the AR of the left side. A and P indicate the anterior and posterior sides of the brain, and T1w is used as a reference.

spinal tract (CST) and other large tracts. This could mean that the intersubject variability is higher for the AR than for large fiber bundles. We performed t-tests along the tract that were corrected for multiple comparisons to account for family-wise errors. With this procedure, we did not find any statistically significant difference between the two groups at any point along the tract.

4. Discussion

Previous studies have shown that extracting the AR is possible *in vivo* on data from the MGH adult diffusion dataset of HCP with ultra-high b-values up to 10,000 s/mm^2 (16). In this study, we showed that extracting the core of the AR in high-quality dMRI data with lower b-values (b = 1,000, 2,000, and 3,000 s/mm^2) from the HCP young adult dataset by using masks of neighboring fiber bundles is also possible. One issue of our approach is that our strategy is very restrictive and time-consuming.

Thus, in order to reduce the computation time, we trained the neural network of TractSeg (29) with the segmentation masks of the core of the AR created from HCP data. There are two main advantages of using TractSeg for segmenting the AR compared to using atlases: (a) that the resulting masks are subject-specific, and (b) it is not necessary to do registration to a template. Regarding the former, subject-specific masks can tackle the anatomical variability of the AR, HG, and MGN. As for the latter, misregistrations can generate errors that are not a problem for TractSeg. An alternative to using TractSeg is to generate the core of the AR as proposed in Section 2.3. The main gain of using TractSeg is that the segmentation mask is obtained in a few seconds instead of several hours of the proposed methodology from Section 2.3.

The trained neural network of TractSeg was able to segment the core of the AR in HCP data in a few seconds instead of several hours. We used a workstation equipped with an Intel Xeon CPU E5-2630 v3 with 8 cores at 2.40 GHz, and a GPU NVIDIA GeForce GTX 1070. The processing of one HCP subject using the methodology described in Section 2.3 was 8–10 h in this workstation. Computing the peaks of the fODFs took

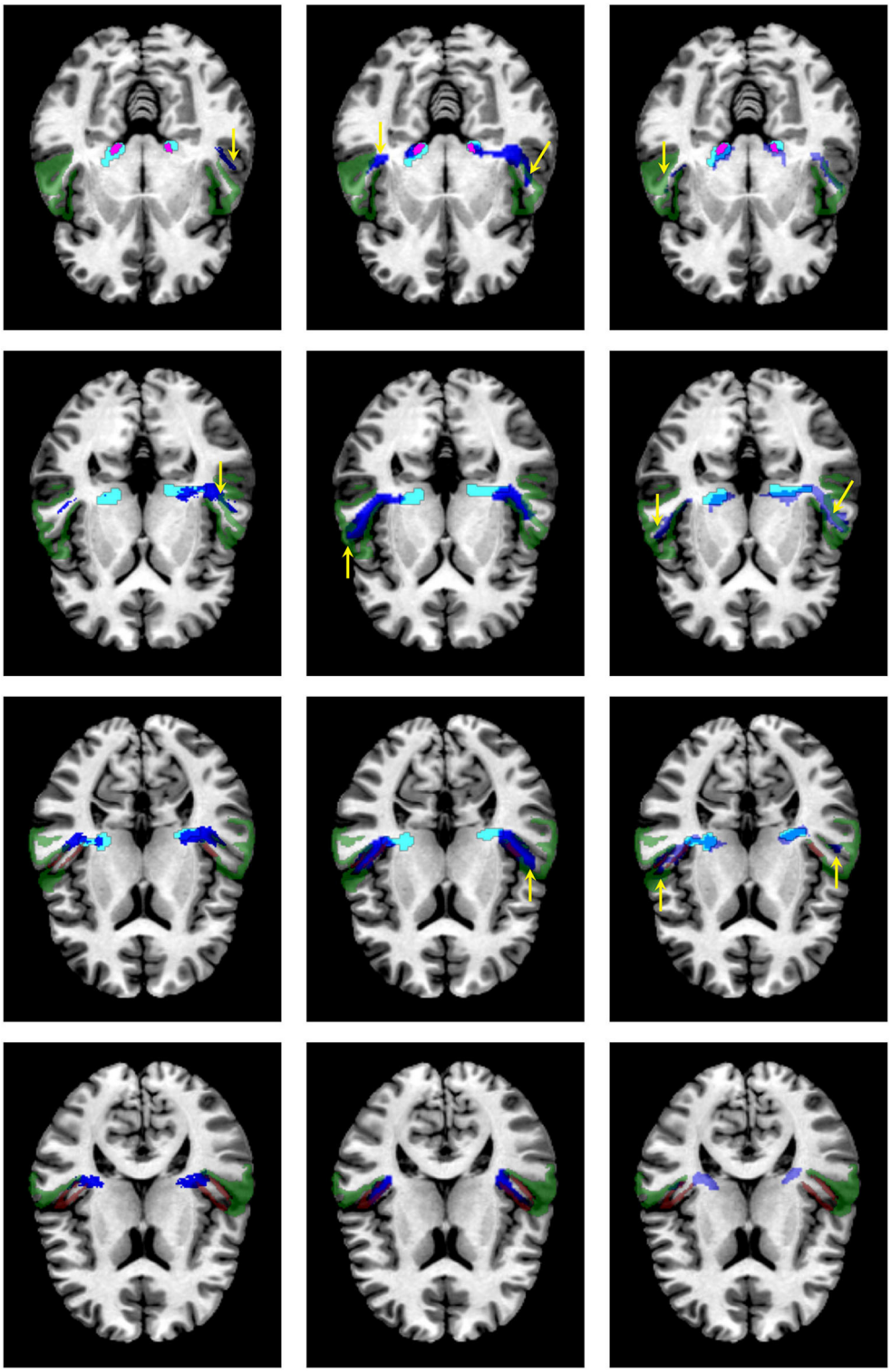


FIGURE 8
Visual comparison of the segmentation masks on one subject of the clinical dataset. First column: segmentation mask of the proposed methodology (in cyan) and the atlas by Maffei et al. (16) (in blue). Second column. segmentation mask of the proposed methodology (in cyan) and the atlas by Bürgel et al. (2) (in blue). Third column: segmentation mask of the proposed methodology (in cyan) vs. the result from XTRACT (in blue). Every row corresponds to a different axial slice. The superior temporal gyrus (STG), medial geniculate nucleus, and Heschl's gyrus are depicted in green, magenta, and brown, respectively. Yellow arrows indicate where the segmentation masks reach the STG.

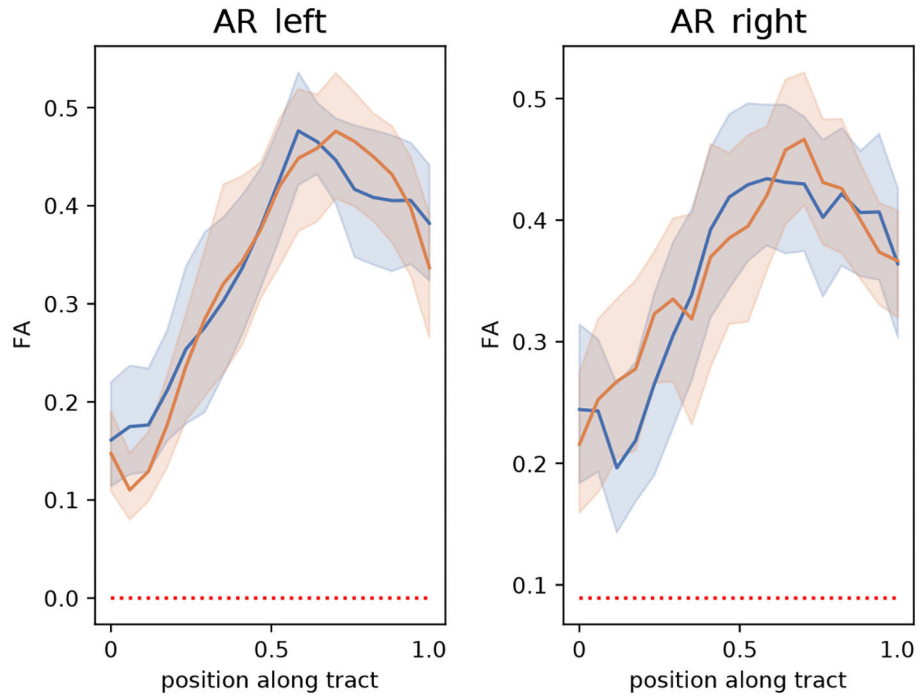


FIGURE 9
Bundle analysis of the fractional anisotropy (FA) applied to the left and right acoustic radiations (AR) for the clinical dataset used in this paper. The mean FA of patients and controls along the tracts are shown with lines in blue and orange, respectively. The 95% CIs are shown in light blue and light orange bands, respectively for the two groups. Position 0 and 1 along the tract are located at the medial geniculate nucleus and the Heschl's gyrus, respectively.

approximately 1 min and applying the trained neural network took around 40 s for both the HCP data and the clinical data. The segmentations generated by the trained neural network were anatomically plausible when applied to an independent set of subjects from HCP. The methodology proposed in Section 2.3 is conservative. Thus, the segmentation masks obtained with the neural network are also conservative compared to the publicly available atlases of the AR. We argue that it is important to have a conservative approach to extracting the core of AR. This way, the downstream conclusions drawn from group analyses of the AR will become more meaningful.

The trained neural network had more problems with data acquired in a clinical setting. Still, it was able to completely segment the core of the ARs in 77.9% of the cases, yielded fragmented masks in 20.6% of the cases, and only failed in a single subject. The performance was very similar in patients and controls. The neural network tended to reconstruct the core of the left AR better than the core of the right AR.

As shown in some cases, the neural network yields a fragmented segmentation. Such fragments can be used as seeds for tractography, which has the advantage of reducing the high cost of running tractography to extract the core of the AR.

We compared the proposed methodology with the segmentation generated by TractSeg (29) trained with masks created with XTRACT (20, 21). From the results, an important difference between our methodology and XTRACT is that the latter included tracts that reached the STG in the segmentation masks. It is important to differentiate the fibers connecting only the MGN and the HG from those that can get the STG, as they can have different purposes in the human brain (9). For example, Ito et al. (36) reported that the STG might be involved in the joint processing of visual and auditory stimuli. Unlike XTRACT, the proposed methodology actively removes the fibers reaching the STG to target the core of the AR. At this stage, it is not possible to know if the fibers covered by XTRACT and not covered by our methodology belong to the belt of the AR. The STG is a structure that is larger compared to the HG. Thus, it is not clear which substructures of the STG might be part of the AR. Such information is crucial to assess whether the voxels reaching the STG by the masks of XTRACT belong to the AR or are artifacts.

Unlike our methodology, XTRACT was able to generate the AR in all cases. Since XTRACT uses less restrictive rules for generating the masks, they cover more voxels, which makes TractSeg increase its robustness at the cost of being less specific. In some cases, the XTRACT masks covered parts of the ventricles and the most posterior parts of the STG, almost reaching the medial temporal gyrus. Thus, we

recommend a manual review of these masks before any further analysis.

Previously, Bertó et al. (37) added prior information for improving the segmentation of fiber bundles. Our results are in line with that study since we show that adding the segmentation masks of other bundles is needed for the segmentation of small fiber bundles like the AR.

We showcased the use of segmentation masks by performing a bundle analysis on the clinical dataset to assess differences in FA between patients and control in the AR. We did not find any statistically significant difference between the groups. The 95% CI was larger than other bundles (e.g., the CST). This suggests that the intersubject variability is higher for the AR.

The results of this study are encouraging but also show that more research is needed toward a fully automatic segmentation of the AR from images acquired in clinical settings. For example, as mentioned, TractSeg uses three peaks of the fODFs (29). Recently, it has been argued that up to seven fiber bundles might appear in certain brain regions (38). Thus, it is possible that more peaks could be helpful for extracting the AR. However, enlarging the number of inputs to the neural network has the disadvantage of needing more training data or changing the neural network architecture, which is beyond the scope of this article. Although TractSeg (29) can still be considered state-of-the-art for fiber bundle segmentation, new AI-based segmentation methods have recently been proposed [e.g., (39–42)]. It is interesting to assess if adapting these methods can yield better results for segmenting the AR. Plans for the future also include the analysis of the AR for other diseases affecting the auditory system and datasets acquired in different clinical settings.

This study has many limitations. One of the main issues is that there is not possible to have a personalized ground truth that can be used to assess the accuracy. This is a general limitation of any method based on tractography. The atlas by Bürgel et al. (2) was created from histology and is expected to depict the anatomy of AR better. However, the variability of the HG, MGN, and the AR among subjects, makes it less appropriate for group analyses. A second limitation is that although FreeSurfer is relatively accurate for segmenting the HG [Desikan et al. (43) reported intraclass correlations between automatic and manual segmentations of 0.712 and 0.719 for the left and right HG, respectively], it can be inaccurate in cases where the HG has duplications. Marie et al. (44) found in a cohort with 430 participants that 36.6 and 48.8% of the right-handed subjects and 30.8 and 39.4% of the left-handed subjects had duplications on the left and right side, respectively. Considering duplications of the HG in the pipeline is clinically relevant since they have been associated with neurological conditions (45). In order to account for this anatomical variability of the HG, it would be necessary not only to use during training more accurate segmentation tools tailored explicitly for the HG [e.g., TASH (46)] but also to train independent TractSeg models for subjects with and without duplications in the HG. The most appropriate TractSeg model for a specific subject could be chosen once the type of HG is detected. Still, it is uncertain whether such an approach could lead to differences in AR.

5. Conclusion

In this study, we proposed a methodology to extract the core of the AR in subjects from the HCP young adult dataset by using masks of neighboring fiber bundles obtained with TractSeg. Since the procedure is expensive, we trained TractSeg to extract the AR automatically. For this, we used the masks of the AR extracted from a set of subjects from the HCP young adult dataset. The trained neural network was applied both to unseen subjects of the HCP young adult dataset and a clinical dataset.

The main conclusion of this study is that it is possible to segment the core of the AR in most cases, even in images acquired in clinical settings in a few seconds with the trained network. In case it is not possible to reconstruct the core of the AR, the results can be used as masks for tractography.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: We used data from the Human Connectome Project (HCP), WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The training data and the trained neural network is available at https://doi.org/10.5281/zenodo.7052849. The data from the Karolinska Institute cannot be shared. Requests to access these datasets should be directed to RM, rodmore@kth.se.

Ethics statement

The Human Connnectome Project (HCP) data is publicly available and all authors have accepted the HCP Open Access Data Use Terms. The acquisition of the dataset from Karolinska Institute was reviewed and approved by the Swedish Ethical Board (Etikprövningsmyndigheten) Dnr 2012/1661-31/3. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MS: conceptualization, data curation, investigation, methodology, and writing—review and editing. CE:

conceptualization, methodology, resources, writing—review and editing, supervision, and funding acquisition. RM: conceptualization, methodology, visualization, resources, project administration, writing—original draft, review and editing, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was partially supported by VINNOVA, through AIDA, the Center for Innovative Medicine (CIMED), Region Stockholm, and Digital Futures, Project dBrain.

Acknowledgments

We thank Blanca Bastardés Climent for performing the initial tests for generating the training data and Chiara Maffei for her advice in using the atlas of the AR from Maffei et al. (16).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Maffei C, Jovicich J, De Benedictis A, Corsini F, Barbareschi M, Chioffi F, et al. Topography of the human acoustic radiation as revealed by *ex vivo* fibers microdissection and *in vivo* diffusion-based tractography. *Brain Struct Func.* (2018) 223:449–59. doi: 10.1007/s00429-017-1471-6

2. Bürgel U, Amunts K, Hoemke L, Mohlberg H, Gilsbach JM, Zilles K. White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *Neuroimage.* (2006) 29:1092–105. doi: 10.1016/j.neuroimage.2005.08.040

3. Cusack R, Wild CJ, Zubiaurre-Elorza L, Linke AC. Why does language not emerge until the second year? *Hear Res.* (2018) 366:75–81. doi: 10.1016/j.heares.2018.05.004

4. Jaroszynski C, Attyé A, Job A, Delon-Martin C. Tracking white-matter brain modifications in chronic non-bothersome acoustic trauma tinnitus. *Neuroimage Clin.* (2021) 31:213–8. doi: 10.1016/j.nicl.2021.102696

5. Koops EA, Haykal S, Van Dijk P. Macrostructural changes of the acoustic radiation in humans with hearing loss and tinnitus revealed with fixel-based analysis. *J Neurosci.* (2021) 41:3858–965. doi: 10.1523/JNEUROSCI.2996-20.2021

6. Rueckriegel SM, Homola GA, Hummel M, Willner N, Ernestus RI, Matthies C. Probabilistic fiber-tracking reveals degeneration of the contralateral auditory pathway in patients with vestibular schwannoma. *Am J Neuroradiol.* (2016) 37:1610–6. doi: 10.3174/ajnr.A4833

7. Tokida H, Kanaya Y, Shimoe Y, Imagawa M, Fukunaga S, Kuriyama M. Auditory agnosia associated with bilateral putaminal hemorrhage: a case report of clinical course of recovery. *Clin Neurol.* (2017) 57:441–5. doi: 10.5692/clinicalneurol.cn-001046

8. Koyama T, Domen K. A case of hearing loss after bilateral putaminal hemorrhage: a diffusion-tensor imaging study. *Prog Rehabil Med.* (2016) 1:20160003. doi: 10.2490/prm.20160003

9. Maffei C, Sarubbo S, Jovicich J. A missing connection: a review of the macrostructural anatomy and tractography of the acoustic radiation. *Front Neuroanat.* (2019) 13:27. doi: 10.3389/fnana.2019.00027

10. Fernández L, Velásquez C, García Porrero JA, de Lucas EM, Martino J. Heschl's gyrus fiber intersection area: a new insight on the connectivity of the auditory-language hub. *Neurosurg Focus.* (2020) 48:E7. doi: 10.3171/2019.11.FOCUS19778

11. Javad F, Warren JD, Micallef C, Thornton JS, Golay X, Yousry T, et al. Auditory tracts identified with combined fMRI and diffusion tractography. *Neuroimage.* (2014) 84:562–74. doi: 10.1016/j.neuroimage.2013.09.007

12. Latini F, Trevisi G, Fahlström M, Jemstedt M, Alberius Munkhammar Å, Zetterling M, et al. New insights into the anatomy, connectivity and clinical implications of the middle longitudinal fasciculus. *Front Neuroanat.* (2021) 14:106. doi: 10.3389/fnana.2020.610324

13. Rademacher J, Bürgel U, Zilles K. Stereotaxic localization, intersubject variability, and interhemispheric differences of the human auditory thalamocortical system. *Neuroimage.* (2002) 17:142–60. doi: 10.1006/nimg.2002.1178

14. Hackett TA, Stepniewska I, Kaas JH. Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol.* (1998) 400:271–86. doi: 10.1002/(SICI)1096-9861(19981019)400:2<271::AID-CNE8>3.0.CO;2-6

15. Kaas JH, Hackett TA. Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci USA.* (2000) 97:11793–9. doi: 10.1073/pnas.97.22.11793

16. Maffei C, Sarubbo S, Jovicich J. Diffusion-based tractography atlas of the human acoustic radiation. *Sci Rep.* (2019) 9:1–13. doi: 10.1038/s41598-019-40666-8

17. Fan Q, Witzel T, Nummenmaa A, Van Dijk KRA, Van Horn JD, Drews MK, et al. MGH-USC human connectome project datasets with ultra-high b-value diffusion MRI. *Neuroimage.* (2016) 124:1108. doi: 10.1016/j.neuroimage.2015.08.075

18. Setsompop K, Kimmlingen R, Eberlein E, Witzel T, Cohen-Adad J, McNab JA, et al. Pushing the limits of *in vivo* diffusion MRI for the Human Connectome Project. *Neuroimage.* (2013) 80:220–33. doi: 10.1016/j.neuroimage.2013.05.078

19. Forkel SJ, Friedrich P, Thiebaut de Schotten M, Howells H. White matter variability, cognition, and disorders: a systematic review. *Brain Struct Funct.* (2022) 227:529–44. doi: 10.1007/s00429-021-02382-w

20. Warrington S, Bryant KL, Khrapitchev AA, Sallet J, Charquero-Ballester M, Douaud G, et al. XTRACT - Standardised protocols for automated tractography in the human and macaque brain. *Neuroimage.* (2020) 217:116923. doi: 10.1016/j.neuroimage.2020.116923

21. De Groot M, Vernooij MW, Klein S, Ikram MA, Vos FM, Smith SM, et al. Improving alignment in Tract-based spatial statistics: evaluation and optimization of image registration. *Neuroimage.* (2013) 76:400–11. doi: 10.1016/j.neuroimage.2013.03.015

22. Maffei C, Lee C, Planich M, Ramprasad M, Ravi N, Trainor D, et al. Using diffusion MRI data acquired with ultra-high gradient strength to improve tractography in routine-quality data. *Neuroimage.* (2021) 245:118706. doi: 10.1016/j.neuroimage.2021.118706

23. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage.* (2012) 62:782–90. doi: 10.1016/j.neuroimage.2011.09.015

24. Van Essen DC, Smith SM, Barch DM, Behrens TEJ, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: an overview. *Neuroimage*. (2013) 80:62–79. doi: 10.1016/j.neuroimage.2013.05.041

25. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*. (2013) 80:105–24. doi: 10.1016/j.neuroimage.2013.04.127

26. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. (2016) 19:1523–36. doi: 10.1038/nn.4393

27. Yendiki A, Panneck P, Srinivasan P, Stevens A, Zöllei L, Augustinack J, et al. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front Neuroinform*. (2011) 5:23. doi: 10.3389/fninf.2011.00023

28. Fischl B. FreeSurfer. *Neuroimage*. (2012) 62:774–81. doi: 10.1016/j.neuroimage.2012.01.021

29. Wasserthal J, Neher P, Maier-Hein KH. TractSeg - fast and accurate white matter tract segmentation. *Neuroimage*. (2018) 183:239–53. doi: 10.1016/j.neuroimage.2018.07.070

30. Tournier JD, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*. (2019) 202:116137. doi: 10.1016/j.neuroimage.2019.116137

31. Tournier JD, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage*. (2004) 23:1176–85. doi: 10.1016/j.neuroimage.2004.07.037

32. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lecture Notes Comput Sci*. (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28

33. Srikrishna M, Heckemann RA, Pereira JB, Volpe G, Zettergren A, Kern S, et al. Comparison of two-dimensional- and three-dimensional-based U-Net architectures for brain tissue classification in one-dimensional brain CT. *Front Comput Neurosci*. (2022) 15:785244. doi: 10.3389/fncom.2021.785244

34. Smith RE, Tournier JD, Calamante F, Connelly A. Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage*. (2012) 62:1924–938. doi: 10.1016/j.neuroimage.2012.06.005

35. Chandio BQ, Risacher SL, Pestilli F, Bullock D, Yeh FC, Koudoro S, et al. Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations. *Sci Rep*. (2020) 10:1–18. doi: 10.1038/s41598-020-74054-4

36. Ito T, Ohashi H, Gracco VL. Somatosensory contribution to audio-visual speech processing. *Cortex*. (2021) 143:195–204. doi: 10.1016/j.cortex.2021.07.013

37. Bertó G, Bullock D, Astolfi P, Hayashi S, Zigiotto L, Annicchiarico L, et al. Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation. *Neuroimage*. (2021) 224:117402. doi: 10.1016/j.neuroimage.2020.117402

38. Schilling KG, Tax CMW, Rheault F, Landman BA, Anderson AW, Descoteaux M, et al. Prevalence of white matter pathways coming into a single white matter voxel orientation: the bottleneck issue in tractography. *Hum Brain Mapp*. (2022) 3;43:1196–213. doi: 10.1002/hbm.25697

39. Lu Q, Liu W, Zhuo Z, Li Y, Duan Y, Yu P, et al. A transfer learning approach to few-shot segmentation of novel white matter tracts. *Med Image Anal*. (2022) 79:102454. doi: 10.1016/j.media.2022.102454

40. Lu Q, Li Y, Ye C. Volumetric white matter tract segmentation with nested self-supervised learning using sequential pretext tasks. *Med Image Anal*. (2021) 8;72:102094. doi: 10.1016/j.media.2021.102094

41. Yang Q, Hansen CB, Cai LY, Rheault F, Lee HH, Bao S, et al. Learning white matter subject-specific segmentation from structural MRI. *Med Phys*. (2022) 49:2502–13. doi: 10.1002/mp.15495

42. Liu W, Lu Q, Zhuo Z, Li Y, Duan Y, Yu P, et al. Volumetric segmentation of white matter tracts with label embedding. *Neuroimage*. (2022) 250:118934. doi: 10.1016/j.neuroimage.2022.118934

43. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. (2006) 31:968–80. doi: 10.1016/j.neuroimage.2006.01.021

44. Marie D, Jobard G, Crivello F, Perchey G, Petit L, Mellet E, et al. Descriptive anatomy of Heschl's gyri in 430 healthy volunteers, including 198 left-handers. *Brain Struct Func*. (2015) 220:729–43. doi: 10.1007/s00429-013-0680-x

45. Takahashi T, Sasabayashi D, Takayanagi Y, Higuchi Y, Mizukami Y, Nishiyama S, et al. Heschl's gyrus duplication pattern in individuals at risk of developing psychosis and patients with schizophrenia. *Front Behav Neurosci*. (2021) 15:647069. doi: 10.3389/fnbeh.2021.647069

46. Dalboni da Rocha JL, Schneider P, Benner J, Santoro R, Atanasova T, Van De Ville D, et al. TASH: toolbox for the automated segmentation of Heschl's gyrus. *Sci Rep*. (2020) 10:1–15. doi: 10.1038/s41598-020-60609-y

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 28 September 2022
DOI 10.3389/fninf.2022.934433

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
National Research Council (CNR), Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Sophie Brice,
Swinburne University of Technology, Australia
Eleftheria Iliadou,
National and Kapodistrian University of Athens, Greece

*CORRESPONDENCE
Katarzyna A. Tarnowska
k.tarnowska@unf.edu

RECEIVED 02 May 2022
ACCEPTED 15 August 2022
PUBLISHED 28 September 2022

CITATION
Tarnowska KA, Ras ZW and Jastreboff PJ (2022) A data-driven approach to clinical decision support in tinnitus retraining therapy. *Front. Neuroinform.* 16:934433. doi: 10.3389/fninf.2022.934433

A data-driven approach to clinical decision support in tinnitus retraining therapy

Katarzyna A. Tarnowska[1]*, Zbigniew W. Ras[2,3] and Pawel J. Jastreboff[4]

[1]School of Computing, University of North Florida, Jacksonville, FL, United States, [2]Computer Science Department, University of North Carolina, Charlotte, NC, United States, [3]Polish-Japanese Academy of Information Technology, Warsaw, Poland, [4]Department of Otolaryngology—Head & Neck Surgery, School of Medicine Emory University, Atlanta, GA, United States

Background: Tinnitus, known as "ringing in the ears", is a widespread and frequently disabling hearing disorder. No pharmacological treatment exists, but clinical management techniques, such as tinnitus retraining therapy (TRT), prove effective in helping patients. Although effective, TRT is not widely offered, due to scarcity of expertise and complexity because of a high level of personalization. Within this study, a data-driven clinical decision support tool is proposed to guide clinicians in the delivery of TRT.

Methods: This research proposes the formulation of data analytics models, based on supervised machine learning (ML) techniques, such as classification models and decision rules for diagnosis, and action rules for treatment to support the delivery of TRT. A knowledge-based framework for clinical decision support system (CDSS) is proposed as a UI-based Java application with embedded WEKA predictive models and Java Expert System Shell (JESS) rule engine with a pattern-matching algorithm for inference (Rete). The knowledge base is evaluated by the accuracy, coverage, and explainability of diagnostics predictions and treatment recommendations.

Results: The ML methods were applied to a clinical dataset of tinnitus patients from the Tinnitus and Hyperacusis Center at Emory University School of Medicine, which describes 555 patients and 3,000 visits. The validated ML classification models for diagnosis and rules: association and actionable treatment patterns were embedded into the knowledge base of CDSS. The CDSS prototype was tested for accuracy and explainability of the decision support, with preliminary testing resulting in an average of 80% accuracy, satisfactory coverage, and explainability.

Conclusions: The outcome is a validated prototype CDS system that is expected to facilitate the TRT practice.

KEYWORDS

clinical decision support systems, tinnitus, knowledge-based systems, knowledge discovery, action rules, tinnitus retraining therapy

1. Introduction

1.1. Tinnitus

Tinnitus is a highly prevalent and frequently severely impairing hearing disorder with a worldwide impact. Often described as "ringing in the ears", tinnitus is the sensation of sound perception without an external sound source—"phantom auditory perception" (Jastreboff, 1990). The U.S. Centers for Disease Control estimates that nearly 15% of the general public—over 50 million Americans—experience a form of tinnitus. In addition, close to 90% have experienced at least temporary tinnitus, making it one of the most common health conditions in the United States. While about 20 million people struggle with burdensome chronic tinnitus, 2 million have extreme and debilitating cases (American Tinnitus Association, 2018). There are millions of general practice consultations every year where the primary complaint is tinnitus, equating to a major burden on healthcare services. Tinnitus has been the #1 claimed service-related disability for the American Veteran Administration for more than a decade (US Department of Veterans Affairs, 2019). Chronic disabling tinnitus has a devastating impact on the quality of life and psychosocial aspects of those affected (Makar et al., 2017). The disorder has a considerable heterogeneity and no single mechanism is likely to explain the presence of tinnitus in all those affected. Tinnitus can be associated with head and neck injuries, hearing loss, acoustic neuromas, drug toxicity, ear disease, and depression (Savage and Waddell, 2014).

1.2. Treatment of tinnitus

The heterogeneity and current limited knowledge about the pathophysiology of the different forms of tinnitus are reasons that hamper the identification of good candidates for an effective pharmacological treatment for tinnitus. Despite its growing prevalence and often-devastating effects, tinnitus remains a severely underfunded condition. There are no Food and Drug Administration (FDA) approved drugs available, and the quest for a new treatment option for tinnitus focuses on important challenges in tinnitus management (Swain et al., 2016). Clinical management strategies include counseling (education and advice), sound enrichment using ear-level sound generators or hearing aids, tinnitus masking, relaxation therapy, cognitive behavior therapy (CBT), and tinnitus retraining therapy (TRT) (Makar et al., 2017). Although a variety of therapeutic interventions are available, the complexity of tinnitus makes the management of the condition challenging. Evaluating results in the field of tinnitus is a difficult task, as no objective tinnitus measurement exists. It means there is no objective method for detecting the presence and the extent of tinnitus.

1.3. Tinnitus retraining therapy

During the last decades, advances in neuroimaging methods and the development of an animal model of tinnitus have contributed to an increasing understanding of the neuronal correlates of tinnitus (Langguth, 2015). TRT is a clinical implementation of the neurophysiological model of tinnitus (Jastreboff and Hazell, 2004). It is the habituation therapy used for the management of chronic subjective tinnitus. It includes counseling (TC) during structured sessions in combination with sound therapy (ST) to reduce the patient's tinnitus-evoked negative reaction to, and awareness of, tinnitus. ST sound stimulation is performed with low-level broadband sound generators and aims to mask tinnitus at the sound perception level. By reducing the tinnitus perception, TRT successfully helps patients to achieve control over their tinnitus, live a normal life, and participate in everyday activities (Reddy et al., 2019). Clinical studies confirm that TRT is an effective and robust treatment for chronic decompensated tinnitus (Zhao and Jiang, 2018; Nemade and Shinde, 2019). The majority of published clinical studies indicate TRT offers notable help for about 80% of patients and the severity of tinnitus decreases in a clinically significant and persistent manner. Furthermore, TRT offers an approach to treat other hearing disorders: hyperacusis, which is reduced tolerance to sounds, phonophobia, which is the fear of sound, and misophonia, increased sound sensitivity (Jastreboff and Jastreboff, 2000). TRT, although effective, is a complex treatment and must be highly individualized. Counseling and teaching are tailored to the needs of the patient, and therefore, they cannot be performed as group therapy (Jastreboff and Jastreboff, 2006). Sound therapy involves different types/models of instruments, and they must be fitted optimally at the "mixing point" to achieve habituation in the most effective manner (Jastreboff and Jastreboff, 2006). Because TRT has to focus on the individual needs and profile of a patient, it consequently requires significant time involvement of the personnel. Although promising, it is expensive and spans from several months to a couple of years. Despite its high effectiveness and international recognition, the therapy is not widely offered, mainly due to a lack of expertise and experience in its delivery. The main obstacle to the widespread adoption of this technique is a lack of trained and experienced audiologists.

1.4. Tinnitus data analytics

Data-driven approaches have the potential to reveal novel insights into tinnitus heterogeneity. However, there are limitations in data-driven studies for tinnitus management proposed so far. Most efforts involve applying traditional statistical methods, such as correlation and regression (Langguth et al., 2017). New forms of discovery *via* machine learning and big data methods have not been widely investigated.

Data mining/machine learning methods proposed on tinnitus data were mostly confined to association analysis, predictive modeling, and clustering analysis. However, these studies were limited in terms of analyzed variables or provided inconclusive results (Anwar, 2013; van den Berge et al., 2017). The status quo of tinnitus data analytics lacks the application of discovery methods for actionable and personalized knowledge needed by medical practitioners. The outcomes are not analyzed with regard to treatment methods in order to seek actions leading to improvement. Also, the temporality of data is not considered. So far, data analytics efforts focus on variables describing psychoacoustic measures of tinnitus. These measures, although routinely obtained in many clinics and as part of research studies, have not been validated for being diagnostic, prognostic, discriminative, or responsive (Henry, 2016; Watts et al., 2018). Medical history and evaluation, review of the patient's medications, and assessment of an individual's distress or handicaps are also crucial for effective diagnosis and treatment (Kari et al., 2010). Finally, most research efforts conclude by presenting analytics without any further developments in the decision support tool. No integration into health IT systems nor plans on how to utilize the findings in clinical decision-making is currently being proposed. To date, this research is the first to propose a decision support system for TRT.

1.5. Technological perspectives on tinnitus

The postal survey of general practitioners (GPs) concluded that there was a substantial discrepancy between the scientific and technological perspectives on the management of tinnitus and the actual day-to-day practice in the primary care setting (Hall et al., 2011). Many GPs expressed an unmet need for a specific and concise training on tinnitus management. Low satisfaction with available treatment options was unequivocally mentioned by both GPs and ENTs (ear-nose-throat specialists) from all developed countries investigated by Hall et al. (2011). The results of that survey highlight the need for an effective therapy option, particularly for chronic subjective tinnitus. Despite a variety of options, the low success of the available tinnitus treatment options leads to the frustration of physicians and patients alike. Effective therapeutic options with guidelines about key diagnostic criteria are urgently needed.

2. Materials and methods

Clinical decision support (CDS) is a process for enhancing health-related decisions and actions with pertinent, organized, clinical knowledge, and patient information to improve health and healthcare delivery. Systems, known as clinical decision support systems (CDSS), offer intelligent support for human-oriented diagnosis and treatment of patients. "CDS provides clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare" (Osheroff et al., 2007). They were proposed for various diseases, including traumatic brain injury, diabetes, Parkinson's disease, and other health-related decisions such as drug dosing (Ciecierski, 2013; Nielsen et al., 2014; Fartoumi et al., 2016; Torrent-Fontbona and López, 2019). Yet, nobody developed a clinical decision support system for tinnitus management. It was hypothesized that DSS can improve the accuracy and time efficiency of tinnitus management, but a design or implementation of such a system was not attempted (Thompson et al., 2007; Anwar, 2013). Within this research, we proposed a knowledge-based clinical decision support system (refer to Figure 1). The knowledge base is developed with validated models extracted from data mining experiments.

2.1. Knowledge discovery methods

The proposed knowledge discovery from tinnitus data, as opposed to previous research in this area, provides multidimensional evaluation beyond the psychoacoustic characteristics of tinnitus. Since the clinical data available describes temporal changes in tinnitus score and particular areas of a patient's life affected, it is possible to perform an analysis of changes and increased involvement in life activities that were previously prevented or interfered with by tinnitus (or hyperacusis). This approach will help in a better understanding of complex auditory, psychological, and medical conditions and aid in selecting the most significant variables to consider in TRT. We propose a variety of data mining methods to extract novel knowledge about TRT diagnosis and treatment. This includes predictive and descriptive models, which are extracted from the pre-processed and transformed data. The following software was used for knowledge discovery:

- WEKA—Open-source data mining software which offers a wide choice of algorithms for feature selection and for prediction as well as a user-friendly interface and feature to build a complete "knowledge flow" (Bouckaert et al., 2014). It also allows using Java API to embed machine learning models into a Java program.
- LISp-Miner—An academic system that offers exploratory data analysis, including modules for association rule discovery (4ft-Miner) and action rule discovery (Ac4ft-Miner) (Simunek, 2014).

2.1.1. Dataset

To evaluate our data-driven approach to building CDSS, we use clinical data collected at the Tinnitus and Hyperacusis

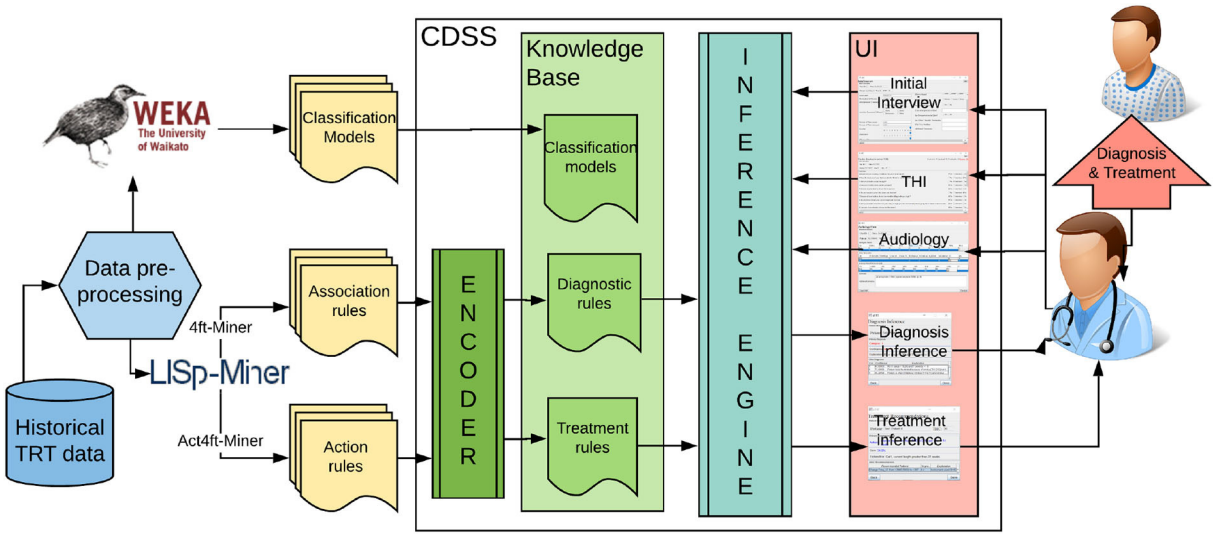


FIGURE 1
The architecture of the proposed data-driven clinical decision support system for tinnitus diagnosis and treatment.

Center of Emory University School of Medicine. The dataset contains records of tinnitus patients and the records for their sequential visits to the clinic. The dataset was collected over a period of several years and describes 555 unique patients and 3,000 visits in total. The raw data resides in 11 separate tables describing demographics, interview response, audiological measurements, pharmacology, additional medical evaluation, and visits. The visit data contains treatment methods applied by the physician at the visit (sound therapy with instruments/counseling, real ear measurements to help fit the instruments) along with the measure of the treatment progress using the Tinnitus Handicap Inventory (THI). The raw data were exported to the relational database system to ensure the structure, consistency, and integrity of the data (refer to Figure 2).

2.1.2. Data pre-processing and feature selection

Various data-preprocessing techniques were applied to cleanse the data and handle real-life data issues, such as inconsistencies, incompleteness, duplication, and other problems. Data cleansing removed all inconsistencies, such as missing values, outliers, and duplicate data (e.g., duplicate visit numbers for the same patient). To handle missing data in the total score of the tinnitus handicap inventory (THI), an algorithm for data imputation was developed and validated. Additional transformations were applied, such as alphanumeric to numeric encoding, aggregation, and handling data temporality. Feature selection was proposed to reduce the data to a manageable and relevant size. Only the most relevant variables were involved in developing an analytical model. A more detailed description of the challenges with the real-world data and applied data-preprocessing methods to mitigate those can be found in our previous publication (Tarnowska et al., 2017).

2.1.3. Feature extraction

Additional features describing the patient and characteristics of tinnitus were developed from the text attributes to make the dataset more suitable for machine learning:

- Tinnitus background: *STI* (stress-induced), *NTI* (noise-induced), *HLTI* (hearing-loss-induced), *DETI* (depression-related), *AATI* (auto accident-related), *OTI* (surgery-related), and *OMTI* (induced as a symptom of another medical condition).
- Temporal features: *DTI* (date tinnitus induced), *AgeInd* (the patient's age when tinnitus induced), *AgeBeg* (the patient's age when treatment began), binary features denoting how many days/weeks/months/years ago the hearing problem started.
- Binary attributes that represent the intake of medication.
- Attributes that keep track of a patient's improvement over time: *ChTsc* (change in the THI's total score from the previous visit) and *PerChTsc* (relative change measuring the percentage change in the THI's total score from the previous visit).

A comprehensive list of the attributes from the clinical database, as well as extracted features, can be found in our previous publication by Tarnowska et al. (2017).

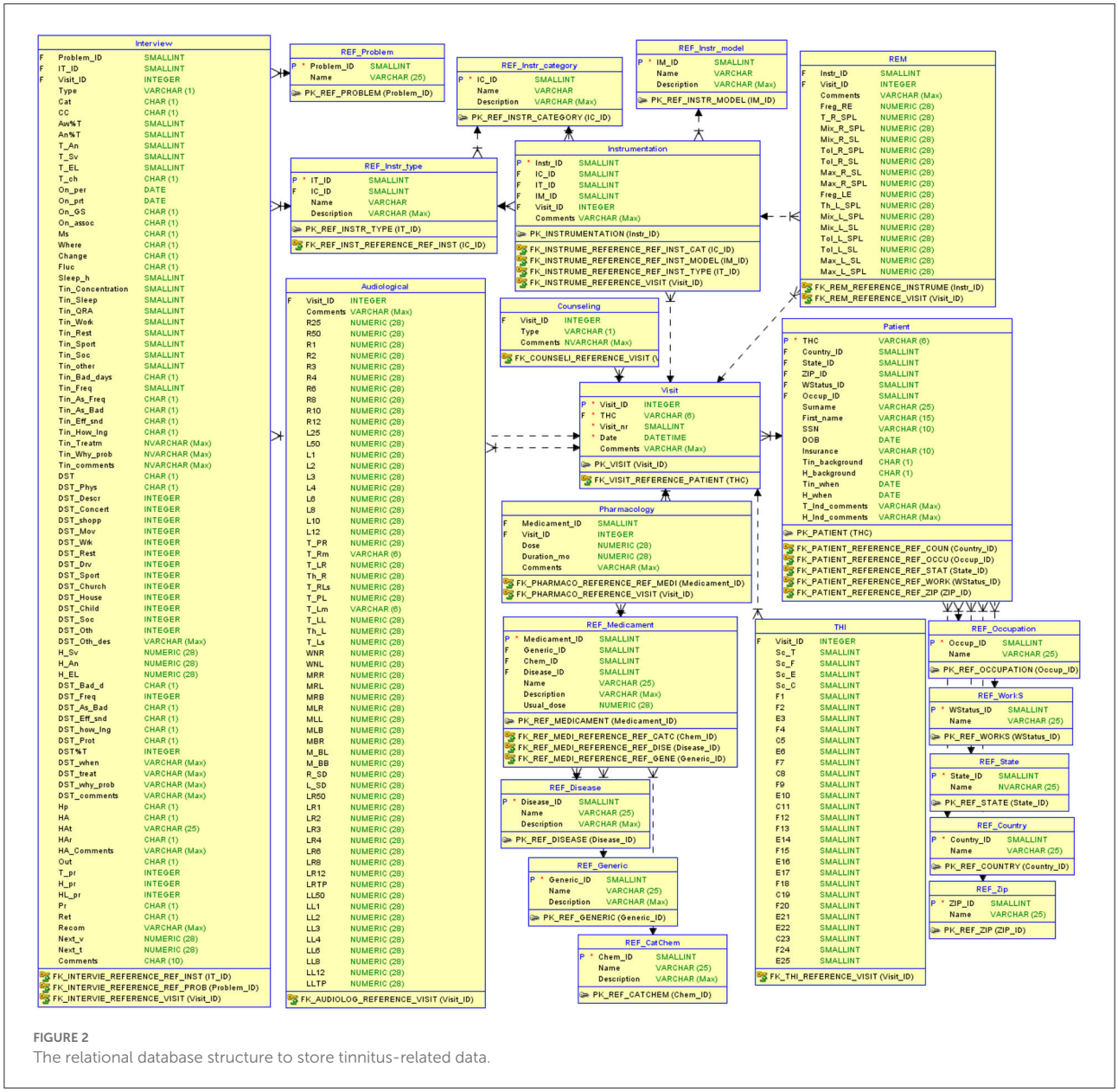


FIGURE 2
The relational database structure to store tinnitus-related data.

2.1.4. Predictive models

The first type of machine learning applied is supervised machine learning to build predictive models. The goal is to build an analytical model predicting a target measure of interest. In our domain, it is the category of the hearing problem, which determines the TRT treatment protocol. TRT protocol differentiates the following five categories, which differ in further treatment protocol: C0 (tinnitus minimal problem), C1 (tinnitus significant problem), C2 (tinnitus significant and hearing loss present), C3 (tinnitus irrelevant, hyperacusis significant), and C4 (prolonged tinnitus/prolonged exacerbation of hyperacusis). The proposed classification model, built using supervised machine learning methods, is used to predict the category of an unseen patient under consideration. The following ML algorithms in WEKA are used for classification models: tree-based J48, random forest, and probabilistic-based Naive Bayes.

2.1.5. Descriptive models

The goal is to extract valid and useful medical patterns in tinnitus diagnosis and treatment. The patterns describe patients' diagnosis/treatment and are used to develop the domain knowledge for TRT. The descriptive methods used in this research include association rules and action

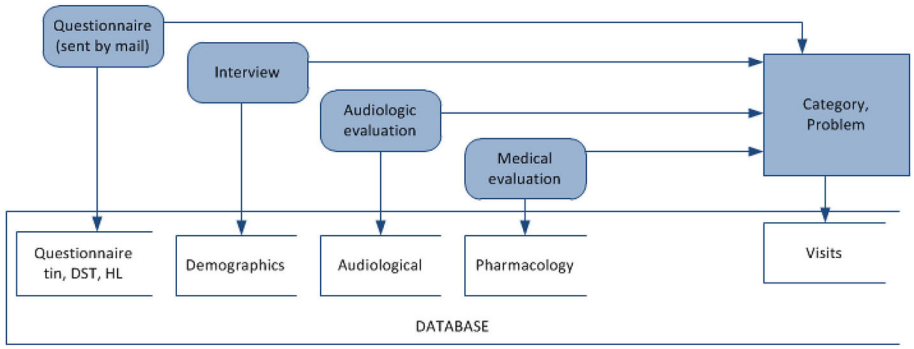


FIGURE 3
Mining associations between questionnaire and interview answers, audiology variables, medications, and category of a hearing problem for decision support in diagnosis in TRT.

rules. Rules are characterized by statistical measures quantifying their strength. Support and confidence are two key measures to quantify the strength and relevance of a rule. The support reflects the usefulness of a rule and confidence—its certainty. To find the significant associations, support and confidence must be set at a certain minimum threshold value (usually 1% for support, and 80% for confidence).

2.1.5.1. Association rules for diagnosis in TRT

The TRT diagnosis is to be supported by the descriptive models based on the association (decision) rule discovery, as supplemental to predictive models.

A **decision rule** is a rule r in the form $(\phi \Rightarrow \delta)$, where ϕ is called *antecedent* (or assumption), and δ is called *descendant* (or thesis). Each rule is characterized by *support* and *confidence*. *Support(r)* is defined as the number of objects matching the rule's antecedent. *Confidence(r)* is the relative number of objects matching both the rule's antecedent and descendant of the rule. The data mining experiments for decision rule discovery were modeled after the TRT diagnosis process, which involves an initial interview, audiology and medical evaluation (refer to Figure 3). Association rules mining aims at detecting frequently occurring associations between variables in TRT. Accordingly, associations between audiological measurements, demographics, questionnaire responses, pharmacology, and the category of tinnitus were extracted using LISP-Miner software for data mining (Simunek, 2014).

2.1.5.2. Action rules for treatment in TRT

The concept of an action rule was first proposed by Ras and Wieczorkowska in 2000 (Ras and Wieczorkowska, 2000), and since then, its application was proposed, among others, for business, medicine, and music indexing (Ras and Wieczorkowska, 2000, 2010; Wasyluk et al., 2008; Tarnowska et al., 2020). Action rules are especially promising in the field of medical data, as a doctor can examine the effect of treatment decisions on a patient's improved state. This technique is also particularly useful for building knowledge-based decision support systems.

Action rule r is a term $[(\omega) \wedge (\alpha \rightarrow \beta) \Rightarrow (\theta \rightarrow \psi)]$, where $(\omega \wedge \alpha) \Rightarrow \theta$ and $(\omega \wedge \beta) \Rightarrow \psi$ are classification rules, ω is a conjunction of stable attribute values, $(\alpha \rightarrow \beta)$ shows changes in flexible attribute values, and $(\theta \rightarrow \psi)$ shows the desired effect of the action. In this domain, it is proposed to apply action rules to recommend effective methods of treatment in TRT (refer to Figure 4). Such rules, extracted from large sets of data, represent actions to undertake (e.g., treatment methods) to improve the defined state (e.g., tinnitus awareness) when specified conditions hold (e.g., the current patient's state and profile). Action is understood as changing certain ("flexible") variables to achieve the desired results. The purpose is to analyze data to seek specific actions to enhance the decision-making process. Action rules applied for TRT will suggest, with a certain confidence, the most effective treatment method for an individualized profile of a patient (defined by "stable" attributes), at a particular time (considering temporal variables).

2.2. Knowledge base

Within this research, we propose a knowledge-based clinical decision support system. The knowledge is developed with models extracted from knowledge discovery experiments. These experiments yield a vast number of diagnostic and treatment patterns. These are general clinical rules already known by experts, or they represent novel and unknown patterns useful for future diagnosis/treatment. We propose interpreting and validating the results from analytical modeling with clinical expertise and including only validated patterns

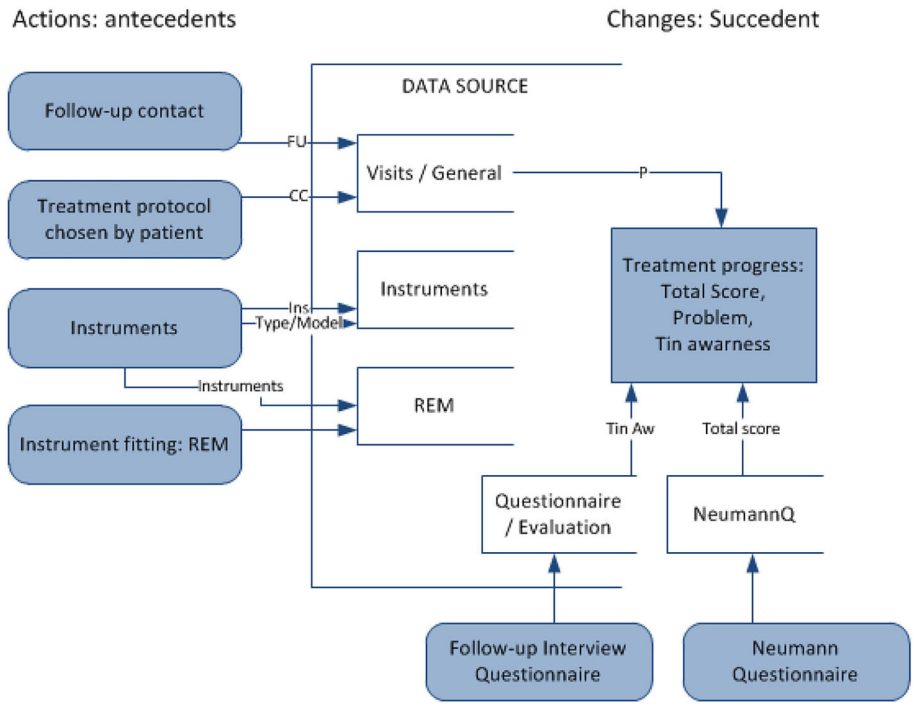


FIGURE 4
Mining action rules for changes in the type of counseling and tuning sound generators for decision support in TRT treatment.

with the highest confidence in the framework of the built system.

2.2.1. Knowledge translator (encoder)

The goal of the Knowledge Translator ("Encoder" in Figure 1) component is to automatically encode the rule-based knowledge from the output files of data mining software into the CDSS knowledge base. The knowledge translating procedure reads rules one by one, parses, interprets them, adds the explanation in natural language, and encodes them into the syntax used in the knowledge base of CDSS. The rules encoded in KB are "if-then" like statements and each rule encodes a small piece of the expert's knowledge available through the dataset. The pseudocode for the knowledge translating procedure is depicted in Table 1.

2.2.2. Inference engine

With knowledge encoded in the form of "if-then" rules, an automatic inference component is used to control the application of the rules. Each rule has a left-hand side ("if" statement—the antecedent of a rule) and a right-hand side ("then" statement—consequent of a rule). The left-hand side contains information about certain facts about the patient. If the left-hand side of the rule (antecedent) is matched, its right-hand side (consequent) is executed. Once a new patient's characteristics are entered into the system, the inference module will fire the matching rules from its knowledge base. Consequent clauses decide on the diagnosis/treatment decision suggested to the physician. The JESS library for Java-based programs was used to implement an inference engine based on the efficient pattern-matching algorithm, called Rete (Forgy, 1982).

2.3. Graphical user interface

The user interface (UI) of the system constitutes the mode of interaction between the physician and the underlying CDS model. The prototype GUI was developed in Java Swing, with customary component extensions for screen development. The developed UI supports clinical processes in

- Storing and managing the data related to
  - Tinnitus patients—demographics, medical history, audiology evaluations, and structured interviews (Jastreboff and Jastreboff, 1999);
  - TRT visits—diagnoses, treatment applied (sound therapy and counseling), and the outcome evaluation with standardized forms, such as *Tinnitus Handicap*

TABLE 1 Steps in the Knowledge Translator procedure.

Step #	Step description
1	Read a rule.
2	Extract confidence, category, and other components from the rule.
3	Split the rule's hypothesis into partial cedents.
4	Parse each partial cedent and create an object representing the cedent.
5	Develop an explanation for each partial cedent.
6	Create a rule object containing the cedent objects and explanations.
7	Encode that rule object to a file in KB.

Inventory (THI) (Newman et al., 1995) and *Tinnitus Functional Index* (TFI) (Meikle et al., 2011).

- Providing evidence-based diagnostic and treatment decision support with explanations and quantifiable predictive outcomes.

Prior to designing the appearance of the user interface, several factors were taken into account. For this process, several ideas explored in Carroll et al. (2002) were considered. The article proposes the following guidelines for designing a user interface for clinical decision support systems, which provided a basis for research methods for effective GUI design for CDSS for tinnitus:

- "All clinical data should be represented clearly in a format familiar to clinicians and easily understood by patients."
- "The system should be easy to learn and navigate around."
- "All information processing should be 'invisible' to the user."
- Consider both the physician and the patient as primary stakeholders.
- Use visual aids to describe data such as sliding bars and color codes, where applicable.

3. Results

Within this section, results on feature selection, machine learning experiments, and the evaluation of KB and CDSS are described.

3.1. Feature selection

The feature selector used in WEKA was used based on a chi-square measure to identify a subset of most predictive attributes. Table 2 shows the results of feature selection, from around 603 attributes that describe the TRT visits dataset (questionnaires, interviews, audiology, and pharmacology).

TABLE 2 Feature selection results for categorizing patients based on chi-squared ranking in WEKA.

Feature	Feature description	Ranking score
LR4	LDL (RE) at 4 kHz	725.1
Th L	Hearing threshold (LE)	712.7
LR3	LDL (RE) at 3 kHz	688.0
LR2	LDL (RE) at 2 kHz	683.4
LR1	LDL (RE) at 1 kHz	683.1
T LR	Tinnitus Loudness Match (RE)	672.6
LR8	LDL (RE) at 8 kHz	670.57
LL3	LDL (LE) at 3 kHz	667.47
Th R	Hearing threshold (RE)	618.94
LL2	LDL (LE) at 2 kHz	617.06

Audiological measurements were indicated as the most relevant factors in the TRT categorization process. The results point out various audiological measurements, such as loudness discomfort level (LDL), the threshold of hearing (Th), and loudness match as primary in relation to classifying (diagnosing) patients into categories.

3.2. Machine learning models

WEKA was used to test different classification algorithms and determine the classification model with the highest accuracy. The evaluation was carried out by splitting the dataset into training and test subsets using cross-validation with 10 folds. Performance measures for predictive models include classification accuracy (the percentage of correctly classified patients) and precision (how many of the predicted categories are actually in that category). Preliminary results of predictive models with different algorithms are presented in Table 3. The tests were performed on different types of datasets and using different feature selection methods. *Pat-vis* is a dataset with each visit of a patient as a separate instance. *Pat-vis-med* dataset additionally includes binary attributes for all types of medications, that is, each visit instance is repeated for a medication that a patient is taking. *Pat-vis0* includes only initial visits (Visits with ordering number 0 or 1), that is when the diagnosis and categorization of a patient are decided by a clinician. Depending on the feature selection method chosen, the dimensionality of datasets (# features) was reduced accordingly. ML algorithms tested included tree-based (J48), random forests, and Naive Bayes. The most reliable results were obtained using the dataset with the initial visit only, but due to the reduction in the number of data instances, the best accuracy was 57.4% with the Naive Bayes. It is expected that once more data on initial visits is collected, the more precise the trained models become.

TABLE 3 Results on patient classification using WEKA using different data pre-processing, feature selection, and algorithms.

Dataset	# instances	# features	J48 (%)	Naive Bayes (%)	Random forest (%)
Pat-vis-med	6,991	80	88.5	75.2	**89.3**
Pat-vis-med	6,991	20	**87.5**	81.5	87.1
Pat-vis	3,125	603	70.2	55.4	**71**
Pat-vis	3,125	488	69.7		
Pat-vis01	1,090	603	**52.1**	46	49.2
Pat-vis0	599	603	43.2	52	**53.4**
Pat-vis0	599	100	41.0	**57.4**	49.2

The best results are in bold.

3.3. Rule mining

The results described in this section include results from rule mining with LISp-Miner 4ft-Miner (association rules) and Act4ft-Miner (action rules) (Simunek, 2014).

3.3.1. Association rules for diagnosis

Experiments on decision rule discovery were carried out to complement results on predictive models for diagnostic decision-making. The variables investigated included 593 variables describing the patient and their tinnitus. Audiology variables include a pure-tone audiogram (up to 12kHz) and the determination of pure tone loudness discomfort levels (LDL) measured for all frequencies in the audiogram. For example, R6 describes the right ear (R) pure-tone threshold for 6kHz. LDL is the audiological measure crucial for TRT diagnosis. For example, LR1/LL1 describes LDL for the right ear/the left ear tests with 1 kHz. Patients' responses to initial/follow-up questionnaires are another important source of information for determining the category in TRT. The questions provide a structure for the interview with a patient and allow physicians to track the progress of the treatment. Variables describing subjective tinnitus are measured on a Likert scale (0–10) and patients are asked to assess them "on average over the last month".

Table 4 shows examples of extracted associations between audiological measurements, questionnaire responses, and a category of a hearing problem.

These rules are interpreted as follows:

- *If* an audiometric value of R_3 (audiogram at 3 kHz for the right ear) is in the range $< 15; 20)$ and annoyance over tinnitus T_{An} is greater than or equal to 8, *then* a patient falls under Category 1 with 94% confidence.
- *If* hyperacusis H_{pr} and hearing loss HL_{pr} are not indicated as problems, but tinnitus T_{pr} indicated a problem—*then* a patient falls under Category 1 with 85% confidence.
- *If* an audiometric value of L_2 (audiogram at 2kHz for the left ear) is greater or equal to 50 and R_6 (audiogram at 6kHz for the right ear) is less or equal to 75, *then* a patient falls under Category 2 with 87% confidence.
- *If* a patient was taking *Norvasc* and tinnitus was its side effect, then a patient falls under Category 2 with 67% confidence.
- If the score for tinnitus as a problem T_{pr} was in the range $< 0, 2.5)$, annoyance over hyperacusis H_{An} in range $< 1.5; 3.5)$ and severity of hyperacusis H_{Sv} in range $< 1.5; 3.5)$, then a patient is categorized into Category 3 with 83% confidence.

In general, many such rules are generated and each rule represents a small chunk of knowledge available through a clinical dataset. For example, patients in Category 1 have a significant tinnitus problem (T_{Pr}—Tinnitus as a Problem) but without hyperacusis (H) and there is no significant hearing loss (HL). Category 2 is characterized by a significant hearing loss, as indicated by lower values of the pure-tone audiogram (L_2 and R_6). Patients in Category 3 are on the other hand characterized by a significant hyperacusis problem (H_{An}—Hyperacusis annoyance and H_{Sv}—Hyperacusis severity). The experiments also yield novel and unknown patterns such as dependencies between certain medications and their side effects (T *side*) being tinnitus symptoms. Experiments between demographics of patients and a TRT category indicated, that tinnitus in elderly patients was frequently related to hearing loss and was affected by many other medical conditions, such as hypertension and age-related afflictions, and associated with Category 2. Patients in Category 1 (C1) were middle-aged, and their tinnitus was associated with psychological disorders, such as depression, anxiety, and panic. Category 3 was frequent in the younger group (30–38 years) and association rules indicate, for men: background in noise exposure, occupation, type of work; and for women: background in stress and hormonal therapy. These findings lead to a hypothesis that a personalized approach to tinnitus treatment based on a patient's profile could be effective. For example, for C1-patients personalized counseling is expected to be more effective, as it is frequently associated with psychological disorders. C2 would be most effectively treated

TABLE 4 Examples of discovered decision rules for the category of a hearing problem determined based on the interview and audiometric values.

Sample association rule for diagnostics in TRT	Confidence (%)
$R3(< 15; 20)) \wedge TAn \geq 8 \Rightarrow Category(1)$	94
$H_{pr}(< 0; 0.5)) \wedge HL_{pr}(< 0; 0.5)) \wedge T_{pr}(< 6; 8)) \Rightarrow Category(1)$	85
$L_2 \geq 50 \wedge R_6 \leq 75 \Rightarrow Category(2)$	87
$Norvasc(yes) \wedge Tside(yes) \Rightarrow Category(2)$	67
$T_{pr}(< 0; 2.5)) \wedge H_{An}(< 1.5; 3.5)) \wedge H_{Sv}(< 1.5; 3.5)) \Rightarrow Category(3)$	83

TABLE 5 Results on actionable knowledge discovery for recommending treatment in TRT.

A sample treatment action rule	Conf. (%)
$G(m) \wedge NTI(yes) : (Ins_{vis(01)}(GHH) \rightarrow Inst_{vi(01)}(GHS)) => Ch(better)$	80
$T_{side}(yes) \wedge OMTI(yes) : (Ins_{vis(01)}GHH \rightarrow V) \wedge FU(0 \rightarrow T) \Rightarrow Ch(better)$	82
$Ins(SG) : (Mix_{RSL}(< 11; 12) \rightarrow < 9; 10)) \Rightarrow Ch(better)$	100
$FU(A) \wedge Ins_{vis(01)}(GHI) \wedge Freq_{LE}(< 3000; 3150)) : (treat(< 5; 6) \rightarrow < 6; 8)) \Rightarrow Ch(better)$	88

with hearing aids and instrument fitting, as it is frequently associated with hearing loss.

3.3.2. Action rules for recommending treatment

Action rules are methods proposed within this research to support treatment within TRT protocols. The attribute used as a decision attribute is THI's total score (*T sc*), which keeps track of the treatment progress. In case the total score is missing in the data, the tinnitus awareness score (T_{aw}) was used instead. The action rule mining was set up to extract patterns that bring changes in THI's total score/tinnitus awareness for the better (*ChTsc*/*ChTaw*-change in the total score/awareness from the previous visit and *PerChTsc*/*PerChTaw*—percentage change of the previous). The action rule mining experiments involved checking variables related to changeable ("flexible") treatment methods within TRT and setting other attributes as "stable" (patient demographics, tinnitus characteristics), with the goal to improve metrics measuring the severity of tinnitus. Sound therapy with instruments involves choosing the right instrument and fitting the instrument with the optimal setting over time at subsequent visits. There are different types of instruments, as described by the category variable (*Ins*): hearing aid (*HA*), sound generator (*SG*), and combination instrument. There are different SG models, e.g., General Hearing Instrument (*GHI*): soft/hard, Viennatone (*V*), and many others. A specific fitting of instruments is a significant aspect and real-ear measurements (REM) assist in instrument fitting. Sound therapy is accompanied by counseling. The variable FU describes the types of follow-up contact: audiology and counseling (A), counseling (*C*), telephone-based (*T*), and e-mail based (*E*). The results of the sample extracted patterns are presented in Table 5.

The rules present different actions in treatment leading to a change in patients for the better, as measured by the total score of THI and tinnitus awareness. These rules are interpreted as follows:

- *If* a patient is a male and tinnitus is noise-induced *then* changing sound therapy from the instrument model of GH hard (*GHH*) to GH soft (*GHS*) at the first visit improves a patient with 80% confidence.
- *If* tinnitus was induced by another medical condition (*OMTI*) and as a side effect of taking medications (T_{side}), *then* changing the sound generator model GH hard (*GHH*) to the Viennatone model (*V*) at the first visit and changing the follow-up contact to the telephone-based (*T*) improves patient with 82% confidence.
- *If* the current treatment involves sound generator *SG*, *then* changing the mixing point for the right ear Mix_{RSL} from $< 11; 12)$ to $< 9; 10)$ improves a patient's state with 100% confidence.
- *If* the current treatment involves audiology ($FU(A)$) with the GHI instrument and frequency in the left ear measured by REM -$Freq_{LE}$—in the range of $< 3000; 3150)$ *then* prolonging that treatment from 5–6 weeks to 6–8 weeks brings improvement with 88% confidence.

The extracted rules offer high precision, e.g., how to fit a particular model of a particular type of instrument ($Ins(SG) : (Mix_{RSL}(< 11; 12) \rightarrow < 9; 10))$) or how to change the length of treatment with a specific method [e.g., $treat(< 5; 6) \rightarrow < 6; 8]$: change the length of treatment from 5–6 to 6–8 weeks. This approach also offers high personalization: the treatment actions leading to improvement are extracted for the individual patients' profiles, as described by demographics [e.g., *G(m)* - gender: male] and the tinnitus

TABLE 6 Runtimes for encoding and parsing diagnosis (total of 2,192 rules) and action rules (total of 1,348).

Rule type	Total encoding time (s)	Total parsing time (s)	Time to parse a rule (ms)
Diagnosis rule	0.29	0.22	0.098
Treatment rule	0.24	0.13	0.094

background (e.g., *NTI* - noise-induced tinnitus, *OMTI* - other medical-induced tinnitus, T_{side} - tinnitus as a side effect of pharmacology).

3.4. Knowledge translator

The Knowledge Translator was tested within CDSS for efficiency and scalability. The runtimes of various steps of the Knowledge Translator are depicted in Table 6. The encoding step encompasses all operations from reading the files with rules to writing the rules into KB. Parsing, in this case, only refers to parsing the rule and creating an object in memory, but it does not include any I/O operations. The tests were run using 2,192 diagnosis rules and 1,348 treatment rules. An average from running one test 5 times is presented in Table 6.

As one can see from the results in Table 6, the developed Knowledge Translator encodes and parses a massive amount of extracted rules in a relatively very short time. This provides an important step in the future scalability of the CDS system. When comparing the time to parse a single rule by the Knowledge Translator (less than 0.1 ms) vs. the same task performed manually (manual encoding by a human, which takes approximately 2 min at least to read, interpret and encode a rule in a correct syntax), the time gain is enormous. Additionally, the developed Knowledge Translator encodes the human-understandable explanations, which are critical for clinical use and support in the accurate diagnosis of the category of a hearing problem and treatment actions recommendations (refer to Figure 5).

3.5. CDSS evaluation

The evaluation study was to determine whether the built CDS system does what it was intended to and at an adequate level of accuracy. The expectation from the proposed CDSS is to generate accurate, patient-specific, and interpretable clinical suggestions. This will encourage efficient and effective use of tinnitus retraining therapy for the management of hearing disorders. The evaluation study involved:

1. Developing a user-friendly interface to input the patient's data.
2. Identifying a set of representative test cases of patients from the dataset not used for building the model.
3. Running inference on the chosen test cases entered into the system (refer to Figures 5–7).
4. Performing quantitative and qualitative evaluation of the system based on the results from the above.

The metrics used for this evaluation of the system include:

- Accuracy—The number of correct predictions vs. the total number of predictions. To compute the accuracy we compare the system's recommendations with the actual diagnosis/treatment decision made by a physician.
- Coverage—The number of test cases matched against the knowledge base.
- Interpretability—If the recommendations are explainable and understandable by humans.

3.5.1. Test cases

The representative cases from each of the 5 categories, were identified. Future testing will involve identifying more cases per category. The chosen test cases reflect the heterogeneity of the hearing problem and patient profile; a test patient for each etiology and each category of the hearing problem was identified from the tinnitus patient database (refer to Table 7).

Tables 8, 9 provide the diagnostic and treatment inference results for all test cases.

3.5.2. Diagnostic decision support

The diagnosis prediction was 80% accurate and covered 100% of cases (refer to Table 8). The average confidence in the primary diagnosis inference was 83.51%. The only incorrect prediction was for test case 5. After closer investigation, this case was annotated by the physician as a "discrepancy in information" in interview data, and "inconsistent results" in audiological evaluation, which are the reasons that misled the predictive model (as an "outlier" data point). Moreover, the actual protocol followed was the same as for the category predicted by the system.

3.5.3. Treatment decision support

The treatment recommendations were generated for 3 out of 5 patient test cases (refer to Table 9). The other two cases were not covered, that is, no action rule was matched with the patient profile, due to a limited number of rules encoded manually in KB at the time of testing.

For all the tested cases, both the diagnostic and the treatment recommendations were explained with a human-comprehensible message/reason. The explanations were provided by means of the premises of the rules in KB that were matched against the current patient's

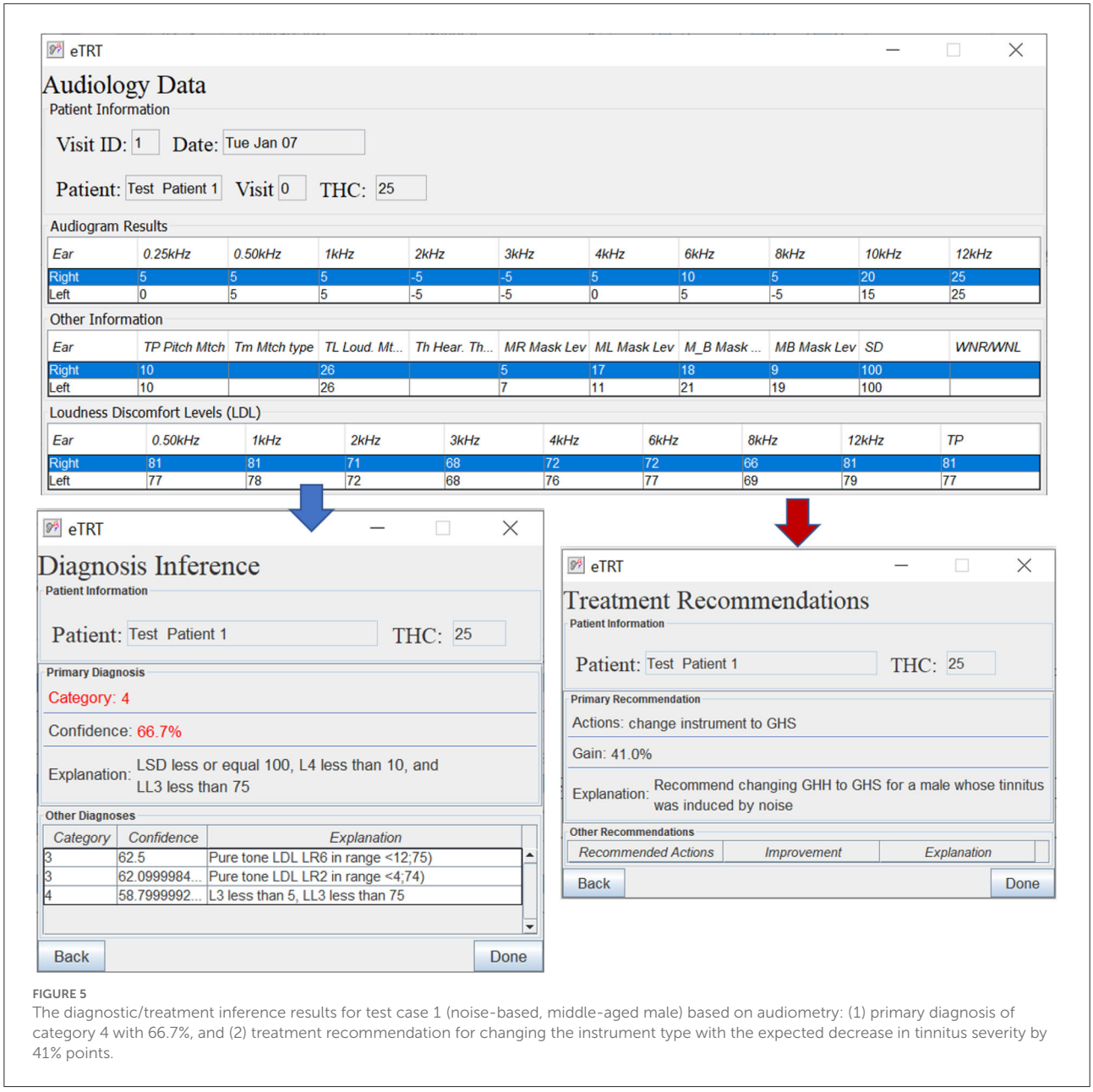


FIGURE 5
The diagnostic/treatment inference results for test case 1 (noise-based, middle-aged male) based on audiometry: (1) primary diagnosis of category 4 with 66.7%, and (2) treatment recommendation for changing the instrument type with the expected decrease in tinnitus severity by 41% points.

profile/visit data. The predictions' probabilities were quantified by means of the matched rules' confidence metric.

4. Discussion

Data mining is an active field and helps uncover links between variables, with the goal to develop optimal strategies for tinnitus management. This will open new horizons for TRT, which does not have a stagnant protocol but continues to evolve based on information gathered from treatments of patients (Jastreboff, 2015). TRT has successfully been used in a clinical setting to help patients with tinnitus and decreased sound tolerance since 1988, but the method of TRT underwent many modifications since its first description.

4.1. Strengths

The proposed analytical models applied to a clinical dataset detected both trivial and known patterns in TRT, but also unexpected, unknown, and potentially interesting patterns. The extracted knowledge utilizes clinical knowledge available from

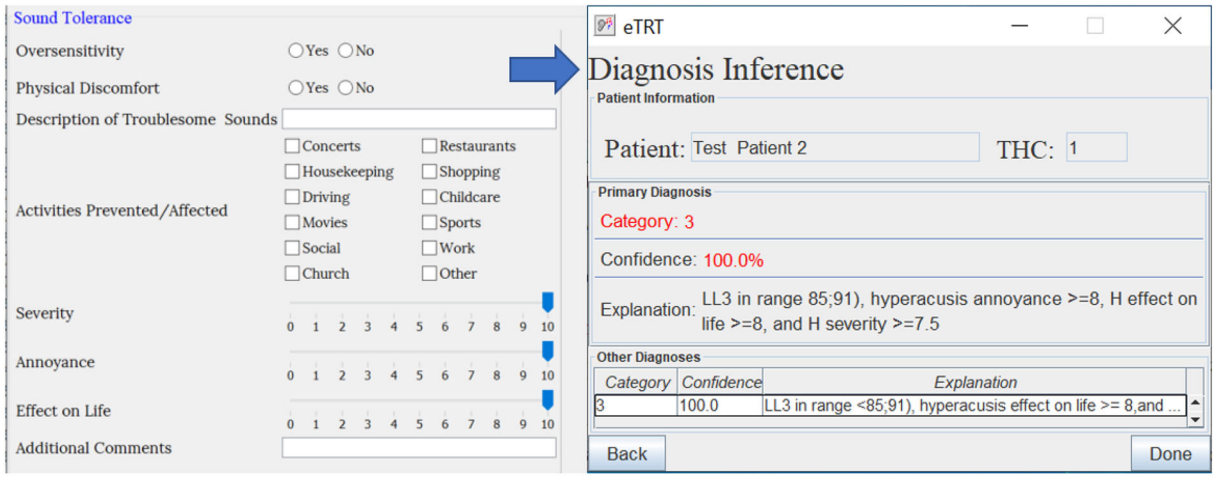


FIGURE 6
The diagnostic inference for test case 2 based on audiometry/initial interview. The explanation for Category 3 with 100% confidence included a high score for hyperacusis as a problem.

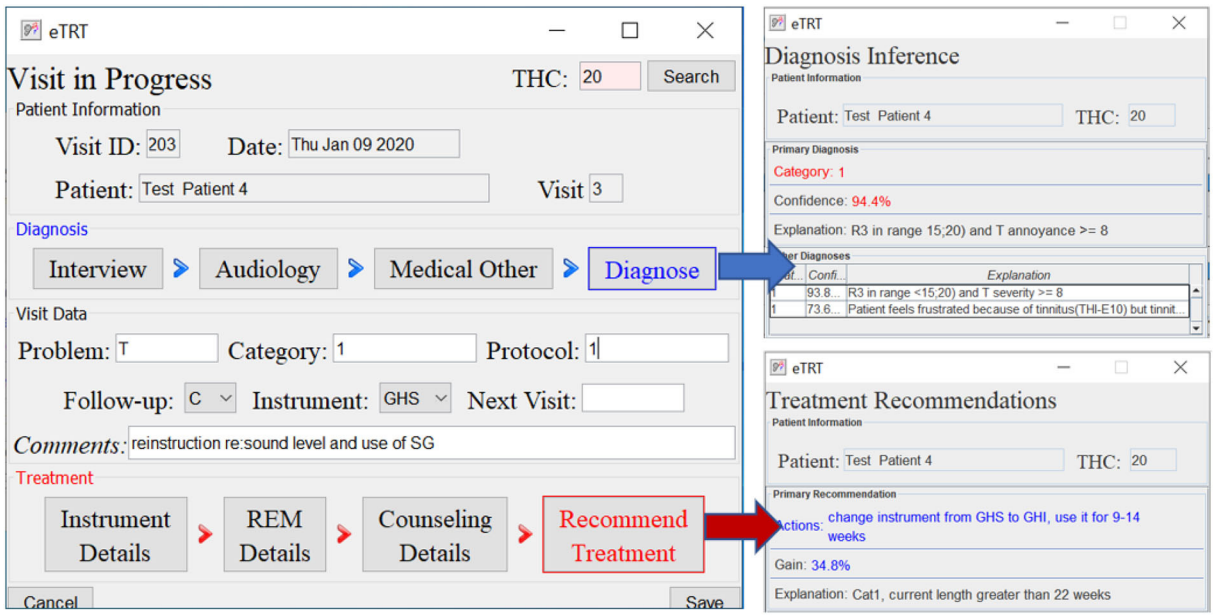


FIGURE 7
The diagnostic/treatment inference results for test case 4: (1) category 1 was inferred based on the audiometry results and initial interview (annoyance over tinnitus high); (2) recommendation included the change of the sound instrument from GH soft and shorten its application time to 9–14 weeks with an expected gain of 34.4% points.

the TRT expert through the dataset. The prime characteristic of the approach is its capability of expressing knowledge in a linguistic way allowing a system to be described by simple "human-friendly" rules. Knowledge represented in form of rules is closely related to human thinking and can be explained natively. It also offers an approach for modeling the uncertainty and the imprecision typical of human reasoning. As knowledge is created based on feedback from an expert, users can also rely on it. It can be used for educational purposes as a training tool to spread expertise in TRT. The knowledge once encoded will be preserved permanently and utilized on any hardware. New patient data can be analyzed by inferring from rules with the goal of providing a prediction of optimal diagnosis and treatment. This computer-based approach will help clinicians reveal the

TABLE 7 Patient test cases—patient profile, etiology of their hearing problem, the diagnosed category, and the treatment protocol determined by the physician.

Test case	Patient profile	Etiology	Diagnosis	Treatment protocol
1	Male, age 38, KY	Noise exposure	Category 4	Category 4
2	Male, age 49, GA	Ear surgery	Category 3	Category 3
3	Female, age 77, FL	Hearing loss	Category 2	Category 2
4	Male, age 53, GA	Stress-related	Category 1	Category 1
5	Male, age 36, GA	Car accident	Category 0	Category 1

mysteries of tinnitus heterogeneity and decrease the impact of this major health problem on the patient and society.

4.2. Limitations

The currently identified limitations include limited access to clinical TRT datasets and a lack of standardization in data management techniques among clinics offering TRT. The more data, including from many providers, and the more recent the data (since TRT is still evolving) the more accurate and useful the CDSS tool. The key to the successful adoption of the system is a collaboration between the system's developers and clinicians/clinics. At the current stage, mostly the prototype version was designed and developed, which nevertheless has to be tested for usability in real-world environments and undergo rigorous testing.

The main so-far identified problem is the quality of the data, and particularly its sparsity. Data is available from only one expert, which provides consistency of knowledge, but also limited records of data. Generally, the more data, the more accurate the results. Missing values and a limited amount of data results in extracting rules characterized by relatively low support and confidence. Additional strategies will be investigated to develop algorithms for the reliable imputation of missing values. Another potential problem is the computational complexity of learning analytical models. The strategy to handle this issue is to investigate alternative efficient algorithms for mining that make use of multiple cores and distributed processing. Additionally, the hardware platform will be scaled adding RAM and CPU power.

The validation of CDSS accuracy is a challenge. System reliability and trustworthiness depend on the quality of the rules. In the current testing design, retrospective patient data will be used to design and test the system. CDS logic may not precisely fit the patient. Therefore, if coverage results prove to be unsatisfactory, more rules will be extracted/added. Another potential problem is that the user interface proves to be unsatisfactory. In that case additional, alternative UI designs will be proposed, evaluated, and compared. If the attractiveness of the interface will be insufficient, alternative technologies for UI development will be investigated, e.g., Java FX.

4.3. Future study

In the future, the system is to be used as a TRT assistant by medical professionals to support both the efficient and effective management of tinnitus.

The tasks to be performed in the longer term, related to the development of the CDS system include:

- Expanding the knowledge base with new clinics and new/updated treatment methods. Adding new data sources to the system, such as patient and treatment datasets from other clinics, to expand the knowledge base of the system. Both clinics in TRT as well as other tinnitus treatment methods should be included. The goal central repository should be made available to participating TRT clinicians and researchers (Landgrebe et al., 2010). Additional approaches for tinnitus treatment will be investigated, such as music therapy, brain stimulation, or cognitive therapy. It is expected more data available from more than one TRT expert will improve the accuracy and coverage of the knowledge base in TRT.
- Applying machine learning methods to investigate additional factors and variables that may help understand and treat tinnitus. Some of the considered factors include magnetic resonance imaging (MRI) data, e.g., to understand how sound therapy changes neuronal activity and modulates the brain network (Han et al., 2019); or associations between gene variants and tinnitus states (Pulley et al., 2012).
- Expanding AI-based methods to provide decision support, i.e., natural language processing from the clinical text data, i.e., doctor's comments (Tarnowska and Ras., 2019, 2021). Another potential is to utilize natural language understanding to develop conversational agents, that can help in delivering counseling to tinnitus patients. Additionally, machine learning methods based on clustering techniques to develop new models for more personalized treatment can be investigated.
- Integration with other software used in audiology (Rajkumar et al., 2017), i.e., software for sound generators' tuning; investigating computer methods to generate personalized sound used in tinnitus habituation (Barozzi et al., 2017); integrating music therapy and music recommendation into the system (Tarnowska, 2021).
- Expanding modes for the system—i.e., publicly-available touch-screen stations or developing mobile applications

TABLE 8 Results on predicting diagnosis by the system on the chosen patient test cases—actual category vs. category predicted by the system, characterized by confidence, and explanation.

Test case	Actual	Prediction	Conf.	Explanation
1	Cat 4	Cat 4	66.7%	LSD <= 100, L4 < 10, and LL3 < 75
2	Cat 3	Cat 3	100	LL3 in < 85; 91), Hyper. Annoy ≥8, H Eff on Lif ≥8, and H Sev ≥7.5
3	Cat 2	Cat 2	96.2	LR8 ≥ 999, R6 ≥ 75, and T_{sv} ≥ 8
4	Cat 1	Cat 1	94.4	LL3 in < 15; 20) and Tin. annoy. ≥ 8
5	Cat 0	Cat 1	60.3	A patient often irritable by tinnitus (E14) and tinnitus makes him anxious (E22)

TABLE 9 Results on recommending treatment actions, characterized by an expected improvement gain in percentage points and explanation(s) for the patients' test cases.

Test case	Recommended action(s)	Gain	Explanation
1	Change instrument from GHH to GHS	41 pp	A male whose tinnitus was induced by noise
4	Change instrument from GHS to GHI, use it for 9–14 weeks	34.8 pp	Cat1, instr. duration greater than 22 weeks
5	Change Freq LE from <2,800; 3,000) to <2,670; 2,800) in REM	8.4 pp	Instrument used GHS

to improve personalization and streamline data collection from the patients (Blome, 2015).

The long-term goal of this research is to deploy such a system in a clinical setting to enhance health-related decisions in TRT delivery. This step will be preceded by testing the system in the clinical environment and testing its usability within real physician-patient consultation. More extensive testing involving more test cases and new patients will be conducted in the future. Usability evaluation with actual clinical users should be performed to determine its acceptability. In the future, the system should be integrated with health IT systems and electronic health records (EHR) to fit into the workflow of clinical decision-making. The electronic health record (EHR) with embedded clinical decision support is recognized as an important component in providing improvement in patient safety, healthcare quality, and efficiency, as promised by HITECH (Health Information Technology for Economic and Clinical Health) policy initiatives (Blumenthal and Glaser, 2007). The project is intended to connect primary care providers and TRT specialists using a knowledge-driven computational engine that aids in diagnosing and planning treatment for tinnitus patients. Decision support, delivered using an information system with the electronic medical record as the platform, will provide decision-makers with tools making it possible to achieve large gains in performance, narrow gaps between knowledge and practice, and improve tinnitus habituation rates. The proposed novel and efficient approach to developing a data-driven CDSS can be applied to various other medical domains. The results are replicable by others, and useful to tinnitus researchers and other medical practitioners.

5. Conclusion

The main contribution of this study is proposing and evaluating a data-driven clinical decision support system to assist audiologists in the diagnosis and treatment of hearing disorders, such as tinnitus, hyperacusis, and misophonia. Up to date, no CDSS specialized in tinnitus diagnosis and therapy has been designed and implemented. Collaboration between experts in the fields of both data analysis and tinnitus is of utmost importance to prepare and validate optimal CDSS that will be reliable and efficient. Such decision support will bring advantages such as speed, accuracy, and long-term storage of information. Medical users will receive rapid and synchronous advice. With the user-friendly interface, non-computer professionals will be able to easily operate the system and interpret its results. Documented knowledge can be used for future training and educational purposes. This type of research is expected to provide an important step toward the widespread and effective use of TRT knowledge in clinical practice. This is significant because the diagnosis and treatment of TRT is a complex task. It requires a very high level of expertise to operate accurately and efficiently. Data and information being used in tinnitus management are becoming heterogeneous and large in volume, and therefore,

they are overwhelming. A CDSS needs to be developed once and customized locally to the clinic's needs. It can be used frequently in many places by many people without location restrictions. The system offers a scalable architecture that can be extended by new knowledge.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZR and KT: study conception and design. PJ: acquisition of data. KT, ZR, and PJ: analysis and interpretation of data. KT: drafting of the manuscript. ZR and PJ: critical revision. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher. All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

American Tinnitus Association (2018). *Understanding the Facts*.

Anwar, M. N. (2013). Mining and analysis of audiology data to find significant factors associated with tinnitus masker. *SpringerPlus* 2, 595. doi: 10.1186/2193-1801-2-595

Barozzi, S., Ambrosetti, U., Callaway, S., Behrens, T., Passoni, S., and Bo, L. (2017). Effects of tinnitus retraining therapy with different colours of sound. *Int. Tinnitus J.* 21, 139–143. doi: 10.5935/0946-5448.20170026

Blome, J. (2015). *Implementation and evaluation of a mobile Android application for auditory stimulation of chronic tinnitus patients* (Ph.D. thesis). Ulm University, Ulm, Germany.

Blumenthal, D., and Glaser, J. (2007). Information technology comes to medicine. *N. Engl. J. Med.* 356, 2527–2534. doi: 10.1056/NEJMhpr066212

Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., et al. (2014). *WEKA Manual for Version 3-6-12*. The University of Waikato.

Carroll, C., Marsden, P., Soden, P., Naylor, E., New, J., and Dornan, T. (2002). Involving users in the design and usability evaluation of a clinical decision support system. *Comput. Methods Prog. Biomed.* 69, 123–135. doi: 10.1016/S0169-2607(02)00036-6

Ciecierski, K. (2013). *Decision Support System for surgical treatment of Parkinson's disease* (Ph.D. thesis). Warsaw University of Technology, Warsaw, Poland.

Fartoumi, S., Emeriaud, G., Roumeliotis, N., Brossier, D., and Sawan, M. (2016). Computerized decision support system for traumatic brain injury management. *J. Pediatr. Intensive Care* 5, 101–107. doi: 10.1055/s-0035-1569997

Forgy, C. L. (1982). Rete: a fast algorithm for the many pattern/many object pattern match problem. *Artif. Intell.* 19, 17–37. doi: 10.1016/0004-3702(82)90020-0

Hall, D. A., Lainez, M. J., Newman, C. W., Sanchez, T. G., Egler, M., Tennigkeit, F., et al. (2011). Treatment options for subjective tinnitus: self reports from a sample of general practitioners and ent physicians within Europe and the USA. *BMC Health Serv Res.* 11, 302. doi: 10.1186/1472-6963-11-302

Han, L., Yawen, L., Hao, W., Chunil, L., Pengfei, Z., Zhengyu, Z., et al. (2019). Effects of sound therapy on resting-state functional brain networks in patients with tinnitus: a graph-theoretical-based study. *J Mang Reson Imaging* 50, 1731–1741. doi: 10.1002/jmri.26796

Henry, J. A. (2016). "Measurement" of tinnitus. *Otol. Neurotol.* 37, 276–285. doi: 10.1097/MAO.0000000000001070

Jastreboff, M., and Jastreboff, P. (1999). "Questionnaires for assessment of the patients and treatment outcome," in *Sixth International Tinnitus Seminar* (Cambridge).

Jastreboff, P. J. (1990). Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neurosci. Res.* 8, 221–254. doi: 10.1016/0168-0102(90)90031-9

Jastreboff, P. J. (2015). 25 years of tinnitus retraining therapy. *HNO* 63, 307–311. doi: 10.1007/s00106-014-2979-1

Jastreboff, P. J., and Hazell, J. W. P. (2004). *Tinnitus Retraining Therapy: Implementing the Neurophysiological Model*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511544989

Jastreboff, P. J., and Jastreboff, M. M. (2000). Tinnitus retraining therapy (TRT) as a method for treatment of tinnitus and hyperacusis patients. *J. Am. Acad. Audiol.* 11, 156–161. doi: 10.1055/s-0042-1748042

Jastreboff, P. J., and Jastreboff, M. M. (2006). Tinnitus retraining therapy: a different view on tinnitus. *ORL J. Otorhinolaryngol. Relat. Spec.* 68, 23–29. doi: 10.1159/000090487

Kari, E., Mattox, D. E., and Jastreboff, P. J. (2010). "Tinnitus," in *Glasscock-Shambaugh Surgery of the Ear*, 6th Edn., eds J. Aina and L. B. Gulya (Shelton: People?s Medical Publishing House), 293–306.

Landgrebe, M., Zeman, F., Koller, M., Eberl, Y., Mohr, M., Reiter, J., et al. (2010). The tinnitus research initiative (TRI) database: a new approach for delineation of tinnitus subtypes and generation of predictors for treatment outcome. *BMC Med. Inform. Decis Mak.* 10, 42. doi: 10.1186/1472-6947-10-42

Langguth, B. (2015). Treatment of tinnitus. *Curr. Opin. Otolaryngol. Head Neck Surg.* 23, 361–368. doi: 10.1097/MOO.0000000000000185

Langguth, B., Landgrebe, M., Schlee, W., Schecklmann, M., Vielsmeier, V., Steffens, T., et al. (2017). Different patterns of hearing loss among tinnitus patients: a latent class analysis of a large sample. *Front. Neurol.* 8, 46. doi: 10.3389/fneur.2017.00046

Makar, S. K., Mukundan, G., and Gore, G. (2017). Treatment of tinnitus: a scoping review. *Int. Tinnitus J.* 21, 144–156. doi: 10.5935/0946-5448.20170027

Meikle, M., Henry, J., Griest, S., Stewart, B., Abrams, H., McArdle, R., et al. (2011). The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. *Ear Hear.* 33, 153–176. doi: 10.1097/AUD.0b013e31822f67c0

Nemade, S. V., and Shinde, K. J. (2019). Clinical efficacy of tinnitus retraining therapy based on tinnitus questionnaire score and visual analogue scale score in patients with subjective tinnitus. *Turk. Arch. Otorhinolaryngol.* 57, 34–38. doi: 10.5152/tao.2019.3091

Newman, C. W., Wharton, J. A., and Jacobson, G. P. (1995). Retest stability of the tinnitus handicap questionnaire. *Ann. Otol. Rhinol. Laryngol.* 104, 718–723. doi: 10.1177/000348949510400910

Nielsen, A. L., Henriksen, D. P., Marinakis, C., Hellebek, A., Birn, H., Nybo, M., et al. (2014). Drug dosing in patients with renal insufficiency in a hospital setting using electronic prescribing and automated reporting of estimated glomerular filtration rate. *Basic Clin. Pharmacol. Toxicol.* 114, 407–413. doi: 10.1111/bcpt.12185

Osheroff, J. A., Teich, J. M., Middleton, B., Steen, E. B., Wright, A., and Detmer, D. E. (2007). A roadmap for national action on clinical decision support. *J. Am. Med. Inform. Assoc.* 14, 141–145. doi: 10.1197/jamia.M2334

Pulley, J., Denny, J., Peterson, J., Bernard, G., Vnencak-Jones, C., Ramirez, A., et al. (2012). Operational implementation of prospective genotyping for personalized medicine: the design of the vanderbilt predict project. *Clin. Pharmacol. Ther.* 92, 87–95. doi: 10.1038/clpt.2011.371

Rajkumar, S., Muttan, S., Sapthagirivasan, V., Jaya, V., and Vignesh, S. (2017). Software intelligent system for effective solutions for hearing impaired subjects. *Int. J. Med. Inform.* 97, 152–162. doi: 10.1016/J.IJMEDINF.2016.10.009

Ras, Z. W., and Wieczorkowska, A. (2000). "Action-rules: how to increase profit of a company," in *Principles of Data Mining and Knowledge Discovery*, eds D. A. Zighed, J. Komorowski, and J. Zytkow (Berlin; Heidelberg: Springer), 587–592. doi: 10.1007/3-540-45372-5_70

Ras, Z. W., and Wieczorkowska, A. (2010). *Advances in Music Information Retrieval*. Heidelberg: Springer. doi: 10.1007/978-3-642-11674-2

Reddy, K. V. K., Chaitanya, V. K., and Babu, G. R. (2019). Efficacy of tinnitus retraining therapy, a modish management of tinnitus: Our experience. *Indian J. Otolaryngol. Head Neck Surg.* 71, 95–98. doi: 10.1007/s12070-018-1392-6

Savage, J., and Waddell, A. (2014). Tinnitus. *BMJ Clin. Evid.* 2014, 0506.

Simunek, M. (2014). Lisp-miner control language description of scripting language implementation. *J. Syst. Integr.* 5, 28–44. doi: 10.20470/jsi.v5i2.193

Swain, S. K., Nayak, S., Ravan, J. R., and Sahu, M. C. (2016). Tinnitus and its current treatment - still an enigma in medicine. *J. Formos. Med. Assoc.* 115, 139–144. doi: 10.1016/j.jfma.2015.11.011

Tarnowska, K. A. (2021). "Emotion-based music recommender system for tinnitus patients (EMOTIN)," in *Recommender Systems for Medicine and Music*, Vol. 946, eds Z. W. Ras, A. Wieczorkowska, and S. Tsumoto (Springer), 197–221. doi: 10.1007/978-3-030-66450-3_13

Tarnowska, K. A., and Ras, Z. W. (2019). Sentiment analysis of customer data. *Web Intell. J.* 17, 343–363. doi: 10.3233/WEB-190423

Tarnowska, K. A., and Ras, Z. W. (2021). NLP-based customer loyalty improvement recommender system (CLIRS2). *Big Data Cogn. Comput.* 5, 4. doi: 10.3390/bdcc5010004

Tarnowska, K. A., Ras, Z. W., and Daniel, L. (2020). *Recommender System for Improving Customer Loyalty*. Springer Nature. doi: 10.1007/978-3-030-13438-9

Tarnowska, K. A., Ras, Z. W., and Jastreboff, P. J. (2017). "Mining for actionable knowledge in tinnitus datasets," in *Thriving Rough Sets*, Vol. 708, 6th Edn., eds G. Wang, A. Skowron, Y. Yau, D. Slezak, and L. Polkowski (Springer), 367–396. doi: 10.1007/978-3-319-54966-8_18

Thompson, P. L., Zhang, C. X., Jiang, W., and Ras, Z. W. (2007). "From mining tinnitus database to tinnitus decision-support system, initial study," in IAT (Washington, DC: IEEE), 203–206. doi: 10.1109/IAT.2007.88

Torrent-Fontbona, F., and López, B. (2019). Personalized adaptive cbr bolus recommender system for type 1 diabetes. *IEEE J. Biomed. Health Inform.* 23, 387–394. doi: 10.1109/JBHI.2018.2813424

US Department of Veterans Affairs (2019). *Veterans Benefits Administration Reports*.

van den Berge, M. J. C., Free, R. H., Arnold, R., de Kleine, E., Hofman, R., van Dijk, J. M. C., et al. (2017). Cluster analysis to identify possible subgroups in tinnitus patients. *Front. Neurol.* 8, 115. doi: 10.3389/fneur.2017.00115

Wasyluk, H., Ras, Z. W., and Wyrzykowska, E. (2008). Application of action rules to HEPAR clinical decision support system. *Exp. Clin. Hepatol. Bd* 4, 46–48.

Watts, E. J., Fackrell, K., Smith, S., Sheldrake, J., Haider, H., and Hoare, D. J. (2018). Why is tinnitus a problem? A qualitative analysis of problems reported by tinnitus patients. *Trends Hear.* 22. doi: 10.1177/2331216518812250

Zhao, D., and Jiang, Z. G. (2018). Observation of effect of retraining therapy in patients with chronic tinnitus. *J. Clin. Otolaryngol. Head Neck Surg.* 32, 583–586. doi: 10.13201/j.issn.1001-1781.2018.08.006

frontiers | Frontiers in Neuroinformatics

TYPE Methods
PUBLISHED 13 October 2022
DOI 10.3389/fninf.2022.970372

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
National Research Council (CNR), Italy

REVIEWED BY
Kirill Vadimovich Nourski,
The University of Iowa, United States
Eduardo Aubert,
Cuban Neuroscience Center, Cuba
Yinchen Song,
Dartmouth College, United States

*CORRESPONDENCE
Karolina Ignatiadis
karolina.ignatiadis@oeaw.ac.at

RECEIVED 15 June 2022
ACCEPTED 05 September 2022
PUBLISHED 13 October 2022

CITATION
Ignatiadis K, Barumerli R, Tóth B and Baumgartner R (2022) Effects of individualized brain anatomies and EEG electrode positions on inferred activity of the primary auditory cortex. *Front. Neuroinform.* 16:970372. doi: 10.3389/fninf.2022.970372

Effects of individualized brain anatomies and EEG electrode positions on inferred activity of the primary auditory cortex

Karolina Ignatiadis[1]*, Roberto Barumerli[1], Brigitta Tóth[2] and Robert Baumgartner[1]

[1]Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria, [2]Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary

Due to its high temporal resolution and non-invasive nature, electroencephalography (EEG) is considered a method of great value for the field of auditory cognitive neuroscience. In performing source space analyses, localization accuracy poses a bottleneck, which precise forward models based on individualized attributes such as subject anatomy or electrode locations aim to overcome. Yet acquiring anatomical images or localizing EEG electrodes requires significant additional funds and processing time, making it an oftentimes inaccessible asset. Neuroscientific software offers template solutions, on which analyses can be based. For localizing the source of auditory evoked responses, we here compared the results of employing such template anatomies and electrode positions versus the subject-specific ones, as well as combinations of the two. All considered cases represented approaches commonly used in electrophysiological studies. We considered differences between two commonly used inverse solutions (dSPM, sLORETA) and targeted the primary auditory cortex; a notoriously small cortical region that is located within the lateral sulcus, thus particularly prone to errors in localization. Through systematical comparison of early evoked component metrics and spatial leakage, we assessed how the individualization steps impacted the analyses outcomes. Both electrode locations as well as subject anatomies were found to have an effect, which though varied based on the configuration considered. When comparing the inverse solutions, we moreover found that dSPM more consistently benefited from individualization of subject morphologies compared to sLORETA, suggesting it to be the better choice for auditory cortex localization.

KEYWORDS

electroencephalography, template anatomy, individual anatomical data, electrode locations, inverse problem, human, hearing

1. Introduction

Being inexpensive and non-invasive, electroencephalography (EEG) is a widely used neuroimaging method. Due to its high temporal resolution it can reflect fast processes, which is especially advantageous in hearing research as well as objective audiometry (Somers et al., 2021). Its spatial resolution is limited by the number of electrodes that can be positioned on the scalp and the indirect measurement of neural activity through electric fields. If the goal is to postulate about the function of specific brain regions, inferring the sought activity requires knowledge about the underlying brain structures and the exact electrode locations. However, subject brain and general anatomical characteristics present a significant variability (Bartley et al., 1997; Yang et al., 2019; Haładaj, 2020). Individual head shapes also cause differences in the relative placement and studies show that the considered electrode locations can greatly affect analyses results (Schwartz et al., 1996; Van Hoey et al., 2000; Wang and Gotman, 2001; Dalal et al., 2014; Hirth et al., 2020).

While individualization of those measures is possible, every step comes at a further cost: recording the individual electrode locations after an experiment requires the appropriate hardware and additional processing time, as, depending on the method used, laborious manual processing might be necessary (Koessler et al., 2007; Taberna et al., 2019). Acquiring individual brain anatomies can come in expensive for research institutions, as neither the facilities nor the necessary resources might be available. Moreover, it is oftentimes the case that prior medical procedures or implantations prevent individuals from procedures such as magnetic resonance imaging. For hearing research specifically, cochlear implants frequently fall under the latter category (Leinung et al., 2020; Holtmann et al., 2021). In either case, a better understanding of the effects of the individualization steps may help the planning and uncertainty assessment of EEG studies.

Activity recorded by scalp-EEG sensors comprises a superposition of various brain sources, making it non-trivial to uncover underlying mechanisms. To expose specific information about the auditory functions in the brain, it is often relevant to move from the sensor- to the source space, spatially separating the signals and attributing them to their original generators. Source estimation is a complex task, involving several modeling steps. Localizing where the recorded activity actually originated from requires in the first place a representation of the elements of the subject's head (Vorwerk et al., 2014). The scalp, skull, gray, and white matter and cerebrospinal fluid have different conductivity characteristics, requiring an appropriate model accounting for them. This information is incorporated in the forward model, which describes how the electric field generated by a cortical source is picked up as an electric potential by a sensor. Source estimation based on EEG is especially subject to errors in the forward modeling; as it is based on electric fields and the sensors are positioned directly on the skin, it is heavily influenced by the differences in conductivity estimates (Leahy et al., 1998; von Ellenrieder et al., 2009). Various solutions have been developed, and the forward-model choice mainly relies on the available computational resources and chosen measurement modality (Baillet et al., 2001; Hallez et al., 2007). For EEG research, using the boundary element method is considered an appropriate solution, offering a realistic representation of the head model (Wang and Gotman, 2001; Adde et al., 2003; Akalin-Acar and Gençer, 2004; Kybic et al., 2005). For the cases where individualization steps cannot be included, relevant software offers the option for template anatomies and electrode locations. Those can be used on the acquired experimental data, to approximate actual head characteristics.

Given the forward model, the activity of the brain regions can be estimated *via* the inverse solution; the sensor data is combined to create an estimate of the activity at the various brain locations. This constitutes an ill-posed problem because the number of sources is typically much larger than the number of sensors. Hence, the inverse solution is not unique and requires additional assumptions or constraints to become so (Baillet et al., 2001). Various approaches have been developed toward tackling this problem (Grech et al., 2008); among those, minimum-norm solutions fall under the category of distributed inverse solvers (Ou et al., 2009). They rely on minimal prior assumptions, and are therefore well-suited in data driven approaches, where data is too noisy or no prior knowledge about source activity can be reliable (Hauk, 2004). Each grid point is considered to be the location of one or a set of equivalent current dipoles, subject to specific constraints regarding their degrees of freedom. Those algorithms look for a fitting solution to the data at each grid location simultaneously, under the restriction of a minimum overall activity amplitude. As most cognitive processing relies on distributed sources rather than isolated sources, such approaches offer an ecologically plausible solution, suited for mapping complex function in the perceptual field (Komssi et al., 2004). In the implementations of dynamic statistical parametric mapping (dSPM; Dale et al., 2000) and standardized low-resolution electromagnetic tomography (sLORETA; Pascual-Marqui, 2002), noise statistics information derived either from data or separate recordings is used to standardize the source maps, in order to compensate for depth current-orientation inhomogeneity (Hauk et al., 2011). Generally, the choice of the inverse method relies on parameters such as the sensory modality or experimental paradigm; there are, though, no

precise guidelines on selecting a method, rendering the option to frequently depend on common practice and preference. Meanwhile, toolboxes offer direct implementations of multiple inverse solutions, thereby facilitating comparative studies on the same dataset, an oftentimes suggested approach (Nawel et al., 2019). In auditory research, dSPM and sLORETA are frequently applied algorithms toward solving the inverse problem (e.g., Jaworska et al., 2012; Raghavan et al., 2017; Justen and Herbert, 2018; Stropahl et al., 2018; Hsu et al., 2020; Mohan et al., 2020). Based on anecdotal evidence, dSPM has been deemed to be specifically good for modeling auditory cortex sources (Stropahl et al., 2018). Instead, various method comparisons demonstrated how sLORETA can return most satisfactory results for single source localization (Grech et al., 2008).

As the first relay of auditory information, the primary auditory cortex (PAC) is essential in auditory research. It is the main generator of early evoked activities denoted as the excitatory, more exogenic P1 component and the inhibitory, more endogenic N1 component (Picton et al., 1999; Kudela et al., 2018), typically assessed by their peak amplitudes and latencies. Yet, with its small size and intricate placement on the superior temporal lobe (i.e., within the lateral sulcus and its non-orthogonal orientation to the scalp), the correct extraction of its activity constitutes a difficult matter (Hari and Puce, 2017). Our aim in the current study was to examine the effect of those individualization steps on inferred PAC activity. For that reason we considered two main factors that play a crucial role in the source localization process: the electrode positions and the subject anatomy. We combined those in pairs of two, yielding four different and commonly used approaches in EEG experiments (template anatomy with template electrode positions, template anatomy with individual electrode positions, individual anatomy with template electrode positions, individual anatomy with individual electrode positions). Our basic assumption was that a fully individualized configuration should lead to the most likely precise and valid source localization (Akhtari et al., 1994; Buchner et al., 1995; Van Hoey et al., 2000; Darvas et al., 2006; Dalal et al., 2014) and more focal activity to elicit larger P1 and N1 component amplitudes (Picton et al., 2000). Effects on component latencies may also occur but we had no prior expectations on those. In addition, to more directly assess how much the evoked activity is restricted to the PAC, we defined a metric sensitive to spatial leakage by evaluating the power ratio between the PAC and the surrounding region for each component. Because the two components reflect different postsynaptic activities with known hemispheric asymmetries (e.g., Hine and Debener, 2007), we analyzed all metrics in a within-subject manner and for each hemisphere separately. To control for robustness or interaction with regard to the specific inverse solutions used, we studied and compared each combination of electrode and anatomical configurations with the two inverse solutions dSPM and sLORETA.

2. Materials and equipment

For the current study we analyzed data originally collected for an auditory spatial perception experiment (Baier et al., 2022). The auditory stimuli used were complex harmonic tones (Schroeder, 1970; $F_0 = 100$ Hz, bandwidth 1–16 kHz). They were presented through earphones (Etymotic Research, ER-2) and were filtered with listener-specific head-related transfer functions to sound as coming from either the right or left direction on the interaural axis. The duration of every stimulus was 1.2 s with an inter-stimulus interval of 500 ms. Onset and offset ramps with raised-cosine shape had a duration of 10 ms. The stimuli were presented at a sound pressure level of about 70 dB (all three intensity offsets of $2.5, 0$, and -2.5 dB from the original study were pooled together). The experiment consisted of an initial passive listening part, during which subjects were watching a silent subtitled movie while being exposed to the 600 trials. In a second part, subjects performed a spatial discrimination task on those stimuli. For our current study we only considered the EEG data during passive listening, in order to avoid any task-related effects of attention and arousal.

Our dataset was recorded with a 128-channel EEG system (actiCAP with actiCHamp; Brain Products GmbH, Gilching, Germany) at a sampling rate of 1 kHz. We initially measured participants' hearing thresholds using pure tone audiometry between 1 and 12.5 kHz (Sennheiser HDA200, AGRA Expsuite application, https://www.oeaw.ac.at/isf/das-institut/software/expsuite) to ensure that they deviated not more than 20 dB from their age mean. Further exclusion criteria included neurological disorders. For 23 participants we acquired individual anatomical structures and electrode positions. Our later event-related component analyses yielded missing values for three of our subjects with template attributes, hence we restricted our set to the remaining 20 subjects (9 female: $\text{mean}_{age} = 25.4$; $\text{SD}_{age} = 2.51$; 11 male: $\text{mean}_{age} = 25.4$; $\text{SD}_{age} = 3.04$). For those three subjects, comparisons of the evoked PAC activity time courses for the fully individualized vs. fully default conditions are provided as Supplementary material.

3. Methods

3.1. EEG data preprocessing

EEG data were manually inspected to detect potential noisy channels, which were then spherically interpolated. The data were subsequently bandpass-filtered between 0.5 and 100 Hz (Kaiser window, $\beta = 7.2$, $n = 462$) and epoched ($[-200, 1500]$ ms) relative to stimulus onset. We applied hard thresholds at

−200 and 800 μV to remove extremely noisy trials. Undetected bad channels were further identified through an automatic channel rejection step; if found, they would be visually inspected and interpolated. No additional noisy channels were detected for any of the subjects. We performed independent component analysis (ICA) and followed up with a manual artifact inspection and rejection of oculomotor artifacts (removal of up to three components per subject). The data were thereafter re-referenced to their average. Trials were equalized within each subject by pseudo-selection, in order to match the minimum amount within the subject after trial rejection and maintain an equal distribution across the recordings. On average, this resulted in 569 clean trials (SD = 27.7) per subject. All preprocessing steps were undertaken on the EEGLAB free software (Delorme and Makeig, 2004; RRID:SCR_007292).

3.2. Source estimation

We investigated the effects of subject anatomy, electrode locations and inverse solution on the estimated source activity, as detailed below. These analyses were implemented in the Brainstorm free software (Tadel et al., 2011; RRID:SCR_001761).

3.2.1. Subject anatomy

For subject anatomy we considered two conditions, namely a template anatomy and an individual anatomy. The template anatomy used was the standard ICBM152 brain template as implemented in Brainstorm. For the individual subject anatomies, a structural T1-weighted magnetic resonance (MR) scan for each subject was recorded at the MR center of the SCAN-Unit (Faculty of Psychology, University of Vienna) with a Siemens MAGNETOM Skyra 3 Tesla MR scanner (32-channel head coil; Siemens-Healthinieers, Erlangen, Germany). Anatomical MR scans for all subjects were subsequently segmented *via* Freesurfer (Fischl, 2012; RRID:SCR_001847), and loaded in Brainstorm. Fifteen thousand vertices were calculated for the generated surfaces, in line with the segmentation of the template anatomies used. They were then used as the basis for each subject's head model in the corresponding cases comprising individual subject anatomies. We created the anatomical models using the boundary element method in OpenMEEG (Gramfort et al., 2010; RRID:SCR_002510). Boundary surfaces were constructed by Brainstorm with 1922 vertices per layer for scalp, outer skull and inner skull, and a skull thickness of 4 mm was considered. In line with the default settings, the relative conductivity of the outer skull was set to 0.0125 and to 1 for the remaining layers. We kept the adaptive integration selected, to increase accuracy of our results.

Overall, obtaining the individual structural MRIs required about 30 min from the participant and an additional hour from the experimenter in order to schedule the session and post-process the data.

3.2.2. Electrode locations

Regarding electrode locations, we compared template-based against individually tagged ones. As template electrode locations we used the ICBM 152 BrainProducts Acticap 128 default EEG cap as implemented in Brainstorm, thereby matching our experimental setup. Individual electrode positions were acquired through the following scanning process:

An optical 3D scan (Structure Sensor with Skanect Pro, Occipital Inc., Boulder, Colorado) of each subject's head was made, after data collection and while still wearing the EEG cap. The scanner was placed on a nearby surface at the level of their ears. The seat was then being steadily rotated counter-clockwise until a 360° full turn was completed. After returning to the initial position, another quarter of a turn was done with the scanner upwards, to record the information of the upper surface of the head. The scans were visually checked immediately after their recording; if their resolution or overall quality was deemed inadequate the process was repeated. Two experimenters were present during scanning the electrode cap and thereby assessing the quality or need for repeating the measurement. Each electrode scan of the subjects' heads required at least 15 min, including setting up the system and correctly positioning the participant. Depending on each individual case, scans had to be retaken until satisfactory 3D models could be created. The individual electrodes were subsequently manually tagged on the 3D scans while we additionally added the three fiducial points (LPA/RPA/NAS). This manual procedure lasted ∼25 min per participant. Two experimenters were involved in the process of electrode tagging and MRI co-registration.

An electrode file was created as an input for the upcoming head model creation steps (Fieldtrip; Oostenveld et al., 2010; RRID:SCR_004849). The electrode order in the channel file was modified to match the corresponding order of the channels in our collected data. The default channel positions were overwritten by the individual ones in the cases comprising individual electrode positions.

3.2.3. Co-registration

For each subject and condition (combination of subject anatomy and electrode locations) we performed a manual co-registration between the head models and the channel locations using the fiducial points as an initial reference. In this process, the electrode cap was manually adjusted on either the template or the individual anatomy, using the translation, rotation, or resizing options through the graphical user interface. To minimize individual intervention and therefore inconsistencies in reproducibility, the cap was always adjusted as a whole; no channels were fine-tuned individually. After concluding the

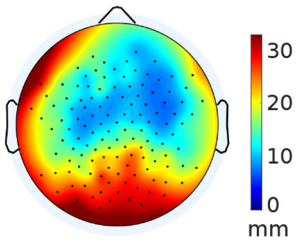


FIGURE 1
Average electrode distance between template and individual electrode locations. The distances have been averaged between anatomies, as the values show minimal deviation from each other.

realignment and in cases of offsets between electrodes and head surface, the "project to surface" functionality was used, to make sure no electrodes were placed inappropriately. This projection process was being monitored to make sure the projections did not significantly deviate from the initial tagged position. In none of the subjects were any electrodes greatly deviating from the head surface, such that a projection would alter the tagged position. The entirely non-individualized condition with templates used for both the electrode locations and subject anatomies needed no manual co-registration, as it was already accounted for by the template models.

After co-registration, the default cap locations differed from the individually tagged locations by a Euclidean distance of 17 ± 3.3 mm (mean $\pm$ SD), after adjustment to the default anatomy. Similarly, after adjustment to the individual anatomy, the locations of the default cap differed from the individually tagged locations by 16 ± 3.5 mm. Given this high similarity between the two variants of subject anatomy, we pooled the distances to further investigate their topographic distribution (Figure 1). Distances are largest at occipital channels and smallest at frontal channels. There is also a slight asymmetric bias in frontal distances that may have been caused by our scanning routine. The counter-clockwise rotation starting with the subject facing the scanner may have led to an accumulation of errors toward the end of the scanning procedure.

3.2.4. Inverse solution

With dSPM and sLORETA we selected two distributed source solutions widely used and implemented in Brainstorm. Both aim for a minimum norm estimate with implicit depth weighting to improve localization accuracy of deep sources (Lin et al., 2006), but differ in the normalization approach (Hauk et al., 2011; Nawel et al., 2019). In dSPM (Dale et al., 2000), the current density normalization is done based on the noise covariance information. In sLORETA (Pascual-Marqui, 2002; RRID:SCR_013829), the current density normalization is based on the data covariance, which is a combination of the noise covariance and a modeled brain signal covariance estimate. For the calculation of the covariances, we here considered a single-trial pre-stimulus baseline interval of $[-200, 0]$ ms. For both solutions, the source orientations were considered constrained; in that case, a dipole, that is assumed to be placed perpendicular to the cortical surface, is considered for each vertex location (Tadel et al., 2011). Noise covariance regularization was done with a factor of $\lambda^2 = 0.1$. Depth weighting and regularization parameters were selected as motivated and recommended by Brainstorm (depth weighting order = 0.5, SNR= 3 dB). Generator signals were reconstructed at 15,000 vertices describing the pial surface, representing the interface between gray matter and cerebrospinal fluid, for all configurations.

3.3. Evaluation

We focused our study on the evoked activity of the right and left PAC, defined as trasverse temporal regions by the Desikan-Killiany parcellation scheme (Desikan et al., 2006). There, we evaluated the effects of the considered individualization factors with respect to both spatial and temporal aspects.

3.3.1. Metrics

For each subject we extracted the evoked PAC activity for each hemisphere. These time series of current source densities were then low-pass filtered at 20 Hz (Hamming-based FIR, $n = 150$) with ERPLAB (Lopez-Calderon and Luck, 2014; RRID:SCR_009574) and baseline-corrected by a 100-ms-pre-event interval; the average across trials for each subject was subsequently calculated. Based on literature (Hari and Puce, 2017) as well as the grand average profiles, we defined a time interval for each of the signature components P1 and N1: $[10 - 90]$ ms was defined for P1 and $[50 - 150]$ ms for N1. In those windows, the peak amplitude (maximum for P1 and minimum for N1) and peak latency values were extracted for each component from the individual subject trial-averages, based on the `findpeaks` function as implemented in MATLAB 2018b. The single-subject data were plotted and inspected for accuracy. They were then analyzed and statistically compared based on the factors electrode location (template or individual) and subject anatomy (template or individual), individually for each hemisphere (left or right) and inverse solution (dSPM and sLORETA).

In the present study we assumed that the generating sources of P1 and N1 components are linked to focal activity in the PAC yielding maximal current source density values within it; yet localization errors likely arise, especially at neighboring vertices, due to the probability-based approach of the minimum norm estimate methods. The considered inverse solutions weigh

spatially neighboring vertices higher in order to yield a smooth distribution of current source densities (Michel and Brunet, 2019). This artifact is often referred to as spatial leakage. In order to assess the leakage of localized source activity from within the PAC toward the neighboring regions, we specified a region on the cortical surface, spatially surrounding the atlas-defined PAC for each hemisphere. In order to aid reproducibility, we decided to expand the region by recruiting additional atlas-defined surrounding regions, which the PAC activity might have leaked into. As the Desikan-Killiany parcellation was deemed too coarse, we based our new region selection on the finer Destrieux atlas (Destrieux et al., 2010). We hence constructed an extended region of interest (ROI) by merging the regions of the planum temporale, fissure, transverse temporal sulcus, circular sulcus as well as our initially defined PAC region. On the right hemisphere, the area covered by the extended ROI spanned 24.67 cm^2 vs. the 4.56 cm^2 of the initially defined PAC region (factor of 5.4). On the left side, the original PAC surface of 6.16 cm^2 was expanded to 28.03 cm^2 (factor of 4.6). For each of the two components (P1 and N1) we considered the previously extracted peak latency found for each subject average; for the exact same time points we extracted the activity of the extended ROI. We then calculated the squared amplitude ratio between the two (squared sum over PAC vertices divided by squared sum over extended ROI vertices), denoted as "ROI power ratio". This metric quantifies the proportion of evoked power contained in the PAC relative to that occurring in the extended ROI and is thus assumed to reflect the leakage to the neighboring regions in the sense that higher ratios indicate less leakage.

All aforementioned analysis procedures were implemented in MATLAB 2018b (RRID:SCR_001622).

3.3.2. Statistical analysis

Statistical analyses on the source localized time series and the extracted data relied on a mixed-model design with a multi-way ANOVA, considering a within-subject design. In particular, the analysis of the peak amplitude, peak latency, and ROI power ratio included two factors with two levels each: subject anatomy (template or individual) and electrode positions (template or individual). All ANOVAs were performed separately for each inverse solution and hemisphere.

Before each test, data were z-scored within subject and transformed according to the Box-Cox transformation (Hawkins and Weisberg, 2017). Furthermore, we ran Levene's test assessing violations in the homogeneity of variance and inspected the ANOVA residuals verifying the normality assumption. *Post-hoc* analyses of interactions/contrast have been done with Bonferroni correction. Finally, for the metrics that violated these assumptions, non-parametric aligned ranks transformation ANOVA (Wobbrock et al., 2011) was performed and the Wilcoxon test was used in the *post-hoc* analysis. Effect sizes are only reported if the parametric ANOVA was applied.

All statistical analyses were performed in R Project for Statistical Computing (RRID:SCR_001905). In addition to the standard environment, we relied on the following packages for the analysis: `afex` for the ANOVA tests (Barr et al., 2013), `emmeans` for the *post-hoc* comparison (RRID:SCR_018734), `ARTool` for the non-parametric ANOVA (Wobbrock et al., 2011), and `ggplot2` for data visualization (RRID:SCR_014601).

4. Results

4.1. Evoked PAC activity

We compared average event-related PAC responses to sounds locked to the stimulus onset. Figure 2 shows time courses comparing the different individualization levels for the two hemispheres and inverse solutions. For all source localization conditions we reconstructed stereotypical auditory-evoked responses in the PAC with a prominent positive deflection between 10 and 90 ms, denoted as the P1 component, followed by a negative deflection between 50 and 150 ms, denoting the N1 component. The left hemisphere (Figures 2A,C) was more clearly affected by the different configurations than the right hemisphere (Figures 2B,D). Later components were moreover more susceptible to latency differences than earlier components. Additionally, time series profiles differed depending on the inverse solution used; dSPM curves (Figures 2A,B) appeared more pronounced for the fully individualized configurations, while template peaks were rather more salient among the sLORETA curves (Figures 2C,D). We statistically analyzed the extracted source activation profiles for each hemisphere and inverse solution separately. Detailed information regarding the extracted values can be found in Figure 3.

Based on dSPM, the left PAC (Figure 2A) shows a differentiation depending on the degree of individualization. At P1, peak latencies were shorter for individual- than template electrode locations ($F = 6.59$, $p = 0.01$) and peak amplitudes increased with individual electrodes locations only within template anatomies ($F = 12.16$, $p < 0.001$). At later time points, the characteristics appear to be driven by the inclusion or not of a template or individual brain anatomy. Concordantly, only the use of individual anatomy yielded a significant increase of the N1 amplitude [$F_{(1, 19)} = 5.31$, $p = 0.03$, $\eta^2 = 0.22$]. N1 latencies were significantly longer for individual anatomies [$F_{(1, 19)} = 16.17$, $p < 0.001$, $\eta^2 = 0.46$] and individual electrode locations [$F_{(1, 19)} = 17.19$, $p < 0.001$, $\eta^2 = 0.47$].

In the right PAC (Figure 2B) the curves of all four individualization conditions are highly overlapping, suggesting no strong impact of any of the individualization steps.

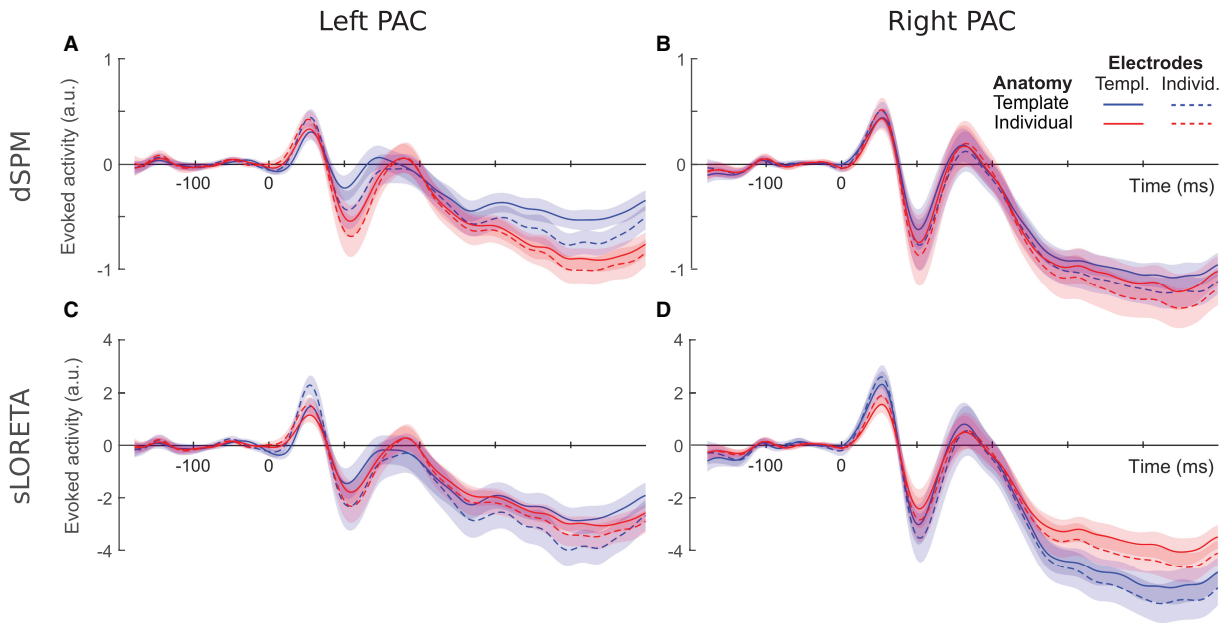


FIGURE 2
Evoked PAC activity inferred *via* dSPM (top) and sLORETA (bottom) for the left **(A,C)** and right **(B,D)** PAC. Shaded areas denote the standard error of the mean (SEM) values.

Nevertheless, individual electrode locations resulted consistently in slightly larger P1 amplitudes ($F = 16.69$, $p < 0.001$).

For the N1 component, electrode locations were again a significant factor ($F = 6.99$, $p < 0.01$), yielding more pronounced peaks with individualization. No significant effect was found on either the P1 or N1 latencies.

sLORETA applied to the left hemisphere (Figure 2C) revealed a significant interaction between anatomy and electrode locations on the P1 amplitude ($F = 11.62$, $p < 0.001$); individual electrode locations generated highest values in particular when combined with template anatomies. Significant differences were neither found on P1 latencies nor on N1 amplitudes. The N1 latency, though, was affected by both anatomy [$F_{(1, 18)} = 15.45$, $p < 0.001$, $\eta^2 = 0.46$] and electrode locations [$F_{(1, 18)} = 11.86$, $p < 0.01$, $\eta^2 = 0.40$]; overall individual anatomies produced later N1 peaks than the corresponding template configurations and individual electrode locations yielded later peaks.

On the right hemisphere (Figure 2D), anatomy [$F_{(1, 19)} = 10.31$, $p < 0.001$, $\eta^2 = 0.35$] and electrode locations [$F_{(1, 19)} = 20.38$, $p < 0.001$, $\eta^2 = 0.52$] showed significant main effects on P1 amplitude for sLORETA. Template anatomies produced higher peak values, more so in combination with individual electrode locations. The interaction between anatomy and electrode locations was found significant for P1 latency [$F_{(1, 19)} = 5.22$, $p < 0.05$, $\eta^2 = 0.22$]; the combination of individual electrode locations and template anatomies yielded the shortest values. Anatomy caused a differentiation on the N1 peak values ($F = 7.05$, $p < 0.01$), where template anatomies led to slightly more pronounced peaks. No significant effects were found on N1 latency.

4.2. Spatial leakage

We inferred the brain activity not only to the PAC but to the entire cortical surface. The corresponding brainmaps are shown in Figure 4 for dSPM and Figure 5 for sLORETA. There was a clear evoked activation in the temporal region, yet that activation differed in its precise location and spread, depending on the particular configuration considered. Generally, the activation was attributed to regions extending rather more superior and posterior than the atlas-defined PAC (cyan outline) and the overall pattern seemed to be dominated by the subject anatomies. The configurations comprising individual subject anatomies exhibited more constrained activation patterns within the atlas-defined PAC, whereas those with the template anatomies showed a higher spread of activation. This was most pronounced within the extended ROI area surrounding and including the PAC but also reached further to other parts of the temporal and parietal lobes. When regarding the general brain profiles, electrode locations appeared to play a secondary

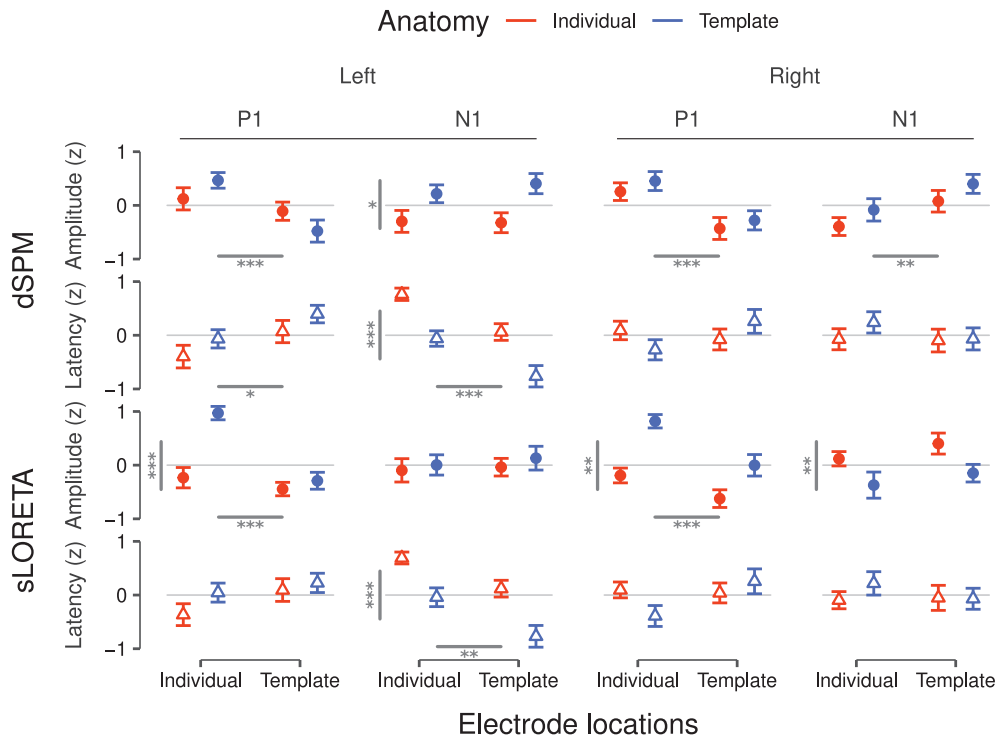


FIGURE 3
Peak amplitudes and latencies for P1 and N1 components inferred *via* dSPM **(top)** and sLORETA **(bottom)** for the left and the right hemisphere. Values correspond to the average over subjects after z-scoring per condition and within subject. Error bars denote the SEM. Statistically significant main effects are reported as gray bars—horizontal bars for electrode position and vertical bars for brain anatomy—and from one to three asterisks indicating the significance levels (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Significant interactions are reported in Table 1.

role; individual electrode locations did not consistently seem to constrain the activation further. Between the two inverse solutions, the spread was noticeably more distributed in the use of sLORETA, compared to dSPM.

Complementary to the brainmaps, we evaluated the P1 and N1 power ratios between the PAC and the extended ROI (Figure 6). When considering the dSPM inverse solution and the left hemisphere, individualization of the electrode locations led to higher P1 ratio [$F_{(1, 19)} = 5.27$, $p < 0.05$, $\eta^2 = 0.22$]. No significant effects were found on the corresponding ratio for the N1 component.

In the right hemisphere under the dSPM inverse solution, anatomy had a significant main effect on both the P1 ($F = 71.14$, $p < 0.001$) and N1 [$F_{(1, 19)} = 21.84$, $p < 0.001$, $\eta^2 = 0.53$] power ratios. Individual anatomies generated higher power ratios at the peak of both components.

When considering the sLORETA inverse solution in the left hemisphere, anatomy [$F_{(1, 19)} = 7.05$, $p < 0.05$, $\eta^2 = 0.27$] and electrode locations [$F_{(1, 19)} = 9.51$, $p < 0.01$, $\eta^2 = 0.33$] were main effects for the P1 power ratio. Individual subject anatomies led to lower power ratio values, while individual electrode locations ameliorated the result. The corresponding N1 power ratios were significantly affected only by anatomy ($F = 4.94$, $p < 0.05$), with individualized subject anatomies leading to a deterioration of the value, hence denoting higher leakage.

In the corresponding right hemisphere, anatomy was a significant main effect for both P1 ($F = 69.92$, $p < 0.001$) and N1 [$F_{(1, 19)} = 17.64$, $p < 0.001$, $\eta^2 = 0.48$]. In both cases individualization benefited the localization accuracy.

5. Discussion

In the present study our aim was to single out the effects of different individualization steps on the accuracy of inferring PAC activity from EEG data. We compared combinations of template or individualized electrode locations and subject anatomies while using two different inverse solutions (dSPM and sLORETA). Through that we reconstructed and characterized the evoked PAC time series and assessed the spatial leakage around the PAC in each hemisphere. Table 1 summarizes the significant effects for all configurations and their consistency with individualization benefit. As evident, both the factors of electrode location as well as subject anatomy were found to have an impact on the defined current source

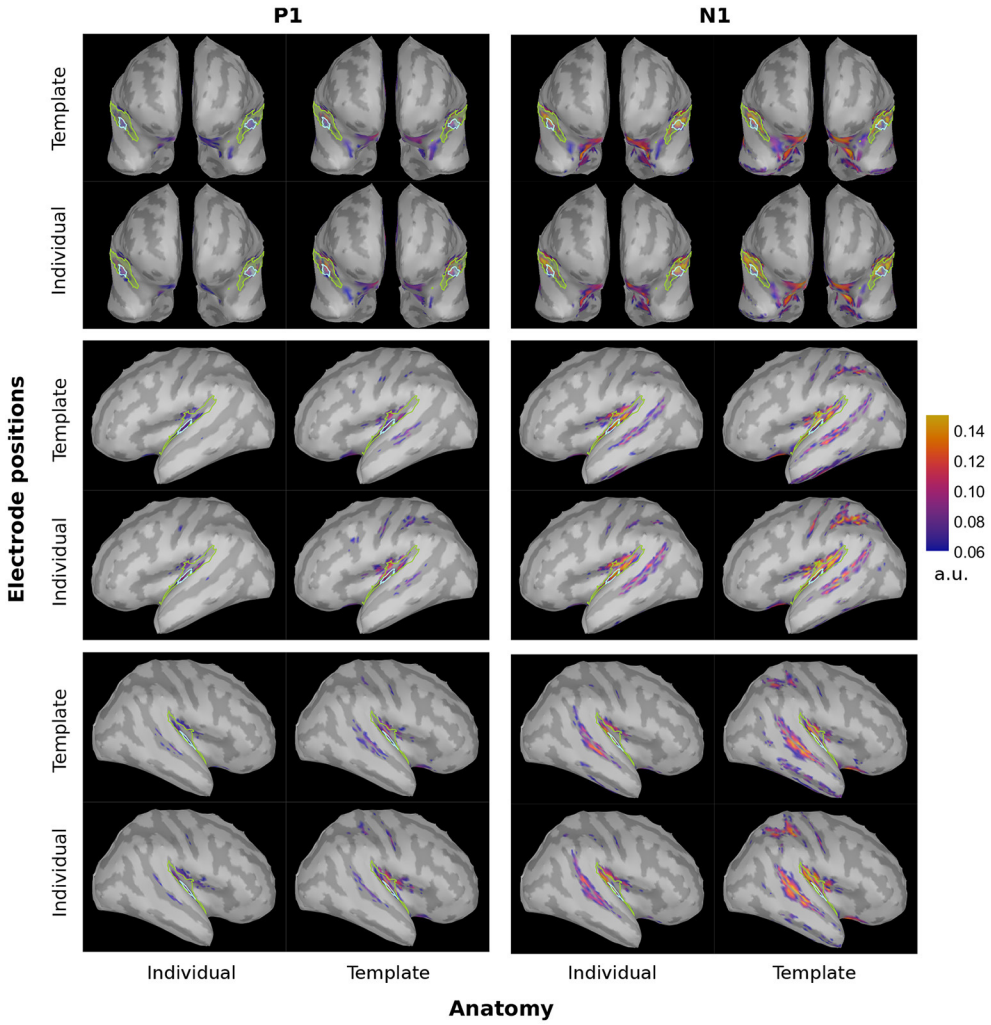


FIGURE 4
dSPM brainmaps depicting the PAC activation at the P1 (~55 ms, top) and N1 (~100 ms, bottom) timing for both hemispheres and all considered configurations. For display, no minimal cluster size was set and the minimum amplitude was set at 20% of the maximum activation across all four configurations of the inverse solution. Cyan: atlas-defined PAC; Green: extended ROI.

density metrics; yet their effect varied depending on the target brain area (PAC in the left or right hemisphere), the evoked component characteristic (amplitude or latency of P1 or N1), and also the type of inverse solution (dSPM or sLORETA) considered.

5.1. Individualization factors

The considered individualization factors influenced the two hemispheres quite differently. In the right hemisphere, anatomy affected mainly the power ratio indicating spatial leakage, while electrode positions had an impact on peak amplitudes. Contrary to that, we found a more complex pattern of effects in the left hemisphere: individualized solutions gave earlier peaks for P1 and later ones for N1, individual electrode locations increased both the P1 amplitude and power ratio, and individual anatomies interacted with that effect on P1 amplitude and independently enlarged N1 amplitudes.

We speculate that such a regional variance in source reconstruction could be resulting from either state or trait effects. On the one hand, source localization estimates could show higher variance across subjects because of brain morphology (i.e., cerebral size, Bartley et al., 1997; handedness, Good et al., 2001) that, on a group level, may result in higher or lower uncertainty for different regions, especially for template solutions. Higher inter-individual variability is also generally found in the left auditory cortex (Ren et al., 2021). On the other hand, the observed asymmetry might be related to stimulus features; auditory stimulation

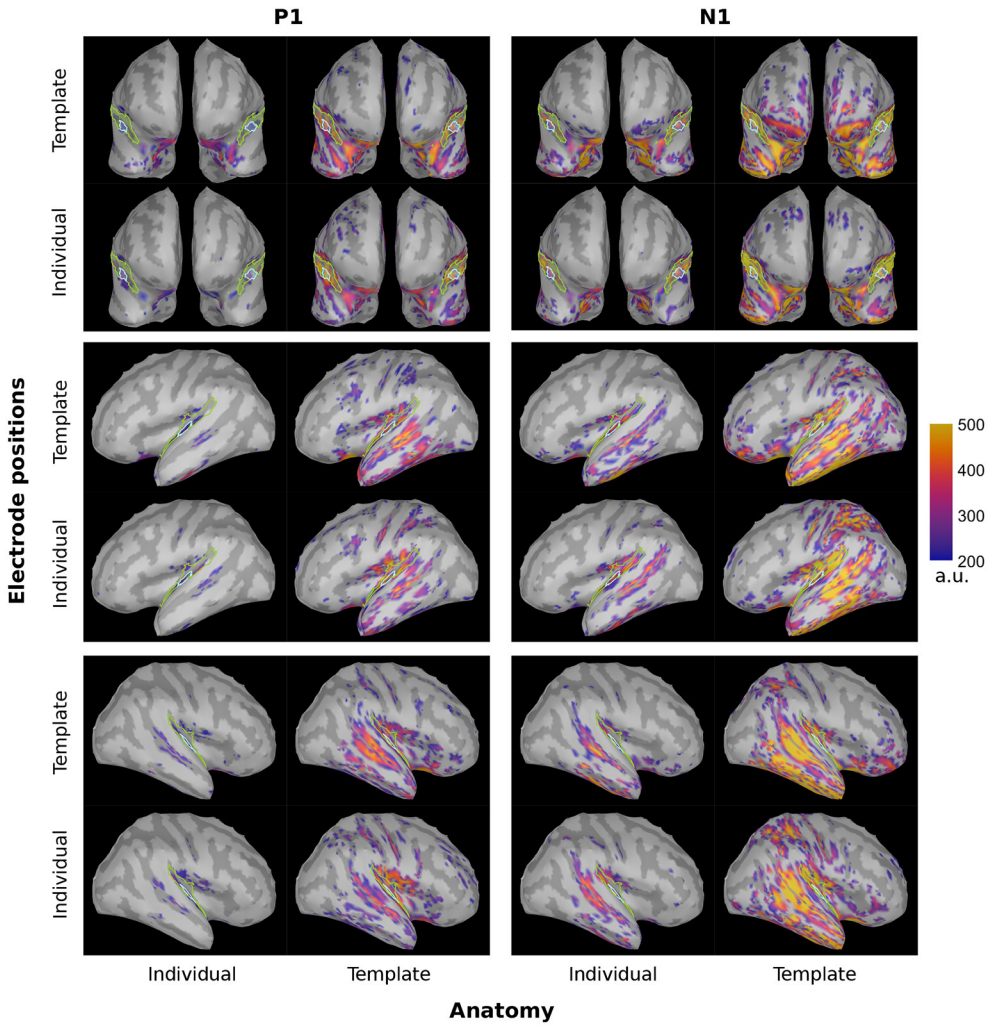


FIGURE 5
sLORETA brainmaps depicting the PAC activation at the P1 (~55 ms, top) and N1 (~100 ms, bottom) timing for both hemispheres and all considered configurations. For display, no minimal cluster size was set and the minimum amplitude was set at 20% of the maximum activation across all four configurations of the inverse solution. Cyan: atlas-defined PAC; Green: extended ROI.

characteristics preferentially processed on either hemisphere, such as unattended automatic change detection of spectral or temporal features, may influence the variance of the source reconstruction estimates (Schönwiesner et al., 2007; Okamoto et al., 2009). Our stimuli moreover carry spatial characteristics, as they are presented from either the left or right side of- and at different distances from the listener. Spatial processing has been shown to exhibit right-hemispheric dominance, possibly explaining the differences we observe (Kaiser et al., 2000; Middlebrooks, 2015; Deng et al., 2020). From a more technical perspective, it could also be that the slight asymmetry we observed in distances between default and individual electrode positions may have contributed to hemispheric asymmetries in inferred source activity, but only when using individual electrode locations.

Individual anatomies and electrode locations allow for a more precise attribution of the recorded activity to the corresponding regions, thereby likely accounting for the various inter-individual variability characteristics. Acquiring individual electrode locations, though, usually comes with considerable measurement uncertainty; to some degree this also depends on the acquisition strategy. With our procedure, a considerable amount of the experimenter's individual intervention is necessary in obtaining the 3D scan as well as tagging the electrodes. The extent of it might differ, when more automatized—and therefore also more reproducible—methods are used (Koessler et al., 2011; Hirth et al., 2020), potentially yielding different effects regarding the choice between template or individual electrode locations.

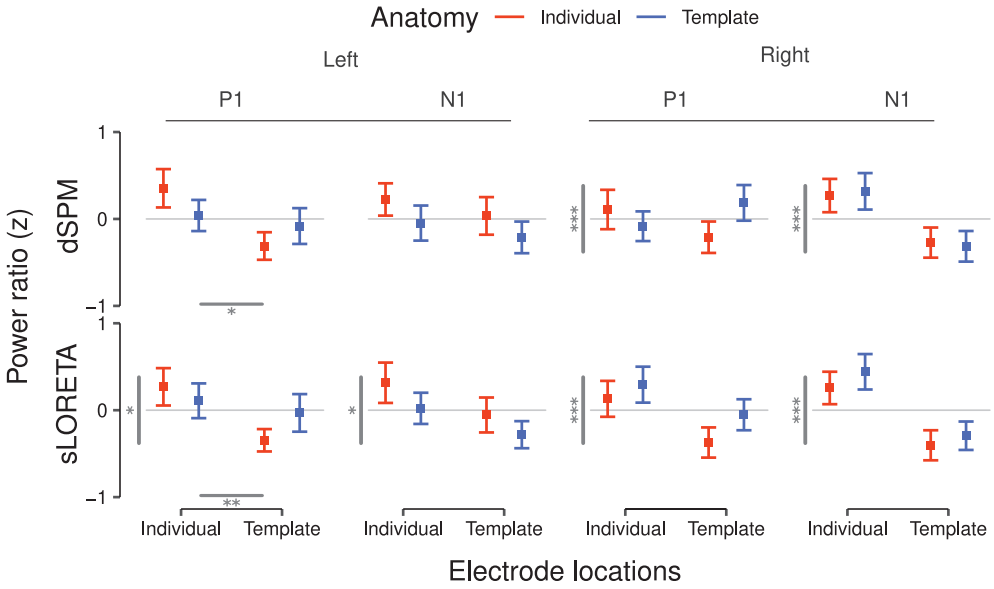


FIGURE 6
ROI power ratios for P1 and N1 components inferred *via* dSPM **(top)** and sLORETA **(bottom)** for the left and right hemisphere. Higher ratios are interpreted as indication for lower spatial leakage of inferred PAC source activity. Values correspond to the average over subjects after z-scoring per hemisphere, inverse solution, and subject. Error bars denote the SEM. Statistically significant differences between main effects are reported as gray bars—horizontal bars for electrode positions and vertical bars for brain anatomies—and from one to three asterisks indicating the significance levels (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Significant interactions are reported in Table 1.

TABLE 1 Summary of the effects of anatomy (A) and electrode locations (E) as well as their interaction (A:E) resulting from the statistical tests.

	Hemisphere	Left						Right					
	Component	P1			N1			P1			N1		
	Factor	A	E	A:E	A	E	A:E	A	E	A:E	A	E	A:E
dSPM	Amplitude		↑	-	↓				↑			↓	
	Latency		←		→	→							
	Power ratio		↑					↑			↑		
sLORETA	Amplitude	↓	↑	-				↓	↑		↑		
	Latency				→	→				-			
	Power ratio	↓	↑		↓			↑			↑		
Inverse solution	Attributes												

Only significant results are reported ($p < 0.05$). Arrows indicate the direction of change from the template factor level to the individual factor level. Dashes indicate opposing interactions. Color coding denotes the assessment of every such change as individualization benefit (green) or degradation (red). If this interpretation on the direction of change seemed ambiguous, as is the case for latency changes and opposing interactions, the cell has not been colored.

As we were interested in localizing the PAC we restricted our search on the cortical surface and this is where our results apply. Our choice of constrained sources in the brain might be an essential contributor to our outcomes: when individual anatomies are unavailable, selecting fixed sources might be too restricting and introduce errors in the considered orientations (Hillebrand and Barnes, 2003; Westner et al., 2022); therefore a different setting might be more suited for the case of template anatomies.

In order to extract the targeted cortical activity, we focused on the PAC region as defined by the Desikan-Killiany atlas. Yet, as seen on Figures 4, 5, none of the configurations seem to perfectly capture the core of the PAC activation. Different atlases vary in their parcellation; as a result, using a different parcellation scheme for such an investigation might capture the activation differently and hence lead to deviating results regarding the accuracy of localization. Another possibility could be to move away from an atlas-based- and toward a functional

ROI definition. Extending the study by manually defining the PAC based on the observed, and therefore actual source activation, could increase the effect sizes and give further insight into the localization precision offered by the individualization steps.

Given the aforementioned limitations and despite the clear loss in spatial acuity obtained without individualization, we found the stereotypical auditory evoked response elicited within the PAC in all configurations (Figure 2). In that regard the EEG based source localization of early evoked activity may be considered satisfactory in all cases. This is important especially in occasions where no individualization steps can be taken, as could happen with infants, implantees, or situations where the corresponding resources (time, personnel, or funds) are not available. Contrarily, there are cases where individualization is indispensable. Such can be investigations with known underlying structural differences, as could be in the case of hearing loss (Alfandari et al., 2018; Chen et al., 2021; Manno et al., 2021). In a general setting, though, where no such restrictions apply, our results can aid in the direction of designing the aspects of an experimental study: depending on the effect examined and available resources, decisions can be made about whether template configurations would be sufficient or a further individualization, whether electrode locations or subject anatomy, would be in order.

5.2. Differences between inverse solutions

Regarding the choice of an inverse solution itself, different algorithms are based on different prior assumptions (Grech et al., 2008). We here restricted our study to two widely used methods falling under the same algorithmic category (minimum-norm solutions); an informative and oftentimes suggested way is to compare different algorithms before drawing conclusions on the plausibility of the results (Nawel et al., 2019).

When comparing the analyses outcomes of our considered source localization configurations as shown in Table 1, the differences between inverse solutions become noticeable. With sLORETA individualization steps yielded rather inconsistent main effects: we found some benefit of individual electrode positions, yet anatomy seemed to work in the opposite direction than what was expected. Inclusion of individual subject anatomies had incongruent effects on our metrics, oftentimes leading to a deterioration of the accuracy with higher individualization (Table 1, red cells). Contrary to that, dSPM showed consistent results: all main effects contributed toward an amelioration of the considered values with individualization of either the subject anatomy or the electrode positions.

Though not reflected in the metrics of Table 1, there is a considerable difference in the overall activity spread between dSPM and sLORETA. The activity is largely distributed over the temporal and parietal lobes using sLORETA, while with dSPM it remains rather spatially focused. As such, dSPM seems to be more suitable for capturing more focal auditory processes targeting specific regions implicated in them.

In sum, our findings demonstrate the benefit of using additional individualized information regarding brain anatomy and electrode positioning; they further support previous notions toward using dSPM for investigating auditory processes (Stropahl et al., 2018). A restricted activation profile can be especially beneficial, for instance, when considering a differentiation between the ventral and dorsal auditory stream, both of which also comprise relatively small areas (Bizley and Cohen, 2013).

Data availability statement

The data used for the present study are available *via* the OSF platform under https://doi.org/10.17605/OSF.IO/YM26X.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KI and RBau designed the study. KI performed the EEG analyses. RBar performed the statistical analyses. KI, RBau, and BT interpreted the results. KI and RBar wrote the first draft of the manuscript and all authors revised it. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Austrian Science Fund (FWF) within the projects Born2Hear (I4294-B) and Dynamates (ZK66).

Acknowledgments

We would like to thank Dr. Ronald Sladky for his assistance during the collection of the MRI data at the SCAN-Unit of the Faculty of Psychology, University of Vienna (https://scan-psy.univie.ac.at). A prior version of the current manuscript has been uploaded as a preprint on the BioRxiv server (Ignatiadis et al., 2022).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2022.970372/full#supplementary-material

References

Adde, G., Clerc, M., Faugeras, O., Keriven, R., Kybic, J., and Papadopoulo, T. (2003). Symmetric BEM formulation for the M/EEG forward problem. *Inform. Process. Med. Imaging* 18, 524–535. doi: 10.1007/978-3-540-45087-0_44

Akalin-Acar, Z., and Gençer, N. G. (2004). An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging. *Phys. Med. Biol.* 49, 5011–5028. doi: 10.1088/0031-9155/49/21/012

Akhtari, M., McNay, D., Mandelkern, M., Teeter, B., Cline, H. E., Mallick, J., et al. (1994). Somatosensory evoked response source localization using actual cortical surface as the spatial constraint. *Brain Topogr.* 7, 63–69. doi: 10.1007/BF01184838

Alfandari, D., Vriend, C., Heslenfeld, D. J., Versfeld, N. J., Kramer, S. E., and Zekveld, A. A. (2018). Brain volume differences associated with hearing impairment in adults. *Trends Hear.* 22:2331216518763689. doi: 10.1177/2331216518763689

Baier, D., Ignatiadis, K., Tóth, B., and Baumgartner, R. (2022). *Attentional Modulation and Cue-Specificity of Cortical Biases in Favour of Looming Sounds*. Technical report, Zenodo.

Baillet, S., Mosher, J., and Leahy, R. (2001). Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 18, 14–30. doi: 10.1109/79.962275

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bartley, A. J., Jones, D. W., and Weinberger, D. R. (1997). Genetic variability of human brain size and cortical gyral patterns. *Brain* 120, 257–269. doi: 10.1093/brain/120.2.257

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci.* 14, 693–707. doi: 10.1038/nrn3565

Buchner, H., Waberski, T. D., Fuchs, M., Wischmann, H. A., Wagner, M., and Drenckhahn, R. (1995). Comparison of realistically shaped boundary-element and spherical head models in source localization of early somatosensory evoked potentials. *Brain Topogr.* 8, 137–143. doi: 10.1007/BF01199777

Chen, Q., Lv, H., Wang, Z., Wei, X., Zhao, P., Yang, Z., Gong, S., and Wang, Z. (2021). Brain structural and functional reorganization in tinnitus patients without hearing loss after sound therapy: a preliminary longitudinal study. *Front. Neurosci.* 15:573858. doi: 10.3389/fnins.2021.573858

Dalal, S., Rampp, S., Willomitzer, F., and Ettl, S. (2014). Consequences of EEG electrode position error on ultimate beamformer source reconstruction performance. *Front. Neurosci.* 8:42. doi: 10.3389/fnins.2014.00042

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *Neuron* 26, 55–67. doi: 10.1016/S0896-6273(00)81138-1

Darvas, F., Ermer, J. J., Mosher, J. C., and Leahy, R. M. (2006). Generic head models for atlas-based EEG source analysis. *Hum. Brain Mapp.* 27, 129–143. doi: 10.1002/hbm.20171

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Deng, Y., Choi, I., and Shinn-Cunningham, B. (2020). Topographic specificity of alpha power during auditory spatial attention. *NeuroImage* 207:116360. doi: 10.1016/j.neuroimage.2019.116360

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Fischl, B. (2012). FreeSurfer. *NeuroImage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Good, C. D., Johnsrude, I., Ashburner, J., Henson, R. N., Friston, K. J., and Frackowiak, R. S. (2001). Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage* 14, 685–700. doi: 10.1006/nimg.2001.0857

Gramfort, A., Papadopoulo, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *BioMed. Eng. OnLine* 9:45. doi: 10.1186/1475-925X-9-45

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *J. NeuroEng. Rehabil.* 5:25. doi: 10.1186/1743-0003-5-25

Haładaj, R. (2020). Anatomical variations of the dentate gyrus in normal adult brain. *Surg. Radiol. Anat.* 42, 193–199. doi: 10.1007/s00276-019-02298-5

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., et al. (2007). Review on solving the forward problem in EEG source analysis. *J. Neuroeng. Rehabil.* 4:46. doi: 10.1186/1743-0003-4-46

Hari, M. D., and Puce, P. (2017). *MEG-EEG Primer*. Oxford: Oxford University Press.

Hauk, O. (2004). Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *NeuroImage* 21, 1612–1621. doi: 10.1016/j.neuroimage.2003.12.018

Hauk, O., Wakeman, D. G., and Henson, R. (2011). Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *Neuroimage* 54, 1966–1974. doi: 10.1016/j.neuroimage.2010.09.053

Hawkins, D. M., and Weisberg, S. (2017). Combining the box-cox power and generalised log transformations to accommodate nonpositive responses in linear and mixed-effects linear models. *South Afr. Stat. J.* 51, 317–328. doi: 10.37920/sasj.2017.51.2.5

Hillebrand, A., and Barnes, G. R. (2003). The use of anatomical constraints with MEG beamformers. *NeuroImage* 20, 2302–2313. doi: 10.1016/j.neuroimage.2003.07.031

Hine, J., and Debener, S. (2007). Late auditory evoked potentials asymmetry revisited. *Clin. Neurophysiol.* 118, 1274–1285. doi: 10.1016/j.clinph.2007.03.012

Hirth, L. N., Stanley, C. J., Damiano, D. L., and Bulea, T. C. (2020). Algorithmic localization of high-density EEG electrode positions using motion capture. *J. Neurosci. Methods* 346:108919. doi: 10.1016/j.jneumeth.2020.108919

Holtmann, L., Hans, S., Kaster, F., Müller, V., Lang, S., Göricke, S., et al. (2021). Magnet dislocation following magnetic resonance imaging in cochlear implant

users: diagnostic pathways and managment. *Cochlear Implants Int*. 22, 195–202. doi: 10.1080/14670100.2021.1872906

Hsu, Y. -F., Xu, W., Parviainen, T., and Hämäläinen, J. A. (2020). Context-dependent minimisation of prediction errors involves temporal-frontal activation. *NeuroImage* 207:116355. doi: 10.1016/j.neuroimage.2019.116355

Ignatiadis, K., Barumerli, R., Tóth, B., and Baumgartner, R. (2022). Benefits of individualized brain anatomies and EEG electrode positions for auditory cortex localization. *bioRxiv [preprint]*. doi: 10.1101/2022.06.15.496307

Jaworska, N., Blier, P., Fusee, W., and Knott, V. (2012). Scalp- and sLORETA-derived loudness dependence of auditory evoked potentials (LDAEPs) in unmedicated depressed males and females and healthy controls. *Clin. Neurophysiol*. 123, 1769–1778. doi: 10.1016/j.clinph.2012.02.076

Justen, C., and Herbert, C. (2018). The spatio-temporal dynamics of deviance and target detection in the passive and active auditory oddball paradigm: a sLORETA study. *BMC Neurosci*. 19:25. doi: 10.1186/s12868-018-0422-3

Kaiser, J., Lutzenberger, W., Preissl, H., Ackermann, H., and Birbaumer, N. (2000). Right-hemisphere dominance for the processing of sound-source lateralization. *J. Neurosci*. 20, 6631–6639. doi: 10.1523/JNEUROSCI.20-17-06631.2000

Koessler, L., Cecchin, T., Caspary, O., Benhadid, A., Vespignani, H., and Maillard, L. (2011). EEG-MRI Co-registration and sensor labeling using a 3D laser scanner. *Ann. Biomed. Eng*. 39, 983–995. doi: 10.1007/s10439-010-0230-0

Koessler, L., Maillard, L., Benhadid, A., Vignal, J. -P., Braun, M., and Vespignani, H. (2007). Spatial localization of EEG electrodes. *Clin. Neurophysiol*. 37, 97–102. doi: 10.1016/j.neucli.2007.03.002

Komssi, S., Huttunen, J., Aronen, H. J., and Ilmoniemi, R. J. (2004). EEG minimum-norm estimation compared with MEG dipole fitting in the localization of somatosensory sources at S1. *Clin. Neurophysiol*. 115, 534–542. doi: 10.1016/j.clinph.2003.10.034

Kudela, P., Boatman-Reich, D., Beeman, D., and Anderson, W. S. (2018). Modeling neural adaptation in auditory cortex. *Front. Neural Circuits* 12:72. doi: 10.3389/fncir.2018.00072

Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., and Papadopoulo, T. (2005). A common formalism for the integral formulations of the forward EEG problem. *IEEE Trans. Med. Imaging* 24, 12–28. doi: 10.1109/TMI.2004.837363

Leahy, R. M., Mosher, J. C., Spencer, M. E., Huang, M. X., and Lewine, J. D. (1998). A study of dipole localization accuracy for MEG and EEG using a human skull phantom. *Electroencephalogr. Clin. Neurophysiol*. 107, 159–173. doi: 10.1016/S0013-4694(98)00057-1

Leinung, M., Loth, A., Gröger, M., Burck, I., Vogl, T., Stöver, T., et al. (2020). Cochlear implant magnet dislocation after MRI: surgical management and outcome. *Eur. Arch. Oto-Rhino-Laryngol*. 277, 1297–1304. doi: 10.1007/s00405-020-05826-x

Lin, F. -H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., and Hämäläinen, M. S. (2006). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage* 31, 160–171. doi: 10.1016/j.neuroimage.2005.11.054

Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci*. 8:213. doi: 10.3389/fnhum.2014.00213

Manno, F. A. M., Rodríguez-Cruces, R., Kumar, R., Ratnanather, J. T., and Lau, C. (2021). Hearing loss impacts gray and white matter across the lifespan: systematic review, meta-analysis and meta-regression. *NeuroImage* 231:117826. doi: 10.1016/j.neuroimage.2021.117826

Michel, C. M., and Brunet, D. (2019). EEG source imaging: a practical review of the analysis steps. *Front. Neurol*. 10:325. doi: 10.3389/fneur.2019.00325

Middlebrooks, J. C. (2015). "Sound localization," in *Handbook of Clinical Neurology, Vol*. 129, eds M. J. Aminoff, F. Boller, and D. F. Swaab (Amsterdam: Elsevier B.V.), 99–116. doi: 10.1016/B978-0-444-62630-1.00006-8

Mohan, A., Bhamoo, N., Riquelme, J. S., Long, S., Norena, A., and Vanneste, S. (2020). Investigating functional changes in the brain to intermittently induced auditory illusions and its relevance to chronic tinnitus. *Hum. Brain Mapp*. 41, 1819–1832. doi: 10.1002/hbm.24914

Nawel, J., Abir, H., Ichrak, B., Amal, N., and Chokri, B. A. (2019). "A comparison of inverse problem methods for source localization of epileptic MEG spikes," in *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)* (Athens), 867–870. doi: 10.1109/BIBE.2019.00161

Okamoto, H., Stracke, H., Draganova, R., and Pantev, C. (2009). Hemispheric asymmetry of auditory evoked fields elicited by spectral versus temporal stimulus change. *Cereb. Cortex* 19, 2290–2297. doi: 10.1093/cercor/bhn245

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. -M. (2010). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci*. 2011:e156869. doi: 10.1155/2011/156869

Ou, W., Hämäläinen, M. S., and Golland, P. (2009). A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* 44, 932–946. doi: 10.1016/j.neuroimage.2008.05.063

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol*. 24(Suppl. D), 5–12.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiol. Neuro-Otol*. 4, 64–79. doi: 10.1159/000013823

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi: 10.1111/1469-8986.3720127

Raghavan, M., Li, Z., Carlson, C., Anderson, C. T., Stout, J., Sabsevitz, D. S., et al. (2017). MEG language lateralization in partial epilepsy using dSPM of auditory event-related fields. *Epilepsy Behav*. 73, 247–255. doi: 10.1016/j.yebeh.2017.06.002

Ren, J., Xu, T., Wang, D., Li, M., Lin, Y., Schoeppe, F., et al. (2021). Individual variability in functional organization of the human and monkey auditory cortex. *Cereb. Cortex* 31, 2450–2465. doi: 10.1093/cercor/bhaa366

Schönwiesner, M., Krumbholz, K., Rübsamen, R., Fink, G. R., and von Cramon, D. Y. (2007). Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. *Cereb. Cortex* 17, 492–499. doi: 10.1093/cercor/bhj165

Schroeder, M. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (Corresp.). *IEEE Trans. Inform. Theory* 16, 85–89. doi: 10.1109/TIT.1970.1054411

Schwartz, D., Lemoine, D., Poiseau, E., and Barillot, C. (1996). Registration of MEG/EEG data with 3D MRI: methodology and precision issues. *Brain Topogr*. 9, 101–116. doi: 10.1007/BF01200710

Somers, B., Long, C. J., and Francart, T. (2021). EEG-based diagnostics of the auditory system using cochlear implant electrodes as sensors. *Sci. Rep*. 11:5383. doi: 10.1038/s41598-021-84829-y

Stropahl, M., Bauer, A. -K. R., Debener, S., and Bleichner, M. G. (2018). Source-modeling auditory processes of EEG data using EEGLAB and brainstorm. *Front. Neurosci*. 12:309. doi: 10.3389/fnins.2018.00309

Taberna, G. A., Marino, M., Ganzetti, M., and Mantini, D. (2019). Spatial localization of EEG electrodes using 3D scanning. *J. Neural Eng*. 16, 026020. doi: 10.1088/1741-2552/aafdd1

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci*. 2011:879716. doi: 10.1155/2011/879716

Van Hoey, G., Vanrumste, B., D'Havé, M., Van de Walle, R., Lemahieu, I., and Boon, P. (2000). Influence of measurement noise and electrode mislocalisation on EEG dipole-source localisation. *Med. Biol. Eng. Comput*. 38, 287–296. doi: 10.1007/BF02347049

von Ellenrieder, N., Muravchik, C. H., Wagner, M., and Nehorai, A. (2009). Effect of head shape variations among individuals on the EEG/MEG forward and inverse problems. *IEEE Trans. Bio-Med. Eng*. 56, 587–597. doi: 10.1109/TBME.2009.2008445

Vorwerk, J., Cho, J. -H., Rampp, S., Hamer, H., Knösche, T. R., and Wolters, C. H. (2014). A guideline for head volume conductor modeling in EEG and MEG. *NeuroImage* 100, 590–607. doi: 10.1016/j.neuroimage.2014.06.040

Wang, Y., and Gotman, J. (2001). The influence of electrode location errors on EEG dipole source localization with a realistic head model. *Clin. Neurophysiol*. 112, 1777–1780. doi: 10.1016/S1388-2457(01)00594-6

Westner, B. U., Dalal, S. S., Gramfort, A., Litvak, V., Mosher, J. C., Oostenveld, R., et al. (2022). A unified view on beamformers for M/EEG source reconstruction. *NeuroImage* 246:118789. doi: 10.1016/j.neuroimage.2021.118789

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11* (New York, NY: Association for Computing Machinery), 143–146. doi: 10.1145/1978942.1978963

Yang, S., Zhao, Z., Cui, H., Zhang, T., Zhao, L., He, Z., et al. (2019). Temporal variability of cortical gyral-sulcal resting state functional activity correlates with fluid intelligence. *Front. Neural Circuits* 13:36. doi: 10.3389/fncir.2019.00036

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 16 November 2022
DOI 10.3389/fninf.2022.942978

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
National Research Council (CNR), Italy

REVIEWED BY
Milutin Stanacevic,
Stony Brook University, United States
Norbert Dillier,
University of Zurich, Switzerland

*CORRESPONDENCE
Duowei Tang
duowei.tang@kuleuven.be

†PRESENT ADDRESS
Maja Taseska,
Microsoft Applied Sciences, Munich, Germany

RECEIVED 13 May 2022
ACCEPTED 21 October 2022
PUBLISHED 16 November 2022

CITATION
Tang D, Taseska M and van Waterschoot T (2022) Toward learning robust contrastive embeddings for binaural sound source localization. *Front. Neuroinform.* 16:942978. doi: 10.3389/fninf.2022.942978

Toward learning robust contrastive embeddings for binaural sound source localization

Duowei Tang*, Maja Taseska† and Toon van Waterschoot

Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Leuven, Belgium

Recent deep neural network based methods provide accurate binaural source localization performance. These data-driven models map measured binaural cues directly to source locations hence their performance highly depend on the training data distribution. In this paper, we propose a parametric embedding that maps the binaural cues to a low-dimensional space where localization can be done with a nearest-neighbor regression. We implement the embedding using a neural network, optimized to map points that are close to each other in the latent space (the space of source azimuths or elevations) to nearby points in the embedding space, thus the Euclidean distances between the embeddings reflect their source proximities, and the structure of the embeddings forms a manifold, which provides interpretability to the embeddings. We show that the proposed embedding generalizes well in various acoustic conditions (with reverberation) different from those encountered during training, and provides better performance than unsupervised embeddings previously used for binaural localization. In addition, the proposed method performs better than or equally well as a feed-forward neural network based model that directly estimates the source locations from the binaural cues, and it has better results than the feed-forward model when a small amount of training data is used. Moreover, we also compare the proposed embedding using both supervised and weakly supervised learning, and show that in both conditions, the resulting embeddings perform similarly well, but the weakly supervised embedding allows to estimate source azimuth and elevation simultaneously.

KEYWORDS

manifold learning, non-linear dimension reduction, *siamese* neural network, binaural sound source localization, deep learning

1. Introduction

Sound source localization is aiming to estimate a sound source position in terms of azimuth, elevation, and distance. A large part of the source localization literature focuses on the azimuth and elevation estimation only, hence this is also the scope we adopt in this paper. The human auditory system is capable of localizing acoustic signals using binaural cues

such as the Interaural Phase Differences (IPDs) and Interaural Level Differences (ILDs) (Blauert, 1997). Computational localization algorithms in robot audition (Argentieri et al., 2015), hearing aid (Farmani et al., 2018), virtual reality (Keyrouz and Diepold, 2007), etc., aim at mimicking this process and therefore estimate the binaural cues from binaural microphone signals. The binaural microphones are typically two identical microphones that are mounted at the entries of two ear canals of an artificial head. In a sound source localization scenario, the human/artificial head together with the pinna and the torso act as filters that modify the incident sound waves. This filter effect is crucial for sound source localization, especially vertical sound source localization (i.e., elevation estimation), and can be characterized by the Head-related Transfer Function (HRTF) (Risoud et al., 2018).

Acoustic artifacts such as noise and reverberation, introduce uncertainties in the binaural cues. Although the existence of reverberation can aid distance localization (Risoud et al., 2018), the resulting noisy and reverberant binaural cues make sound source localization challenging. Traditionally, robustness to reverberation has been tackled with statistical model-based approaches (Mandel et al., 2010; May et al., 2011; Woodruff and Wang, 2012), which outperform lookup tables and template matching methods that rely on an anechoic assumption (Raspaud et al., 2010; Karthik and Ghosh, 2018). Some works propose to estimate the direct-path relative transfer function, which encodes the source azimuth information, in order to avoid the contamination of audio from reverberation noise, however, this type of methods highly rely on the onset of the source acoustic events (Li et al., 2016).

In contrast, data-driven approaches are able to learn the non-linear functions that map binaural cues to source locations (Datum et al., 1996). Recently, Deep Neural Networks (DNNs) has been used to learn the relationship between azimuth and binaural cues, by exploiting head movements to resolve the front-back ambiguity (Ma et al., 2017), and by combining spectral source models to robustly localize the target source in a multiple sources scenario (Ma et al., 2018). Additionally, a few works use DNNs to enhance the binaural features so that they can eliminate reverberation and additive noise (Pak and Shin, 2019; Yang et al., 2021). In Yalta et al. (2017) and Vecchiotti et al. (2019), the authors utilize DNNs to directly map the audio spectrogram or its raw waveform to the source azimuth in an end-to-end manner, which is also applicable to reverberant and noisy environments. However, those works only consider source azimuth estimation and the localization is done by classification (i.e., the predictions can only be in a pre-defined grid).

A different data-driven approach was used in Deleforge and Horaud (2012) and Deleforge et al. (2015), where the relationship between source locations and binaural cues was modeled with a probabilistic piecewise linear function. By learning the function parameters, sources can be localized by probabilistic inversion. An implicit assumption of the piecewise linear model in Deleforge and Horaud (2012) and Deleforge et al. (2015) is that similar source locations result in similar binaural cues. The same assumption is also used in non-parametric source localization algorithms based on manifold learning in Laufer et al. (2013) and Laufer-Goldshtein et al. (2015). In this paper, we focus on data-driven source localization approaches, inspired by low-dimensional manifold learning (Laufer et al., 2013; Laufer-Goldshtein et al., 2015).

Manifold learning in sound source localization is aiming to find a non-linear transformation that transforms acoustic measurements to a low-dimensional representation that preserves the source locality information. Manifold learning methods in Laufer et al. (2013) and Laufer-Goldshtein et al. (2015) rely on smoothness in the measurement space with respect to the underlying source locations, an assumption that might generalize poorly to varying acoustic conditions. The uncertainties in the binaural cue measurements introduced by reverberation, introduce variations in the measurement space neighborhoods that might not be consistent with their source locations. To preserve neighborhoods in term of the source location, we are inspired by the *"siamese"* neural network in the machine learning community that is optimized with a contrastive loss function (Hadsell et al., 2006). This particular model learns a similarity metric defined in the latent space (i.e., written digit classes and orientation of air plane pictures in Hadsell et al., 2006). This paradigm, which doesn't rely on an explicit neighborhoods definition in the measurement space, is suitable for problems that have a large amount of classes and in each class there are only a few training examples, such as face verification (Chopra et al., 2005; Taigman et al., 2014) and signature verification (Bromley et al., 1993), and can also be used in sound source localization. We have proposed and published earlier a regression method for binaural sound source localization based on the *"siamese"* neural network and contrastive loss in Tang et al. (2019). This method converts binaural cues into a low-dimensional embedding, and there is a small Euclidean distance between the embeddings obtained from binaural cues of similar source locations. A similar work using triplet loss somewhat resembles our idea (Opochinsky et al., 2019), but in their work, a model directly maps the binaural cues to source location predictions, and pre-defined proximity for both positive and negative cases (i.e., points with similar and dissimilar source locations) have to be present at the same time for the triplet loss.

In this paper, we first propose an update on the model architecture introduced in Tang et al. (2019), and then validate its robustness with respect to three aspects:

1. mismatched audio content between the training and testing sets,
2. the presence of unknown reverberation and noise,
3. and the availability of only a small amount of annotated training data,

through abundant experiments in fixed and varying acoustic scenarios, respectively. Afterwards, we extend our method to a weakly supervised learning scheme, where the annotation of source directions (i.e., azimuth and elevation) is no longer required for training the embeddings, but only the relative source position proximity is needed for any pair of training examples. Unlike the supervised approach proposed in Tang et al. (2019), which treats azimuth and elevation estimation in two separate tasks, this weakly supervised embedding can be used to estimate both the source azimuth and elevation at the same time, and providing a good visualization of the manifold.

The proposed methods have potential in a number of practical applications where the location of a sound source is to be identified, for example in signal processing front-ends for hearing aids, and in an intelligent interactive dialogue systems, to localize the speaker for denoizing beamformers, or for a synthesizer to render stereo sounds. Note that the proposed methods start from binaural signal features, which implies that binaural rather than bilateral hearing aids are required when using these methods for sound source localization in hearing aid systems, and the issue of binaural hearing aids that need to transmit and synchronize the binaural features needed for this model is beyond the scope of this paper. Yet there is a large body of research literature that addresses this issue, and the reader is referred to Kreisman et al. (2010), Ibrahim et al. (2013), Wei et al. (2014), and Geetha et al. (2017).

The paper is organized as follows. In Section 2, we first revise the binaural cue extraction and formulate the source localization problem. Then, in Section 3, we provide a brief overview of the related manifold learning work that has been applied in binaural sound source localization. Next, the proposed method is presented in Section 4 and finally, experimental results are shown in Section 5.

2. Data model and problem formulation

2.1. Binaural cue extraction

Let $s_1(\tau)$ and $s_2(\tau)$ denote the signals captured at the left and right microphones in a binaural recording setup in a noisy and reverberant environment. In this work, we extract the binaural cues in the Short-time Fourier transform (STFT) domain, as in Raspaud et al. (2010) and Deleforge et al. (2015).

Let $S_1(t,k)$ and $S_2(t,k)$ denote the STFT coefficients of $s_1(\tau)$ and $s_2(\tau)$, where t and k are the time frame and frequency index, respectively. At a time-frequency bin (t,k) an ILD α_{tk} and an IPD ϕ_{tk} are defined as

$$\alpha_{tk} = 20\log_{10}\frac{|S_1(t,k)|}{|S_2(t,k)|}, \quad \phi_{tk} = \angle\frac{S_1(t,k)}{S_2(t,k)}. \tag{1}$$

Assuming that a single sound source is active, we follow the binaural feature extraction approach from Deleforge et al. (2015), and compute time-averaged ILDs and IPDs across T frames as follows

$$a_k = T^{-1}\sum_{t=1}^{T}\alpha_{tk}, \quad p_k = T^{-1}\sum_{t=1}^{T}\exp(j\phi_{tk}). \tag{2}$$

By concatenating the ILDs, and the real and imaginary parts of the IPDs in selected frequency ranges $[k_1,k_2]$ and $[k_3,k_4]$, the binaural information is summarized in a measurement vector $\boldsymbol{x}\in\mathcal{X}\subset\mathbb{R}^D$,

$$\boldsymbol{x} = [a_{k_1},\ldots,a_{k_2},\mathcal{R}\{p_{k_3}\},\mathcal{I}\{p_{k_3}\},\ldots,\mathcal{R}\{p_{k_4}\},\mathcal{I}\{p_{k_4}\}]^{\mathrm{T}} \tag{3}$$

with dimensionality $D = k_2 - k_1 + 2(k_4 - k_3)$.

It is known that IPDs carry reliable location cues below 2 kHz (Blauert, 1997), while ILDs contribute to localization at higher frequencies as well (Deleforge et al., 2015). Hence, we used the ranges $\frac{f_s}{K}[\tilde{k_1},\tilde{k_2}] = [200; 7,000]$ Hz for ILDs and $\frac{f_s}{K}[\tilde{k_3},\tilde{k_4}] = [200; 2,500]$ Hz for IPDs, where f_s denotes the sampling frequency and K is the Discrete Fourier transform (DFT) size used in the STFT, and $k_i = \mathrm{round}(\tilde{k_i}), i = 1,2,3,4$, where the round() operation rounds $\tilde{k_i}$ to the closest integer. For a typical audio recording with sampling rate $f_s = 16$ kHz, and the DFT size $K = 1,024$, the dimensionality D is equal to 729 (i.e., a 729-dimensional feature vector $\boldsymbol{x}$).

2.2. Measurement to embedding transformation

From the above binaural cue extraction process, a pair of signals $s_1(\tau)$ and $s_2(\tau)$ is associated to a vector $\boldsymbol{x}\in\mathcal{X}$. We refer to $\mathcal{X}$ as the *measurement* space. Let the unknown source location be denoted by $\boldsymbol{u}\in\mathcal{U}$. We refer to $\mathcal{U}$ as the *latent space*. $\mathcal{U}$ is one-dimensional if one considers azimuth or elevation separately, or two-dimensional if the localization angles are considered simultaneously. Given a training set of N pairs $\mathcal{T} = \{(\boldsymbol{x}_i,\boldsymbol{u}_i)\}_{i=1}^{N}$, the localization problem consists of finding a function h

$$\hat{\boldsymbol{u}} = h(\boldsymbol{x}), \quad h:\mathcal{X}\to\mathcal{U}. \tag{4}$$

that accurately maps measurements to latent variables. Although, one can implement h with a powerful non-linear model (e.g., a DNN), the proposed approach of first transforming the measurement space to an embedding space and then performing the localization in the embedding space comes with several advantages:

1. Learning the transformation from measurement space to embedding space does not necessarily require the latent space annotation information, thus enables the possibility of semi-supervised learning and weakly supervised learning.

2. The low-dimensional embedding can preserve the latent space neighborhood relationships (in which the Euclidean distance in the embedding space roughly corresponds to the latent space "semantic" relationship) and the embedding eliminates useless information, which can be used to study or visualize the latent space structure. A vanilla example of this is the Principal Component Analysis (PCA).
3. By learning the structure of the latent space, the training of the model will be less dependent on the distribution of the training data. In contrast, if the mapping from measurement space to latent space is learned directly, the model is more likely to over-fit to the dense part of the training data and its generalization capability decreases when there is not enough annotated training data.

Therefore, our main objective in this work is to learn an embedding function f that maps the vectors $\boldsymbol{x}$ to a low-dimensional space which preserves latent space neighborhoods, i.e.,

$$\boldsymbol{z} = f(\boldsymbol{x}), \quad f : \mathcal{X} \to \mathcal{Z} \subset \mathbb{R}^d, \quad d << D. \tag{5}$$

We propose a neural network framework to learn a parametric function f that satisfies these properties both in a supervised and weakly supervised manner. A nearest-neighbor regression function $h : \mathcal{Z} \to \mathcal{U}$ is then used for localization.

3. Baseline manifold learning method

If the microphone location in a given room is fixed, the authors in Laufer-Goldshtein et al. (2015) showed that features extracted from binaural signals can be embedded in a low-dimensional space $\mathcal{Z}$, in a way that recovers source locations. The framework in Laufer-Goldshtein et al. (2015) is based on unsupervised manifold learning, in particular, *Laplacian eigenmaps* (LEM) (Belkin and Niyogi, 2003).

The Laplacian Eigenmaps (LEM) method defines the neighborhood relationships of the data using a similarity matrix $\boldsymbol{K} \in \mathbb{R}^{N \times N}$, with entries $\boldsymbol{K}[i,j]$ related to the Euclidean distances $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$ between feature vectors x_i and x_j, with $i, j \in [1, N]$. One way to compute $\boldsymbol{K}$ is using nearest-neighbors, i.e., $\boldsymbol{K}[i,j] = \boldsymbol{K}[j,i] = 1$ if $\boldsymbol{x}_i$ is among the M nearest neighbors of $\boldsymbol{x}_j$, or if $\boldsymbol{x}_j$ is among the M nearest neighbors of $\boldsymbol{x}_i$ (in Euclidean distance). A second way is using an exponentially decaying kernel function, such as the Gaussian kernel

$$\boldsymbol{K}[i,j] = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{\varepsilon}\right), \tag{6}$$

where ε is the kernel bandwidth. Such kernel is used for source localization in Laufer-Goldshtein et al. (2015).

Given the similarity matrix $\boldsymbol{K}$, the neighborhood-preserving optimization problem of LEM to find the embeddings $z_1, z_2, \ldots, z_N$ is given by (Belkin and Niyogi, 2003)

$$\begin{aligned} &\underset{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N}{\arg\min} \quad \sum_{i,j=1}^{N} \|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2^2 \, \boldsymbol{K}[i,j], \\ &\text{subject to} \quad \boldsymbol{Z}^T \boldsymbol{D} \boldsymbol{Z} = \boldsymbol{I} \end{aligned} \tag{7}$$

which enforces that points x_i, x_j with large similarity $\boldsymbol{K}[i,j]$, are to be mapped to points z_i, z_j with a small Euclidean distance $\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$ where $\boldsymbol{D}$ is a diagonal matrix with entries $\boldsymbol{D}[i,i] = \sum_{j=1}^{N} \boldsymbol{K}[i,j]$.

The optimization problem (7) has a closed-form solution, given by the eigenvectors of $\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{K}$ corresponding to the largest eigenvectors. If $\{\boldsymbol{\psi}_i\}_{i=1}^N$ denote the eigenvectors of $\boldsymbol{P}$, with eigenvalues $1 = \lambda_1 > \lambda_2 \geq \ldots, \geq \lambda_N$, the d-dimensional LEM embedding is given by (Belkin and Niyogi, 2003)

$$\boldsymbol{z}_i = f(\boldsymbol{x}_i) = \left[\, \boldsymbol{\psi}_2[i], \; \boldsymbol{\psi}_3[i], \; \ldots, \; \boldsymbol{\psi}_{d+1}[i] \right]^{\mathrm{T}}, \tag{8}$$

where the constant eigenvector $\boldsymbol{\psi}_1$ is not included (Chung, 1997; Belkin and Niyogi, 2003) and $[i]$ denotes the vector element index. The LEM embedding f is non-parametric, and the low-dimensional representation $\boldsymbol{z}$ of a new measurement $\boldsymbol{x}$ is obtained as a linear combination of the training points $\{\boldsymbol{z}_i\}_{i=1}^N$ (Bengio et al., 2003). However, this procedure is often insufficiently accurate and represents a disadvantage of LEM and of spectral embeddings in general. One can include every new testing data and re-run the unsupervised training to get a more accurate representation for the new testing data, however, this may prolong the training time, especially for large datasets, and due to the fact that the kernel matrix $\boldsymbol{K}$ is $N \times N$, the computation of eigenvectors will dramatically increase for a large N.

Besides the promising performance of spectral embeddings for localization (Laufer et al., 2013; Laufer-Goldshtein et al., 2015; Taseska and van Waterschoot, 2019), their major drawback is the assumption that the neighborhoods in the measurement space are consistent with the source locations. Although the assumption is shown to hold when all signals are recorded in one room for fixed microphone locations (Deleforge and Horaud, 2012; Laufer-Goldshtein et al., 2015; Taseska and van Waterschoot, 2019), this is not the case when the signals are filtered by various acoustic channels in different enclosures.

4. Contrastive embedding for localization

We propose a parametric embedding, designed to preserve neighborhoods in terms of sound source locations. Such embeddings are robust to unseen room reverberation and small training set size (e.g., when the training set does not contain the complete latent space annotations). The proposed framework

firstly includes the definition of the neighborhoods, which can be supervised (Section 4.1) or weakly supervised (Section 4.2) depending on whether one uses the azimuth/elevation label or the source relative proximity. Secondly it includes the transformation from the measurement space to the embedding space by training a DNN which is optimized on a contrastive loss function (Sections 4.3 and 4.4). Finally the sound source localization will be performed in the embedding space using nearest-neighbor regression (Section 4.5).

4.1. Supervised neighborhoods definition

Consider two labeled measurements $(\boldsymbol{x}_i, u_i)$ and $(\boldsymbol{x}_j, u_j)$ where u_i and u_j are denoted as scalars since we estimate azimuth and elevation separately. To avoid the phase wrapping ambiguity, we define $d_u(u_i, u_j) = \min(|u_i - u_j|, 360^\circ - |u_i - u_j|)$ denote the shortest possible distance in the latent space $\mathcal{U}$, where u_i, u_j corresponds to the source azimuth or elevation angles in degree. A neighborhood indicator $y_{ij} \in \{0, 1\}$ is defined as

$$y_{ij} = \begin{cases} 0, & \text{if} \quad d_u(u_i, u_j) > \epsilon_u \\ 1, & \text{if} \quad d_u(u_i, u_j) \leq \epsilon_u, \end{cases} \tag{9}$$

for a user-defined threshold angle ϵ_u.

4.2. Weakly supervised neighborhoods definition

As an alternative to directly using the latent space label information to define the neighborhoods, we can also use the relative proximity between sound sources. Here, we only consider the sound sources at the ball with radius Φ and centered at the receiver, or sources whose relative position to the receiver can be found (then the source locations can be firstly projected onto a ball with radius Φ around the receiver by distance normalization).

In order to define the weakly supervised neighborhoods, we can use the physical distance $d_s(S_i, S_j)$ between two sound sources S_i and S_j which corresponds to the Euclidean distance between the Cartesian coordinate vectors of S_i and S_j. Similarly,

$$y'_{ij} = \begin{cases} 0, & \text{if} \quad d_s(S_i, S_j) > \epsilon_s \\ 1, & \text{if} \quad d_s(S_i, S_j) \leq \epsilon_s, \end{cases} \tag{10}$$

for a user-defined threshold distance ϵ_s. The threshold ϵ_s and ϵ_u are related as ϵ_s represents the arc length of the angle ϵ_u on a circle with radius Φ and hence,

$$\epsilon_s \approx \epsilon_u \cdot \Phi \cdot \pi / 180^\circ \tag{11}$$

In particular, in our proposed method, one can also implicitly define the similarity indicator y'_{ij} by using it as a training data label. For example, consider a scenario when multiple recordings are acquired from excitations at each of the pre-defined sound source locations, then y'_{ij} equals to 1 for recordings acquired at the same or at close source locations, and y'_{ij} equals to 0 for recordings acquired at different or far source locations.

4.3. Contrastive loss

We seek to learn a parametric function $f_W : \mathcal{X} \to \mathcal{Z} \subset \mathbb{R}^d$, with parameters W, that maps $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ to their low-dimensional embeddings $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. If $y_{ij} = 1$, the Euclidean distance $\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$ should be small, and if $y_{ij} = 0$, then $\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2$ should be large. For a given embedding function f_W, we have

$$\|\boldsymbol{z}_i - \boldsymbol{z}_j\|_2 = \|f_W(\boldsymbol{x}_i) - f_W(\boldsymbol{x}_j)\|_2. \tag{12}$$

A *contrastive loss function* over the parameters W, tailored for neighborhood preservation has been proposed in Hadsell et al. (2006) for non-linear dimensionality reduction, and is given by

$$L(W) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Big(y_{ij} \|f_W(\boldsymbol{x}_i) - f_W(\boldsymbol{x}_j)\|_2^2 + (1 - y_{ij}) \max(0, \mu_{ij} - \|f_W(\boldsymbol{x}_i) - f_W(\boldsymbol{x}_j)\|_2)^2 \Big). \tag{13}$$

The parameter μ_{ij} is a positive real-valued margin, such that $\mu_{ij}/2$ can be interpreted as the same radius of circles centered on $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. If the circles intersect and $y_{ij} = 0$, the two dissimilar pairs are too close in the embedding space, thus increasing the *contrastive loss* in (13). On the other hand, if $y_{ij} = 1$, large distances are penalized, enforcing f_W to preserve neighborhoods.

Intuitively speaking, during the training, each example in a mini-batch is subjected to two "forces." One force is between the similar pairs, pulls them closer to each other in the embedding space. The other force between dissimilar pairs is repulsive and it pushes the dissimilar pair away from each other in the embedding space (if they are too close when $\|f_W(\boldsymbol{x}_i) - f_W(\boldsymbol{x}_j)\|_2 < \mu_{ij}$). During training, the embeddings are moving according to the forces they encounter, and thus will eventually lead to an equilibrium (i.e., convergence). Globally, the embedding space convergences to a manifold. Since the forces are subjected to latent space similarities, this will result in meaningful distances between each pair of embeddings (i.e., the distance between a pair of embeddings somewhat indicates the proximity of their corresponding sound sources).

It is important to note that in Hadsell et al. (2006), where the *contrastive loss* was first proposed for classification, $\mu_{ij} \equiv \mu$

is a constant margin. In our application, the latent space of azimuths and elevations is continuous. To accurately preserve its geometry, we propose an adaptive margin as follows,

$$\mu_{ij} = \frac{\exp\left(d_{ij}\right)}{\exp\left(d_{ij}\right) + 1}. \tag{14}$$

As d_{ij} decreases, the margin μ_{ij} decreases as well. One can compute d_{ij} either in a supervised manner using the azimuth/elevation, thus $d_{ij} = d_u(u_i, u_j)$, or in a weakly supervised manner, where $d_{ij} = d_s(S_i, S_j)$. In the case that there is no quantitative measure in the latent space, a constant margin can be used (e.g., $\mu = 1$).

4.4. Learning the embedding

We implement f_W with a DNN as shown in Figure 1A. The DNN architecture consists of two fully-connected hidden layers with D neurons in each layer. Between the fully connected layers, we add batch-normalization layers (Ioffe and Szegedy, 2015) to speed up the convergence and dropout layers to prevent the model from over-fitting (Srivastava et al., 2014). The output layer has three neurons, corresponding to a three-dimensional embedding space, i.e., $d = 3$. The hidden neurons have *Sigmoid* non-linear activations, and the output neurons have linear activations. In order to train the DNN model to minimize the cost function in (13), we use the *siamese* architecture that was proposed in Bromley et al. (1993) and used for various tasks in Chopra et al. (2005) and Hadsell et al. (2006). This special DNN architecture consists of two identical branches that are sharing the same model parameters. Taking a pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ as an input, the measurements $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are passed through the branches (one per branch) and hence produce their corresponding embeddings $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. Then the cost is evaluated in (13) using the neighborhood indicator y_{ij} and the outputs $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ of the branches. Finally, the gradient per model parameter is calculated and back-propagated to update the model parameters. Depending on which definition for the neighborhood indicator is used, we call the corresponding embedding Supervised Contrastive Embedding (SCE) if the supervised neighborhoods definition is used, or Weakly-supervised Contrastive Embedding (WSCE) if the weakly supervised neighborhoods definition is used.

A key aspect of the proposed framework is the selection of pairs $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for training. For small datasets, one could consider all pairs and proceed with training on all training data pairs. However the polynomial growth of the number of pairs results in memory problems even for moderately large datasets. To solve this problem, we use mini-batches and calculate the neighborhood indicator y_{ij} for every pair of examples in each mini-batch. To be noted, we suggest to choose a large enough batch size so that there are both similar pairs and dissimilar pairs in one batch. Because a randomly selected mini-batch generally contains examples from sources of different locations (i.e., those examples will be defined as dissimilar pairs), if the batch size is too small, the probability of having similar pairs in a batch will be very low, so that the loss will be inaccurately evaluated and thus slow down the convergence rate. Intuitively, if there is no similar pair in a batch, the embeddings will not be subjected to a pulling force to their similar points. This would lead to the embeddings that just randomly reside in the embedding space and form local clusters.

4.5. Nearest-neighbor localization

Once the weights of f_W are optimized, we compute the embedding of a new $\boldsymbol{x}$ by a forward-pass through the DNN model. Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ denote the K nearest-neighbors of $\boldsymbol{z}$ in the training set. The latent variable (azimuth or elevation) is then estimated as

$$\hat{u} = \sum_{i=1}^{K} w_i u_i, \quad \text{with} \quad w_i = \frac{\exp\left(-\frac{\|\boldsymbol{z}-\boldsymbol{z}_i\|_2^2}{\varepsilon}\right)}{\sum_{j=1}^{K} \exp\left(-\frac{\|\boldsymbol{z}-\boldsymbol{z}_j\|_2^2}{\varepsilon}\right)}. \tag{15}$$

The bandwidth ε of the exponential kernel is obtained as the median of the squared distances from the K neighbors, i.e.,

$$\varepsilon = \text{median}\left(\|\boldsymbol{z} - \boldsymbol{z}_1\|_2^2, \ldots, \|\boldsymbol{z} - \boldsymbol{z}_K\|_2^2\right). \tag{16}$$

Note that if the embedding is accurately preserving neighborhoods, the choice of regression weights is not critical. For instance w_i can be inversely proportional to $\|\boldsymbol{z} - \boldsymbol{z}_i\|_2^2$. However, in our experiments, the latter generally leads to less accurate location estimates than exponentially decaying weights.

5. Experiments

5.1. Experimental settings

To evaluate the proposed SCE in terms of the localization error and robustness, we compare the SCE with two baseline methods:

1. The LEM embeddings (Laufer et al., 2013; Laufer-Goldshtein et al., 2015) with nearest neighbor localization.
2. A feed-forward neural network which is optimized with the Mean Squared Error (MSE) loss. This feed-forward neural network has the same structure as one of the branches in the proposed *siamese* structure except for an additional output layer with *tanh* activation functions that outputs the source location predictions, shown in Figure 1B. Since the *tanh* activation function has a range of $(-1, 1)$, we normalize the training labels also to the same range by $\hat{u}_i = u_i/180^\circ$,

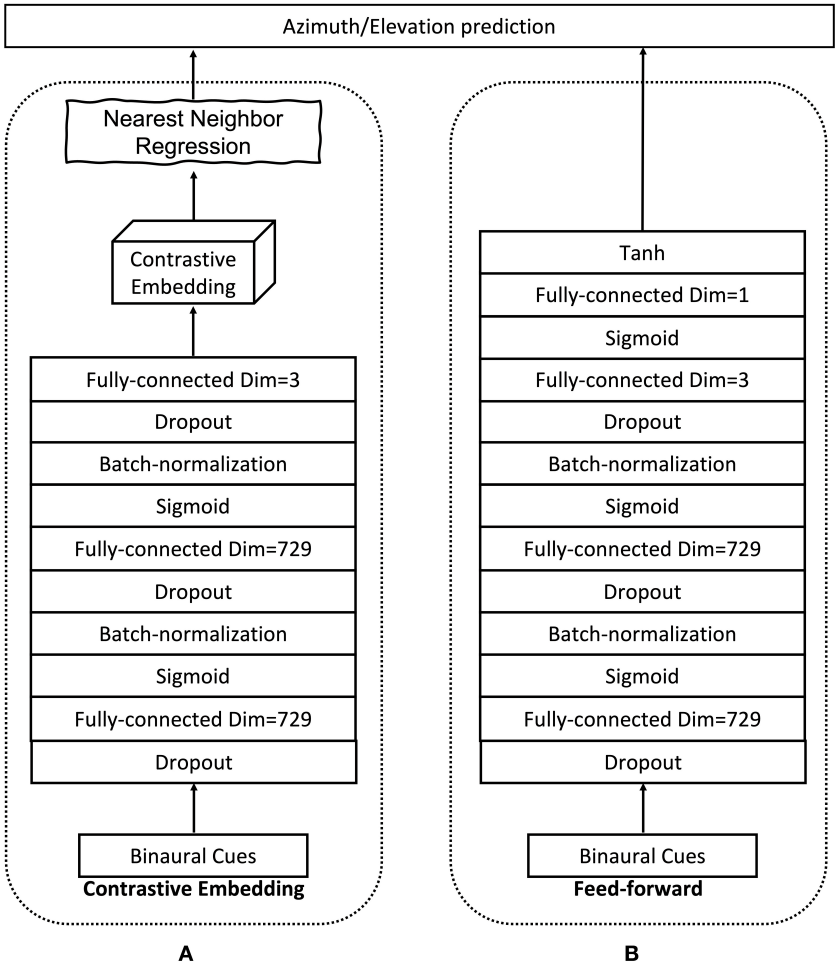


FIGURE 1
Model architecture. The proposed contrastive embedding model in **(A)**, and the feed-forward model in **(B)**.

and $i = 1 \ldots N$. Note that, the original labels have a range $[-180, 180^\circ]$. During testing, the feed-forward predictions are firstly converted back to degree before calculating the localization errors.

As the neighborhoods for LEM are defined in the input space, a single embedding is used to estimate both azimuth and elevation. Similarly, our proposed method can be trained to estimate azimuth and elevation simultaneously as well by using the weakly supervised neighborhoods definition introduced in Section 4.2. However, a system with two separately trained embeddings might provide better results for the same amount of data, which we will compare for SCE and WSCE in the later experiments.

For the nearest neighbor regression in (15), $K = 5$ neighbors are used in all localization experiments. A few threshold values ϵ_u in (9) and (11) are tested for both azimuth and elevation. We choose ϵ_u in $\{5, 15, 30^\circ\}$ to have a big span so that we can evaluate its impact on the localization results. Essentially, ϵ_u is a hyper-parameter that can be tuned with a validation set. We implemented the LEM using a nearest neighbor kernel $\boldsymbol{K}$ with $M = 10$ nearest neighbors, which in our experiments, provided better results than the Gaussian kernel used in Laufer-Goldshtein et al. (2015) and Taseska and van Waterschoot (2019).

For DNN training, we use the Adam optimizer (Kingma and Ba, 2015) with a learning rate equal to 10^{-3} that is automatically halved if the validation performance does not improve after 20 epochs. The mini-batch size is set to 128, and this will result 8,128 pairs of measurements per mini-batch for training. We select the model based on the best validation performance, and then the selected model is used to calculate the testing set predictions.

All audio files are sampled at 16 kHz. To extract the ILD and IPD features, we use the STFT with a cosine window of 1,024 samples at 16 kHz, 75% overlapping.

5.2. Datasets

5.2.1. Fixed acoustic conditions

With the first dataset, we want to verify the effectiveness of our proposed methods for preserving the locality information of the audio source when the training and the testing set have different audio content (and different spectral distribution). We employ the CAMIL dataset which consists of binaural recordings and was gathered using a Sennheiser MKE 2002 dummy head in a real-life reverberant room (i.e., a room with a few furnitures and background noise; Deleforge et al., 2015). To generate recordings that have different azimuth and elevation angles, a loudspeaker (i.e., the source) is placed at a fixed position, 2.7 m from the dummy head (i.e., the receiver). The dummy head is mounted on a step-motor which generates 10,800 pan-tilt states. This results in source azimuth and elevation angle in the range $[-180, 180^\circ]$ and $[-60, 60^\circ]$, respectively (with 2°resolution). To only evaluate the methods in localizing frontal sources, we select the recordings that have source azimuth and elevation angle in the range $[-90, 90^\circ]$ and $[-45, 45^\circ]$, respectively. The CAMIL dataset consists of a training set made using white noise (1 s per recording), and a testing set made using 1–5 s speech samples from the TIMIT corpus (Garofolo et al., 1993). We further randomly divide the whole training set into a smaller training set (consisting of 70% samples from the original training set), and a validation set (consisting of the remaining 30% samples from the original training set). Finally, spatially uncorrelated white noise with a Signal to Noise Ratio (SNR) of 15 dB is added to the testing set.

5.2.2. Varying acoustic conditions

With the second dataset, we want to verify the robustness of the proposed methods for varying acoustic conditions. We use the VAST dataset (Gaultier et al., 2017) of simulated binaural room impulse responses of a KEMAR dummy head (Gardner and Martin, 1995; Schimmel et al., 2009). The training set consists of 16 different rooms with reverberation time 0.1–0.4 s. For each room we select spherical grids of source positions with radii 1, 1.5, and 2 m, centered at nine predefined receiver positions (inside each room). Similarly to the fixed acoustic conditions in Section 5.2.1, we use 70% of randomly selected data as the training set, and the remaining 30% as the validation set. The receiver's height is fixed at 1.7 m. Then two testing sets are provided:

- *Testing-set-1*: The source and receiver are placed at random positions in the same 16 rooms as the training set.
- *Testing-set-2*: The source and receiver are placed in shoebox rooms of random width and length between 3×2 and 10×4 m, with absorption profiles randomly picked from those of the training rooms. Those rooms have reverberation time 0.1–0.4 s.

All the training set's and testing sets' Head-related impulse responss (HRTFs) are simulated using the image source method (Allen and Berkley, 1979) and provided by the VAST dataset (Gaultier et al., 2017).

As in Section 5.2.1, we have only selected recordings that have frontal angles. To focus on the influence of the varying room acoustics while exciting all frequencies, 2 s white noise source signals were considered in this experiment.

5.3. SCE for unidimensional source localization

5.3.1. Tuning the dropout rate

We first determine an optimal dropout rate for both the SCE method and the feed-forward model by line search. We test dropout rate values in {0.0, 0.2, 0.5, 0.8}, and similarity threshold values ϵ_u for SCE equals to 5 and 15° (denoted by "_sim5" and "_sim15," respectively). The azimuth/elevation localization error of the validation sets for both the CAMIL dataset and the VAST dataset are plotted in Figure 2.

In Figures 2A,B, the azimuth and elevation estimation results for the CAMIL dataset are illustrated, respectively. We can observe that the SCE has better validation performance than the feed-forward model for all testing dropout rates, and its localization error is essentially equal to zero when using either similarity threshold value, i.e., 5 or 15°. The feed-forward model exhibits a clear concave curve in median localization error and has the lowest localization error at the dropout rate value of 0.2, thus indicating that a dropout rate equal to 0.2 is an optimal value for the feed-forward method.

In Figures 2C,D, the median azimuth and elevation localization error for the VAST dataset are illustrated respectively. Both the SCE and the feed-forward model in this case exhibit a concave curve in median localization error and they both exhibit an optimal dropout rate of 0.2. We also observe that, in the VAST azimuth validation performance, the SCE_sim5 performs equally well as the feed-forward model when dropout rate is 0.2, which is slightly better than SCE_sim15. In the elevation estimation, SCE_sim5 performs the best over the feed-forward model and SCE_sim15.

Based on the validation results, we choose the dropout rate equal to 0.2 for both the SCE and the feed-forward methods for the next experiments.

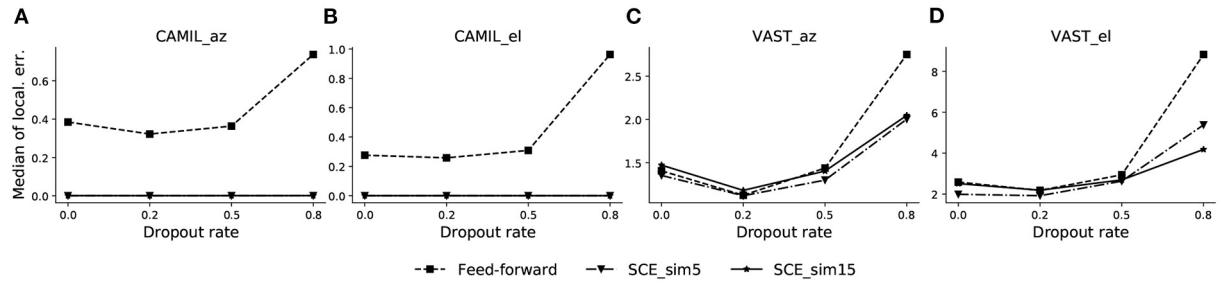


FIGURE 2
Validation performance across different dropout rates for CAMIL validation set **(A,B)**, and VAST validation set **(C,D)**. The dropout rate is tested for both the proposed SCE with similarity threshold angle ϵ u equal to 5 and 15 (denoted by_sim5 and _sim15, respectively), and the baseline feed-forward method. Both methods are tested for source azimuth (denoted by _az) and elevation (denoted by _el) estimation. Localization errors are in degrees.

5.3.2. Comparison with the baseline

In this experiment, we compare the localization performance of the proposed SCE with the baseline LEM embedding and the feed-forward model. For the proposed SCE, we evaluate a small threshold angle (i.e., $\epsilon_u = 5\ ^\circ$) and a large threshold angle (i.e., $\epsilon_u = 15\ ^\circ$), denoted by "_sim5" and "_sim15," respectively.

The testing set results are illustrated in Figure 3. It can be seen that in the fixed acoustic condition with the CAMIL dataset, the proposed SCE performs better than the LEM embedding and the feed-forward model in terms of median error and maximum error. Especially when using the small similarity threshold, the SCE performs excellent, as the SCE_sim5 has almost zero median error in azimuth and elevation estimations. It can also be noted that the feed-forward model performs slightly better than the LEM embedding, with a median error equal to 0.61 and 0.29° for azimuth and elevation respectively, whereas the LEM model has median errors equal to 0.72 and 0.49° for azimuth and elevation, respectively. In summary, in the fixed acoustic condition, the proposed SCE can almost perfectly preserve the source location information even when reverberation and additive white noise are present, while the feed-forward model performs better than the LEM embedding, but both exhibit some estimation error. This could be due to the fact that the feed-forward model highly depends on the training data, and due to the presence of audio content mismatch between the training and testing sets, the feed-forward model has some difficulty to generalize to unseen audio contents, thus negatively influencing the localization performance.

In the varying acoustic conditions with the VAST testing sets, the proposed SCE_sim5 performs slightly better than the SCE_sim15 and equally well as the feed-forward model. The SCE_sim5 and feed-forward model achieve VAST testing-set-1 azimuth median errors equal to 1.96 and 1.95°, VAST testing-set-1 elevation median errors equal to 3.32 and 3.24°, VAST testing-set-2 azimuth median errors equal to 2.1 and 2.01°, and VAST testing-set-2 elevation median errors equal to 3.94 and 3.99°, respectively. Since in the various acoustic conditions, the source excitations are white noise in both the training and testing set, the SCE and the feed-forward model can both generalize well to unseen acoustic environments, and show robustness toward reverberation and noise.

The LEM embedding performs the worst in the presence of various reverberations. It achieves median errors equal to 3.3 and 11.7 for azimuth and elevation in VAST test-set-1, respectively, and 3.1 and 13.3 for azimuth and elevation in VAST test-set-2, respectively. This may indicate that the LEM, which is easily affected by geometric distortion in the measurements, is not robust to reverberation.

5.3.3. Reduced training-set

A common problem related to data-driven methods is the model generalizability, or in other words, how can a trained model generalize to unseen data. In the source localization scenario, the training set may not include training recordings from every pair of azimuth/elevation angles, hence it is desirable that the model can somehow interpolate the predictions that lie in-between the training points. In this experiment, we are aiming to evaluate the robustness of the proposed SCE toward the training size. With a smaller training size, there will be more source locations that are not included in the training. We use a similarity threshold angle $\epsilon_u = 5\ ^\circ$ for SCE in this experiment and all methods are conducted with 10, 25, 50, and 70% randomly selected training sets. The median localization errors are illustrated in Figure 4.

As illustrated by these results, all methods show a decreasing trend in localization error when a larger training set is used,

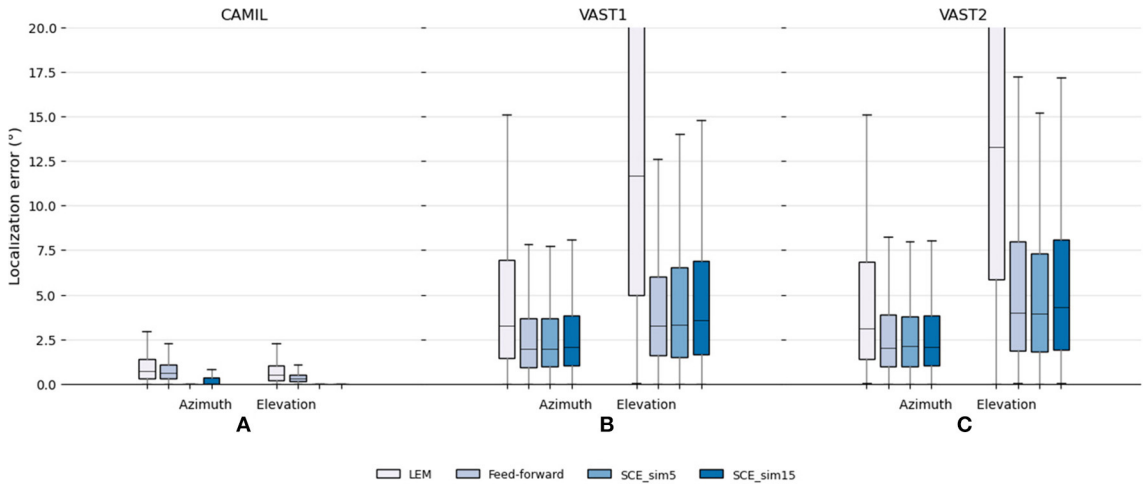


FIGURE 3
Testing set localization performance for the baseline LEM, the baseline feed-forward and the proposed SCE methods. Localization errors in **(A)** the fixed acoustic condition using the CAMIL testing set, and **(B,C)** the localization performance in the varying acoustic condition using the VAST testing sets. "_sim5" and "_sim15" denote the use of similarity threshold angles ϵ_u equal to 5 and 15° , respectively.

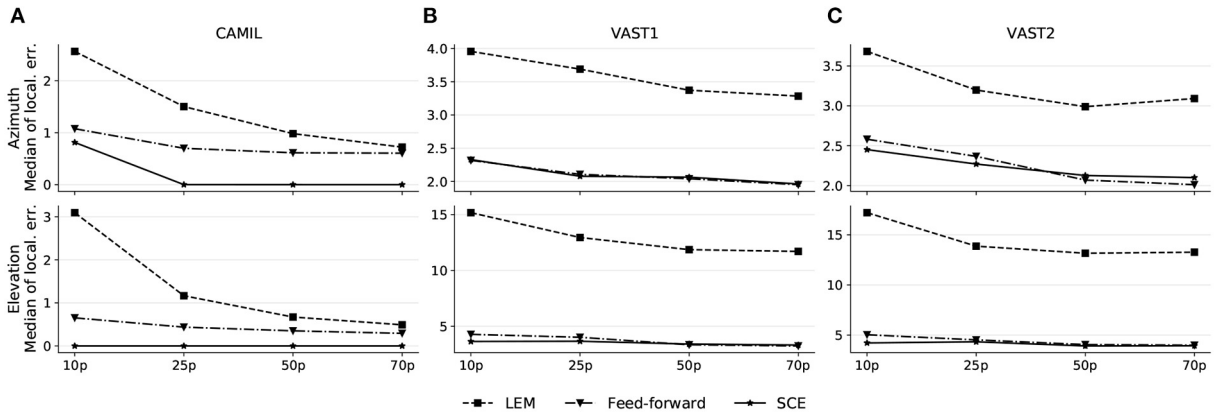


FIGURE 4
Localization performance on the CAMIL testing set **(A)**, the VAST testing-set-1 **(B)**, and the VAST testing-set-2 **(C)**. 10p, 25p, 50p, and 70p denote the cases when 10, 25, 50, and 70% of the original training data is used, respectively.

however, the median localization error of the proposed SCE does not vary much with the changing size of the training set, and shows a flatter pattern. Although in the fixed acoustic condition, SCE results a in a higher median error when 10% of the training set is used (median azimuth error equal to 0.81°) than when a larger training set is used, the error is still lower than for the other two methods (as feed-forward and LEM achieve median azimuth errors equal to 1.08° and 2.56°respectively when 10% of the training data is used). This allows to conclude that the SCE is more robust to the use of training data that not cover the entire latent space.

The results allow us to hypothesize that the proposed SCE, leveraged by the contrastive loss and the adaptive margin (see Section 4.3), is aiming to learn a similarity metric between input binaural cues from the latent space. This similarity metric implies that the underlining structure in the latent space is robust to unseen source locations. In contrast, the feed-forward model tends to transform the measurement space to an abstract high-level space in which the Euclidean distance between embeddings is not necessarily a similarity metric, and thus it is difficult to infer the unseen source locations from this embedding space.

5.4. WSCE for multidimensional source localization

The LEM embedding as well as the proposed WSCE are capable of estimating the sound source azimuth and elevation simultaneously. It should be noted that both the proposed WSCE and the LEM need source annotations in order to localize new examples under the nearest-neighbor localization framework, thus the localization phase is still a supervised learning task for both methods.

To explore the learned latent space structure, we test several similarity threshold angles $\epsilon_u \in \{5^\circ, 15^\circ, 30^\circ\}$, indicated as "_sim5," "_sim15," and "_sim30," respectively. Since when calculating the similarity labels, we first normalize the relative source location coordinates to have unit norm (i.e., source coordinates are relocated to have unit distance to the receiver), chosen the similarity threshold angles yield the following similarity threshold for the physical source distance: $\epsilon_s \in \{0.09, 0.26, 0.52\,\text{m}\}$. Figure 5 shows the training set embeddings and the testing set embeddings for the CAMIL testing set and the VAST testing-set-1. Firstly, it can be observed that the proposed WSCE method learns a manifold from the binaural cues that can reflect the sound source location without any azimuth/elevation annotations. This manifold has a clear structure and a similar structure is obtained in both the CAMIL dataset (with reverberant speech) and the VAST dataset (with varying reverberation). Secondly, when using smaller similarity threshold angles (i.e., $\epsilon_u = 5^\circ$), the structure of the manifold tends to become irregular and folded, and when using larger threshold angles (i.e., $\epsilon_u = 15^\circ$ and $\epsilon_u = 30^\circ$), the structure of the manifold tends to become smooth and unfolded. Elaborating the intuition introduced in Section 4.3, this may be due to the fact that when the similarity threshold angle is small, the contrastive loss has a small range of action on penalizing mislocated dissimilar pairs, resulting in many dissimilar pairs not being subject to repulsive forces, and instead, similar pairs are attracted and clustered in local areas. When a large similarity threshold angle is used, each embedding is subject to both attractive and repulsive forces from a large number of other embeddings, thus maintaining an overall uniformly equilibrium state in the global perspective.

In addition to the above mentioned qualitative experiments, we also conduct quantitative experiments to use the WSCE for source localization and compare the results to the LEM embeddings and the SCE_sim5. The localization results are shown in Figure 6. In the fixed acoustic condition with the CAMIL dataset, the SCE_sim5 still performs the best but it trains separate embeddings for azimuth and elevation. In contrast, both the proposed WSCE and the LEM embedding train one embedding for both azimuth and elevation estimation and show a strong source localization ability as well. In azimuth estimation, the WSCE_sim15 performs slightly better than the WSCE_sim5, then followed by LEM and WSCE_sim30 (achieving median errors equal to 0.64, 0.69, 0.72, and 1.16°, respectively). In elevation estimation, LEM exhibits a median error equal to 0.49° and performs slightly better than the WSCE_sim15 and WSCE_sim5, which have the same median error equal to 0.58°. WSCE_sim30 performs worst in elevation estimation and achieves a median error equal to 0.82°. Nevertheless, the WSCE shows a comparable localization ability to the LEM in the fixed acoustic condition.

In varying acoustic conditions with the VAST dataset, instead, the WSCE shows a much lower localization error than the LEM embeddings and it is even approaching the SCE_sim5 performance. Firstly, with the VAST testing-set-1, the WSCE_sim5, WSCE_sim15, and WSCE_sim30 perform equally well (azimuth median errors equal to 1.96, 1.94, and 1.98°, respectively, and elevation median errors equal to 3.69, 3.64, and 3.69°, respectively), and the SCE_sim5 has slightly better elevation estimation than either WSCE method (achieving azimuth and elevation median errors equal to 1.96 and 3.32°, respectively). For the VAST testing-set-2, similarly, the WSCE_sim5, WSCE_sim15, WSCE_sim30, and SCE_sim5 perform somewhat equally well (achieving azimuth median errors equal to 1.93, 2.1, 2.1, and 2.1°, respectively, and elevation median errors equal to 3.99, 4.34, 4.44, and 3.94°, respectively). Although the unidimensional SCE_sim5 and the WSCE with a small similarity threshold show narrower interquartile range than other methods, we do suggest to use a similarity threshold angle $\epsilon_u = 15^\circ$ for WSCE to achieve both good visualization and localization performance.

Secondly, the WSCE largely outperforms the LEM embeddings in varying acoustic conditions where LEM only obtains an azimuth median error of 3.28°and an elevation median error of 11.7° for VAST testing-set-1, and an azimuth median error of 3.09°and an elevation median error of 13.27° for VAST testing-set-2, respectively. Also, the WSCE has a much narrower interquartile range than the LEM, which may indicate that the proposed WSCE is more robust to reverberation than the LEM embeddings.

5.5. WSCE with unseen HRTFs

To further verify the generalization capability of the proposed WSCE, we test the WSCE with different HRTFs that are not seen during the training. To create simulated binaural recordings, we use the CIPIC dataset (Algazi et al., 2001), which consists of 45 real-life measured HRTFs. There are in total 45 subjects (43 human subjects and 2 dummy head subjects), and for each subject, 1250 HRTFs are measured for each ear and from different azimuth and elevation angles. We select azimuth and elevation angles in range the $[-90, 90^\circ]$ and $[-45, 45^\circ]$, respectively, corresponding to the other datasets mentioned in the former sections. The HRTFs are then convoluted with

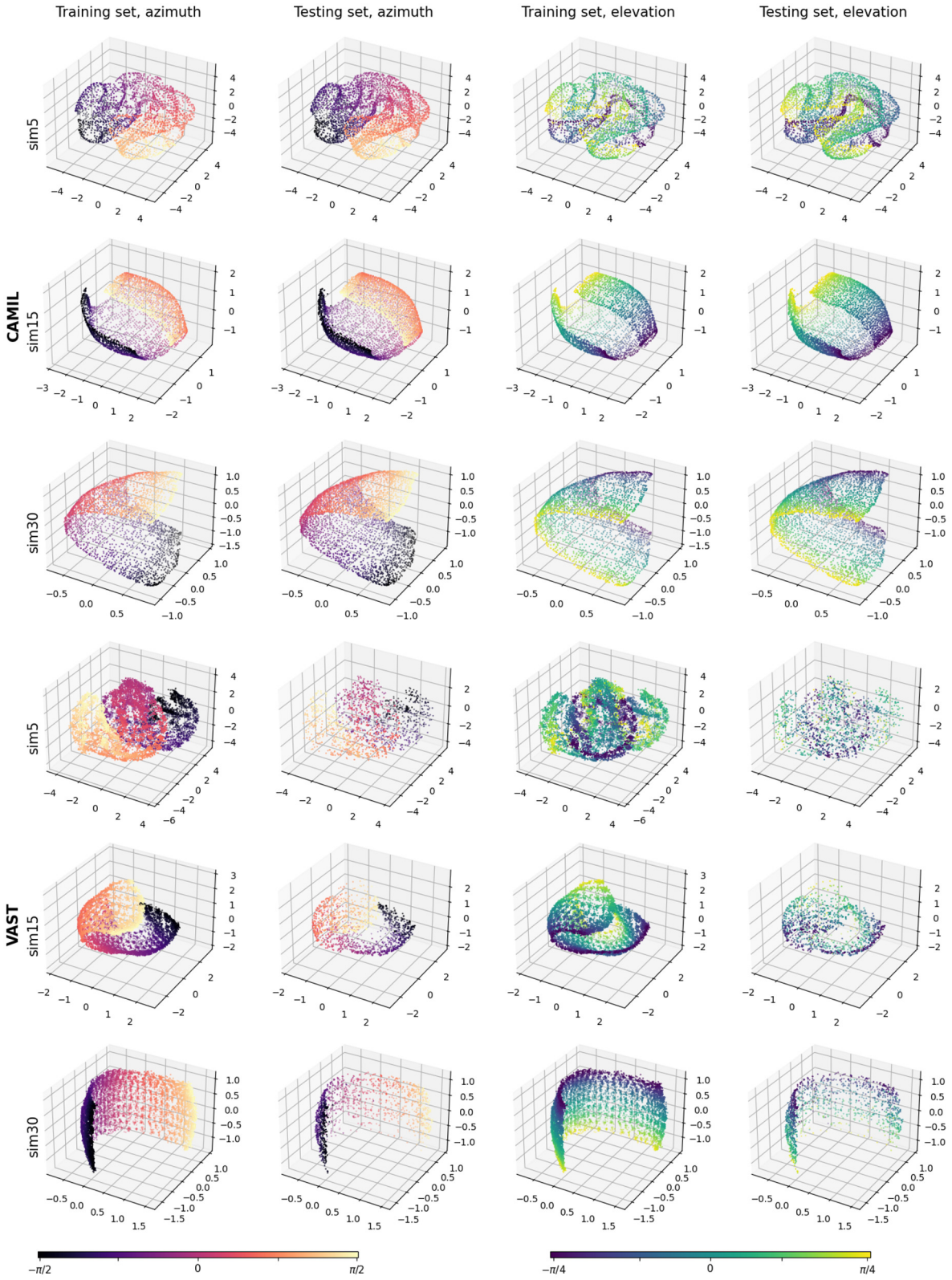


FIGURE 5
Visualizations of the WSCE embeddings. Column 1 and 2 are azimuth training and testing embeddings. Column 3 and 4 are elevation training and testing embeddings.

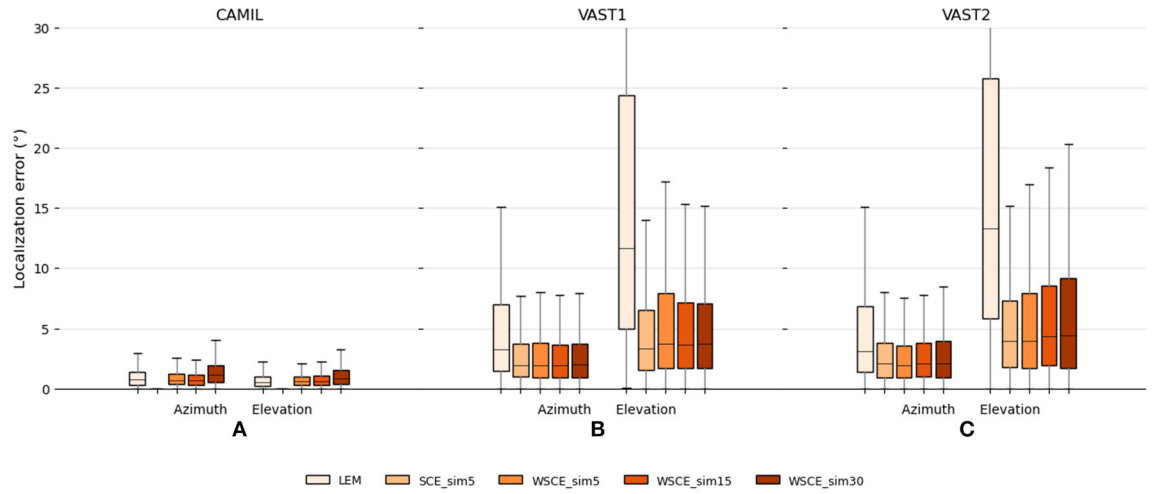


FIGURE 6
Localization performance on the CAMIL testing set **(A)**, the VAST testing-set-1 **(B)**, and the VAST testing-set-2 **(C)** for the LEM, the SCE and the WSCE methods.

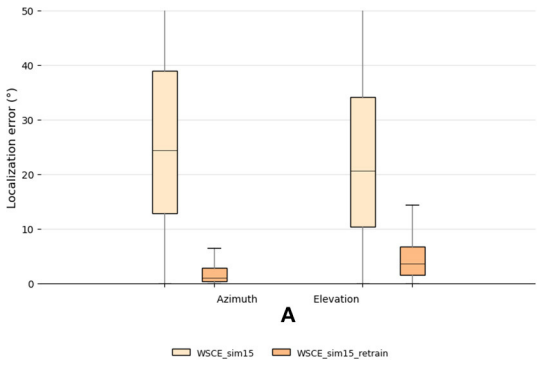


FIGURE 7
Testing set localization performance for the WSCE_sim15 and WSCE_sim15 retrained with 33 different HRTFs other than the once used in generating test recordings.

simulated reverberant recordings (excited by 2 s white noise). Those recordings are generated using the image method (Allen and Berkley, 1979), in a shoebox room that has dimension 3.5 × 5 × 2.8 m, and reverberation time equal to 0.3 s.

We randomly select recordings from 10 subjects for testing, and use WSCE_sim15 for estimating their source locations. The localization results are plotted in Figure 7. Since we train the WSCE_sim15 only using one HRTF, the model could not generalize well to recordings made with unseen HRTFs. Therefore, we observe a dramatic performance degradation, in which the median errors of azimuth and elevation localization are 24.3 and 20.1°, respectively.

To overcome the performance degradation, we propose two approaches:

1. Personalized training (user-dependent): this approach is especially interesting for hearing-aid applications since the hearing-aid is designed for a specific user, and it is not shared with different people. Therefore, the HRTF of the designated user can be measured and be used in the model training or fine-tuning process to create a user-dependent model.
2. Increase training data variety (user-independent): another solution consists in using more HRTFs to create the training data for training the WSCE. Then, the trained model can generalize to people with different HRTF than the ones in training data. A rule of thumb is that the higher the variety of the training data (with annotation), the better the generalization capability of the model.

We adopt the second approach to retrain the WSCE_sim15 and use the rest of the HRTFs from the CIPIC dataset, which are different from the data used in the testing (i.e., user-independent). This results in 33 HRTFs that are used for training, 2 for validation and 10 for testing. We also simulate random shoebox rooms that have reverberation time between 0.1 and 0.4 s. The localization error of the retrained model is shown in Figure 7 with name "WSCE_sim15_retrain." The azimuth and elevation median errors of the retrained model have been largely reduced from 24.3 to 1.1°and 20.1 to 3.6°, respectively, showing the effectiveness of this approach.

We further analyse the relationship between the CIPIC testing set embeddings and their respective nearest training set neighbors and illustrate the results in Figure 8. The X-axis is the true azimuth or elevation angle of the testing embedding, and the Y-axis is the location of the corresponding nearest training set neighbor predicted by the WSCE_sim15 which is

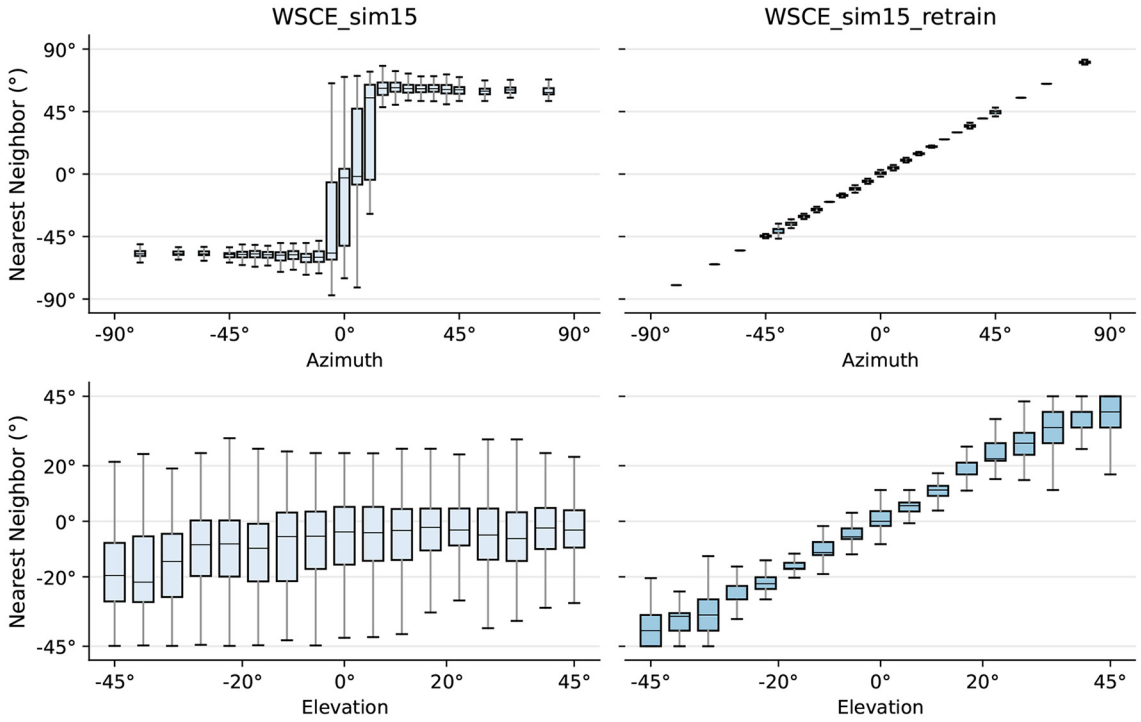


FIGURE 8
True locations of sources in the testing set versus the locations of their respective nearest neighbor in the training set for WSCE_sim15 and WSCE_sim15 retrained with 33 different HRTFs other than the once used in generating test recordings (i.e., WSCE_sim15_retrain).

summarized using box plots. For the original WSCE_sim15, the median nearest neighbor location angles are shifted compared to the true testing location in the case of both azimuth and elevation angles, and the interquartile ranges of the nearest neighbor location angles are large, indicating that the neighbors of the original WSCE_sim15 are poorly preserved, which also suggests that the original WSCE_sim15 trained with only one HRTF cannot be generalized to the unseen HRTFs. In contrast, WSCE_sim15_retrain preserves the neighborhoods much better because the median of the location angles of the nearest neighbors predicted by the WSCE_sim15_retrain is close to the true location angle of the test embedding. In addition, compared to the original WSCE_sim15 model, the location angle of the nearest neighbor predicted by the WSCE_sim15_retrain has a smaller interquartile range. In summary, the generalization to unseen HRTFs is much better after retraining WSCE_sim15 with 33 real-life HRTFs.

However, a limitation of our simulations is that we use synthetic rooms with slightly different acoustic properties than real-life rooms. In addition, we always excite the sound source with white noise, which has a broadband spectrum, while real-life sounds may not have the same characteristics. We propose to increase the variety of training data covering real-life conditions, using more HRTFs recorded at finer azimuth/elevation angles, and using Room Impulse Responses (RIRs) from more complex rooms, which we believe will further improve the generalization capability of the proposed WSCE model.

6. Conclusions

We proposed a DNN framework for supervised dimensionality reduction of binaural cue measurements, followed by a nearest-neighbor regression method for source localization. Our manifold-learning-based method has better binaural sound source localization performance than the baseline manifold learning method in both know and unknown reverberant conditions and in a small training set condition. In comparison with a feed-forward learning method, our proposed method not only provides a better visualization ability, but also achieves a similar or better performance in binaural sound source localization. Moreover, our proposed method can capture a smooth manifold structure for low data density regions and outperforms the baseline manifold learning method and the feed-forward method in case of a small amount of training data.

In addition to the supervised dimensionality reduction method, we also proposed a weakly supervised embedding, i.e., WSCE, that only requires implicit latent space proximity labels for training. This WSCE can simultaneously estimate the

azimuth and elevation of the sound source, and is also robust to unknown reverberation. Quantitative experimental results demonstrate that this WSCE has almost similar localization performance as the supervised method, and it performs much better than the traditional unsupervised embedding in varying acoustic conditions.

To further increase the generalization capability of the proposed model, we hope to learn the SCE and WSCE embeddings with big variety of training data covering more real-life conditions, such as using more HRTFs recorded at finer azimuth/elevation angles and using RIRs from more complex rooms. In addition, we also aim to further investigate how to apply the proposed SCE and WSCE in data synthesis. When combining these methods with a generative model, we speculate that the embeddings can be used to synthesize binaural features or even audio waveforms to aid data-driven binaural source localization models.

Since potentially applicable systems for the proposed model (e.g., hearing aids) often have limited computational resources, reducing the model complexity and the number of model parameters is therefore a relevant direction for future research. Possible approaches to achieve this include model pruning (i.e., removing the DNN neurons that are associated with very small weights), model information distillation (Hinton et al., 2015) and model parameter quantization.

Data availability statement

The datasets generated and/or analysed during the current study are available in the CAMIL repository (https://team.inria.fr/perception/the-camil-dataset), and the VAST repository (http://thevastproject.inria.fr/dataset).

Author contributions

DT and MT invented the concept of estimating binaural sound source location in reverberant acoustic conditions using the *siamese* neural network with a contrastive loss, designed and interpreted the computer simulations. DT, MT, and TW jointly developed the research methodology to turn this concept into a usable and effective algorithm. DT implemented the computer simulations and MT managed to the dataset used. All authors contributed in writing the manuscript, and further read and approved the final manuscript.

Funding

DT was sponsored by the Chinese Scholarship Council (CSC) (no. 201707650021). MT was a Postdoctoral Fellow of the Research Foundation Flanders—FWO-Vlaanderen (no. 12X6719N). This research work was carried out at the ESAT Laboratory of KU Leuven. The research leading to these results has received funding from the KU Leuven Internal Funds C2-16-00449 and VES/19/004, and the European Research Council under the European Union's Horizon 2020 research and innovation program/ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Algazi, V., Duda, R., Thompson, D., and Avendano, C. (2001). "The CIPIC HRTF database," in *Proceedings of IEEE Applications of Signal Processing to Audio Acoustics (WASPAA 2001)* (New Platz, NY), 99–102. doi: 10.1109/ASPAA.2001.969552

Allen, J. B., and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics prediction of energy decay in room impulse responses simulated with an image-source model image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* 65, 943–950. doi: 10.1121/1.382599

Argentieri, S., Danès, P., and Souères, P. (2015). A survey on sound source localization in robotics: from binaural to array processing methods. *Comput. Speech Lang.* 34, 87–112. doi: 10.1016/j.csl.2015.03.003

Belkin, M., and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 6, 1373–1396. doi: 10.1162/0899766033217 80317

Bengio, Y., Paiement, J.-F., Vincent, P., Delalleau, O., Le Roux, N., and Ouimet, M. (2003). "Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps and spectral clustering," in *Proceedings of IEEE Conference on Advances in Neural Information Processing Systems (NeurIPS 2003).* (Barcelona), 177–184.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press. doi: 10.7551/mitpress/6391.001.0001

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1993). "Signature verification using a "siamese" time delay neural network," in *Advances in Neural Information Processing Systems, Vol. 6* (Denver, CO), 737–744. doi: 10.1142/S0218001493000339

Chopra, S., Hadsell, R., and LeCun, Y. (2005). "Learning a similarity metric discriminatively, with application to face verification," in *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)* (San Diego, CA), *Vol. 1*, 539–546. doi: 10.1109/CVPR.2005.202

Chung, F. R. K. (1997). *Spectral Graph Theory*. Providence, RI: American Mathematical Society.

Datum, M. S., Palmieri, F., and Moiseff, A. (1996). An artificial neural network for sound localization using binaural cues. *J. Acoust. Soc. Am.* 100, 372–383. doi: 10.1121/1.415854

Deleforge, A., Forbes, F., and Horaud, R. (2015). Acoustic space learning for sound source separation and localization on binaural manifolds. *Int. J. Neural Syst.* 25:1440003. doi: 10.1142/S0129065714400036

Deleforge, A., and Horaud, R. (2012). "2D sound-source localization on the binaural manifold," in *2012 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2012)* (Santander). doi: 10.1109/MLSP.2012.6349784

Farmani, M., Pedersen, M. S., and Jensen, J. (2018). "Sound source localization for hearing aid applications using wireless microphones," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2018)* (Sheffield), 455–459. doi: 10.1109/SAM.2018.8448967

Gardner, W. G., and Martin, K. D. (1995). HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.* 97, 3907–3908. doi: 10.1121/1.412407

Garofolo, J. S.,Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, J. S., Dahlgren, N.L, et al. (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1*. Web Download. Philadelphia, PA: Linguistic Data Consortium. doi: 10.35111/17gk-bn40

Gaultier, C., Kataria, S., and Deleforge, A. (2017). "VAST: the virtual acoustic space traveler dataset," in *Proceedings of International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)* (Grenoble), 68–79, doi: 10.1007/978-3-319-53547-0_7

Geetha, C., Tanniru, K., and Rajan, R. R. (2017). Efficacy of directional microphones in hearing aids equipped with wireless synchronization technology. *J. Int. Adv. Otol.* 13:113. doi: 10.5152/iao.2017.2820

Hadsell, R., Chopra, S., and LeCun, Y. (2006). "Dimensionality reduction by learning an invariant mapping," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006)* (New York, NY), *Vol. 2*, 1735–1742. doi: 10.1109/CVPR.2006.100

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*. doi: 10.48550/arXiv.1503.02531

Ibrahim, I., Parsa, V., Macpherson, E., and Cheesman, M. (2013). Evaluation of speech intelligibility and sound localization abilities with hearing aids using binaural wireless technology. *Audiol. Res.* 3:e1. doi: 10.4081/audiores.2013.e1

Ioffe, S., and Szegedy, C. (2015). "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning, Vol. 37* (Lille), 448–456.

Karthik, G. R., and Ghosh, P. K. (2018). "Binaural speech source localization using template matching of interaural time difference patterns," in *2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '18)* (Calgary), 5164–5168. doi: 10.1109/ICASSP.2018.8462586

Keyrouz, F., and Diepold, K. (2007). Binaural source localization and spatial audio reproduction for telepresence applications. *Presence Teleoper. Virt. Environ.* 16, 509–522. doi: 10.1162/pres.16.5.509

Kingma, D. P., and Ba, J. (2015). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. doi: 10.48550/arXiv.1412.6980

Kreisman, B. M., Mazevski, A. G., Schum, D. J., and Sockalingam, R. (2010). Improvements in speech understanding with wireless binaural broadband digital hearing instruments in adults with sensorineural hearing loss. *Trends Amplif.* 14, 3–11. doi: 10.1177/1084713810364396

Laufer, B., Talmon, R., and Gannot, S. (2013). "Relative transfer function modeling for supervised source localization," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2013)* (New Platz, NY), 1–4. doi: 10.1109/WASPAA.2013.6701829

Laufer-Goldshtein, B., Talmon, R., and Gannot, S. (2015). "A study on manifolds of acoustic responses," in *Proceedings of the International Conference on Latent Variable Analysis and Signal Separation* (Liberec: LVA/ICA 2015), 203–210. doi: 10.1007/978-3-319-22482-4_23

Li, X., Girin, L., Horaud, R., and Gannot, S. (2016). Estimation of the direct-path relative transfer function for supervised sound-source localization. *IEEE/ACM Trans. Audio Speech Lang. Process.* 24, 2171–2186. doi: 10.1109/TASLP.2016.2598319

Ma, N., Gonzalez, J. A., and Brown, G. J. (2018). Robust binaural localization of a target sound source by combining spectral source models and deep neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* 26, 2122–2131. doi: 10.1109/TASLP.2018.2855960

Ma, N., May, T., and Brown, G. J. (2017). Exploiting deep neural networks and head movements for robust binaural localization of multiple sources in reverberant environments. *IEEE/ACM Trans. Audio Speech Lang. Process.* 25, 2444–2453. doi: 10.1109/TASLP.2017.2750760

Mandel, M., Weiss, R., and Ellis, D. (2010). Model-based expectation-maximization source separation and localization. *IEEE Trans. Audio Speech Lang. Process.* 18, 382–394. doi: 10.1109/TASL.2009.2029711

May, T., Van De Par, S., and Kohlrausch, A. (2011). A probabilistic model for robust localization based on a binaural auditory front-end. *IEEE Trans. Audio Speech Lang. Process.* 19, 1–13. doi: 10.1109/TASL.2010.2042128

Opochinsky, R., Laufer-Goldshtein, B., Gannot, S., and Chechik, G. (2019). "Deep ranking-based sound source localization," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2019)* (New Paltz, NY), 283–287. doi: 10.1109/WASPAA.2019.8937159

Pak, J., and Shin, J. W. (2019). Sound localization based on phase difference enhancement using deep neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* 27, 1335–1345. doi: 10.1109/TASLP.2019.2919378

Raspaud, M., Viste, H., and Evangelista, G. (2010). Binaural source localization by joint estimation of ILD and ITD. *IEEE Trans. Audio Speech Lang. Process.* 18, 68–77. doi: 10.1109/TASL.2009.2023644

Risoud, M., Hanson, J. N., Gauvrit, F., Renard, C., Lemesre, P. E., Bonne, N. X., et al. (2018). Sound source localization. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 135, 259–264. doi: 10.1016/j.anorl.2018.04.009

Schimmel, S. M., Muller, M. F., and Dillier, N. (2009). "A fast and accurate shoebox room acoustics simulator," in *Procedings of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '09)* (Taipei), 241–244.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). "DeepFace: closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)* (Columbus, OH), 1701–1708. doi: 10.1109/CVPR.2014.220

Tang, D., Taseska, M., and Van Waterschoot, T. (2019). "Supervised contrastive embeddings for binaural source localization," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2019)* (New Paltz, NY), 358–362. doi: 10.1109/WASPAA.2019.8937177

Taseska, M. and van Waterschoot, T. (2019). "On spectral embeddings for supervised binaural source localization," in *Proceedings of the 27th European Signal Processing Conference (EUSIPCO '27)* (Coru na), 1–5. doi: 10.23919/EUSIPCO.2019.8902761

Vecchiotti, P., Ma, N., Squartini, S., and Brown, G. J. (2019). "End-to-end binaural sound localisation from the raw waveform," in *Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '19)* (Brighton), 451–455. doi: 10.1109/ICASSP.2019.8683732

Wei, Q., Park, Y. S., Lee, S. B., Kim, D. W., Seong, K. W., Lee, J. H., et al. (2014). Novel design for non-latency wireless binaural hearing aids. *IEEE Trans. Electr. Electron. Eng.* 9, 566–568. doi: 10.1002/tee.22007

Woodruff, J., and Wang, D. L. (2012). Binaural localization of multiple sources in reverberant and noisy environments. *IEEE Trans. Audio Speech Lang. Process.* 20, 1503–1512. doi: 10.1109/TASL.2012.2183869

Yalta, N., Nakadai, K., and Ogata, T. (2017). Sound source localization using deep learning models. *J. Robot. Mechatron.* 29, 37–48. doi: 10.20965/jrm.2017.p0037

Yang, B., Li, X., and Liu, H. (2021). "Supervised direct-path relative transfer function learning for binaural sound source localization," in *Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '21)* (Toronto, ON), 825–829. doi: 10.1109/ICASSP39728.2021.9413923

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 09 March 2023
DOI 10.3389/fninf.2023.934472

OPEN ACCESS

EDITED BY
Alessia Paglialonga,
Institute of Electronics, Information Engineering and Telecommunications (IEIIT), Italy

REVIEWED BY
Tobias Weissgerber,
University Hospital Frankfurt, Germany
Etienne Gaudrain,
INSERM U1028 Centre de Recherche en Neurosciences de Lyon, France

*CORRESPONDENCE
Waldo Nogueira
✉ NogueiraVazquez.Waldo@mh-hannover.de

RECEIVED 02 May 2022
ACCEPTED 15 February 2023
PUBLISHED 09 March 2023

CITATION
Alvarez F, Kipping D and Nogueira W (2023) A computational model to simulate spectral modulation and speech perception experiments of cochlear implant users. *Front. Neuroinform.* 17:934472. doi: 10.3389/fninf.2023.934472

A computational model to simulate spectral modulation and speech perception experiments of cochlear implant users

Franklin Alvarez[1,2], Daniel Kipping[1,2] and Waldo Nogueira[1,2]*

[1]Medizinische Hochschule Hannover, Hannover, Germany, [2]Cluster of Excellence "Hearing4All", Hannover, Germany

Speech understanding in cochlear implant (CI) users presents large intersubject variability that may be related to different aspects of the peripheral auditory system, such as the electrode–nerve interface and neural health conditions. This variability makes it more challenging to proof differences in performance between different CI sound coding strategies in regular clinical studies, nevertheless, computational models can be helpful to assess the speech performance of CI users in an environment where all these physiological aspects can be controlled. In this study, differences in performance between three variants of the HiRes Fidelity 120 (F120) sound coding strategy are studied with a computational model. The computational model consists of (i) a processing stage with the sound coding strategy, (ii) a three-dimensional electrode-nerve interface that accounts for auditory nerve fiber (ANF) degeneration, (iii) a population of phenomenological ANF models, and (iv) a feature extractor algorithm to obtain the internal representation (IR) of the neural activity. As the back-end, the simulation framework for auditory discrimination experiments (FADE) was chosen. Two experiments relevant to speech understanding were performed: one related to spectral modulation threshold (SMT), and the other one related to speech reception threshold (SRT). These experiments included three different neural health conditions (healthy ANFs, and moderate and severe ANF degeneration). The F120 was configured to use sequential stimulation (F120-S), and simultaneous stimulation with two (F120-P) and three (F120-T) simultaneously active channels. Simultaneous stimulation causes electric interaction that smears the spectrotemporal information transmitted to the ANFs, and it has been hypothesized to lead to even worse information transmission in poor neural health conditions. In general, worse neural health conditions led to worse predicted performance; nevertheless, the detriment was small compared to clinical data. Results in SRT experiments indicated that performance with simultaneous stimulation, especially F120-T, were more affected by neural degeneration than with sequential stimulation. Results in SMT experiments showed no significant difference in performance. Although the proposed model in its current state is able to perform SMT and SRT experiments, it is not reliable to predict real CI users' performance yet. Nevertheless, improvements related to the ANF model, feature extraction, and predictor algorithm are discussed.

KEYWORDS

computational model, cochlear implant, neural health, sound coding strategies, speech-in-noise recognition, spectral modulation detection, speech understanding prediction

1. Introduction

People diagnosed with severe or profound sensorineural hearing loss that keep some healthy auditory nerve fibers (ANFs) are good candidates to receive a cochlear implant (CI) and recover to some extent their sense of hearing. A CI consists of an electrode array implanted in the cochlea, and a wearable sound processor usually located behind the ear. The sound processor is responsible for converting acoustic signals into electric stimulation patterns that are delivered to the ANFs *via* the intracochlear electrodes (Wouters et al., 2015). In many auditory tasks, there is a big gap in performance between normal hearing (NH) and CI listeners (Nelson et al., 2003; Nelson and Jin, 2004). Electric stimulation has its limitations to convey the necessary information for the proper coding of sounds in the auditory system (Moore, 2003). To reduce this gap, researchers are dedicated to find better CI sound coding strategies (Nogueira et al., 2005, 2009; Landsberger and Srinivasan, 2009; Dillon et al., 2016; Langner et al., 2020a; Gajecki and Nogueira, 2021), but the evaluation of the potential benefits of new ideas usually requires extensive testing procedures with implanted volunteers. In addition, there is high variability in the performance among CI users (Moberly et al., 2016), which makes it more difficult to generalize from the results.

CI sound coding strategies using current steering aim at providing an increased number of stimulation places in the implanted cochlea (Landsberger and Srinivasan, 2009; Nogueira et al., 2009). The general idea is to create virtual channels by "steering" the electrical field between two adjacent electrodes, balancing their output current at different ratios. The commercial sound coding strategy HiRes with Fidelity120 (F120), from Advanced Bionics, offers up to 120 virtual channels using 16 electrodes because every electrode pair is able to steer the electrical field to eight different locations. Furthermore, power savings can be achieved by stimulating various virtual channels simultaneously. Simultaneous stimulation allows to increase the pulse duration and consequently decrease the maximum current needed (Langner et al., 2017). The drawback is that simultaneous stimulation produces electric interaction that causes spectral smearing across channels, which also causes temporal smearing since temporal modulations may be reduced (Nogueira et al., 2021). The balance between power savings and CI users' performance was investigated by Langner et al. (2017) using three variations of the F120. Sequential stimulation (F120-S), where one virtual channel was active at a time, was compared to paired (F120-P) and triplet (F120-T) stimulation, where two and three virtual channels were active at the same time, respectively. They found out that the channel interaction that occurs with the simultaneous stimulation in F120-P and F120-T has a negative impact on performance, with F120-T obtaining the worst score. Nevertheless, high inter-subject variability was found in speech intelligibility and spectral modulation detection threshold.

It has been shown that peripheral aspects such as neural health condition (Nadol, 1997), insertion depth and position of the electrode array (Dorman et al., 1997), along with more central aspects such as neural plasticity (Han et al., 2019) may account for an important part of the inter-subject variability observed in CI users. However, it is not possible to estimate the degree of ANF degeneration without invasive methods unless the individual is already implanted with a CI (Prado-Gutierrez et al., 2006; Ramekers et al., 2014; Imsiecke et al., 2021; Langner et al., 2021). Langner et al. (2021) investigated the hypothesis that individuals with good neural health and electrode positioning will show a lower difference in performance when using simultaneous stimulation strategies (F120-P and F120-T) compared to sequential stimulation (F120-S). Healthy conditions lead to lower focused thresholds and less channel interaction between virtual channels; therefore, healthy neural conditions could lead to less detriment in performance when comparing sequential and simultaneous stimulation. The performance was evaluated using the Hochmair–Schulz–Moser (HSM) sentence test (Hochmair-Desoyer et al., 1997) at a signal-to-noise ratio (SNR) where participants roughly understood 50% of the words with F120-S. The results showed no correlation between any measure intended to estimate the neural health and difference in performance, arguably, because of the small number of individuals measured. On the contrary, computational models that simulate the electrode–nerve interface in CI users can assess the relation between neural health and performance. These computational models can isolate the parameter of study to remove the inter-subject variability, i.e., nerve count, nerve degeneration, electrode position, or insertion depth.

Computational models of the electrode–nerve interface for CIs have been proposed at different levels of complexity. Fredelake and Hohmann (2012) presented a one-dimensional interface model where the ANFs are equally distributed along a cochlear axis with the electrode array positioned in the center of this ANF population with equidistant electrodes. The spatial spread of stimulation was calculated depending on the distance between electrodes and ANFs with an exponential decay function. Neural health conditions with this model were assessed by changing the ANF density while maintaining the total neural activity constant. Lower ANF density requires higher current levels; therefore, the excitation from a single electrode reaches further ANFs in the cochlear axis causing channel interaction and spectral smearing. However, this electrode–nerve interface is very limited when representing physical aspects that occur in real implantation. From clinical imaging data, a patient-specific three-dimensional model of the implanted cochlea can be constructed (Rattay et al., 2001; Stadler and Leijon, 2009; Kalkman et al., 2014; Malherbe et al., 2016; Nogueira et al., 2016; Heshmat et al., 2020, 2021; Croner et al., 2022). These models fit a population of ANFs (type 1 spiral ganglion neurons) that extend from the organ of Corti to the central axons. Also, it is possible to control the positioning of the electrode array inside the scala tympani. The voltage spread produced by the electric stimulation from the electrode array can be calculated using a homogeneous model of the extracellular medium (Rattay et al., 2001; Litvak et al., 2007a; Nogueira et al., 2016), or using a finite element method (FEM) to account for the different electrical properties of all structures between the stimulating electrode and the ANF population (Nogueira et al., 2016). Such an electrode–nerve interface can be coupled with an ANF model capable of simulating action potentials (also called spikes) from the electrical stimulation (Ashida and Nogueira, 2018).

Regarding the ANF models, there are two different approaches. The "physiological" approach aims to simulate processes on

a microscopic level. An example is the multi-compartment Hodgkin–Huxley model (Rattay, 1999; Rattay et al., 2001; Smit et al., 2008) that offers a very precise electrical behavior of the ANF segments when transmitting the action potentials throughout the peripheral axon, the soma, and the central axon. The drawback of this approach is the high demand for computational resources. The "phenomenological" approach tries to reproduce the effective outcome without detailed simulation of the involved intermediate processes. The spike generation algorithm does not consider how the spike travels through the nervous system, hence there is no geometric information involved. Some models depend completely on a probabilistic function (Bruce et al., 1999), while others are based on a leaky integrate-and-fire electrical circuit where the membrane voltage is calculated for every time step, and when it reaches a threshold, the ANF produces an action potential (Hamacher, 2004; Joshi et al., 2017). The membrane voltage depends on many other parameters like feedback currents, refractory periods, and membrane noise to introduce stochasticity. These parameters can be adjusted to fit data measured in humans or animal models to account for physiological aspects. The output of an ANF model is the "spike train", which consists of a binary array indicating the time frames where an action potential (spike) is produced. With a population of ANFs, it is possible to integrate the spike trains, in time and cochlear place, to obtain features that are representations of sound at higher levels in the auditory system. Integration allows to reduce the amount of data while preserving the information that reaches, for example, a speech recognition algorithm (Fredelake and Hohmann, 2012; Jürgens et al., 2018).

The simulation framework for auditory discrimination experiments (FADE) (Schädler et al., 2016) is a computational tool capable of performing speech recognition tasks and psychoacoustic experiments simulating human performance. Originally, FADE was used to simulate the performance of NH and hearing aided people (Kollmeier et al., 2016; Schädler et al., 2018). Then, a CI sound coding strategy and a CI auditory model were incorporated into FADE to perform simulations of speech reception thresholds (SRTs) using data from different CI users (Jürgens et al., 2018). The SRT is defined as the signal-to-noise ratio (SNR) where 50% of the words in a sentence are correctly identified (Wagener et al., 1999) and it is a direct indicator of the CI user performance in speech understanding. However, Jürgens et al. (2018) used the same peripheral auditory model as Fredelake and Hohmann (2012), which is a simplified one-dimensional representation of the electrode–nerve interface. The incorporation of a more complex peripheral auditory model with a three-dimensional representation of the electrode–nerve interface should turn FADE into a powerful framework to assess studies related to neural health conditions in CI users. It can also be useful to assess the benefits of novel sound coding strategies. Objective instrumental measures commonly used for this purpose rely on vocoders to simulate the degraded sound delivered by the CI (Chen and Loizou, 2011; Santos et al., 2013; El Boghdady et al., 2016), not accounting for physiological aspects of the implantation.

Performance with a CI may be also predicted using simpler behavioral measurements than the SRT. The spectral modulation threshold (SMT) is defined as the ripple depth in dBs at which 79.4% of spectral rippled noise is differentiated from flat noise. SMT has been used alongside speech recognition experiments because it is a good indicator of how well the spectral cues in speech signals were perceived (Litvak et al., 2007b; Langner et al., 2017). It was used by Langner et al. (2017) as an indicator of how these spectral cues are affected by the channel interaction occurring in simultaneous stimulation (F120-P and F120-T), compared to sequential stimulation (F120-S). Their results showed similar performance between F120-S and F120-P but a clear lowering of performance with F120-T.

In this study, a computational model that simulates the performance of real CI users in SRT and SMT experiments is presented. The goal is to show the effects of parallel stimulation and neural degeneration in CI outcome performance. This model was tested with the three sound coding strategies used by Langner et al. (2021) (F120-S, F120-P and F120-T), and the hypothesis that channel interaction affects individuals with poorer neural health conditions to a larger extent is assessed. In the next section, the different parts composing this computational model and how the SRT and SMT experiments were implemented with FADE are described. A further section presents the results obtained, and the last section contains the discussion and conclusions of this study.

2. Materials and methods

2.1. The computational model

The proposed computational model consists of (i) a "front-end" containing the CI sound coding strategy, the peripheral auditory model with a three-dimensional electrode–nerve interface, and a feature extraction algorithm; (ii) a "back-end" with a hidden Markov model (HMM) already incorporated in the framework FADE.

2.1.1. Front-end

2.1.1.1. Cochlear implant sound coding strategy

The software BEPS+ from Advanced Bionics was used to create the pulse tables for the F120-S, F120-P, and F120-T sound coding strategies. The pulse tables are defined as the sequence of electrical pulses to create one cycle of stimulation. Figure 1 shows partial pulse tables corresponding to these strategies. A pulse consists of a cathodic-leading biphasic pulse. The electrodes are enumerated from the most apical to the most basal and each virtual channel composed of two simultaneously stimulated electrodes are depicted with a color code. The pulse phase duration was set to 18 μs and the pulse rate across virtual channels was kept constant at 1,852 pps by adding a gap between subsequent pairs, or triplets, of stimulating virtual channels for F120-P and F120-T, respectively.

The HiRes implantable cochlear stimulator (ICS) from Advanced Bionics was used to transform audio signals into electrodograms. The audio signal was calibrated to −49 dB full scale [dB_{FS}], corresponding with an audio signal at 65 dB sound pressure level [dB_{SPL}] captured by the microphone of the CI device. At this value, the signal level was close to the knee point of the adaptive gain control of the CI sound processor.

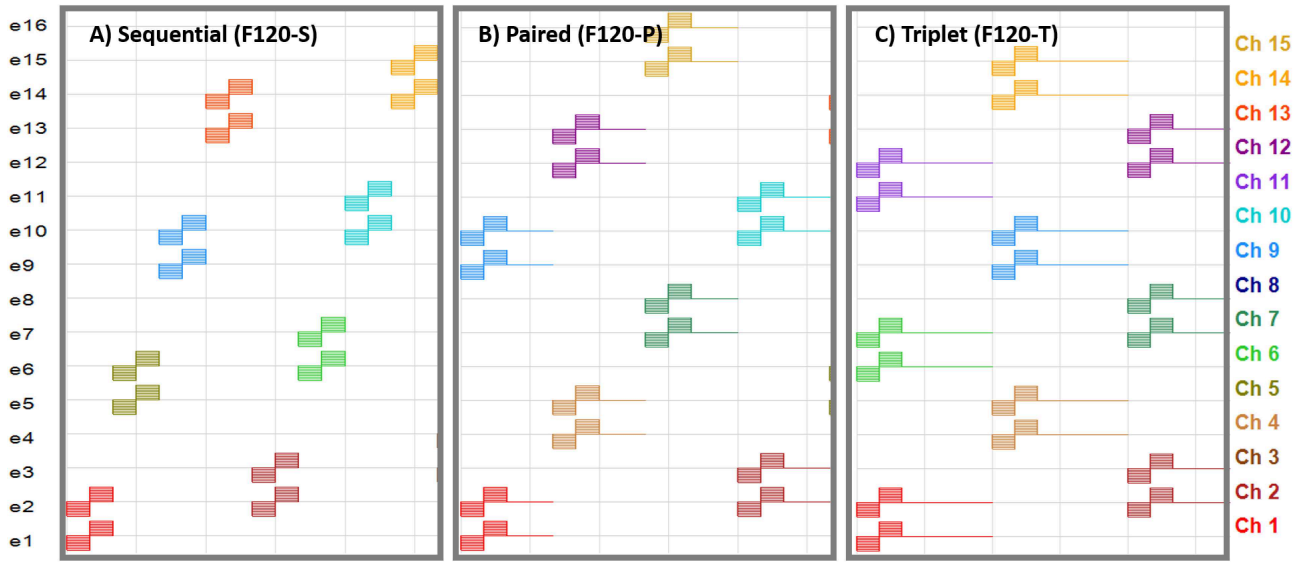


FIGURE 1
Pulse tables for the sequential, paired, and triplet stimulation versions of the HiRes Fidelity120 sound coding strategy. Horizontal axis is the temporal axis. Electrodes are enumerated from e1 (most apical) to e16 (most basal). The virtual channels composed of two simultaneously stimulated electrodes are enumerated from channel 1 (Ch 1; stimulation with e1 and e2) to channel 15 (Ch 15; stimulation with e15 and e16). **(A)** Sequential stimulation, where the virtual channels are activated one after another. **(B)** Paired stimulation, where two virtual channels are activated at the same time followed by a zero-phase gap with an equivalent duration of one biphasic pulse. **(C)** Triplet stimulation, where three virtual channels are activated at the same time followed by a zero-phase gap with an equivalent duration of two biphasic pulses.

Biphasic pulses of the electrodograms were resampled to 1 MHz to guarantee equal anodic and cathodic phases. This sample rate was also needed in the implementation of the peripheral auditory model.

2.1.1.2. Peripheral auditory model

Figure 2 shows the composition of the proposed peripheral auditory model. The electrodograms obtained from the sound coding strategy were transformed to obtain the voltage spread based on a three-dimensional electrode–nerve interface model embedded in a homogeneous medium. The amount of stimulation at every ANF was obtained from this voltage spread and the times when action potentials are elicited in every ANF (spike trains) were simulated with an active nerve fiber model. The spike activity is defined as the collection of spike trains produced in an ANF population.

The electrode–nerve interface used in the proposed model was based on the cochlea model presented in Nogueira et al. (2016). Cochlear geometry, electrode location, and position of the ANF population were taken from a generic version of their model. The number of ANFs was increased from 7,000 to 9,001 and distributed along 900° of insertion angle from base to apex (two turns and a half) with a separation of 0.1°. The ANFs were indexed in order from the base of the cochlea (high frequencies) to the apex of the cochlea (low frequencies). Another adjustment was done to the electrode array. Nogueira et al. (2016) modeled an electrode array of 22 electrodes; therefore, the electrodes 21, 19, 17, 15, 13, and 11 were removed to obtain the 16 electrodes present in advanced bionics CIs. The resulting electrode–nerve interface is shown in Figure 3.

The morphology of the ANFs was modeled after the myelinated fibers presented in Ashida and Nogueira (2018), which is a simplified representation consisting of segments with a constant internodal length (L_i) equal to 200 μm that extends from the location of the peripheral terminal toward the cochlear nerve. In this morphological model, there is no differentiation between the peripheral axon, central axon, or the soma. As in Ashida and Nogueira (2018), the electric stimulation in the myelinated model was calculated at the nodes that join together two adjacent segments. The voltage produced by the stimulation current I_n coming out of the electrode n was calculated for every node a of every fiber f as shown in Equation (1).

$$U_{nf_a} = \frac{\rho_{\text{ext}} I_n}{4\pi d_{nf_a}}. \quad (1)$$

The extracellular resistivity of the homogeneous medium (ρ_{ext}) was set to 3.0 Ωm as in Ashida and Nogueira (2018). The variable d_{nf_a} is the distance between the electrode n and the node a. This approach results in a voltage spread inversely proportional to the distance d_{nf_a} (Litvak et al., 2007a; Nogueira et al., 2016).

The activation function has been proposed by Rattay (1999) to approximate the amount of functional electrical stimulation over an ANF. In this model, it was calculated as shown in Equation (2). The activation function in a node a depends on its external voltage (U_{nf_a}), and the external voltage on its adjacent nodes (a-1 toward the periphery and a+1 toward the central neural system). The axon internal resistance (R_i) was obtained as "$R_i = 4L_i r/\pi D^2$". The axon diameter (D) was set to 2.0 μm, and the axial resistivity (r) to 1.0 Ωm, as mentioned by Ashida and Nogueira (2018). Notice that the activation function in this study has units in Amperes [A]

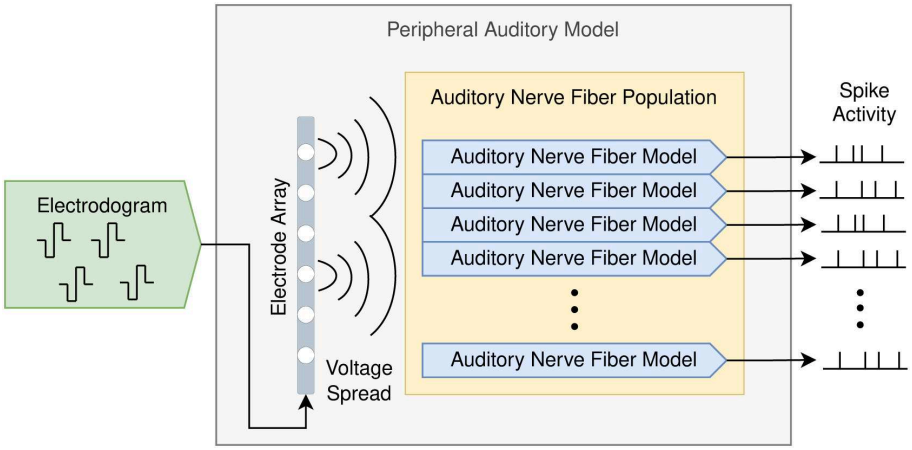


FIGURE 2
Peripheral auditory model for cochlear implants consisting of a population of auditory nerve fibers and an electrode array. The input is the electrodogram generated by the cochlear implant sound coding strategy, and the output is the spike activity produced by the auditory nerve fiber population.

instead of volts per second [V/s] as originally defined by Rattay (1999). This is because the membrane capacitance is not included in Equation (2); however, this membrane capacitance is taken into account in a further stage of the model.

$$A_{nf_a} = \frac{U_{nf_{a-1}} - U_{nf_a}}{R_i} + \frac{U_{nf_{a+1}} - U_{nf_a}}{R_i}. \tag{2}$$

To simulate the spikes generated by each ANF, the neuron model of Joshi et al. (2017) was implemented. This is a "phenomenological" model that represents the peripheral and central axons as two independent adaptative integrate-and-fire circuits that are coupled together by a logical "OR" gate (see Figure 1 in Joshi et al., 2017). This phenomenological model does not convey any geometric information such as the distance between the stimulating electrode and the ANF. Therefore, the induced current I (called stimulation current in Joshi et al., 2017) was adjusted according to the activation function obtained from the electrode–nerve interface model as shown in Equation (3).

$$I = M_C \sum_{n=1}^{N} A_{nf_{a_{max}}}. \tag{3}$$

Notice that the activation function in Equation (2) has a value for every node in an ANF. To simplify the implementation, only the node (a_{max}) with the maximum absolute value of the activation function was taken into account to compute the induced current I. This is based on the fact that this is the node with the highest probability to produce a spike. In addition, a modeling factor (M_C) that allowed to calibrate the peripheral auditory model was added. It was adjusted to reproduce approximately the same spike count reported by Joshi et al. (2017) given different stimulation current levels.

The model of Joshi et al. (2017) assumes that the peripheral and central circuits share the same induced current (I), but they respond differently to the positive (anodic; I^+) and negative (cathodic; I^-) phases of the biphasic pulses. Therefore, in this study, the peripheral axon circuit is referred to as cathodic-excitatory while the central axon circuit as anodic-excitatory. The circuit specific induced current (I_{Stim}) was obtained with Equation (4), where the inhibitory compression (β) was set to 0.75.

$$I_{Stim} = \begin{cases} -(I^- + \beta I^+) & \text{Cathodic-excitatory circuit.} \\ I^+ + \beta I^- & \text{Anodic-excitatory circuit.} \end{cases} \tag{4}$$

The membrane voltage (V) for both circuits is calculated with Equation (5), where the membrane capacitance (C) takes different values for the cathodic-excitatory (856.96 nF) and the anodic-excitatory (1772.4 nF) circuit, $h(V)$, is a passive filter dependent on membrane voltage, I_{Sub} and I_{Supra} are internal subthreshold and suprathreshold adaptation currents, and I_{Noise} is a noise current source with a Gaussian spectral shape that introduces stochastic behavior into the spike trains. The passive filtering, the evolution of the adaption currents and the noise have their own function and can be found in the publication of Joshi et al. (2017). Whenever the membrane voltage of the cathodic-excitatory or the anodic-excitatory circuit reached a threshold, a spike was generated and the ANF entered in an absolute refractory period (ARP) of 500 μs. During the ARP, neither the cathodic- nor anodic-excitatory circuit could produce a spike.

$$C\frac{dV}{dt} = h(V) - I_{Sub} - I_{Supra} + I_{Noise} + I_{Stim}. \tag{5}$$

Another important feature of the proposed peripheral auditory model is the representation of different neural health conditions. A degeneration index (α_f) was assigned to every ANF, which was a natural number from 0 to 20, indicating how many segments were removed from its modeled morphology. The segments were always removed from the most peripheral part resembling the dendritic degeneration that occurs when the inner hair cells in the basilar membrane are damaged (Spoendlin and Schrott, 1988;

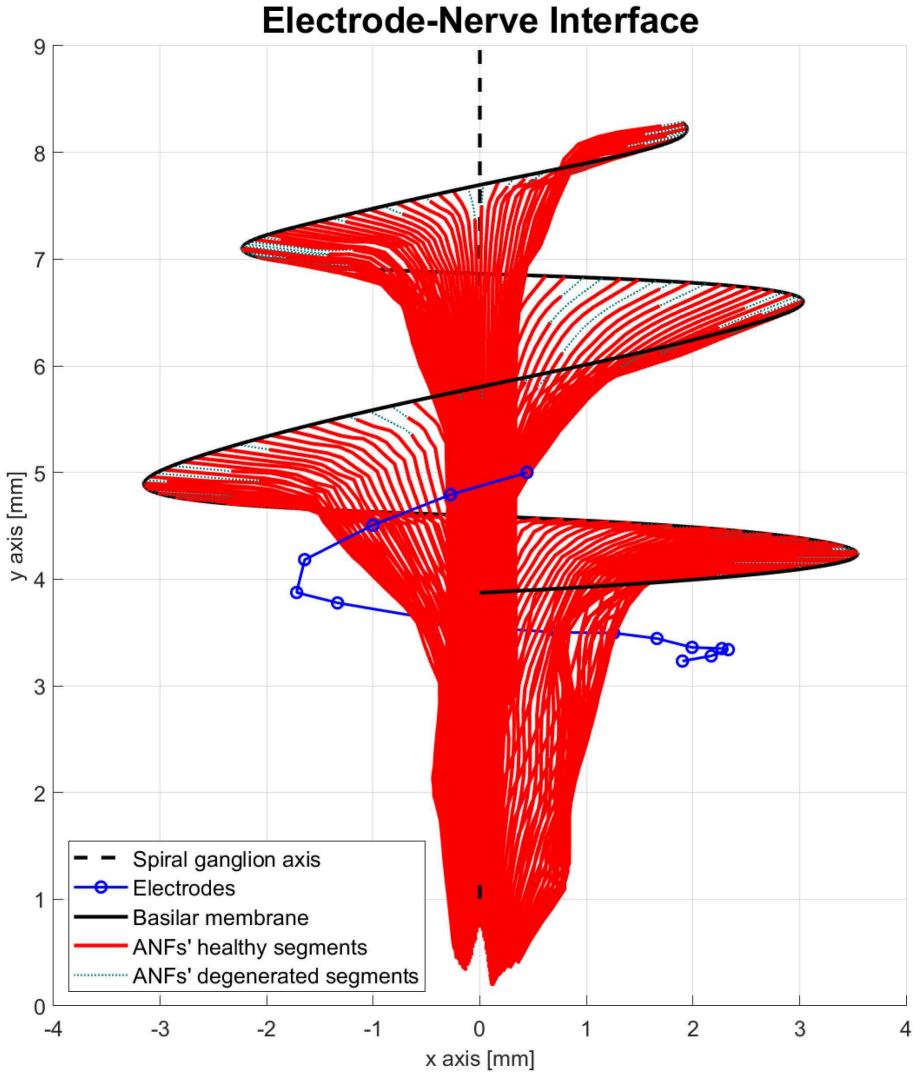


FIGURE 3
Electrode–nerve interface. It is a three-dimensional representation of the auditory nerve fibers (ANFs) (red) arranged in a spiral shape along the basilar membrane (continuous black line) and centered in the spiral ganglion axis (dashed black line). The electrodes (blue) are inserted almost one complete turn into the scala tympani. Some ANFs present a mild degeneration (dotted green).

Nadol, 1997). The nerve degeneration was limited to 20 segments because, beyond this point, the amount of stimulation current required to elicit a spike was excessive compared to real CI users. Figure 4 shows how nerve degeneration was implemented in the proposed electrode–nerve interface.

It is worth mentioning that removing segments may result in situations where the degenerated part surpasses the physical location of the soma, which in real spiral ganglion neurons would be somewhere between the seventh and the twelfth segment. Nevertheless, degeneration of the peripheral axon could also be modeled as the loss of myelin sheets, or by reducing its diameter (Heshmat et al., 2020, 2021; Croner et al., 2022). The effects of this type of degeneration would be that nodes in the central axon will be the ones that produce a spike. Therefore, removing segments in our model was used to investigate excitation at most central locations, rather than to represent the real physical degeneration.

Because it is unknown how the current flows in the most peripheral nodes after degeneration, it was decided to discard them from the activation function calculation. In this regard, Rattay (1999) proposed to remove the first term in Equation (2); nevertheless, in degenerated ANFs, following this proposal could result in a rise of the activation function despite the worst neural health condition and this effect was undesired in our model.

2.1.1.3. Internal representation as features

The feature extraction algorithm was based on the internal representation (IR) presented by Fredelake and Hohmann (2012), which accounts for more central processes in the auditory pathway. The IR consists of a spatial and a temporal integration of the spike activity produced by the ANF population. The first step was to downsample the spike activity to a sample rate of 10 kHz.

To perform the spatial integration, the ANFs were grouped resembling the auditory filters described in Moore (2003). It is

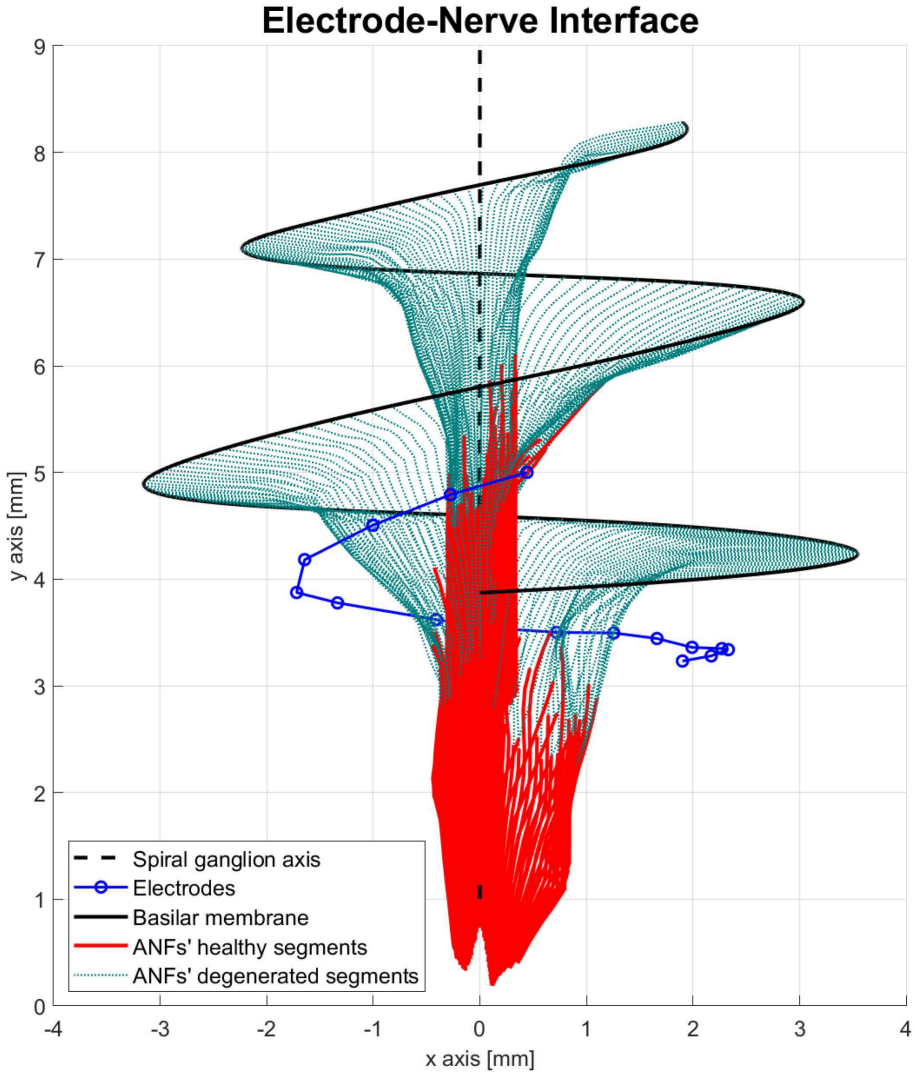


FIGURE 4
Electrode–nerve interface geometrical model with auditory fibers degenerated (dotted green). Mean degeneration (α_f) equal to 20 segments.

mentioned that a maximum of 39 independent auditory filters can be formed at the same time to code sound in NH people, but it is also mentioned that the effective number of channels for CI users could be reduced depending on the number of electrodes implanted. Therefore, in the computational model, the number of auditory filters used was limited between 16 (number of electrodes) and 39. Lengths of the auditory filters ranged between 1.1 and 2.6 mm in the 42 mm basilar membrane of the modeled cochlea. To obtain the number and distribution of these auditory filters along the basilar membrane, adjacent ANFs were grouped by their most likely stimulating electrode (highest absolute value of the activation function) to form auditory filters. If an auditory filter was below the minimum size (1.1 mm), its fibers were merged with the adjacent auditory filter toward the apex. In case an auditory filter size exceeded the maximum value (2.6 mm), its most basal ANFs were used to form a new auditory filter of maximum size while the remaining ANFs were used to form a different auditory filter. Once the auditory filters were formed, the spike trains of their corresponding ANFs were added together to obtain the spike group activity (S_g), where g was the auditory filter index.

The next step was to integrate this spike group activity across time. For each group, the signal was low pass filtered as shown in Equation (6), where $F_g(k)$ is the filtered spike group activity, k is the time frame index, f_s is the sample frequency equal to 10.0 kHz, τ_{LP} is the time constant of the filter set to 1 ms, and the operator "$*$" denotes a convolution.

$$F_g(k) = S_g(k) * \exp\left(-\left(\frac{k}{\sqrt{2} f_s \tau_{LP}}\right)^2\right). \quad (6)$$

From this point, a forward masking effect is implemented. A masker signal Z_g was derived from the filtered spike group activity using a recursive low pass filter (see Section 2.3 in Fredelake and Hohmann, 2012). This masker signal increases exponentially with onsets in the spike group activity and decreases exponentially with

the offsets. The IR is finally the maximum value between the masker signal and the filtered spike group activity. The whole forward masking effect is taken from Fredelake and Hohmann (2012) and it is not detailed in this study. A visual representation is found in Figure 4 of Fredelake and Hohmann (2012).

To meet the requirements of the back-end, the IR was further downsampled to a sample rate of 100 Hz using a moving average low pass filter to diminish the effects of aliasing.

2.1.2. Back-end

The computational model used the framework FADE as the back-end. As a predictor algorithm, FADE uses an HMM that represents the target stimulus with an eight-state Markov chain and a one-component Gaussian mixture model (GMM) to learn, and subsequently predict, the features (each auditory filter in the IR) (Schädler et al., 2016). FADE counts with different "ready-to-use" experiment templates, processing algorithms, and feature extraction algorithms that are intended to predict NH and hearing aid performances (Kollmeier et al., 2016; Schädler et al., 2016, 2018). Hence, two new experiment templates (described in detail in Section 2.4) were developed to perform SRT and SMT experiments. The tasks handled by these new templates are as follows:

1. The generation of the stimulus audio files composing the training and testing corpus.
2. The generation of the electrodograms from these audio files using the respective CI sound coding strategy as the processing algorithm.
3. The generation of the stimuli' IRs using the proposed peripheral auditory model as the feature extraction algorithm.
4. The training of the HMM with the IRs obtained from the training corpus.
5. The predictions over the IRs of the testing corpus with the trained HMM.
6. The evaluation of the performance of the HMM.

In the evaluation stage, FADE generated a file with the score obtained at different training conditions (dB_{SNR} in SRT experiments and $dB_{contrast}$ in SMT experiments). Scores were represented as data points in a scatter plot and a non-linear regression to a psychometric function was performed. This psychometric function is defined in Equation (7), where p_{chance} is the lower horizontal asymptote of the function representing the probability of getting a correct answer with random predictions, p_{max} is the upper asymptote of the function representing the predicted performance in ideal conditions, p_{range} is the difference between the upper and lower asymptotes, s is the slope, or growth rate, at the inflection point of the psychometric function, and x_o is the offset of the inflection point in the x-axis (dB_{SNR} or $dB_{contrast}$). The regression was performed with the MATLAB function "*fitnlm*". The coefficient of determination R^2 was obtained in every experiment, which is a reference of how well the scattered data was represented by the regressed psychometric function.

$$Ps(x) = p_{chance} + \frac{p_{range}}{1 + e^{-s(x-x_o)}}. \quad (7)$$

2.2. Fitting and calibration

The fitting procedure in CI users consists of the adjustment of the stimulation levels of each electrode in the array, or virtual channels in the case of current steering strategies such as F120. Each electrode, or virtual channel, stimulates at levels between threshold (T) and most comfortable loudness (MCL) that are unique for every CI user. By default, the advanced bionics device sets T to 10% of the MCL level resulting in a 20 dB dynamic range. Stimulation levels with the advanced bionics device are given in clinical units (CU), which are integer values from 1 to 471. The equivalent output current (I_n) [μA] was obtained with Equation (8), where T_p is the pulse duration in μs (18 μs in this study), I_{max} is the maximum output current equal to 2,040 μA, and T_{max} is the maximum pulse duration equal to 229 μs (Advanced Bionics, 2020).

$$I_n = \frac{CU}{6000} \frac{I_{max} T_{max}}{T_p}. \quad (8)$$

The process of fitting requires a closed feedback loop, where the CI user indicates the loudness perceived to an audiologist. This loop is virtually closed in the computational model by measuring the spike activity produced by electric pulse trains at different levels of stimulation (from 1 to 471 CU in steps of 30 CU) based on the assumption that the loudness perception is closely related to the neural activity produced by the ANFs (McKay and McDermott, 1998; McKay et al., 2001).

The fitting stimulus used was a cathodic-leading biphasic pulse train with a pulse duration of 18 μs and a periodicity of 540 μs (resulting approximately in 1,852 pps) was consistent with the experimental parameters. The pulse train had a duration of 200 ms with 10 ms of leading and preceding silence. Because the periodicity of this fitting stimulus is just above the ARP, it was expected that ANFs close to the electrodes "fired" with every biphasic pulse of the pulse train.

For each virtual channel, a group of 858 ANFs with the highest absolute activation function was selected. This number corresponds to the number of fibers found in approximately 4 mm section of the modeled basilar membrane, although the selected ANFs were not constrained to be adjacent to each other. The MCL was defined as the CU value where each biphasic pulse elicited a spike in the selected ANF group with a probability of 75%. A similar assumption was used by Kalkman et al. (2014). T level was set to the 10% of the MCL level.

The calibration of the peripheral auditory model refers to the adjustment of the modeling factor (M_C) shown in Equation (3). This process was closely related to the fitting procedure described earlier. It was selected a M_C equal to 89.525×10^6, which guaranteed that MCL levels did not exceed the maximum of 250 cu in any neural health condition used in the experiments. Stimulation above this limit would produce undesired out of compliance stimulation.

2.3. Neural health conditions

In preliminary experiments (not shown in this study) it was found that 9001 fibers introduced a considerable amount of

redundant information, and also, in the majority of the ANFs the node with the highest activation function for any electrode was beyond the fifth node. Therefore, it was decided to use different α_f for every ANF, but defining a mean degeneration value with a standard deviation of three nodes. This diminished the redundant information and was a more realistic representation of how the degeneration gradually occurs (Nadol, 1997).

The peripheral models with a mean ANF degeneration of 5, 10, and 15 nodes were considered to have a "healthy" neural health condition, a "moderate" loss, and a "severe" loss, respectively. Those three cases were assessed in this study and are shown in Figure 5.

2.4. Experiments

A total of nine SMT experiments and nine SRT experiments were performed using the three variants of the F120 sound coding strategies (F120-S, F120-P, and F120-T) and three aforementioned neural health conditions. Each neural health condition in the peripheral auditory model can be considered as an "individual" in experiments with real CI users.

2.4.1. Spectral modulation threshold experiments

The spectral modulation threshold (SMT) experiment was defined by Litvak et al. (2007b). It consists of measuring the smallest detectable spectral contrast in spectral rippled noise. The spectral shape of spectral rippled noise is sinusoidal and it is generated using Equation (9), where $|F(f_r)|$ is the magnitude at the frequency bin f_r, C_t is the spectral contrast in dB, f_{RPO} is the ripple-per-octave, and θ_0 is the ripple phase in the spectrum. Notice that the signal is hard-filtered at values below 350 Hz and above 5,600 Hz.

$$|F(f_r)| = \begin{cases} 10^{\frac{C_t}{2}\sin(2\pi(\log_2(f_r/350))f_{RPO}+\theta_0)/20} & 350 < f_r < 5600 \\ 0 & otherwise \end{cases} \quad (9)$$

With human participants, the SMT is obtained with a three intervals two alternative forced choice procedure consisting of two reference intervals with no ripple (C_t equals to 0), and one target interval with a defined C_t. The first interval is always a reference noise, hence the participant has to indicate if the target interval is presented in the second or third position. This adaptive procedure is detailed by Litvak et al. (2007b), and has an equilibrium point of 79.4% correct answers. Because FADE uses a "training and testing" approach, the adaptive procedure could not be implemented; however, the equilibrium point is kept as the detection threshold.

The corpus was generated using Equation (9) with MATLAB. The ripple phase (θ_0) was randomly selected for every stimulus signal. A ripple per octave (f_{RPO}) equal to 0.5 was selected as in the experiments from Litvak et al. (2007b) and Langner et al. (2017).

The training corpus consisted of 1,000 samples of the spectral ripple noise with random integer values between 2 and 20 dB for contrast depth, and 1,000 samples of reference noise (C_t equals to 0). The testing corpus consisted of 10 sets, each one with 50 samples of reference noise and 50 samples of spectral ripple noise at a target C_t of 2, 3, 4, 5, 7, 9, 11, 14, 17, and 20 dB, respectively. In total 1,000 samples were predicted at 10 different contrast levels.

The sampling frequency of every sample was 17.4 kHz and the stimulus duration was limited to 0.4 s. In all cases, loudness roving was implemented keeping a mean value of −49 dB_{FS}, a roving peak of 5 dB_{FS}, and a roving resolution of 0.5 dB_{FS}.

The spectral modulation detection performance was described by the psychometric function in Equation (7), where p_{chance} was set to 50% because it was a binary decision. The SMT was the x value, where $Ps(x)$ was equal to 79.4%.

2.4.2. Speech reception threshold experiments

The SRT experiments were performed using the Oldenburg sentence test (OLSA). OLSA consists of a matrix sentence test of 50 words that belong to five different categories of 10 words each: name, verb, number, adjective, and noun. The sentences were constructed with one word from each category, following the previously mentioned order, giving a total of 10^5 possible combinations (Wagener et al., 1999). In a closed test procedure, the participants have previous knowledge of the words that may appear. Several sentences, mixed with noise at a specific SNR, are presented to the subject and the subject is asked to repeat them. A score based on the percentage of correctly recognized words is obtained. This is repeated at different SNR conditions and then a psychometric function is fitted to the obtained data points. The SNR value where this psychometric function crosses the 50% mark of correctly recognized words is the SRT result.

For SRT experiments, Schädler et al. (2016) and Jürgens et al. (2018) used a subset of 120 OLSA sentences to generate the training and testing corpus, but in this study, only a subset of 100 OLSA sentences was used to reduce computational resources. In this subset, each of the 50 words in the matrix appears exactly 10 times. A random excerpt of the noise provided by OLSA was added to the sentences at the required level to obtain the different SNRs, while the speech was kept at −49 dB_{FS}.

FADE uses a closed training/testing approach, meaning that the same sentences used in the training are used in the prediction stage. Therefore, the training corpus was generated with all the sentences in the subset mixed with noise at seven different SNR levels, from 0 to 18 dB in steps of 3 dB, and without noise, giving a total of 80 unique instances for each word. The testing corpus was generated with all the sentences mixed at 10 different SNR values, from −9 to 18 dB in steps of 3 dB, giving a total of 5,000 words to be predicted.

Regarding the psychometric function described in Equation (7), p_{chance} was set to 10% because it was a one word out of 10 decisions. The SRT was the x value where $Ps(x)$ was equal to 50%.

3. Results

3.1. Fitting

Figure 6 shows the MCL levels obtained for the computational model with the healthy, moderate degeneration, and severe degeneration condition. It also shows as a reference an ideal case (no degeneration in the ANFs), and the worst case (20 degenerated segments in the ANFs).

The MCL level difference across the 15 virtual channels between ideal and healthy conditions was only on average 2.87 CU.

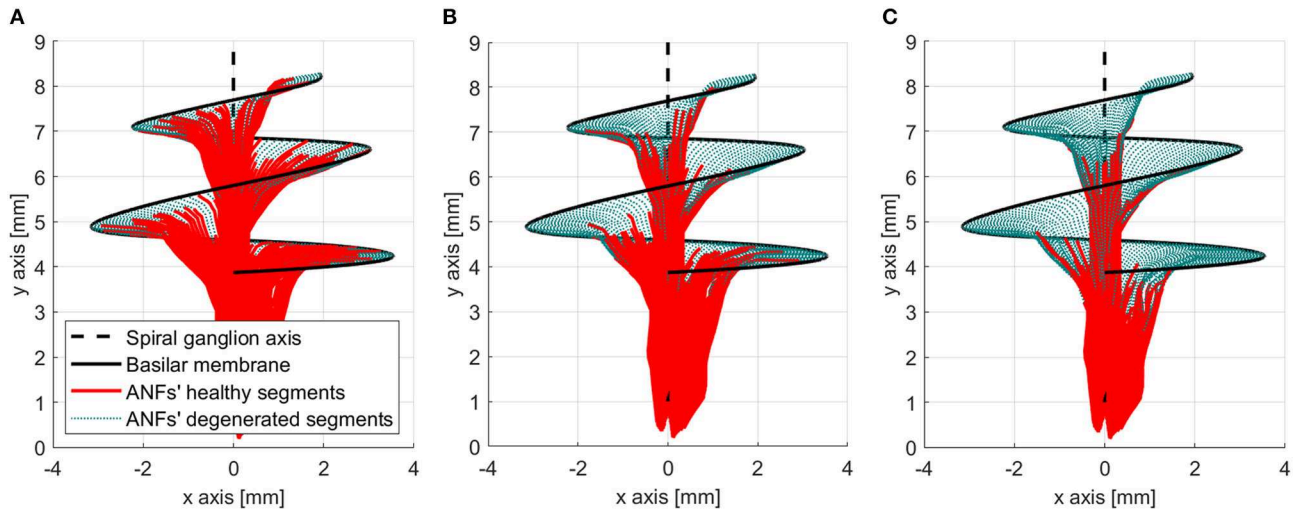


FIGURE 5
Three neural health condition representations. **(A)** Healthy conditions (mean degeneration of 5 nodes). **(B)** Moderate auditory nerve fiber (ANF) degeneration (mean degeneration of 10 nodes). **(C)** Severe ANF degeneration (mean degeneration of 15 nodes).

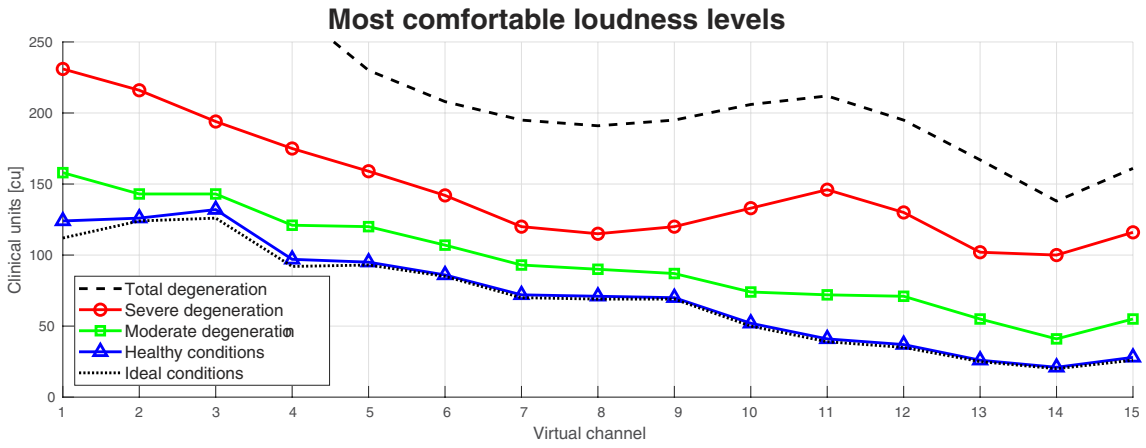


FIGURE 6
Most comfortable levels at different neural health conditions of the auditory peripheral model. Ideal conditions is equivalent to no degeneration in the auditory nerve fibers (ANFs). Healthy conditions, moderate degeneration, and severe degeneration are equivalent to a mean of 5, 10, and 15 segments degenerated in the ANFs, respectively. Total degeneration is equivalent to 20 segments degenerated in almost all the ANFs.

This difference increased with worse neural health conditions. Between healthy conditions and moderate degeneration, the difference was on average 23.47 CU, between moderate and severe degeneration, the difference was 51.27 CU, and between moderate and total degeneration, the difference was 80.53 CU. With total degeneration in the ANFs, the MCL levels of the four most basal electrodes (high frequencies) were above the desired 250 CU.

3.2. Electrical interaction

Figure 7 shows the effects of electrical interaction with simultaneous stimulation after fitting. For this figure, paired biphasic pulses were generated for the 1st, 6th, 11th, and 14th virtual channels to obtain a peak of induced current I across the ANF population of 0.8 mA. The black continuous lines correspond to the induced current with simultaneous stimulation, while the induced currents resulting from each channel individually (sequential stimulation) are shown with different colors. The induced current with simultaneous stimulation in healthy conditions (Figures 7A, B), and with paired stimulation in severe degeneration (Figure 7C), follows the peaks corresponding to the induced current of each individual virtual channel. This is not the case with triplet stimulation in severe degeneration (Figure 7D), where the peaks corresponding to virtual channels 6 and 14 are almost merged together while the peak corresponding to the virtual channel 1 is attenuated. Attenuation occurs when different virtual channels have opposite activation function signs; therefore, they cancel each other in Equation (3). Note

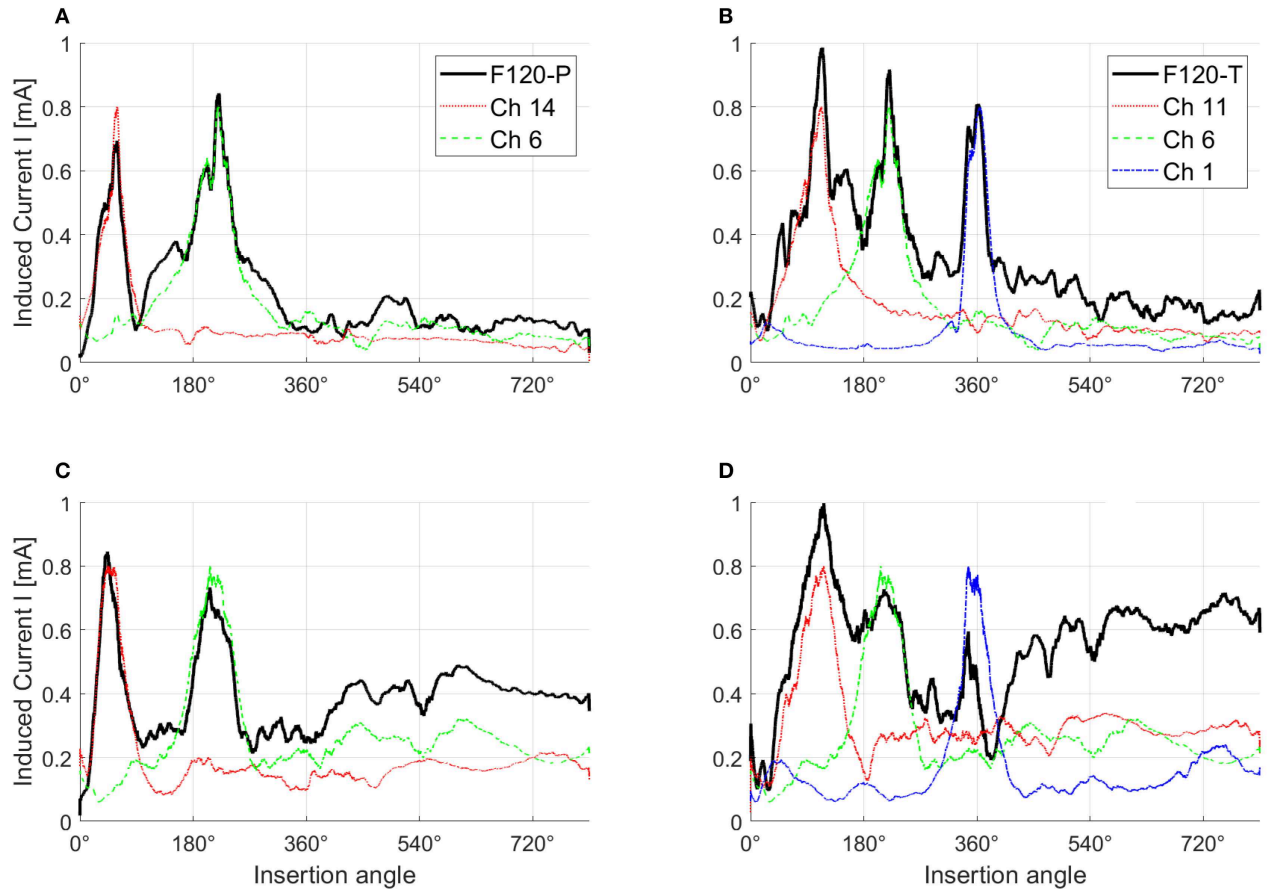


FIGURE 7
Effects of electrical interaction in simultaneous stimulation with F120-P and F120-T sound coding strategies. Panels **(A, B)** show electrical interaction in healthy condition while panels **(C, D)** show electrical interaction in severe degeneration. For this example, the stimulation pulses in every virtual channel were adjusted until they reached a peak of the induced current across the auditory nerve fibers of 0.8 mA. F120-P uses the virtual channels 6 (Ch 6) and 14 (Ch 14) simultaneously as shown in **(A, C)**. F120-T uses the virtual channels 1 (Ch 1), 6 (Ch 6), and 11 (Ch 11) as shown in **(B, D)**. The induced current with simultaneous stimulation is shown with black continuous lines. The y-axis corresponds to the magnitude of the induced current and the x-axis to the insertion angle of the electrode array from base (high frequencies) to apex (low frequencies). The figure was smoothed to provide a better visualization.

that the induced current toward the apex of the cochlea (insertion angles around 540° and 720°) also increases with worse neural health conditions as a consequence of higher stimulation levels.

3.3. Spectral modulation threshold experiments

Figure 8 shows the psychometric functions obtained from the SMT experiments. The upper left box indicates the corresponding SMT, the expected performance with an infinite modulation depth (C_t), and the coefficient of determination (R^2). In general, the psychometric regression obtained an R^2 coefficient ranging from 0.97 to 0.99, which means that the psychometric function represents a good fit for the results obtained.

Nevertheless, Figure 9 shows that there was no significant effect on the performance regarding the sound coding strategy (Figure 9A) or the neural health condition (Figure 9B). As shown in Figure 9B, a small trend toward poorer performance with worse neural health conditions (8.57 dB for healthy conditions, 8.76 dB for moderate degeneration, and 9.63 dB for severe degeneration) is not significant compared to the clinical data obtained by Litvak et al. (2007a) and Langner et al. (2017). Also, as shown in Figure 9A, the expected effect of sound coding strategy on performance is not obtained (e.g., with severe degeneration the SMT is better using F120-P than using F120-S). The p_{max} seems to not be affected either by the neural degeneration or sound coding strategy since it varies from 91.2% (F120-T at healthy conditions) and 93.1% (F120-P at moderate degeneration of the ANFs).

Data from Litvak et al. (2007b) were collected from 25 CI users (Saoji et al., 2005) and are shown in Figure 9, part A, with blue boxes. The CI users were using F120-S (14 participants) and F120-P (11 participants) sound coding strategies. The results from Langner et al. (2017), shown in the same graph with green boxes, included 14

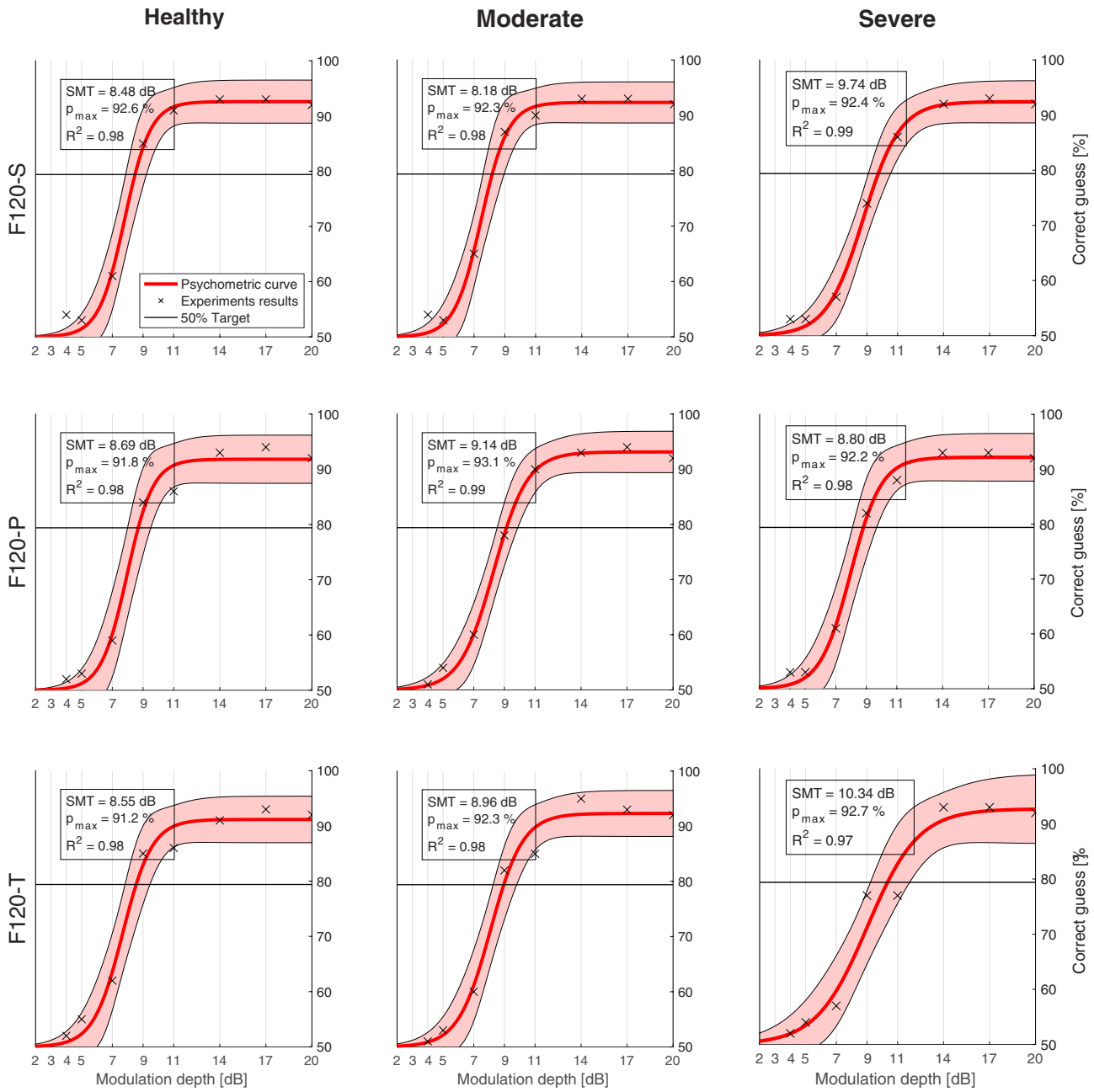


FIGURE 8
Psychometric functions (red) obtained from the computational model results (crosses). The legend in every chart indicates the spectral modulation threshold (SMT), which is the modulation depth where the psychometric function crosses the 79.4% target line (black). p_{max} is the expected performance when the modulation depth tends to infinity. R^2 is the coefficient of determination between the results and the psychometric curve. The shaded area shows 95% confidence interval. Charts are arranged in columns by neural health (**left** to **right**: healthy conditions, moderate degeneration, and severe degeneration), and in rows by sound coding strategy (**top** to **bottom**: F120-S, F120-P, and F120-T).

experiments. Half of them showed the SMT comparison between F120-S and F120-P sound coding strategies, while the other half was between F120-S and F120-T. Their performance showed a large variability but the trend is to get worse with simultaneous stimulation, especially with F120-T.

The performance of the computational model ranged from 8.48 to 10.34 dB, which can be considered as a "good" to "average" performance for CI users, but it does not account for the variability. The difference between the worst and the best performance of the model was only 1.90 dB, while in real CI users, it can be more than 18 dB.

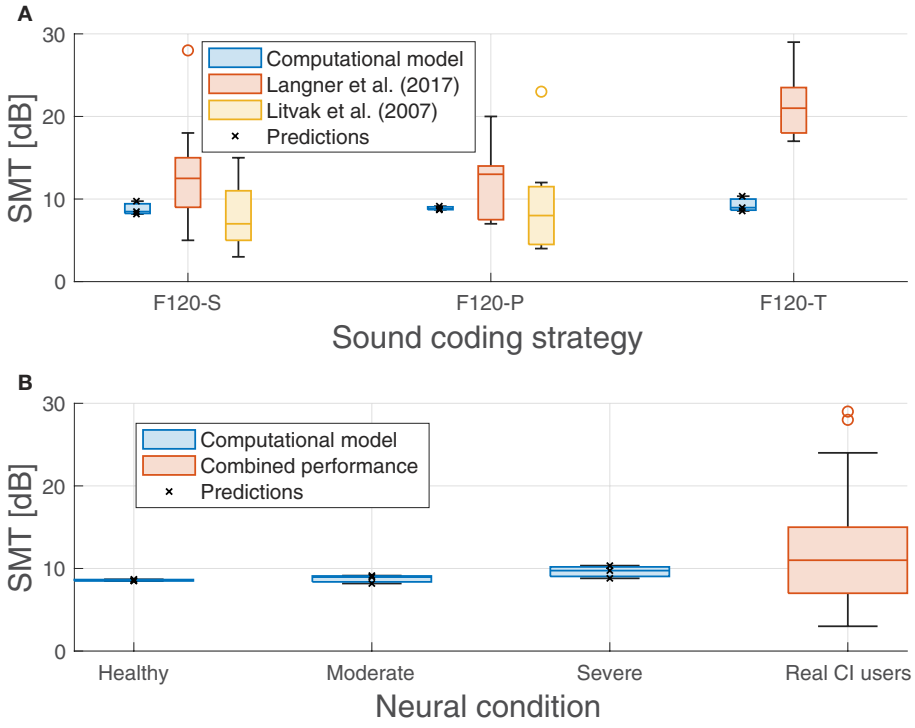


FIGURE 9
Predicted spectral modulation threshold (SMT) obtained in the experiments compared with measurements obtained with real CI users from Litvak et al. (2007a) and Langner et al. (2017). **(A)** Performance of the proposed computational model and the performance from literature grouped by sound coding strategy. **(B)** Performance of the proposed computational model grouped by neural health conditions and the combined performance from Litvak et al. (2007a) and Langner et al. (2017) as a comparison. Predicted values are shown with crosses.

3.4. Speech reception threshold experiments

Figure 10 shows the psychometric functions obtained from the SRT experiments. The upper left box in every chart indicates the corresponding SRT, the expected performance "in quiet" with p_{max}, and the coefficient of determination (R^2) obtained with the regression. In general, the psychometric function represented quite well the data resulted from the experiments, obtaining an R^2 coefficient always equal or greater than 0.99.

The effect of neural health is visible in the overall performance, affecting not only the SRT, but also the predicted performance in quiet. Figure 11 shows the SRT grouped by sound coding strategy (Figure 11A) and neural health condition (Figure 11B). In general, worse neural health conditions led to poorer performance (higher SRTs). The differences in SRT between healthy and severe degeneration conditions for F120-S, F120-P, and F120-T were 1.53, 1.63, and 2.77 dB_{SNR}, respectively. This indicates that simultaneous stimulation with F120-T was more sensitive to the effects of neural health degeneration than the F120-P.

On the contrary, the sound coding strategy that stimulated with two virtual channels simultaneously (F120-P) showed better overall performance, as shown in Figure 11, part A. With better neural health conditions, F120-T obtained better performance than F120-S, but with severe degeneration, the performance with F120-T felt below F120-S performance.

In addition, Figure 11 shows the results obtained by Jürgens et al. (2018). They measured the SRT in a group of 14 CI users using the advanced combinational encoder (ACE) sound coding strategy, which does not use current steering but it is comparable to F120-S because it uses sequential stimulation of biphasic pulses. In addition, they used a computational model to predict the performance of the CI users based on their individualized electrical field spread. Their measured SRTs ranged between −0.1 and 6.2 dB_{SNR}. In contrast, the performance of their computational model ranged between 4.67 and 7.56 dB_{SNR}, which is considered a poor performance according to Jürgens et al. (2018).

4. Discussion

In this study, a novel computational model to simulate the performance of CI users in psychoacoustic experiments was proposed. The proposed model consists of two main parts: (i) a front-end that includes a peripheral auditory model; (ii) a back-end based on the framework FADE. The peripheral auditory model combined a three-dimensional representation of the electrode-nerve interface taken from Nogueira et al. (2016), with a population of "phenomenologically" modeled ANFs taken from Joshi et al. (2017). This combination allowed to overcome the geometric limitations of the

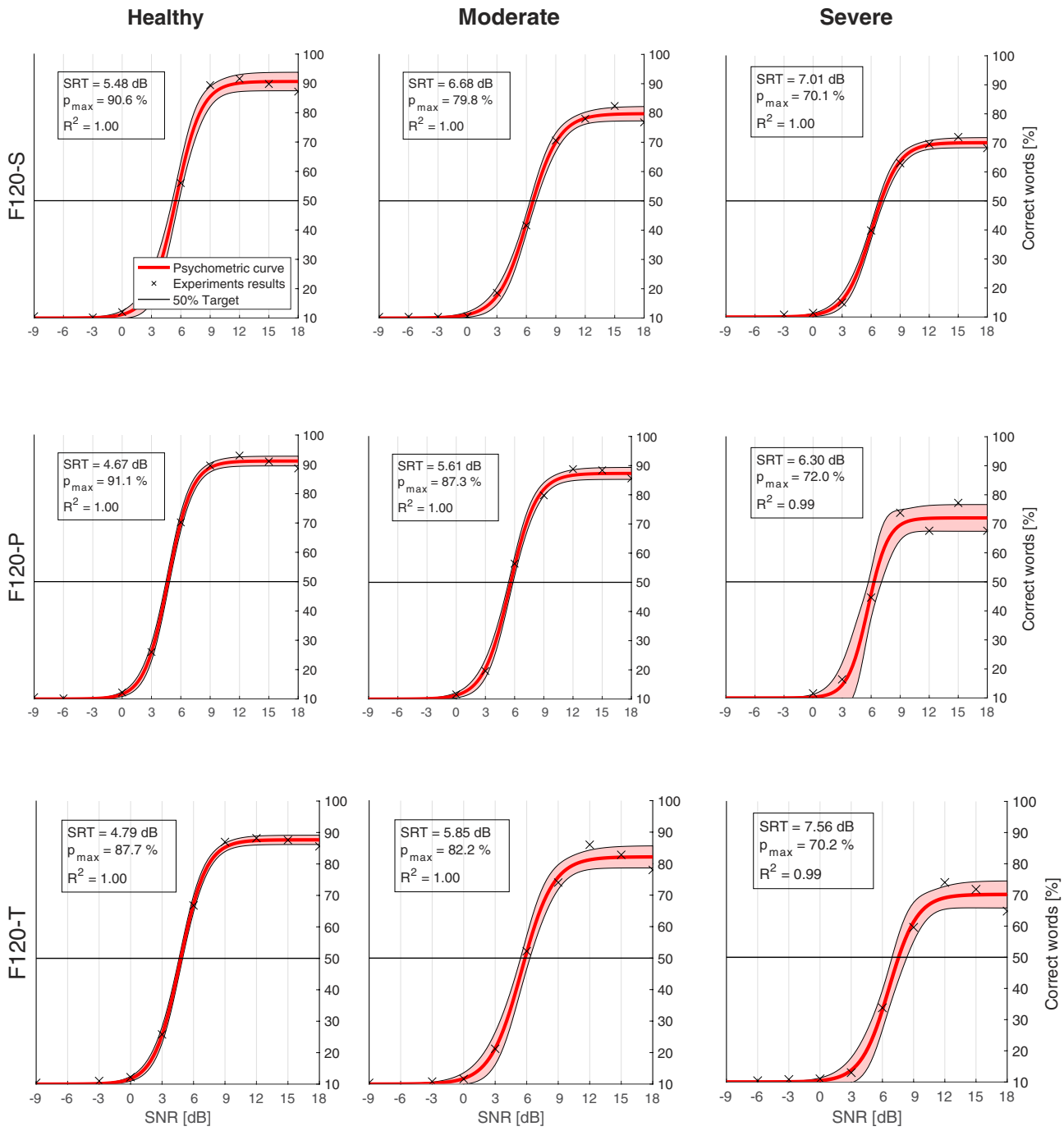


FIGURE 10
Psychometric functions (red) obtained from the computational model results (crosses). The legend in every chart indicates the speech reception threshold (SRT), which is the signal-to-noise ratio (SNR) where the psychometric function crosses the 50% target line (black). p_{max} is the expected performance in quiet. R^2 is the coefficient of determination between the results and the psychometric curve. The shaded area shows 95% confidence interval. Charts are arranged in columns by neural health (**left** to **right**: healthy conditions, moderate degeneration, and severe degeneration), and in rows by sound coding strategy (**top** to **bottom**: F120-S, F120-P, and F120-T).

phenomenological model by making the induced current dependent on the activation function, which was obtained using the "morphological" model of Ashida and Nogueira (2018). In that sense, the proposed peripheral auditory model takes advantage of the benefits of both "phenomenological" and "physiological" approaches.

This study assessed different neural health conditions by gradually removing segments of the ANFs from the periphery toward the spiral ganglion, which is an approximation to the real physiological degeneration (Nadol, 1988; Spoendlin and Schrott, 1988; Nogueira and Ashida, 2018). The activation function decreases and widens with worse neural health

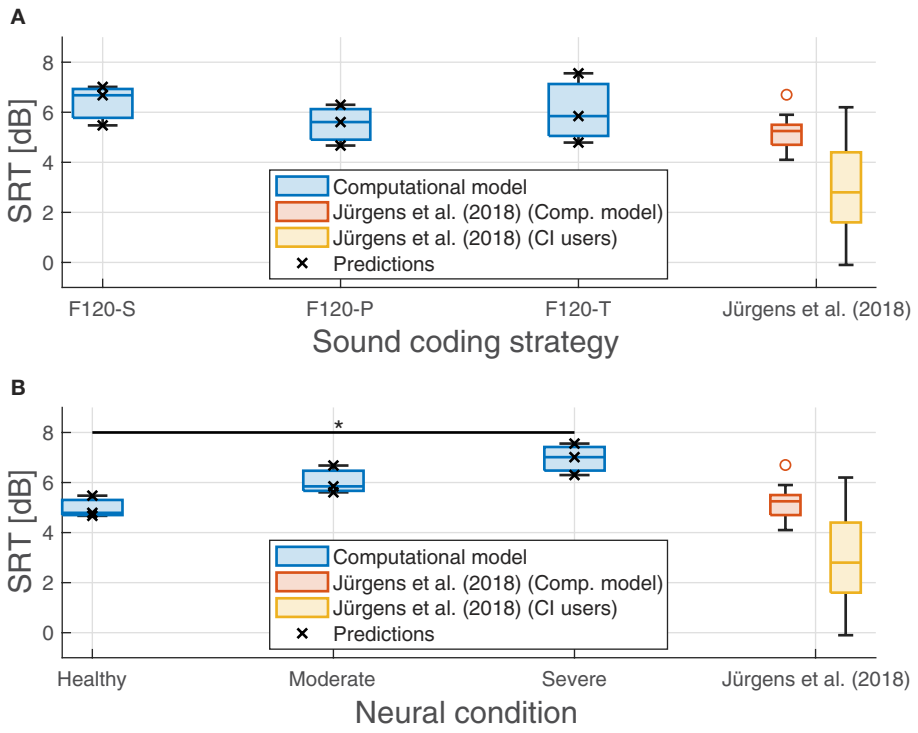


FIGURE 11
Predicted speech reception threshold (SRT) obtained in the experiments. **(A)** Performance of the proposed computational grouped by sound coding strategy. **(B)** Performance of the computational model grouped by neural health conditions. Predicted values are shown with crosses. Results from Jürgens et al. (2018) are shown as a reference in both panels as a reference. The red boxes show their predicted SRTs using an individualized model using the electrical field spread of real CI users and the yellow boxes show their measured SRTs. The symbol "*" denotes a statistical significance with a p-value less than 0.05 between healthy condition and severe degeneration.

conditions having an impact in the MCL levels, as shown in Figure 6. Worst neural health conditions resulted in higher MCL levels, which is consistent with the findings of Langner et al. (2021).

Predictions of the proposed computational model were obtained for two psychoacoustic experiments: SRT and SMT. The results of the SRT experiment show that worse neural health conditions result in poorer speech reception performance. Although this degeneration affected more the performance with simultaneous stimulation sound coding strategies, especially with F120-T, the detriment compared to sequential stimulation (F120-S) was rather small and not as relevant, as shown by Langner et al. (2017, 2021). A more detailed discussion regarding SMT and SRT experiments is presented in the following subsections.

4.1. Spectral modulation threshold experiments

As shown in Figure 9, the computational model performed similarly in SMT despite having different neural health conditions or sound coding strategies. This is because IR is a highly correlated feature along its dimensions (they are not orthogonal) and they cannot convey any relative spectral information that can be used by the HMM to differentiate between spectral rippled noise and flat noise, especially when the phase θ_0 (see Equation 9) of the spectral rippled noise is randomized.

When the phase of the spectral ripple noise is randomized, the spectral peaks and valleys are always located in different auditory filters of the IR. With every realization of spectral ripple noise, the mean value of the IR gets closer to the mean value of the flat noise in every auditory filter. Only the standard deviation of the IR is always greater in spectral ripple noise than in flat noise. Therefore, whenever the neural activity in any auditory filter was greater, or lower, than the activity expected from the flat noise, the HMM classified it as spectral rippled noise. This may also explain why it consistently predicts an SMT around 9 dB since the loudness roving used to generate the training and testing corpus had a dynamic range of 10 dB.

Prediction algorithms based on HMMs work better with decorrelated features such as Mel frequency cepstral coefficients (MFCCs), Garbor filter bank (GFBs), or separable Garbor filter bank (SGFB) that are obtained from the spectral representation of the signal as shown by Kollmeier et al. (2016) and Schädler et al. (2016). On the contrary, IR is somehow equivalent to the bare spectrum of the audio signal, and it is highly correlation to work with HMMs.

4.2. Speech reception threshold experiments

Speech understanding relies on two principal aspects: temporal cues and spectral cues (Xu et al., 2005). Depending on the frequency, the temporal cues can be classified into envelopes (2–50 Hz), periodicity (50–500 Hz), and fine structure (500–10,000 Hz), the envelopes being specially important for speech understanding (Rosen, 1992). In the proposed computational model, the HMM was able to capture these envelopes along the auditory filters using Markov chains of eight states (Schädler et al., 2016); however, periodicity and fine structure is lost. The spectral cues, commonly related to the formants produced by the pronunciation of vowels and some consonants, were not properly captured by the HMM because of the before mentioned limitations of the IR when representing spectral shapes. Thus, words with relatively low SNR resulted in an IR much more similar to a word with a flat spectral shape. This may be the reason why, compared to real CI users, the performance of the computational model was worse than expected (higher SRTs) and did not account for the expected variability of around 6 dB as measured by Jürgens et al. (2018). Figure 11 also shows that this problem can be traced back to their computational model. Jürgens et al. (2018) also used the combination of IR as a feature and FADE as the back-end with an individualized electrical field spread that accounted for differences in the electrode–nerve interface between participants. However, the overall simulated performance resulted in higher SRTs than the one measured in real CI users as in our simulations.

In addition, there is a discrepancy between the effects of channel interaction in simultaneous stimulation (F120-P and F120-T) compared to sequential stimulation (F120-S). According to Langner et al. (2017, 2021), the performance with F120-S is similar to the performance with F120-P, but better than the performance with F120-T. This is not the case in the results shown in Figure 11, part A. This may be caused by the sharp decay of the voltage spread that is inversely proportional to the distance d_{nf_a} (see Equation 1). A sharp decay diminishes the effect of electrical interaction between virtual channels in simultaneous stimulation; therefore, the performance may be more affected by the refractory period of ANFs. Notice in Figure 1 that with F120-S the electrical stimulation is continuous, while with F120-P and F120-T, there are stimulation gaps that may help ANFs to recover from refractoriness.

However, Figure 7 shows that the presented computational model is capable of reproducing the effects of electrical interaction between virtual channels. This effect is larger with the F120-T sound coding strategy because the virtual channels are closer together. Electrical interaction obtained with F120-P is much lower compared to F120-T. This may explain why in real CI users the performance with F120-P is similar to the performance with F120-S, but considerably worse with F120-T (Langner et al., 2017, 2020b). Figures 7, 11, part B, also support the hypothesis assessed in this study since the performance with triplet stimulation was significantly affected by severe degeneration compared to healthy conditions.

4.3. Future improvements

4.3.1. Feature extraction and back-end

As discussed earlier, the IR proposed by Fredelake and Hohmann (2012) may not be the best set of features to use with the HMM already implemented in FADE because IR is highly correlated within its own dimensions. Therefore, it does not carry any spectral shape information that indicates the relative neural activity between the auditory filters. A solution could be to incorporate other decorrelated features that have been shown to improve the performance of automatic speech recognition (ASR) algorithms and that also worked with neural activity (Holmberg et al., 2005, 2007; Nogueira et al., 2007). But, although these algorithms may provide benefits in performance for the computational model, these may not represent any particular physiological aspect of the auditory system, which was the idea behind the proposed computational model.

The central processes that occur in the auditory pathway beyond the peripheral auditory system are not completely understood. The IR is a simple model that makes many assumptions about how spikes are processed centrally to interpret sounds. However, the IR was necessary to accommodate the neural activity to the temporal resolution and number of features adequate for an HMM back-end. Another way to approach this problem would be to substitute FADE as the back-end for an algorithm that can perform predictions directly from the spike activity coming from the peripheral auditory model (Alvarez and Nogueira, 2022), but this is a challenging task due to the amount of ANFs modeled and the sample rate of the spike activity. Artificial neural networks (ANNs) seem to suit well with this approach (Kell et al., 2018; Santana et al., 2018; Wang et al., 2018) since ANNs can manage a large amount of data and there is no intrinsic assumption of any central auditory process. Neural networks could function as a "black box" while the detailed modeling is focused on the peripheral auditory system (Brochier et al., 2022).

4.3.2. Peripheral auditory model

Although the proposed peripheral auditory model can account for many physiological aspects, there is room for improvement.

The voltage spread obtained in this study presents a decay inversely proportional to the distance. This results in a sharper decay compared to the exponential decay measured in laboratories with saline solutions (O'leary et al., 1985; Kral et al., 1998). The sharper the decay, the less electrical interaction between channels in simultaneous stimulation; therefore, the model is less sensitive to the differences between the F120 variants studied. Nevertheless, a more realistic voltage spread could be obtained by using finite element method (FEM) or a boundary element method (BEM) in the three-dimensional electrode–nerve instead of assuming a homogeneous medium (Kalkman et al., 2014, 2022; Nogueira et al., 2016; Croner et al., 2022). This improvement of the peripheral auditory model would take into account the electrical characteristics of the bones, tissue, and other media present in the cochlea.

Regarding the ANF model, the morphological model used in this study considers the ANF as a cable with homogeneously

distributed segments that do not differentiate between peripheral axon, central axon, or the soma. This approach worked because it allowed us to adjust the induced current I for the physiological model of Joshi et al. (2017) depending on different aspects such as electrode and nerve position, the direction of the ANF with respect to the electrode, and the effects of degeneration. A more realistic multi-compartment cable model that represents its morphology and physiology more accurately (Rattay et al., 2013; Bachmaier et al., 2019; Kalkman et al., 2022) could be implemented; however, one must be careful when coupling it with the model of Joshi et al. (2017). If the morphology already takes into account differences between the axons and the soma, the original parameters used in Equation (5) are not validated anymore.

Another improvement could be a new degeneration method that takes into account, not only the progressive inhibition of nodes, but the diameter reduction in the axons (Heshmat et al., 2020; Croner et al., 2022). In the current degeneration method, the number of degenerated nodes is not equal across the ANF population, instead, the number of nodes degenerated in each ANF is governed by a mean value and a standard deviation that is arbitrarily set to three nodes. This is an assumption made since there is no available information on how it is in real implanted cochleas. In fact, it is probable that the fibers degenerate in different patterns, for example, more degeneration in the basal turn than in the apical turns. Therefore, in further studies related to degeneration, different patterns should be taken into account including dead regions in the cochlea, which are relevant in electrical stimulation (Moore et al., 2010).

5. Conclusion

The computational model presented in this study was capable of executing simulations of SRT and SMT experiments. It consisted of a peripheral auditory model with a three-dimensional electrode–nerve interface that allows to represent different neural health conditions by applying a systematic degeneration to the modeled ANFs. The neural health condition affected the fitting procedure and speech reception in the expected manner, augmenting the current needed to reach MCLs and obtaining worst (higher) SRTs, respectively.

The computational model could not reproduce quantitatively the expected results in SRT experiments from real CI users where simultaneous stimulation sound coding strategies (F120-P and F120-T) consistently performed worse than sequential stimulation sound coding strategies (F120-S). Nevertheless, the results showed that the qualitative performance detriment due to neural health conditions with simultaneous stimulation (especially with F120-T) was higher than with sequential stimulation.

SMT experiments with the computational model were inconclusive since the results showed no relevant impact neither from neural health conditions nor channel interaction caused by the simultaneous stimulation. This is arguably caused by the selected IR resulting in features that did not convey spectral shape information, together with an HMM-based recognizer.

Future developments of the computational model could offer a reliable tool to assess the effects of different sound coding strategies and different neural health conditions in psychoacoustic experiments without the need for testing in implanted volunteers, especially, in the early development stages of new CI technologies. The improvements should be focused on the physiological model of the ANFs and the feature extraction from the neural activity.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

FA designed the experiments, programmed and integrated the computational model, verified and analyzed the results, and wrote the original manuscript draft. DK developed the auditory nerve fiber model. WN supervised the research project. All authors contributed to conceptualization, contributed to manuscript revision, read, and approved the submitted version.

Funding

This project was funded by the German Academic Exchange Service (DAAD) under the Research Grants – Doctoral Programmes in Germany 2021/2022 program, and was co-funded by the Cluster for Excellence Hearing4All. This work is part of the project that received funding from the European Research Council (ERC) under the European Union's Horizon-ERC Program (Grant agreement READIHEAR No. 101044753-PI: WN).

Acknowledgments

Special thanks to Dr. Florian Langner for his help in the early stage of the project, specially, for providing a better insight of the sound coding strategies used in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Advanced Bionics (2020). *Target CI Fitting Guide*.

Alvarez, F., and Nogueira, W. (2022). "Predicting speech intelligibility using the spike acativity mutual information index," in *Proc. Interspeech 2022* (Incheon), 3503–3507.

Ashida, G., and Nogueira, W. (2018). Spike-conducting integrate-and-fire model. *eNeuro* 5. doi: 10.1523/ENEURO.0112-18.2018

Bachmaier, R., Encke, J., Obando-Leiton, M., Hemmert, W., and Bai, S. (2019). Comparison of multi-compartment cable models of human auditory nerve fibers. *Front. Neurosci.* 13, 1173. doi: 10.3389/fnins.2019.01173

Brochier, T., Schlittenlacher, J., Roberts, I., Goehring, T., Jiang, C., Vickers, D., et al. (2022). From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants. *IEEE Trans. Biomed. Eng.* 69, 3300–3312. doi: 10.1109/TBME.2022.3167113

Bruce, L. C., White, M. W., Irlicht, L. S., O'Leary, S. J., Dynes, S., Javel, E., etal. (1999). A stochastic model of the electrically stimulated auditory nerve: single-pulse response. *IEEE Trans. Biomed. Eng.* 46, 617–629. doi: 10.1109/10.764938

Chen, F., and Loizou, P. C. (2011). Predicting the intelligibility of vocoded speech. *Ear Hear.* 32, 331–338. doi: 10.1097/AUD.0b013e3181ff3515

Croner, A. M., Heshmat, A., Schrott-Fischer, A., Glueckert, R., Hemmert, W., and Bai, S. (2022). Effects of degrees of degeneration on the electrical excitation of human spiral ganglion neurons based on a high-resolution computer model. *Front. Neurosci.* 16, 914876. doi: 10.3389/fnins.2022.914876

Dillon, M. T., Buss, E., King, E. R., Deres, E. J., Obarowski, S. N., Anderson, M. L., et al. (2016). Comparison of two cochlear implant coding strategies on speech perception. *Cochlear Implants Int.* 17, 263–270. doi: 10.1080/14670100.2016.1244033

Dorman, M. F., Loizou, P. C., and Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J. Acoust. Soc. Am.* 102, 2993–2996. doi: 10.1121/1.420354

El Boghdady, N., Kegel, A., Lai, W. K., and Dillier, N. (2016). A neural-based vocoder implementation for evaluating cochlear implant coding strategies. *Hear. Res.* 333, 136–149. doi: 10.1016/j.heares.2016.01.005

Fredelake, S., and Hohmann, V. (2012). Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation. *Hear. Res.* 287, 76–90. doi: 10.1016/j.heares.2012.03.005

Gajecki, T., and Nogueira, W. (2021). "An end-to-end deep learning speech coding and denoising strategy for cochlear implants," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Singapore).

Hamacher, V. (2004). *Signalverarbeitungsmodelle des elektrisch stimulierten Gehors* (Ph.D. thesis). RWTH Aachen, Wissenschaftsverlag Mainz, Aachen, Germany.

Han, J. H., Lee, H. J., Kang, H., Oh, S. H., and Lee, D. S. (2019). Brain plasticity can predict the cochlear implant outcome in adult-onset deafness. *Front. Hum. Neurosci.* 13, 38. doi: 10.3389/fnhum.2019.00038

Heshmat, A., Sajedi, S., Chacko, L. J., Fischer, N., Schrott-Fischer, A., and Rattay, F. (2020). Dendritic degeneration of human auditory nerve fibers and its impact on the spiking pattern under regular conditions and during cochlear implant stimulation. *Front. Neurosci.* 14, 599868. doi: 10.3389/fnins.2020.599868

Heshmat, A., Sajedi, S., Schrott-Fischer, A., and Rattay, F. (2021). Polarity sensitivity of human auditory nerve fibers based on pulse shape, cochlear implant stimulation strategy and array. *Front. Neurosci.* 15, 751599. doi: 10.3389/fnins.2021.751599

Hochmair-Desoyer, I., Schulz, E., Moser, L., and Schmidt, M. (1997). The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. *Am. J. Otol.* 18, S83–S86.

Holmberg, M., Gelbart, D., and Hemmert, W. (2007). Speech encoding in a model of peripheral auditory processing: quantitative assessment by means of automatic speech recognition. *Speech Commun.* 49, 917–932. doi: 10.1016/j.specom.2007.05.009

Holmberg, M., Gelbart, D., Ramacher, U., and Hemmert, W. (2005). "Automatic speech recognition with neural spike trains," in *9th European Conference on Speech Communication and Technology* (Lisbon), 1253–1256.

Imsiecke, M., Büchner, A., Lenarz, T., and Nogueira, W. (2021). Amplitude growth functions of auditory nerve responses to electric pulse stimulation with varied interphase gaps in cochlear implant users with ipsilateral residual hearing. *Trends Hear.* 25. doi: 10.1177/23312165211014137

Joshi, S. N., Dau, T., and Epp, B. (2017). A model of electrically stimulated auditory nerve fiber responses with peripheral and central sites of spike generation. *J. Assoc. Res. Otolaryngol.* 18, 323–342. doi: 10.1007/s10162-016-0608-2

Jürgens, T., Hohmann, V., Büchner, A., and Nogueira, W. (2018). The effects of electrical field spatial spread and some cognitive factors on speech-in-noise performance of individual cochlear implant users–A computer model study. *PLoS ONE* 13, e0193842. doi: 10.1371/journal.pone.0193842

Kalkman, R. K., Briaire, J. J., Dekker, D. M., and Frijns, J. H. (2014). Place pitch versus electrode location in a realistic computational model of the implanted human cochlea. *Hear. Res.* 315, 10–24. doi: 10.1016/j.heares.2014.06.003

Kalkman, R. K., Briaire, J. J., Dekker, D. M., and Frijns, J. H. (2022). The relation between polarity sensitivity and neural degeneration in a computational model of cochlear implant stimulation. *Hear. Res.* 415, 108413. doi: 10.1016/j.heares.2021.108413

Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 98, 630.e16–644.e16. doi: 10.1016/j.neuron.2018.03.044

Kollmeier, B., Schädler, M. R., Warzybok, A., Meyer, B. T., and Brand, T. (2016). Sentence recognition prediction for hearing-impaired listeners in stationary and fluctuation noise with FADE. *Trends Hear.* 20, 1–17. doi: 10.1177/2331216516655795

Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (1998). Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents. *Hear Res.* 121, 11–28. doi: 10.1016/S0378-5955(98)00061-6

Landsberger, D. M., and Srinivasan, A. G. (2009). Virtual channel discrimination is improved by current focusing in cochlear implant recipients. *Hear. Res.* 254, 34–41. doi: 10.1016/j.heares.2009.04.007

Langner, F., Arenberg, J. G., Buchner, A., and Nogueira, W. (2021). Assessing the relationship between neural health measures and speech performance with simultaneous electric stimulation in cochlear implant listeners. *PLoS ONE* 16, e0261295. doi: 10.1371/journal.pone.0261295

Langner, F., Buchner, A., and Nogueira, W. (2020a). Evaluation of an adaptive dynamic compensation system in cochlear implant listeners. *Trends Hear.* 24. doi: 10.1177/2331216520970349

Langner, F., McKay, C. M., Büchner, A., and Nogueira, W. (2020b). Perception and prediction of loudness in sound coding strategies using simultaneous electric stimulation. *Hear. Res.* 398. doi: 10.1016/j.heares.2020.108091

Langner, F., Saoji, A. A., Büchner, A., and Nogueira, W. (2017). Adding simultaneous stimulating channels to reduce power consumption in cochlear implants. *Hear. Res.* 345, 96–107. doi: 10.1016/j.heares.2017.01.010

Litvak, L. M., Spahr, A. J., and Emadi, G. (2007a). Loudness growth observed under partially tripolar stimulation: model and data from cochlear implant listeners. *J. Acoust. Soc. Am.* 122, 967–981. doi: 10.1121/1.2749414

Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (2007b). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J. Acoust. Soc. Am.* 122, 982–991. doi: 10.1121/1.2749413

Malherbe, T. K., Hanekom, T., and Hanekom, J. J. (2016). Constructing a three-dimensional electrical model of a living cochlear implant user's cochlea. *Int. J. Num. Methods Biomed. Eng.* 32. doi: 10.1002/cnm.2751

McKay, C. M., and McDermott, H. J. (1998). Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals. *J. Acoust. Soc. Am.* 104, 1061–1074.

McKay, C. M., Remine, M. D., and McDermott, H. J. (2001). Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation. *J. Acoust. Soc. Am.* 110, 1514–1524. doi: 10.1121/1.1394222

Moberly, A. C., Lowenstein, J. H., and Nittrouer, S. (2016). Word recognition variability with cochlear implants: "perceptual attention" versus "auditory sensitivity". *Ear Hear.* 37, 14–26. doi: 10.1097/AUD.0000000000000204

Moore, B. C. (2003). Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants. *Otol. Neurotol.* 24, 243–254. doi: 10.1097/00129492-200303000-00019

Moore, B. C. J., Glasberg, B., and Schlueter, A. (2010). Detection of dead regions in the cochlea: relevance for combined electric and acoustic stimulation. *Adv. Otorhinolaryngol.* 67, 43–50. doi: 10.1159/000262595

Nadol, J. (1997). Patterns of neural degeneration in the human cochlea and auditory nerve: implications for cochlear implantation. *Otolaryngol. Head Neck Surg.* 117(3 Pt 1), 220-228. doi: 10.1016/s0194-5998(97)70178-5

Nadol, J. B. (1988). Comparative anatomy of the cochlea and auditory nerve in mammals. *Hear. Res.* 34,253–266.

Nelson, P. B., and Jin, S.-H. (2004). Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearig listeners. *J. Acoust. Soc. Am.* 115, 2286–2294. doi: 10.1121/1.1703538

Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (2003). Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J. Acoust. Soc. Am.* 113, 961–968. doi: 10.1121/1.1531983

Nogueira, W., and Ashida, G. (2018). *Development of a Parametric Model of the Electrically Stimulated Auditory Nerve*. Cham: Springer International Publishing.

Nogueira, W., Büchner, A., Lenarz, T., and Edler, B. (2005). A psychoacoustic "NofM"-type speech coding strategy for cochlear implants. *Eurasip J. Appl. Signal Process.* 2005, 3044–3059. doi: 10.1155/ASP.2005.3044

Nogueira, W., El Boghdady, N., Langner, F., Gaudrain, E., and Bakent, D. (2021). Effect of channel interaction on vocal cue perception in cochlear implant users. *Trends Hear.* 25. doi: 10.1177/23312165211030166

Nogueira, W., Harczos, T., Edler, B., Ostermann, J., and Büchner, A. (2007). "Automatic speech recognition with a cochlear implant front-end," in *International Speech Communication Association - 8th Annual Conference of the International Speech Communication Association, Interspeech 2007* (Antwerp), 2537–2540.

Nogueira, W., Litvak, L., Edler, B., Ostermann, J., and Büchner, A. (2009). Signal processing strategies for cochlear implants using current steering. *Eurasip J. Adv. Signal Process.* 2009. doi: 10.1155/2009/531213

Nogueira, W., Schurzig, D., Büchner, A., Penninger, R. T., and Würfel, W. (2016). Validation of a cochlear implant patient-specific model of the voltage distribution in a clinical setting. *Front. Bioeng. Biotechnol.* 4, 84. doi: 10.3389/fbioe.2016.00084

O'leary, S. J., Black, R. C., and Clark, G. M. (1985). Current distributions in the cat cochlea: a modelling and electrophysiological study. *Hear. Res.* 18, 273–281. doi: 10.1016/0378-5955(85)90044-9

Prado-Gutierrez, P., Fewster, L. M., Heasman, J. M., McKay, C. M., and Shepherd, R. K. (2006). Effect of interphase gap and pulse duration on electrically evoked potentials is correlated with auditory nerve survival. *Hear. Res.* 215, 47–55. doi: 10.1016/j.heares.2006.03.006

Ramekers, D., Versnel, H., Strahl, S. B., Smeets, E. M., Klis, S. F. L., and Grolman, W. (2014). Auditory-nerve responses to varied inter-phase gap and phase duration of the electric pulse stimulus as predictors for neuronal degeneration. *J. Assoc. Res. Otolaryngol.* 15, 187–202. doi: 10.1007/s10162-013-0440-x

Rattay, F. (1999). The basic mechanism for the electrical stimulation of the nervous system. *Neuroscience* 89, 335–346.

Rattay, F., Leao, R. N., and Felix, H. (2001). A model of the electrically excited human cochlear neuron. II. influence of the three-dimensional cochlear structure on neural excitability. *Hear. Res.* 153, 64–79. doi: 10.1016/S0378-5955(00)00257-4

Rattay, F., Potrusil, T., Wenger, C., Wise, A. K., Glueckert, R., and Schrott-Fischer, A. (2013). Impact of morphometry, myelinization and synaptic current strength on spike conduction in human and cat spiral ganglion neurons. *PLoS ONE* 8, e79256. doi: 10.1371/journal.pone.0079256

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. Biol. Sci.* 336, 367–373.

Santana, L. M. Q. D., Santos, R. M., Matos, L. N., and Macedo, H. T. (2018). Deep neural networks for acoustic modeling in the presence of noise. *IEEE Latin Am. Trans.* 16, 918–925. doi: 10.1109/TLA.2018.8358674

Santos, J. F., Consentino, S., Hazrati, O., Loizou, P. C., and Falk, T. H. (2013). Objective speech intelligibility measurement for cochlear implant users in complex listening environments. *Speech Commun.* 55, 815–824. doi: 10.1016/j.specom.2013.04.001

Saoji, A. A., Litvak, L., Emadi, G., and Spahr, A. J. (2005). "Spectral modulation transfer function in cochlear implant listeners," in *Conference on Implantable Auditory Prostheses* (Asilomar, CA).

Schädler, M. R., Warzybok, A., Ewert, S. D., and Kollmeier, B. (2016). A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. *J. Acoust. Soc. Am.* 139, 2708–2722. doi: 10.1121/1.4948772

Schädler, M. R., Warzybok, A., and Kollmeier, B. (2018). Objective prediction of hearing aid benefit across listener groups using machine learning: speech recognition performance with binaural noise-reduction algorithms. *Trends Hear.* 22, 1–21. doi: 10.1177/2331216518768954

Smit, J. E., Hanekom, T., and Hanekom, J. J. (2008). Predicting action potential characteristics of human auditory nerve fibres through modification of the Hodgkin-Huxley equations. *South Afr. J. Sci.* 104, 284–292.

Spoendlin, H., and Schrott, A. (1988). The spiral ganglion and the innervation of the human organ of corti. *Acta Otolaryngol.* 105, 403–410.

Stadler, S., and Leijon, A. (2009). Prediction of speech recognition in cochlear implant users by adapting auditory models to psychophysical data. *Eurasip J. Adv. Signal Process.* 2009. doi: 10.1155/2009/175243

Wagener, K., Kuehnel, V., and Kollmeier, B. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache I: design des oldenburger satztests. *Zeitschrift für Audiologie* 38, 4–15.

Wang, W., Pedretti, G., Milo, V., Carboni, R., Calderoni, A., Ramaswamy, N., et al. (2018). Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses. *Sci. Adv.* 4, 4752–4764. doi: 10.1126/sciadv.aat4752

Wouters, J., McDermott, H. J., and Francart, T. (2015). Sound coding in cochlear implants: from electric pulses to hearing. *IEEE Signal Process. Mag.* 32, 67–80. doi: 10.1109/MSP.2014.2371671

Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405

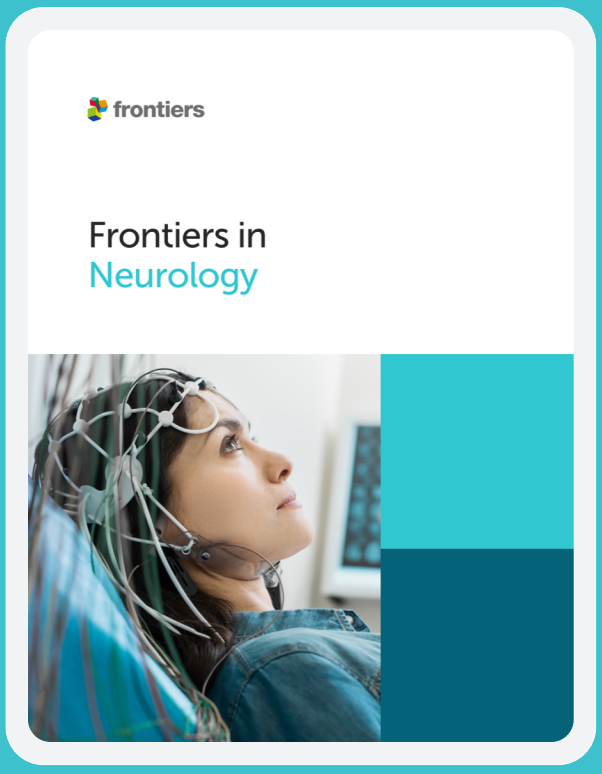
Frontiers in Neurology

Explores neurological illness to improve patient care

The third most-cited clinical neurology journal explores the diagnosis, causes, treatment, and public health aspects of neurological illnesses. Its ultimate aim is to inform improvements in patient care.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact