

At-Home Auditory Assessment Using Portable Automated Rapid Testing (PART) to Understand Self-Reported Hearing Difficulties

Trends in Hearing

Volume 29: 1–14

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23312165251397373

journals.sagepub.com/home/tia



E. Sebastian Lelo de Larrea-Mancera¹ , Tess K. Koerner², William J. Bologna³ , Sara Momtaz⁴ , Katherine N. Menon⁵, Audrey Carrillo¹, Eric C. Hoover⁵ , G. Christopher Stecker⁴ , Frederick J. Gallun² and Aaron R. Seitz¹

Abstract

Previous research has demonstrated that remote testing of suprathreshold auditory function using distributed technologies can produce results that closely match those obtained in laboratory settings with specialized, calibrated equipment. This work has facilitated the validation of various behavioral measures in remote settings that provide valuable insights into auditory function. In the current study, we sought to address whether a broad battery of auditory assessments could explain variance in self-report of hearing handicap. To address this, we used a portable psychophysics assessment tool along with an online recruitment tool (Prolific) to collect auditory task data from participants with ($n = 84$) and without ($n = 108$) self-reported hearing difficulty. Results indicate several measures of auditory processing differentiate participants with and without self-reported hearing difficulty. In addition, we report the factor structure of the test battery to clarify the underlying constructs and the extent to which they individually or jointly inform hearing function. Relationships between measures of auditory processing were found to be largely consistent with a hypothesized construct model that guided task selection. Overall, this study advances our understanding of the relationship between auditory and cognitive processing in those with and without subjective hearing difficulty. More broadly, these results indicate promise that these measures can be used in larger scale research studies in remote settings and have potential to contribute to telehealth approaches to better address people's hearing needs.

Keywords

telehealth, aging, digital health, prevention, auditory assessment, central auditory processing

Received: March 28, 2025; revised: October 25, 2025; accepted: October 28, 2025

Introduction

Although telehealth services and remote testing platforms have been of interest to audiologists and researchers for many years, clinical service limitations and social distancing restrictions imposed during the COVID-19 pandemic led to increased interest in these tools (Almufarrij et al., 2022; Peng et al., 2022). Recent technological advances in consumer grade electronics have also made it possible to use low-cost, personal-owned devices to create and present accurate and controlled audio signals, further advancing the development of new portable and remote auditory testing platforms (Almufarrij et al., 2022; Bright & Pallawela, 2016; Gallun et al., 2018; Irace et al., 2021). Such tools can expand the reach of hearing health services to rural settings with fewer available resources or a lack of options for hearing healthcare, and to patients lacking easy access to traditional clinical

services due to limited mobility or other health conditions.

¹Department of Psychology, Institute for Cognitive and Brain Health, Northeastern University, Boston, MA, USA

²Oregon Hearing Research Center, Oregon Health and Science University, Portland, OR, USA

³Department of Speech-Language Pathology & Audiology, Towson University, Towson, MD, USA

⁴Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, USA

⁵Department of Hearing and Speech Sciences, University of Maryland, Columbia, MD, USA

Corresponding author:

E. Sebastian Lelo de Larrea-Mancera, Department of Psychology, Institute for Cognitive and Brain Health, Northeastern University, Boston, MA 02115-5005, USA.

Email: dr.sebasllm@gmail.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Even mild untreated hearing loss is linked with higher rates of unemployment, social isolation, depression, and anxiety, as well as a higher risk of dementia (Kannan et al., 2024; Livingston et al., 2024). Increasing the ease of access to hearing screening and assessment through remote testing platforms may help address this important public health issue through earlier detection of hearing difficulties and provision of care to at-risk patient populations. Even basic research studies may benefit from remote testing platforms with larger and more diverse participant samples improving the generalizability of research findings to the larger population.

While a number of app- and web-based tools for remote auditory testing currently exist, information on the functionality, validity, and reliability of these tools is often not well reported (Almufarrij et al., 2022; Bright & Pallawela, 2016; Irace et al., 2021; Peng et al., 2022). Our lab has been developing and testing a freely available app called PART (Portable Automated Rapid Testing; https://ucrbraingamecenter.github.io/PART_Uilities/) that supports work in this area with the goal of making assessments that have traditionally been confined to laboratory settings more accessible to researchers and clinicians (Gallun, 2020). To date, a variety of audiological and psychophysical assessments have been created in PART, including suprathreshold measures of speech understanding in quiet and in noise, measures of spectral and temporal processing, measures of cognition, and self-report of experience (e.g., surveys). Recent studies that used PART have aimed to address the validity and reliability of these tests in different settings and with different groups of people. For instance, Lelo de Larrea-Mancera et al. (2020) validated a battery of PART assessments across a range of environmental noise conditions and confirmed consistency with previous laboratory tests (Cherri et al., 2024a; Lelo de Larrea-Mancera et al., 2020). Additionally, several studies confirmed that data collected via PART is in many cases reliable even when administered on participant-owned, uncalibrated devices (Lelo de Larrea-Mancera et al., 2022; Rink et al., 2022). Transferring assessments to languages other than English (Lelo de Larrea-Mancera et al., 2023a; 2023b; Padilla-Bustos et al., 2025) and successful tests of PART in clinical populations such as those living with Parkinson's disease, mild cognitive impairment, or dementia (Lelo de Larrea-Mancera et al., 2024a, 2025) further underscore the capabilities of this platform to investigate the needs of different participant samples. Recently, work has focused on making PART measurements more efficient through methods that collect more data in a shorter amount of time (Lelo de Larrea-Mancera et al., 2023c, 2024b). This research demonstrates functionality and usability of the PART testing platform and indicates a potential opportunity to address the needs of research participants and clinic patients.

While PART allows for accessible and efficient remote auditory testing, to date research with the PART application has mostly addressed participants with normal or near-normal hearing thresholds. The primary aim of the current study was to

establish the validity of a remote testing procedure using PART, identifying measures that can be useful to discriminate between individuals with and without self-reported hearing difficulties. This aim is specific to further establishing PART as a clinically relevant research tool. As a secondary aim, we sought to better understand the factor structure of the testing battery and the extent to which it could inform specific constructs that could individually or in combination help us understand aspects of self-reported hearing difficulties. This aim addresses an unresolved complexity in auditory clinical science (not specific to PART) that concerns the inadequacy of any one measure to fully establish whether someone will experience hearing difficulty (Humes, 2021; Humes et al., 2012). Overall, results of this work speak directly to the feasibility of conducting large-scale studies remotely through an online platform, with a structured hierarchy of auditory processing measures, accommodating individuals with varying ages and hearing sensitivities.

Methods

Participants

One hundred and ninety-two (192) participants were recruited through the online platform Prolific (www.prolific.com). They all provided signed informed consent as approved by the University of California, Riverside Human Research Review Board. Participants were preselected by Prolific to balance the sample based on self-report of hearing status following the prompt: “Do you have any hearing loss or hearing difficulties?”. The accuracy of this preselection question was cross-checked by an additional question specific to the current study: “Do you have hearing loss, reduced hearing, or hearing difficulties (such as difficulty understanding when multiple people are talking or when there is background noise)?”. The Kendall correlation between these two dichotomous variables was of $\tau = .58$ which indicated the responses to these two single questions were not the same. To achieve a more valid approach to classify the sample into hearing difficulty (HD) and no difficulty (ND) groups, final group assignment was based on a cutoff score on the Hearing Handicap Inventory for the Elderly–Screening Version (HHIE-S) of ≥ 8 as suggested by ASHA guidelines (1989). A comparison of the final group assignment to the Prolific preselection and to another single question assessment of hearing status collected for this study is provided in the Supplemental materials (Table S1). Demographic information divided by final group assignment is presented in Table 1. All participants provided informed consent and received \$12 an hour for their participation.

Construct Framework

We used a collection of measures thought to be sensitive to different stages along the hierarchy of auditory processing (see Figure 1). The selection of measures was based upon previous models in cognitive and auditory science (Bronkhorst,

Table 1. Demographic Information Split by Hearing Difficulty Group.

	ND	HD
N	108	84
Gender, female	N = 57 (52.7%)	N = 36 (42.8%)
Age, years	38.5 ± 12.07	45.2 ± 13.63
HHIE-S	1.75 ± 2.09	18.95 ± 7.58
Education (grade)		
High school	N = 22 (20.3%)	N = 28 (33.3%)
Junior college	N = 18 (16.6%)	N = 9 (10.7%)
Bachelor's degree	N = 52 (48.1%)	N = 31 (36.9%)
Professional degree	N = 2 (1.8%)	N = 2 (2.3%)
Graduate school	N = 12 (11.1%)	N = 13 (15.4%)
Hispanic, yes	N = 14 (13%)	N = 10 (11.9%)
Ethnicity		
White	N = 70 (64.8%)	N = 61 (72.6%)
Asian	N = 12 (11.1%)	N = 3 (3.5%)
Black	N = 19 (17.5%)	N = 13 (15.4%)
Native American	N = 3 (2.7%)	N = 2 (2.3%)
Other	N = 3 (2.7%)	N = 5 (6%)

HD=hearing difficulty; HHIE-S=Hearing Handicap Inventory for the Elderly–Screening Version; ND=no difficulty.

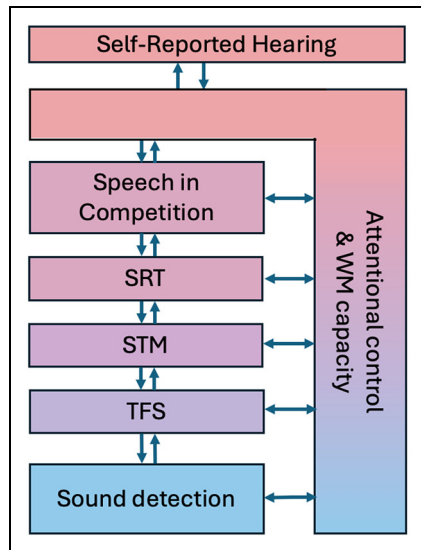


Figure 1. Construct Framework Organizing Instrument Selection and Grouping of Assessments Into Constructs. The Color Indicates Graded Relationships Between Constructs Rather Than Strict Independence. SRT=Speech Reception Threshold; STM=Spectrotemporal Modulation; TFS=Temporal Fine Structure; WM=Working Memory.

2015; Cowan, 1999; Stecker & Gallun, 2012; Vetter et al., 2024). While this framework is not intended as a comprehensive view of all human auditory activity, it is meant to elucidate potential constructs that can inform different dimensions of hearing (Cherri et al., 2024b; Lelo de Larrea-Mancera et al., 2020, 2022). We place at the base of this hierarchy the basic ability to detect sound using a pure tone detection task.

Building upon this, measures of central auditory processing are included, such as temporal fine structure (TFS; guides sound spatialization), or spectrotemporal modulation (STM; guides aspects important for speech intelligibility) (Gallun & Best, 2020; Humes et al., 2012; Stecker & Gallun, 2012). At a higher level we consider tasks that involve speech such as the speech reception threshold (SRT) and the ability to discriminate speech in competition (SiC) with other sources of sound. While all auditory tasks require attentional control and working memory to guide listeners behavior through the task instructions, the speech related tasks are known to rely significantly on these cognitive constructs (Cowan, 1999; Gallun & Jakien, 2019). Our primary hypothesis is that all of these tests can contribute to account for self-reported hearing difficulties. We place self-reported hearing at the top of the hierarchy as it involves a meta-cognitive subjective reflection or an integrated narrative episodic recollection guided by attention and working memory of past perceptual performance. In the following section, we provide more detail on each task utilized to understand these different constructs.

Instruments

The measures employed are detailed below. In all cases, the testing was performed on personally owned devices, following the methods of Lelo de Larrea-Mancera et al. (2022), who demonstrated high test–retest reliability as well as good correspondence between results with calibrated equipment in the laboratory and with personally owned equipment at-home. Because of the lack of precise level measurement in remote conditions with uncalibrated equipment, sound levels are reported as nominal dB throughout (as reported by PART).

Self-Reported Hearing Handicap. Self-reported hearing difficulty was evaluated using the HHIE-S, developed by Newman et al. (1990). This inventory includes 10 questions addressing the social and emotional consequences of hearing-related challenges, with participants responding on a three-point scale: “no” (0 points), “sometimes” (2 points), or “yes” (4 points). Total HHIE-S scores range from 0 to 40, where higher scores indicate a greater self-perceived hearing handicap. According to ASHA guidelines (1989), a cutoff score of 8 or above suggests the presence of hearing handicap.

Adaptive Tracking. The adaptive scan (AS) algorithm that combines progressive and adaptive methods of psychophysical testing (Lelo de Larrea-Mancera et al., 2023c, 2024b) was used to vary the magnitude of an adaptive parameter (e.g., dB) specified for each of the auditory processing tasks described below except pure tone detection and working memory tasks. Every task included two consecutive runs of AS with three scans of nine steps each. Each scan ended after the ninth trial or after three errors. Details about the parameters used in the first

scan are provided below for each task. Placement of the second and third scans was dependent (adaptive) on participant performance and the cumulative estimate of threshold per scan. Threshold was calculated with a simple heuristic $(hits - nTrials * guess\ rate) / (1 - guess\ rate)$ shown to be equivalent to more complex Bayesian threshold estimation procedures (Lelo de Larrea-Mancera et al., 2023c, 2024b). The result of the formula indicates the step number that serves as a threshold estimate in a vector of trials ordered according to the magnitude of the adaptive parameter. Every completed scan was divided into three regions, if the scan threshold fell in either the higher or lower regions, the next scan would slide in that direction by half the range. Otherwise, if the threshold fell in the center region, the next scan would be centered at place and its step-sizes would be halved as described in Lelo de Larrea-Mancera et al. (2023c, 2024b). The average threshold calculated from each of two runs on each test constituted the main outcome at the individual measure level.

Sound Detection. *Pure-tones* were presented to the left and the right ear separately at 1000 Hz, and diotically at frequencies of 500, 2000, and 4000 Hz to estimate detection thresholds, representing an abbreviated digital version of the standard pure-tone audiogram. This measure is intended to provide information about peripheral hearing, not to provide diagnostic measures for hearing loss as the clinical audiogram would. This is due to testing standards (ANSI, 2004) not being easily achieved with uncalibrated participant-owned devices, which can lead to increased variance overall with the greatest difficulty being the high risk of inaccurately elevated thresholds due to external factors such as environmental noise (see Almufarrij et al., 2022).

Participants were presented with a series of three-to-four 100 ms tones initially presented at 70 dB. The question “did you hear the tones” was displayed on screen along with a green button with a check-mark that indicated “yes” and a red button with an “X” on it that indicated “no.” Following three consecutive “yes” responses, the tone level was reduced in steps: first by 20 dB, then by 10 dB down to 10 dB, and finally by 5 dB, using an adapted Hughson-Westlake algorithm (Carhart & Jerger, 1959). If three consecutive “no” responses occurred or 0 dB was reached, the last correctly detected level was recorded as the threshold. Each frequency-specific section of the test lasted about 2 min, totaling approximately 10 min. The primary outcomes were the estimated thresholds at each frequency (average of left and right 1000 Hz). The peripheral hearing construct-level outcome was taken the pure-tone average (PTA) detection threshold across all four frequencies.

Temporal Fine Structure. Sensitivity to the fine structure of sound was assessed with sinusoidal tones set up in three different tasks: (1) the detection of a silent gap placed between two tone bursts; (2) the detection of frequency modulations (FM) in two conditions described below. These assessments

employed a 4-interval 2-cue 2-alternative forced-choice (2AFC henceforth) paradigm where the first and last intervals are presented as cues that have the standard stimuli without the target. It took participants around 12 min to complete both AS runs on all three TFS assessments (described in detail below).

The Gap detection task adapted from Gallun et al. (2014) and Hoover et al. (2015, 2019) was used to measure auditory temporal sensitivity. While both TFS and temporal envelope cues are involved in this task, here it is considered as part of the TFS construct. Each interval contained two diotically presented tone bursts (2 kHz) at 80 dB, and target intervals contained a silent gap placed between the two tone-bursts. We adjusted the duration of the gap using AS (described above). The first scan started at 128 ms and halved its value on an exponential scale until reaching .5 ms on the last step of the first scan unless interrupted.

FM detection tasks (2) adapted from Grose and Mamo (2010, 2012) were used to measure phase-related temporal sensitivity. Target stimuli involved detecting a sinusoidal FM at 2 Hz. Every interval presented pure-tones with center frequencies randomized between 460 and 550 Hz presented at 75 dB. Target intervals contained FM that was either presented diotically to produce a monaural cue (monaural frequency modulation, MFM) or presented dichotically with an antiphasic relationship between the ears to produce a binaural cue (binaural frequency modulation, BFM). The range of the 2 Hz FM was adapted using AS. The MFM started at 256 Hz and halved its value on an exponential scale until reaching 1 Hz unless interrupted. The BFM task started at 64 Hz and halved its value on an exponential scale until reaching .25 Hz unless interrupted.

Spectrotemporal Modulation Sensitivity. *Spectral, temporal, and STM detection tasks (3)* were used to measure sensitivity to spectral, temporal, and spectrotemporal sinusoidal amplitude modulations in broad-band noise through three different tasks that also employed a 2AFC paradigm: (1) the detection of temporal modulation (TM) at 4 Hz; (2) the detection of spectral modulation (SM) at 2 cycles per octave; and (3) the detection of the combined STM at 4 Hz and 2 cycles per octave (see similar stimuli in Bernstein et al., 2013; Lelo de Larrea-Mancera et al., 2020; Sabin et al., 2012). The up/down direction of the modulation was generated randomly for each interval. Stimuli were presented at 65 dB for 500 ms. The adaptive parameter was modulation depth measured mid-to-peak on a logarithmic amplitude scale (dB) as described in Stavropoulos et al. (2021) starting at 6 dB and decreasing .7 dB until reaching .4 dB on the last step of the first scan unless interrupted. These assessments took around 12 min to complete including both AS runs on all three STM assessments (described in detail below).

Speech Reception Threshold. *The Coordinate Response Measure corpus* (Bolia et al., 2000) was used to calculate

SRT. Participants listened to sentences spoken by a single speaker following the structure: “target *code name*, go to *color*, *number*, now.” (e.g., “Ready Charly, go to red one now”). Participants were instructed to identify the color-number combination they heard on a 4×8 grid displayed on the screen with four possible colors (white, green, blue, and red) and eight possible numbers (1–8). The signal level adapted starting at 65 dB and progressed in steps of 5 to 25 dB in the first scan unless interrupted.

Speech in Competition. *The Spatial release from masking task* was used to evaluate the ability to understand speech under competition with highly similar speech maskers (Gallun et al., 2013; Lelo de Larrea-Mancera et al., 2020, 2022; Marrone et al., 2008) and could be considered a measure of auditory selective attention. Masker talkers spoke code-names other than Charly which remained exclusive for target sentences. The target sentence was presented at 65 dB in all trials. Masker sentences were presented either colocated in simulated space or separated by 45 degrees to each side. The maskers in the colocated condition started at 13 dB target-to-masker ratio (TMR) and decreased 2 dB on every step until reaching –5 dB TMR in the first scan unless interrupted. The maskers in the separated condition started at 8 dB TMR and decreased 2 dB on every step until reaching –10 dB TMR unless interrupted. It took about 6 min to complete both runs of both conditions.

The Dichotic sentence identification test, adapted from Fifer et al. (1983), was used to assess the ability to understand speech in dichotic conditions where two speakers need to be attended simultaneously, and could be considered a measure of auditory divided attention. In each trial, two sentences were presented simultaneously at 50 dB, with a different sentence presented to each ear. Sentence pairs were drawn pseudorandomly from a closed set of six seven-word, low-predictability sentences from the Synthetic Sentence Identification test (Speaks & Jerger, 1965), ensuring no repetition within pairs. Participants selected the two sentences they heard by choosing from six written sentence options on a response list. Five trials were presented, and if participants did not achieve 100% accuracy, five additional trials were given. The primary outcome was the percentage of correct responses, with the test taking about 2 min to complete.

Working Memory Capacity and Attentional Control. *The auditory visual divided attention task (AVDAT)* was used to measure auditory and visual attentional control and working memory capacity. This task has been previously shown to account for significant variance in auditory tests (Gallun & Jakien, 2019). Participants started with a cued visual (V) span task where they were briefly presented with an image of an eye followed by a visual sequence of letters and were asked to recall the letters. Then, participants completed a cued auditory span (A) task where they were briefly presented with an image of an ear followed by an auditory sequence of

numbers. In the final block, both auditory and visual stimuli were presented simultaneously, and participants were either cued in advance (Cued_AV) or at the end of the sequence (Divided_AV) whether to respond to the auditory or visual stream. Each section (A, V, Cued_AV, Divided_AV) started at a sequence length of 3 stimuli and advanced to sequence length of 5 stimuli irrespective of performance. Then sequence length adapted to a maximum of 9 stimuli depending on each participant’s performance. Each sequence length contained three trials of each type. A span score was calculated for each condition which was defined as the last sequence length at which there was at least one correct trial. Modality-specific attentional control scores were calculated from the span scores of the first two blocks (single-sensory conditions) minus the span scores obtained in cued conditions of the third block. Higher scores indicate dual streams of sensory information worsen the capacity to store information due to the increased attentional control required to ignore one stream of information. Modality-specific WM capacity scores were taken from the average span score of either visual or auditory target trials in the third block (cued and uncued).

Composite Scores. The primary outcomes at the construct-level of TFS, STM, SRT, SiC, attentional control, and WM capacity were computed by normalizing each of their constituting measures to a z-score (considering the entire sample including both HD and ND groups) and averaging across tasks for each individual. The pure tone tests (four frequencies) were first averaged into a pure tone average and then converted to a z-score (PTA). Visual and auditory scores were averaged together for each of the attentional control and WM capacity constructs.

Procedure

Participants were recruited through Prolific, which provided a pseudorandom preselection of participants into balanced groups of self-reported HD and ND including balanced demographics of sex, age, and education. After initial contact through email, participants were assigned dates for three testing sessions to be supervised through Zoom by a research assistant. In the first session informed consent and demographic forms were completed. Testing proceeded in a between-participant counterbalanced manner across four possible orders of tasks. Participants completed all tests using their own equipment (i.e., headphones and mobile device or tablet) across three remote sessions within 14 days of each other. Participants were instructed to complete their sessions in a quiet area with no distractions, to turn the volume of their devices to the maximum value, and to ensure proper headphone placement. Data was uploaded to Amazon Web Services HIPAA compliant servers. Data were collected as part of a larger study that also included: (1) usability questionnaires about each task; (2) an additional run of each task with alternative stimulus delivery algorithms (up-down staircases and

Table 2. Descriptive Statistics for All Measures Collected for This Study.

Measure	Group	N	Units	Mean	SD	Min	Max
HHIE-S	ND	102	Score	1.74	2.09	0	6
	HD	84		18.95	7.58	8	38
PT .5kHz	ND	104	dB	28.07	11.58	-5	65
	HD	83		33.37	17.48	-15	85
PTL 1kHz	ND	104	dB	26.34	14.28	-20	80
	HD	83		34.21	18.53	-15	80
PTR 1kHz	ND	104	dB	26.15	20.20	-20	80
	HD	83		32.59	26.15	-20	85
PT 2kHz	ND	104	dB	22.01	12.07	5	55
	HD	83		30.06	19.66	-15	80
PT 4kHz	ND	104	dB	21.01	13.04	-5	80
	HD	83		30.78	23.94	-20	85
Gap	ND	102	log2(ms)	0.31	2.07	-3.25	7.25
	HD	81		0.99	2.29	-2.75	6.75
BFM	ND	103	log2(Hz)	-0.86	1.63	-4.50	5.50
	HD	81		0.03	2.17	-3.50	6.50
MFM	ND	103	log2(Hz)	2.20	1.09	-0.25	7.50
	HD	81		2.62	2.21	0	6.50
TM	ND	102	(M) dB	1.63	1.17	0.40	7.80
	HD	82		1.77	1.36	0.25	9.15
SM	ND	102	(M) dB	1.41	0.86	0.10	6.15
	HD	81		2.12	1.80	0.60	10.80
STM	ND	102	(M) dB	1.29	1.06	0.25	7.90
	HD	81		1.77	1.51	0.25	7.85
SRT	ND	102	dB	40.63	7.69	27.5	65.00
	HD	82		45.67	12.17	27.5	87.50
Col	ND	102	TMR dB	3.82	1.61	1.50	12
	HD	82		5.51	5.38	1.00	31
Sep	ND	102	TMR dB	-0.12	4.58	-10.00	18
	HD	82		2.31	5.71	-7.00	20
DSI	ND	103	%	86.57	20.28	0	100
	HD	83		72.60	31.14	0	100
Aud. WM	ND	104	span	6.20	1.22	2.5	9
	HD	82		5.60	1.68	0	9
Vis. WM	ND	104	span	5.47	1.63	0	9
	HD	82		4.89	1.69	0	8
Aud. Att.	ND	104	score	0.64	1.22	-2	6
	HD	82		0.51	1.27	-4	4
Vis. Att	ND	104	score	0.68	1.56	-6	6
	HD	82		0.59	1.53	-7	4

Att=attention; Aud=auditory; BFM=binaural frequency modulation; Col=colocated; DSI=dichotic sentence identification; HHIE-S=Hearing Handicap Inventory for the Elderly–Screening Version; kHz=kilo-Hertz; Max=maximum; MFM=monaural frequency modulation; Min=minimum; ms=milliseconds; n=sample size; PT=pure tone; PTL=pure tone left (monaural); PTR=pure tone right (monaural); Sep=separated; SRT=speech reception threshold; SM=spectral modulation; STM=spectrotemporal modulation; TM=temporal modulation; TMR=target-to-masker ratio; Vis=visual; WM=working memory.

progressive tracking) aimed to further validate the adaptive scan algorithm used here; and (3) Letter + Number WM task to evaluate concurrent validity for the AVDAT task. Evaluation of these additional data is beyond the scope of this study and will be reported elsewhere.

Statistical Analysis

Descriptive statistics for each individual measure (Table 2) are presented and discussed in reference to consistency (mean alignment) with previous studies. Further, to address

the primary aim we evaluated the ability of each construct to discriminate between the self-reported HD and ND groups using robust multiple regression models that tested for the effect of participant group (HD vs. ND) and age on each construct (Table 3). We used a robust regression approach with a Huber weighting function (Huber, 1981) to account for widespread violations of the normality assumption and presence of outlying values in the variables under study. This method downweights the influence of extreme values without reducing them to zero. We deem this adequate as extreme values may be testing failures or real indices of poor perceptual

Table 3. Robust Beta Coefficients for HD and the Covariate of Age on Each Multiple Regression Conducted for Each Construct (z-Scores).

	PTA	TFS	STM	SRT	SiC	Att.	WM	Age (years)
HD	$\beta = \mathbf{0.26}$ $t = \mathbf{2.78}$ $p = \mathbf{.005}$	$\beta = \mathbf{0.25}$ $t = \mathbf{2.86}$ $p = \mathbf{.004}$	$\beta = \mathbf{0.202}$ $t = \mathbf{2.88}$ $p = \mathbf{.004}$	$\beta = 0.24$ $t = 2.01$ $p = .045$	$\beta = 0.108$ $t = 1.38$ $p = .16$	$\beta = -0.11$ $t = -1.72$ $p = .086$	$\beta = -0.01$ $t = -0.17$ $p = .86$	$\beta = \mathbf{6.78}$ $t = \mathbf{3.43}$ $p < \mathbf{.001}$
Age (z)	$\beta = \mathbf{0.29}$ $t = \mathbf{6.28}$ $p < \mathbf{.001}$	$\beta = 0.03$ $t = 0.70$ $p = .48$	$\beta = \mathbf{0.17}$ $t = \mathbf{4.87}$ $p < \mathbf{.001}$	$\beta = \mathbf{0.26}$ $t = \mathbf{4.41}$ $p < \mathbf{.001}$	$\beta = \mathbf{0.13}$ $t = \mathbf{3.32}$ $p = \mathbf{.001}$	$\beta = \mathbf{0.47}$ $t = \mathbf{13.69}$ $p < \mathbf{.001}$	$\beta = -0.04$ $t = -0.82$ $p = .41$	-

Significant effects are indicated in bold and indicate group differences or age effects.

Att=attentional control; HHIE-S=Hearing Handicap Inventory for the Elderly–Screening Version; PTA=pure-tone average; SiC=speech in competition; SRT=speech reception threshold; STM=spectrotemporal modulation; TFS=temporal fine structure; WM=working memory capacity.

performance. Significance on this family-wise set of eight comparisons was controlled for false discovery rate (Benjamini & Hochberg, 1995) as these analyses aimed to confirm the utility of each construct to discriminate self-reported HD. Similar analyses for each individual measure are also provided in the supplemental materials without correction to facilitate exploration and avoid inflating type II error (Supplemental Table S2).

To address the secondary aim of understanding the test battery at the construct level, we examined the bivariate relationship between measures (Table 4) (raw correlations for each individual measure appear in Table S3 of the Supplemental materials). False discovery rate was also considered for this family of 36 correlations. Further, a principal component analysis (PCA) was conducted to evaluate the consistency of our theory-driven dimensionality reduction into constructs with a data-driven approach. All individual measures described above were included in the PCA. The number of components was selected based on a criterion of eigenvalue > 1. To increase interpretability an oblique promax rotation method based on the correlation matrix was employed to allow components to be correlated. This analysis allowed for an evaluation of whether the rotated components and their variable loadings (Table 5) were similar to the construct framework originally proposed (Figure 1).

Finally, to address whether the principled characterization of the test battery at the construct level could discriminate HD, logistic regression was used including participant age and performance on each construct to predict participant group (HD vs. ND). The accuracy, sensitivity, and specificity of the model to discriminate HD as well as the area under the curve (AUC) of a receiver operating characteristic (ROC) curve is also reported. Accuracy, sensitivity, specificity and AUC for the individual constructs is provided in Table S4 of the Supplemental materials.

Results

Auditory processing measures were largely overlapping with the reports of Lelo de Larrea-Mancera et al. (2022), who collected data across both laboratory and remote conditions (see

Table 2). Importantly, the number of outlying cases (> 2 SD) was under 8.5% for every test, which is similar to the reports in Lelo de Larrea-Mancera et al. (2020) for young people without hearing difficulties in laboratory conditions. However, there were a few participants whose pure tone detection and SRT scores were worse than 45 dB (20 participants), raising concern about the audibility of the test stimuli for those participants. Figure S1 in the Supplemental materials shows the distribution of scores for each measure highlighting performance of people with possible audibility issues. The scores of these participants in other measures, importantly on the SiC assessments, is evenly spread across performance levels indicating some of these participants were able to perform well in other auditory and cognitive tasks.

Bivariate Discriminability of Self-Reported Hearing Difficulty

Here the ability of each construct to discriminate between self-reported HD and ND groups is illustrated in Figure 2. Because participant age was distributed unevenly across the groups (mean difference of 6.69 years), and because it is widely known to impact peripheral and central audition (Humes et al., 2012), the effect of age was controlled for in group comparisons. Table 3 shows a summary of the robust beta coefficients for HD (average group differences) and the covariate of age (change in the construct for every SD change in age) on each construct. All lower-level auditory composite measures (PTA, TFS, and STM) were able to discriminate between groups with small-to-medium size decrements for the HD group, significant even after a correction for false-discovery rate (*corrected alpha* = .0051). In contrast, measures of SRT, SiC, attentional control, and WM capacity failed to discriminate between groups after correction for multiple comparisons, although SRT and attentional control showed a trending effect. Analyses were conducted on standardized scores in every construct and the age variable. All constructs except for TFS and WM capacity showed a significant effect of age. Group comparisons for all individual measures can be found in the Supplemental materials (Table S2).

Table 4. Spearman Correlations (*r* and in Parenthesis Below, *p*) Between Construct Level Variables.

Variable	HHIE-S	PTA	TFS	STM	SRT	SiC	WM	Att
PTA	0.343 (< 0.001)	-	-	-	-	-	-	-
TFS	0.218 (0.003)	0.311 (< 0.001)	-	-	-	-	-	-
STM	0.373 (< 0.001)	0.708 (< 0.001)	0.420 (< 0.001)	-	-	-	-	-
SRT	0.235 (< 0.001)	0.297 (< 0.001)	0.392 (< 0.001)	0.472 (< 0.001)	-	-	-	-
SiC	0.202 (0.006)	0.451 (< 0.001)	0.393 (< 0.001)	0.476 (< 0.001)	0.475 (< 0.001)	-	-	-
WM	-0.101 (0.173)	-0.107 (0.146)	-0.015 (0.835)	-0.032 (0.662)	-0.056 (0.453)	-0.352 (< 0.001)	-	-
Att	0.151 (0.039)	0.166 (0.023)	-0.066 (0.369)	0.057 (0.446)	0.015 (0.840)	0.269 (< 0.001)	-0.321 (< 0.001)	-
Age	0.272 (< 0.001)	0.401 (< 0.001)	0.063 (0.396)	0.309 (< 0.001)	0.188 (0.011)	0.244 (< 0.001)	-0.051 (0.492)	0.686 (< 0.001)

Statistical significance is indicated in bold considering a correction for false-discovery rate (corrected $\alpha < .023$).

Att=attentional control; HHIE-S=Hearing Handicap Inventory for the Elderly–Screening Version; PTA=pure-tone average; SiC=speech in competition; SRT=speech reception threshold; STM=spectrotemporal modulation; TFS=temporal fine structure; WM=working memory capacity.

Bivariate and Multivariate Relationships Between Measures

Bivariate Relations at the Construct Level. To gain a first intuition about the multivariate structure of the collected

Table 5. Rotated Components (Oblique Promax) Obtained From the Individual Measures Under Study With Eigenvalues > 1.

Variable	RC1	RC2	RC3	RC4
HHIE-S				
PT500Hz		0.900		
PT1kL		0.762		
PT1kR		0.841		
PT2kHz		0.812		
PT4kHz		0.614		
Gap			0.672	
BFM			0.650	
MFM			0.793	
TM	0.606			
SM	0.883			
STM	0.593			
SRT	0.629			
Col	0.959			
Sep	0.777			
DSI	0.574			
Aud-Att	-0.514		0.686	
Vis-Att			0.525	
Aud-WM			0.706	
Vis-WM			0.744	

Att=attention; Aud=auditory; BFM=binaural frequency modulation; Col=colocated; DSI=dichotic sentence identification; HHIE-S=Hearing Handicap Inventory for the Elderly–Screening Version; MFM=monaural frequency modulation; PT=pure tone; Sep=separated; SRT=speech reception threshold; SM=spectral modulation; STM=spectrotemporal modulation; TM=temporal modulation; Vis=visual; WM=working memory.

measures, Spearman correlations between constructs were conducted and are shown in Table 4. Significant correlations were corrected for false-discovery rate (*corrected alpha* < .023). Auditory performance constructs correlated moderately with one another from sound detection to SiC ($r = 0.29$ – 0.708) such that better scores in one construct relate to better scores in the others. The cognitive measures of attentional control and WM capacity were also correlated ($r = -0.32$). Of the auditory constructs, only SiC correlated with cognitive measures of WM capacity and attentional control ($r = 0.26$ and -0.35 , respectively). The correlations to WM are negative due to the scoring metrics, as better WM capacity is represented by a higher value while better performance on the rest of the measures is represented by lower values. Results showed that all the observed statistically significant correlations followed the expected direction. The relationship between SiC and other auditory tasks was between $r = 0.39$ and 0.47 , as well as an $r = .202$ for the relationship to self-report (HHIE-S). The HHIE-S scores were correlated to all auditory ($r = 0.202$ – 0.37) but not cognitive tasks. STM showed the highest correlations to other auditory constructs ($r = 0.37$ – 0.702) and TFS the lowest ($r = 0.21$ – 0.42). The age of the participants was also correlated significantly with most constructs ($r = 0.11$ – 0.68) except for TFS and WM capacity. The Spearman correlations between all individual measures are presented in Table S3 of the Supplemental materials.

Multivariate Relations and Construct Framework Alignment. A PCA using all individual measures was used to further understand the multivariate relationships between measures and provide a data-driven alternative to theory-driven composite scores. Table 5 shows the rotated component loadings > .5 on each of the four resulting components with eigenvalues >

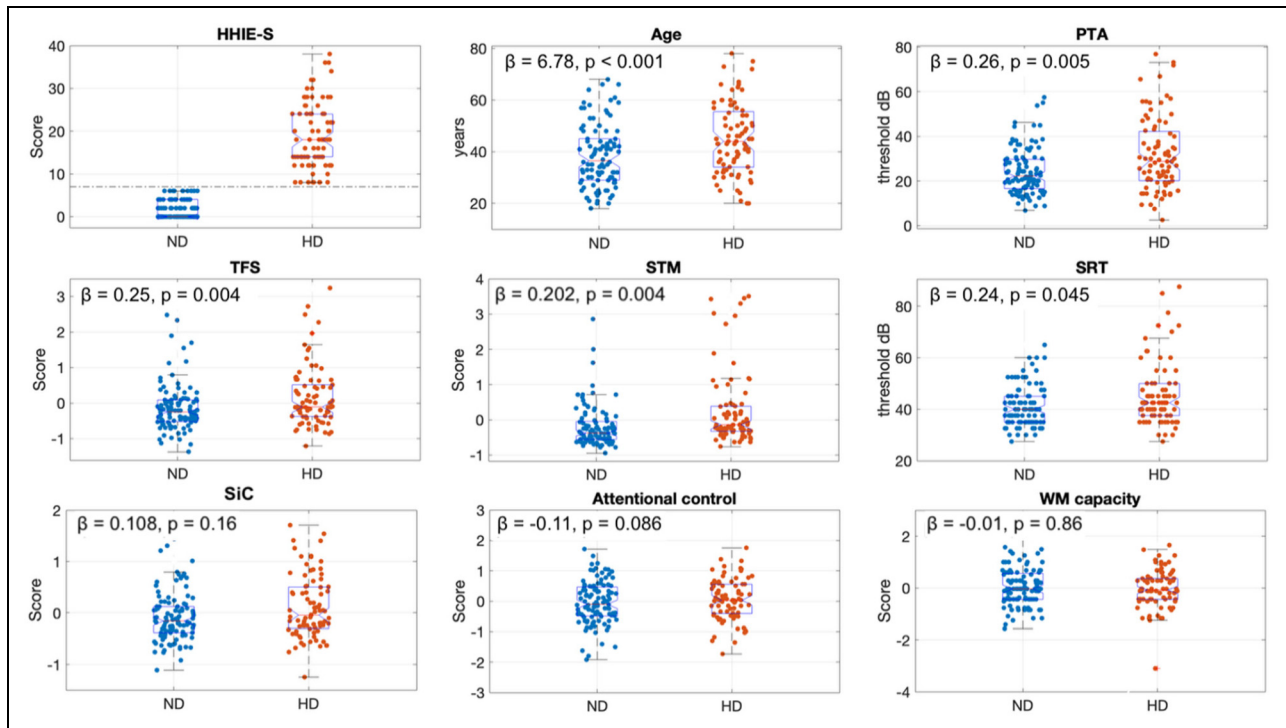


Figure 2. Construct Level Scores for Both the Self-Reported HD and the ND Groups. The First Panel Indicates the Cutoff Value for Group Assignment According to HHIE-S Scores With a Dotted Line. Robust Beta Coefficients of the z-Scores are Provided for Each Construct. The False-Discovery Rate Corrected Alpha Level is 0.0051. HD= Hearing Difficulty; HHIE-S= Hearing Handicap Inventory for the Elderly–Screening Version; ND= No Difficulty.

1. Component 1 consisted of the higher order central auditory processing (CAP) tasks of STM, speech reception with and without competition, and auditory attention. The rest of the (lower-order) CAP tasks that constitute temporal fine structure sensitivity loaded into component 4. Component 2 was constituted by pure-tone detection tasks, and component 3 by cognitive measures of auditory and visual attention and

WM. The self-reported measure (HHIE-S) did not load $>.5$ into any of the selected components ($RC1 = .44$; $RC2 = .26$; $RC3 = -.01$; $RC4 = -.16$), which suggests that it is not correlated with the other variables included and does not fit well into the other constructs under study. In sum, this unsupervised method of dimensionality reduction was able to account for .62 of the total shared variance in the dataset ($RC1 = .239$; $RC2 = .186$; $RC3 = .102$; $RC4 = .102$). This data-driven approach grouped measures in a way that largely confirmed the construct framework under study (see Figure 1).

Table 6. Beta Coefficients for a Multivariate Logistic Model Predicting HD From Each Construct Under Study and the Covariate of Age (z-Scores).

Variable	β Coeff	SE	t Statistic	p Value	VIF
Intercept	−0.177	0.166	−1.067	.286	–
PTA	0.593	0.326	1.821	.069	2.30
TFS	0.389	0.288	1.352	.176	1.37
STM	−0.219	0.395	−0.554	.58	2.53
SRT	0.362	0.352	1.031	.303	1.55
SiC	−0.132	0.439	−0.3	.764	1.95
WM	−0.298	0.263	−1.133	.257	1.33
Att	−0.554	0.415	−1.336	.182	2.57
Age	0.608	0.27	2.257	.024	2.58

Att=attentional control; Coeff=coefficient; HHIE-S=Hearing Handicap Inventory for the Elderly–Screening Version; PTA=pure-tone average; SE=standard error; SiC=speech in competition; SRT=speech reception threshold; STM=spectrotemporal modulation; TFS=temporal fine structure; VIF=variance inflation factor; WM=working memory capacity.

Multivariate Discriminability of Self-Reported Hearing Difficulty

Table 6 summarizes the beta coefficients and associated p -values from the multiple logistic regression model designed to assess the ability of the full battery of tests to discriminate HD. The model's ability to discriminate HD in terms of accuracy, sensitivity, specificity, and ROC AUC was also evaluated. Using a 0.5 threshold on the individual predicted probabilities from the model for binary classification of HD, the model was able to correctly identify 87 participants as ND and 36 as HD (accuracy = 0.64), while also missing 48 cases of HD (sensitivity = 0.42) and falsely identifying 21 cases of ND as HD (specificity = 0.805). Overall, the model showed a ROC AUC = 0.71 and explained a small proportion

of HD variance (adjusted pseudo- $R^2=0.11$). In sum, the multivariate approach does no better than the individual constructs in predicting HD classification. Accuracy, sensitivity, specificity and ROC AUC for each construct is reported in the Supplemental materials (Table S4).

We also assessed multicollinearity in the predictors by computing variance inflation factors (VIFs) from the diagonal of the inverse correlation matrix, as described by Belsley et al. (1980). All VIFs were below 2.58, values below commonly used thresholds of 5 or 10 for problematic multicollinearity (for discussion see Vatcheva et al., 2016). Specific values are included in Table 6.

Discussion

In this study data from a principled set of auditory and cognitive processing measures was collected from an online pool of participants (Prolific) in which half self-identified as having difficulty hearing (HD). The auditory measures used in previous studies on PART (Gap, MFM, BFM, TM, SM, STM, Col, Sep, and spatial release) were within one SD of previous reports at the group level. This indicated that data quality in online testing settings using PART including those that self-report HD is similar to that from laboratory settings. Additionally, measures that were able to discriminate between those that self-report HD and those that do not (ND) at the group level were identified. This included a pure-tone detection average, temporal fine structure, and STM sensitivity tasks. Contrary to our expectations, the speech-based measures such as the SRT and the SiC construct did not show group differences. Within the construct, only the separated condition of the speech-on-speech masking task (Supplemental Table S2) showed significant group differences that would survive a correction for false discovery rate (also see auditory WM capacity). However, even when significant group differences in these tests indicate potential utility to discriminate HD, the small differences observed also indicate that the distributions of these scores are largely overlapping and may lead to poor discrimination.

Although the cognitive measures of attentional control and working memory capacity did not discriminate between groups at a statistically significant level, their usefulness in accounting for variance in other constructs and measures in the correlation and PCA analyses is notable. In particular, the measure of auditory WM capacity showed significant group differences that would survive a correction for false discovery rate. Interestingly, cognitive constructs only correlated with the SiC construct while lower order auditory constructs only correlated amongst themselves, including SiC. In this regard SiC presents an interesting construct with participation of all the levels of processing, yet it was not useful to discriminate HD in this sample (but see Lelo de Larrea-Mancera et al., 2023b). It is possible that the relationship between what is considered perceptual and cognitive is highly dependent on

task selection, sample characteristics, and grouping criteria (for discussion and review see Vetter et al., 2024).

In this study, we selected a hierarchy of auditory and cognitive processing measures following the structure of a construct framework presented in Figure 1. This framework was implicitly present in previous work (Gallun et al., 2022; Lelo de Larrea-Mancera et al., 2020, 2022) but was explicitly under study here. Results indicate that our theory-driven approach for using groupings of measures into constructs is consistent with a data-driven unsupervised dimensionality reduction technique. PCA grouped the individual measures in a very similar manner than the construct framework (Figure 1) in the first four rotated components. Pure tone tasks grouped together in one component, TFS tasks in another, STM and SiC tasks grouped together in another, and cognitive measures grouped together. Auditory attentional control loaded $>.5$ both in the cognitive and the higher order auditory processing component. These results are taken as a confirmation that the observed data distributes in a way consistent with the construct framework.

Self-reported hearing ability stood-out as the most challenging construct to understand as even the simple unitary questions used to divide groups did not produce identical results (Supplemental Table S1). To establish a consistent and validated criterion, final group assignment was based on the HHIE-S cutoff value of 8 for hearing handicap (ASHA, 1989). Although the whole range of measures of auditory processing was significantly correlated to the HHIE-S scores, the latter was not strongly correlated with the first four rotated PCs. Also, the multivariate model was able to predict only a small portion of the variance of the binary HD classification ($r^2=0.11$) an accuracy of 0.64, low sensitivity (0.4) missing more than half the HD cases, and moderate specificity (0.8) mislabeling relatively fewer ND participants as HD. In general terms, this indicates that self-reported measures address a different source of variance that, although complex, remains indispensable to understand an individual hearing profile. In other words, no other measure or combination of measures was able to substitute for self-report.

One aspect of self-report measures that is inherently different from behavioral test results is that self-report requires an individual to reflect on their own abilities and evaluate whether those self-perceptions are positive or negative. The HHIE was initially developed as a tool for assessing “the effects of hearing impairment on the emotional and social adjustment of elderly people” (Ventry & Weinstein, 1982), which was characterized as an introspective dimension of hearing loss that was not inherently reflected in the audiogram. In the case of the simple dichotomous question used here (“Do you have any hearing loss or hearing difficulties?”), responses indicate which participants would describe themselves as a person with hearing difficulties. These self-perceptions were not always reflected in their performance data, which may have contributed to the divergence of self-report measures from the other factors in the multivariate analysis.

In this work we were mainly interested in evaluating the inter-relationships of a hierarchy of auditory and cognitive performance tests as they related to self-reported classification of HD. We did not intend to introduce a battery of tests that would substitute the HHIE, and given the low accuracy attained with the various models for classification of HD, we are not at a stage to refine a discriminatory model and its capacities including cross-validation to avoid overestimates. Overall, there seems to be little relation between the self-report of HD and the performance measures under study here. Future work will be needed to close the gap between “objective measures” of hearing and cognition and self-report of hearing function. One of the main insights of this study is that self-reported and performance measures might be explaining quite different aspects of a person’s ability to hear.

Study Limitations and Future Directions

PART has been proposed as an instrumental development to address the increasing needs for prevention and care in the aging adult populations around the world that may face increased risk of sensory and cognitive decline (see Kannan et al., 2024). Indeed, in this dataset we see significant effects of age on almost all measures under study (except TFS and WM capacity). However, it is important to note areas that must be developed in future studies. For example, despite growing evidence that perceptual and cognitive assessment is the first step to achieve prevention and care (see Humes et al., 2012), PART still lacks large-scale normative datasets that would promote a link to treatment in terms of providing care options or professional advice. To address this, a few steps need to be taken.

As a first step, there is a question of the extent to which the auditory measures in PART provide either convergent or complementary information that informs diagnostics and treatments. While other research groups have made progress in this direction by connecting remote assessment with standard care (see Potgieter et al., 2016; Swanepoel et al., 2019; Wasmann et al., 2022; Zou et al., 2024) further research with PART will be required to create larger scale normative datasets in clinical populations that address both detection and monitoring of conditions. Further, it will be important to understand the extent to which these tests help inform management strategies for auditory and cognitive processing difficulties. To this end, there is growing research indicating that interventions with digital video games may provide benefits in understanding speech-in-competition (Lelo de Larrea-Mancera et al., 2021; Whitton et al., 2014, 2017). Connecting assessment with training and care is an essential future direction of this work.

In addition, it will be important to better determine both the settings and supervisory structures required for test systems like PART to provide informative results. While in the current study we did test people remotely, this was done via supervision during a video-call. While this was appropriate to ensure







that participants understood task instructions and ensure that the testing environment was appropriate to the study (e.g., not too much ambient noise or other distractions), it does potentially limit the applicability of the current set of results. It will be important for future work to evaluate the limits of acceptable testing situations, as well explore opportunities to use technologies such as ambient sound monitoring or online data analysis approaches that flag results that may be unreliable. Of note, for these types of tests it is generally a possibility that poor performance may be the outcome of confusion in self-administered testing. In these cases, it would be appropriate to do follow up in more controlled scenarios to confirm potential indications of clinically relevant conditions.

In sum, this work advances current understanding of the relationship between auditory and cognitive processing in those with and without subjective hearing difficulties. Overall, the study further validates remote testing using online pools of participants as a feasible general methodology. However, at the same time our data indicates that challenges remain in clarifying self-report of hearing difficulties using psychophysical tests of hearing function. It will be important to consider the extent to which other tests will be needed to explain self-report of hearing difficulty and more broadly what testing instruments will be most valuable to understand different hearing conditions. This may inform different intervention approaches that could be beneficial to people with hearing needs.

Acknowledgments

The authors thank Laura Coco and Michelle Molis for their advice in formulating the self-reported hearing difficulty question.

ORCID iDs

E. Sebastian Lelo de Larrea-Mancera  <https://orcid.org/0000-0001-8626-5621>
 William J. Bologna  <https://orcid.org/0000-0003-1851-6892>
 Sara Momtaz  <https://orcid.org/0000-0001-6585-3141>
 Eric C. Hoover  <https://orcid.org/0000-0002-1187-9925>
 G. Christopher Stecker  <https://orcid.org/0000-0003-1442-9839>
 Frederick J. Gallun  <https://orcid.org/0000-0002-4145-2199>

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Institute on Deafness and Other Communication Disorders (grant number R01DC015051, R01DC018166-01A1).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available in a public data repository in the Open Science Framework that can be accessed here: <https://osf.io/q4xzh/>.

Supplemental Material

Supplemental material for this article is available online.

References

- Almufarrij, I., Dillon, H., Dawes, P., Moore, D. R., Yeung, W., Charalambous, A. P., Thodi, C., & Munro, K. J. (2022). Web- and app-based tools for remote hearing assessment: A scoping review. *International Journal of Audiology*, 62(8), 699–712. <https://doi.org/10.1080/14992027.2022.2075798>
- American National Standards Institute. (2004). Methods for manual pure-tone threshold audiometry (ANSI S3.21-2004).
- American Speech-Language-Hearing Association. (1989). Guidelines for the identification of hearing impairment/handicap in adult/elderly persons. *ASHA*, 31(8), 59–63. PMID: 2679569.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek, M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, 24(4), 293–306. <https://doi.org/10.3766/jaaa.24.4.5>
- Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, 107(2), 1065–1066. <https://doi.org/10.1121/1.428288>
- Bright, T., & Pallawela, D. (2016). Validated smartphone-based apps for ear and hearing assessments: A review. *JMIR Rehabilitation and Assistive Technologies*, 3(2), e13–e12. <https://doi.org/10.2196/rehab.6074>
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, 77(5), 1465–1487. <https://doi.org/10.3758/s13414-015-0882-9>
- Carhart, R., & Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders*, 24(4), 330–345. <https://doi.org/10.1044/jshd.2404.330>
- Cherri, D., Eddins, D. A., & Ozmeral, E. J. (2024b). A step toward precision audiology: Individual differences and characteristic profiles from auditory perceptual and cognitive abilities. *Trends in Hearing*, 28, 1–16. <https://doi.org/10.1177/23312165241263485>
- Cherri, D., Ozmeral, E. J., Gallun, F. J., Seitz, A. R., & Eddins, D. A. (2024a). Feasibility and repeatability of an abbreviated auditory perceptual and cognitive test battery. *Journal of Speech, Language, and Hearing Research*, 68(2), 1–21. https://doi.org/10.1044/2024_JSLHR-23-00590
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake P. Shah (Eds.), *Models of working memory* (pp. 62–101). Cambridge University Press.
- Fifer, R. C., Jerger, J. F., Berlin, C. I., Tobey, E. A., & Campbell, J. C. (1983). Development of a dichotic sentence identification test for hearing-impaired adults. *Ear & Hearing*, 4(6), 300–305. <https://doi.org/10.1097/00003446-198311000-00007>
- Gallun, F. J. (2020). Flipping the lab: Using consumer electronics for high-quality data collection. *179th Meeting of the Acoustical Society of America*, 42(2020), 032002. <https://doi.org/10.1121/2.0001418>
- Gallun, F. J., & Best, V. (2020). Age-related changes in segregation of sound sources. *Hearing Research*, 72(1–2), 143–171. https://doi.org/10.1007/978-3-030-49367-7_7
- Gallun, F. J., Coco, L., Koerner, T. K., Larrea-Mancera, E. S. L. d., Molis, M. R., Eddins, D. A., & Seitz, A. R. (2022). Relating suprathreshold auditory processing abilities to speech understanding in competition. *Brain Sciences*, 12(6), 695. <https://doi.org/10.3390/brainsci12060695>
- Gallun, F. J., Diedesch, A. C., Kampel, S. D., & Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Frontiers in Neuroscience*, 7(7 DEC), 1–11. <https://doi.org/10.3389/fnins.2013.00252>
- Gallun, F. J., & Jakien, K. M. (2019). The ability to allocate attentional resources to a memory task predicts speech-on-speech masking for older listeners. *Proceedings of the International Congress on Acoustics, 2019-Sept*(September), 8242–8249. <https://doi.org/10.18154/RWTH-CONV-238789>
- Gallun, F. J., McMillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M., & Konrad-Martin, D. L. (2014). Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity. *Frontiers in Neuroscience*, 8(8 JUN), 1–14. <https://doi.org/10.3389/fnins.2014.00172>
- Gallun, F. J., Seitz, A., Eddins, D. A., Molis, M. R., Stavropoulos, T., Jakien, K. M., Kampel, S. D., Diedesch, A. C., Hoover, E. C., Bell, K., Souza, P. E., Sherman, M., Calandruccio, L., Xue, G., Taleb, N., Sebens, R., & Srinivasan, N. (2018). Development and validation of portable automated rapid testing (PART) measures for auditory research. *175th Meeting of the Acoustical Society of America*, 33(May), 050002. <https://doi.org/10.1121/2.0000878>
- Grose, J. H., & Mamo, S. K. (2010). Processing of temporal fine structure as a function of age. *Ear & Hearing*, 31(6), 755–760. <https://doi.org/10.1097/AUD.0b013e3181e627e7>
- Grose, J. H., & Mamo, S. K. (2012). Frequency modulation detection as a measure of temporal processing: Age-related monaural and binaural effects. *Hearing Research*, 294(1–2), 49–54. <https://doi.org/10.1016/j.heares.2012.09.007>
- Hoover, E., Pasquesi, L., & Souza, P. (2015). Comparison of clinical and traditional gap detection tests. *Journal of the American Academy of Audiology*, 26(06), 540–546. <https://doi.org/10.3766/jaaa.14088>
- Hoover, E. C., Kinney, B. N., Bell, K. L., Gallun, F. J., & Eddins, D. A. (2019). A comparison of behavioral methods for indexing the auditory processing of temporal fine structure cues. *Journal of*

- Speech, Language, and Hearing Research*, 62(6), 2018–2034. https://doi.org/10.1044/2019_JSLHR-H-18-0217
- Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons, Inc.
- Humes, L. E. (2021). An approach to self-assessed auditory wellness in older adults. *Ear & Hearing*, 42(3), 745–761. <https://doi.org/10.1097/AUD.0000000000001001>
- Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., Gates, G. A., Wilson, R. H., & Wingfield, A. (2012). Central presbycusis: A review and evaluation of the evidence. *Journal of the American Academy of Audiology*, 23(08), 635–666. <https://doi.org/10.3766/jaaa.23.8.5>
- Irace, A. L., Sharma, R. K., Reed, N. S., & Golub, J. S. (2021). Smartphone-based applications to detect hearing loss: A review of current technology. *Journal of the American Geriatrics Society*, 69(2), 307–316. <https://doi.org/10.1111/jgs.16985>
- Kannan, L., Lelo de Larrea-Mancera, E. S., Maniglia, M., Vodyanyk, M., Gallun, F. J., Jaeggi, S. M., & Seitz, A. R. (2024). Multidimensional relationships between sensory perception and cognitive aging. *Frontiers in Aging Neuroscience*, 16, 148449. <https://doi.org/10.3389/fnagi.2024.1484494>
- Lelo de Larrea-Mancera, E. S., Correa-Medina, E. I., Padilla-Bustos, K., Romero Terán, D. P., Hernández-Medrano, A. J., Cerda-Hernández, G. I., Cervantes-Arriaga, A., Rodríguez-Violante, M., Gallun, F. J., Seitz, A. R., & Solís-Vivanco, R. (2024a). A characterization of central auditory processing in Parkinson's disease. *Journal of Parkinson's Disease*, 14(5), 999–1013. <https://doi.org/10.3233/JPD-230458>
- Lelo de Larrea-Mancera, E. S., Flores-Medina, Y., Seitz, A. R., & Alcalá-Lozano, R. (2025). Digital screening for cognitive decline in the Spanish language for older adults. *Journal of Alzheimer's Disease*, 108(1 Suppl), S170–S184. <https://doi.org/10.1177/13872877251318205>
- Lelo de Larrea-Mancera, E. S., Koerner, T. K., Hoover, E. C., Stecker, G. C., Gallun, F. J., & Seitz, A. (2024b). Estimating thresholds with the adaptive scan method of psychophysical testing. *The Journal of the Acoustical Society of America*, 155(3_Supplement), A41–A41. <https://doi.org/10.1121/10.0026733>
- Lelo de Larrea-Mancera, E. S., Philipp, M. A., Stavropoulos, T., Carrillo, A. A., Cheung, S., Koerner, T. K., Molis, M. R., Gallun, F. J., & Seitz, A. R. (2021). Training with an auditory perceptual learning game transfers to speech in competition. *Journal of Cognitive Enhancement*, 6(1), 47–66. <https://doi.org/10.1007/s41465-021-00224-5>
- Lelo de Larrea-Mancera, E. S., Solís-Vivanco, R., Sánchez-Jiménez, Y., Coco, L., Gallun, F. J., & Seitz, A. R. (2023a). Development and validation of a Spanish-language spatial release from masking task in a Mexican population. *The Journal of the Acoustical Society of America*, 153(1), 316–327. <https://doi.org/10.1121/10.0016850>
- Lelo de Larrea-Mancera, E. S., Solís-Vivanco, R., Sánchez-Jiménez, Y., Padilla-Bustos, K., Correa-Medina, E. I., Clamage, D. S., Bologna, W. J., Gallun, F. J., & Seitz, A. R. (2023b). Testing the informativeness of diverse measures of auditory processing for clinical audiological practice in middle-aged adults in Mexico. *American Journal of Audiology*, 33(1), 1–11. https://doi.org/10.1044/2023_AJA-23-00135
- Lelo de Larrea-Mancera, E. S., Stavropoulos, T., Carrillo, A. A., Cheung, S., He, Y. J., Eddins, D. A., Molis, M. R., Gallun, F. J., & Seitz, A. R. (2022). Remote auditory assessment using portable automated rapid testing (PART) and participant-owned devices. *The Journal of the Acoustical Society of America*, 152(2), 807–819. <https://doi.org/10.1121/10.0013221>
- Lelo de Larrea-Mancera, E. S., Stavropoulos, T., Carrillo, A. A., Menon, K. N., Hoover, E. C., Eddins, D. A., Gallun, F. J., & Seitz, A. R. (2023c). Validation of the adaptive scan method in the quest for time-efficient methods of testing auditory processes. *Attention, Perception, & Psychophysics*, 85, 2797–2810. <https://doi.org/10.3758/s13414-023-02743-z>
- Lelo de Larrea-Mancera, S., Stavropoulos, T., Hoover, E., Eddins, D., Gallun, F., & Seitz, A. (2020). Portable automated rapid testing (PART) for auditory research: Validation in a normal hearing population. *Journal of the Acoustical Society of America*, 148(4), 1831–1851. <https://doi.org/10.1101/2020.01.08.899088>
- Livingston, G., Huntley, J., Liu, K. Y., Costafreda, S. G., Selbæk, G., Alladi, S., Ames, D., Banerjee, S., Burns, A., Brayne, C., Fox, N. C., Ferri, C. P., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Nakasujja, N., Rockwood, K., & Mukadam, N. (2024). Dementia prevention, intervention, and care: 2024 report of the lancet standing commission. *The Lancet*, 390(24), 2673–2734. [https://doi.org/10.1016/S0140-6736\(24\)01296-0](https://doi.org/10.1016/S0140-6736(24)01296-0)
- Marrone, N., Mason, C. R., & Kidd, G. (2008). Tuning in the spatial dimension: Evidence from a masked speech identification task. *The Journal of the Acoustical Society of America*, 124(2), 1146–1158. <https://doi.org/10.1121/1.2945710>
- Newman, C. W., Weinstein, B. E., Jacobson, G. P., & Hug, G. A. (1990). The hearing handicap inventory for adults: Psycho-metric adequacy and audiometric correlates. *Ear & Hearing*, 11(6), 430–433. <https://doi.org/10.1097/00003446-199012000-00004>
- Padilla-Bustos, K., Gallun, F. J., Seitz, A. R., Solís-Vivanco, R., & Lelo de Larrea-Mancera, E. S. (2025). Exploring the influence of bilingual experience on speech-in-competition measures. *The Journal of the Acoustical Society of America*, 54(1), 1–13. <https://doi.org/10.1121/2.0002016>
- Peng, Z. E., Waz, S., Buss, E., Shen, Y., Richards, V., Bharadwaj, H., Stecker, G. C., Beim, J. A., Bosen, A. K., Braza, M. D., Diedesch, A. C., Dorey, C. M., Dykstra, A. R., Gallun, F. J., Goldsworthy, R. L., Gray, L., Hoover, E. C., Ihlefeld, A., Koelewijn, T., & Venezia, J. H. (2022). FORUM: Remote testing for psychological and physiological acoustics. *The Journal of the Acoustical Society of America*, 151(5), 3116–3128. <https://doi.org/10.1121/10.0010422>
- Potgieter, J. M., Swanepoel, W., Myburgh, H. C., Hopper, T. C., & Smits, C. (2016). Development and validation of a smartphone-based digits-in-noise hearing test in South African English. *International Journal of Audiology*, 55(7), 405–411. <https://dx.doi.org/10.3109/14992027.2016.1172269>
- Rink, K. D., Gallun, F. J., & Diedesch, A. C. (2022). Remote auditory testing: Test-retest and user impressions. *181st Meeting of the Acoustical Society of America*, 45(2021), 050004. <https://doi.org/10.1121/2.0001572>

- Sabin, A. T., Eddins, D. A., & Wright, B. A. (2012). Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system. *The Journal of Neuroscience*, 32(19), 6542–6549. <https://doi.org/10.1523/JNEUROSCI.5732-11.2012>
- Speaks, C., & Jerger, J. (1965). Method for measurement of speech identification. *The Journal of the Acoustical Society of America*, 37(6), 1205–1205. <https://doi.org/10.1121/1.1939554>
- Stavropoulos, T. A., Isarangura, S., Hoover, E. C., Eddins, D. A., Seitz, A. R., & Gallun, F. J. (2021). Exponential spectrotemporal modulation generation. *The Journal of the Acoustical Society of America*, 149(3), 1434–1443. <https://doi.org/10.1121/10.0003604>
- Stecker, G. C., & Gallun, F. J. (2012). Binaural hearing, sound localization, and spatial hearing. In *Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing* (pp. 387–438). San Diego CA: Plural Publishing.
- Swanepoel, D. W., De Sousa, K. C., Smits, C., & Moore, D. R. (2019). Mobile applications to detect hearing impairment: Opportunities and challenges. *Bulletin of the World Health Organization*, 97(10), 717–718. <https://doi.org/10.2471/BLT.18.227728>
- Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology: Open Access*, 6(2), 1000227. <https://doi.org/10.4172/2161-1165.1000227>
- Ventry, I. M., & Weinstein, B. E. (1982). The hearing handicap inventory for the elderly: A new tool. *Ear & Hearing*, 3(3), 128–134. <https://doi.org/10.1097/00003446-198205000-00006>
- Vetter, P., Badde, S., Ferrè, E. R., Seubert, J., & Shinn-Cunningham, B. (2024). Evaluating cognitive penetrability of perception across the senses. *Nature Reviews Psychology*, 3, 804–820. <https://doi.org/10.1038/s44159-024-00382-1>
- Wasmann, J. W., Prag, L., Eikelboom, R., & Swanepoel, D. W. (2022). Digital approaches to automated and machine learning assessments of hearing: Scoping review. *Journal of Medical Internet Research*, 24(2), e32581. <https://doi.org/10.2196/32581>
- Whitton, J. P., Hancock, K. E., & Polley, D. B. (2014). Immersive audiomotor game play enhances neural and perceptual salience of weak signals in noise. *Proceedings of the National Academy of Sciences of the United States of America*, 111(25), E2606–E2615. <https://doi.org/10.1073/pnas.1322184111>
- Whitton, J. P., Hancock, K. E., Shannon, J. M., & Polley, D. B. (2017). Audiomotor perceptual training enhances speech intelligibility in background noise. *Current Biology*, 27(21), 3237–3247.e6. <https://doi.org/10.1016/j.cub.2017.09.014>
- Zou, A., Tang, D., Ferguson, M., Sherman, K., McMahon, C., Gill, L., Lau, A., Lee, J., Williamson, S., Davies, E., Sheng, K., O'Toole, S., Georgiou, A., Mukherjee, P., Wolnizer, P., & Gopinath, B. (2024). Protocol for the co-design of an online support service for adults with hearing loss. *PLoS ONE*, 19(9), 1–10. <https://doi.org/10.1371/journal.pone.0310521>