# OUTCOME MEASURES TO ASSESS THE BENEFIT OF INTERVENTIONS FOR ADULTS WITH HEARING LOSS: FROM RESEARCH TO CLINICAL APPLICATION

EDITED BY: Isabelle Boisvert, Melanie Ferguson, Astrid van Wieringen and Todd Andrew Ricketts

PUBLISHED IN: Frontiers in Neuroscience and Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-104-7
DOI 10.3389/978-2-83250-104-7

## 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

# OUTCOME MEASURES TO ASSESS THE BENEFIT OF INTERVENTIONS FOR ADULTS WITH HEARING LOSS: FROM RESEARCH TO CLINICAL APPLICATION

Topic Editors:
**Isabelle Boisvert,** The University of Sydney, Australia
**Melanie Ferguson,** Curtin University, Australia
**Astrid van Wieringen,** KU Leuven, Belgium
**Todd Andrew Ricketts,** Vanderbilt University, United States

**Citation:** Boisvert, I., Ferguson, M., van Wieringen, A., Ricketts, T. A., eds. (2022). Outcome Measures to Assess the Benefit of Interventions for Adults With Hearing Loss: From Research to Clinical Application. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-104-7

# Table of Contents

frontiers | Frontiers in Neuroscience

EDITORIAL
published: 18 August 2022
doi: 10.3389/fnins.2022.955189

# Editorial: Outcome Measures to Assess the Benefit of Interventions for Adults With Hearing Loss: From Research to Clinical Application

*Isabelle Boisvert[1]*, Melanie Ferguson[2,3], Astrid van Wieringen[4] and Todd Andrew Ricketts[5]*

[1] Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Camperdown, NSW, Australia, [2] Curtin enAble Institute, Faculty of Health Sciences, Curtin University, Perth, WA, Australia, [3] Brain and Hearing Research Group, Ear Science Institute Australia, Perth, WA, Australia, [4] Research Group Experimental Oto-Rino-Laryngologie, Department of Neurosciences, University of Leuven, Leuven, Belgium, [5] Maddox Memorial Hearing Aid Research Laboratory, Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Vanderbilt University, Nashville, TN, United States

**Keywords: outcome measures, hearing loss, adult, clinical application, intervention**

**Editorial on the Research Topic**

**Outcome Measures to Assess the Benefit of Interventions for Adults With Hearing Loss: From Research to Clinical Application**

OPEN ACCESS

***Edited and reviewed by:***
*Isabelle Peretz,*
*Université de Montréal, Canada*

****Correspondence:***
*Isabelle Boisvert*
*isabelle.boisvert@sydney.edu.au*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *28 May 2022*
***Accepted:*** *10 June 2022*
***Published:*** *18 August 2022*

***Citation:***
*Boisvert I, Ferguson M, van Wieringen A and Ricketts TA (2022) Editorial: Outcome Measures to Assess the Benefit of Interventions for Adults With Hearing Loss: From Research to Clinical Application. Front. Neurosci. 16:955189. doi: 10.3389/fnins.2022.955189*

Hearing, listening, communication and participation in the context of hearing loss are complex constructs to measure. This is because those constructs are intertwined with other complex constructs including language, cognition, social engagement, and fatigue. Hearing loss (passive) impacts listening (active) which, for many adults who live with hearing loss, impacts communication (bi- or multi-directional exchange) and participation (everyday life). A plethora of hearing-based interventions are available to support the needs of adults with hearing loss. This includes a range of hearing aid and hearing implant technologies, personal sound amplification products, assistive-listening devices, communication strategies, and auditory/cognitive training. To evaluate the benefits of these interventions, we require valid, relevant and reliable outcome measures before and after the interventions.

A valid outcome measure means that it measures what it intends to measure, a relevant measure is that which taps into the intended mechanism of benefit or outcome domain, while a reliable measure means that the same result would be found if that measure was repeated in the same circumstances (high test-retest reliability). Because of their simplicity and reliability, hearing thresholds and speech recognition tests, typically conducted at a fixed volume in a quiet environment, have dominated clinical practice and research in audiology (Granberg et al., 2014). Although replicable, simple to conduct, and useful in certain contexts, these measures have limited relationship with everyday abilities and needs (Ferguson et al., 2016; Keidser et al., 2020). Developing and selecting valid, relevant and reliable outcome measures remains a challenge in the field of audiology.

We may never succeed in developing a single measure that captures the full relationship between hearing, listening, language, cognition and participation in the context of interactive and sustained communication within the real-world: a dynamic 3D acoustic and visual environment. However, it is important that we explore, and push, the boundaries of this problem in order to benefit the field, and importantly, better support those who live with hearing loss. The collection of articles in this Research Topic highlights current discussions and directions that the field has, and is, taking. This Research Topic begins with a scoping review by Neal et al. that maps the abundance of measures that are used in recently published studies to assess the listening and communication skills of adults with hearing loss. The authors note that these measures mainly target a narrow set of relevant domains. Following from this, Munro et al. discuss how the selection of hearing-related measures that are used in clinical trials have consequences on the outcomes of those trials—and therefore the knowledge that we can derive from these studies. Their article provides guidance about the factors that need to be considered in the development and selection of outcome measures, to increase the value and impact of clinical trials. The article by Allen et al. provides further reflection about the need to carefully consider the choice of self-report outcome measures used, not just for specific clinical trials, but in the development of national databases. For this to be possible, mechanisms are required to standardize the selection, collection and reporting of clinical data. These authors report on a consensus-based approach used to identify a core outcome domain set that is relevant to measure from the perspective of hearing services consumers and clinicians. Aligned with some of the core messages of Munro et al. and Allen et al., Dietz et al. illustrate how selecting the type and the timing of outcome measures impacts a study's outcomes, in particular the additional value that self-report measures provide over and above the conventional speech testing for cochlear implant users. Similarly, Abdel-Latif and Meister highlight how the addition of measures of listening effort can complement routine clinical testing. Hoppe et al. further demonstrate how considering outcome measures for each individual ear, as well as binaurally, impacts the interpretation of study results, in particular for individuals who have asymmetric hearing. In contrast with the numerous studies that use pure-tones and speech-based stimuli, Shafiro et al. conducted a systematic review to showcase the limited evidence base that exists in relation to the perception of environmental sounds with hearing devices. As for the abovementioned studies, inconsistent methodologies limit the potential to compare between studies and to aggregate data from larger datasets, for example for meta-analyses.

To assess outcomes that better reflect real-life situations, the complexity and realism of the stimuli and tasks can also be varied (e.g., using overlapping stimuli types, multiple stimuli locations, or dual-tasks). Historically, the main problem with these types of measures is that they have required larger spaces and more complex and expensive equipment. These measures also need to be designed and evaluated carefully to ensure adequate reliability (Ferguson and Henshaw, 2015). In this Research Topic, Miles et al. assessed new speech intelligibility tasks that are more representative of everyday speech communication outside the laboratory. They show that the more realistic speech task offered a better dynamic range for capturing individual performance and hearing-aid benefit across a range of real-world environments. The article by Salorio-Corbetto et al. describes the assessment of a Virtual Acoustics (VA) version of the Spatial Speech-in-Noise (SSiN) test, the SSiN-VA, for the purpose of evaluating hearing abilities with bilateral hearing aids. This approach can enhance clinical efficiency because testing can be conducted at home. In a similar vein, van Wieringen et al. investigated three different speech perception assessments in the same 40 cochlear implant users in their home environment. Their study showed that home-based speech perception testing is reliable and can be used to complement care in the clinic.

Outcome measures relevant for adults with hearing loss can be categorized in terms of the type of responses collected from participants (behavioral, physiological, or self-reported). In this Research Topic, however, no articles investigating physiological measures were submitted. In contrast, several submissions included self-reported measures, which are easy to conceptualize in terms of validity and relevance. While Neal et al. identified 139 different self-reported measures used with adults with hearing loss in recently published studies, self-reported measures continue being developed. Specific techniques for the development of high-quality Patient-Reported Outcome Measures (PROMs) have been developed, as described in the article by Laplante-Lévesque et al. Modern PROMs are therefore expected to include the rich perspective of people with the lived experience of the construct being measured, and follow good practice guidelines (e.g., COSMIN). Using the Cochlear Implant Quality of Life (CIQOL) instruments as an example, these authors provide useful context and guidance for research groups interested in using existing, or developing new self-reported outcome measures. In terms of new patient-reported measures, Humes' contribution describes the development of a new scale to measure the Subjective Wellbeing of older adults with hearing loss. Tapping into the domains of Life Satisfaction, Acceptance of Hearing Loss, and Social Support, the psychometric analysis of this new scale showed very good reliability and good criterion validity. Another new self-reported measure in this Research Topic and presented by Markodimitraki et al. is the COMPASS PROM that aims to quantify the consciousness of wearing a cochlear implant and how this impacts the daily life of cochlear implant users. This includes sleep disturbances due to the physical sensation of the implant on the head or problems with wearing headgear.

Acknowledging the progress made, as well as the need to select outcome measures that are aligned with specific research questions, more work is required before we can agree on an integrated set of outcome measures that are valid, relevant and reliable to support the everyday communication of adults with hearing loss. Study results based on such sets of outcome measures are critical when policy makers approve and fund new products and services.

Therefore, with the inclusion of hearing benefit claims within the advertising of everyday technologies such as earphones, the development of alternative services delivery models (e.g., remote, automated, over-the counter, direct-to-consumer), the proliferation of hearing-related training programs, and the development of drugs that aim to improve hearing, the need for valid, relevant and reliable outcomes measured cannot be understated, and their selection cannot be overlooked.

## AUTHOR CONTRIBUTIONS

IB wrote the initial draft of this article and all authors listed made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

AW was supported by the KU Leuven grant no. C3/21/046.

## REFERENCES

Ferguson, M., and Henshaw, H. (2015). "How does auditory training work? Joined-up thinking and listening," in *Seminars in Hearing*, Vol. 36. (Thieme Medical Publishers), 237–249.

Ferguson, M. A., Woolley, A., and Munro, K. J. (2016). The impact of self-efficacy, expectations, and readiness on hearing aid outcomes. *Int. J. Audiol.* 55 (Supp. 3), S34–S41. doi: 10.1080/14992027.2016.1177214

Granberg, S., Dahlström, J., Möller, C., Kähäri, K., and Danermark, B. (2014). The ICF core sets for hearing loss–researcher perspective. Part I: systematic review of outcome measures identified in audiological research. *Int. J. Audiol.* 53, 65–76. doi: 10.3109/14992027.2013.851799

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *Ear Hear*. 41 (Supp. 1), S5. doi: 10.1097/AUD.0000000000000944

**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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 June 2021
doi: 10.3389/fpsyg.2021.640165

# Development of the SWB-HL: A Scale of the Subjective Well-Being of Older Adults With Hearing Loss

*Larry E. Humes**

*Department of Speech, Language, and Hearing Sciences, Indiana University, Bloomington, IN, United States*

The objective of this research was to develop and evaluate a self-report measure of subjective well-being (SWB) for use with older adults with hearing loss (HL). A convenience sample of 173 local volunteers between the ages of 60 and 88 years ($M = 74.4$; SD = 7.2 years) participated in this study. The initial 18-item version of the scale was constructed, response characteristics examined, and then subjected to factor analysis, as well as evaluation of the scales' reliability and validity. The analysis of response characteristics and subsequent factor analysis resulted in the elimination of eight of the 18 test items. The SWB-HL Total score was derived from the 10 remaining items. It was shown that the SWB-HL tapped three underlying domains interpreted as: Life Satisfaction (three items); Acceptance of Hearing Loss (Accept HL; four items); and Social Support (three items). Psychometric analysis showed very good reliability and good criterion validity was established for the 10-item SWB-HL Total score. In addition, significant differences were observed between aided and unaided SWB-HL Total scores following 4–6 weeks of hearing aid use. The SWB-HL is a 10-item self-report measure of SWB that shows good reliability and validity when used by older adults with hearing loss and reveals improved SWB following the use of hearing aids.

**Keywords: hearing aid, quality of life, hearing health, aging, subjective well-being**

OPEN ACCESS

***Edited by:***
*Todd Andrew Ricketts,*
*Vanderbilt University, United States*

***Reviewed by:***
*Gurjit Singh,*
*Sonova, Switzerland*
*Peter Busby,*
*The University of Melbourne, Australia*

***Correspondence:***
*Larry E. Humes*
*humes@indiana.edu*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *10 December 2020*
***Accepted:*** *17 May 2021*
***Published:*** *11 June 2021*

***Citation:***
*Humes LE (2021) Development of the SWB-HL: A Scale of the Subjective Well-Being of Older Adults With Hearing Loss. Front. Psychol. 12:640165. doi: 10.3389/fpsyg.2021.640165*

## INTRODUCTION

Recognizing the burgeoning population of older adults worldwide and the importance of facilitating healthy aging, on December 14, 2020, the United Nations adopted a resolution declaring 2021–2030 as the "Decade of Healthy Aging" (United Nations, 2020). Central to heathy aging is the concept of "well-being" (World Health Organization, 2015). Measures of well-being include both objective and subjective measures. Objective measures may include the disabilities experienced, life experiences, such as marriage and divorce, and annual income, among others. Subjective measures are labeled as such because they are self-report measures provided by a respondent. It is the latter, subjective well-being (SWB), that is the focus of this report.

Since the early work of Diener (1984), SWB was conceptualized as having three primary components: (1) life satisfaction; (2) frequent positive affect; and (3) infrequent negative affect (Diener et al., 1999). There have been two general approaches to the measurement of SWB: (1) a cognitive, reflective approach which asks the respondent to think about their life over some period of time, typically days, weeks, or months, and make evaluative ratings of their life satisfaction and affect (Diener, 1984); and (2) application of more time-locked immediate ratings in naturalistic or every-day environments, such as the Experience Sampling Method (Scollon et al., 2003), Ecological Momentary Assessment (Stone et al., 1999), and the Day Reconstruction Method (Kahneman et al., 2004). Perhaps because of the ease of measurement and the subsequent data reduction, most measures used on a widespread basis have been of the cognitive, reflective type. In particular, the most widespread measures make use of 5- to 7-point Likert-scale ratings of various aspects of life satisfaction and happiness or affect. Of SWB measures of this type, it appears that the Satisfaction with Life Scale (Diener et al., 1985) and the Positive and Negative Affect Scale (PANAS; Watson et al., 1988) have been most widely used to capture the three primary components of SWB. More recently, the Scale of Positive and Negative Experience (SPANE; Diener et al., 2010) has been used often as an alternative to the PANAS to capture the affect components of SWB.

A variety of life events are known to impact SWB. The development of hearing loss with advancing age is one such life event, one most commonly occurring with gradual onset. Age-related hearing loss is found in 65% of individuals above 60 years of age worldwide and has significant negative impacts on quality of life or SWB when left untreated (World Health Organization, 2021). The negative impact of hearing loss in older adults on health-related quality of life has been observed whether the impact was measured with general SWB measures or hearing-specific SWB measures (e.g., Bess et al., 1989, 1991; Mulrow et al., 1990; Dalton et al., 2003; Abrams et al., 2005; Chia et al., 2007; Ciorba et al., 2012; Nordvik et al., 2019). On the other hand, for the most part, benefits from intervention with hearing aids for older adults have been easy to document only on hearing-specific measures of SWB (e.g., Bess, 2000; Chisolm et al., 2005, 2007; McArdle et al., 2005; Dawes et al., 2015; Weinstein et al., 2016). The more that the items of the measure were specific to hearing loss, the more likely benefits could be observed. This could lead one to conclude that the benefits of intervention are limited to those within the hearing domain rather than impacting general SWB.

Perhaps a measure of SWB can be primed to be sensitive to the impact of hearing loss on SWB. Self-report surveys in general reveal broad effects of question context (Bertrand and Mullainathan, 2001; Lucas, 2018) and the operation of specific mechanisms such as priming (e.g., Bowling and Windsor, 2008; Garbarski et al., 2015). Similar effects of context and priming have been observed in the measurement of SWB as well (e.g., Schwarz and Strack, 1991; Kahneman et al., 1993). If each type of query is embedded within the same instrument, perhaps the net effect will be to capture variance shared by both hearing-specific and general SWB (e.g., Gehlbach and Barge, 2012). That is, by including queries about hearing loss within a broader measure of SWB can that measure of SWB be primed to be sensitive to the impact of hearing-aid intervention on SWB? Schwarz and Strack (1991) provided an interesting demonstration illustrating possible priming effects on SWB measures using two queries with young adults, one about the frequency of dating and one about life happiness. The responses to each query were moderately correlated ($r = 0.66$) when asked first about dating frequency but uncorrelated ($r = -0.12$) when the order was reversed. When another group of respondents was informed that queries would be made about "two areas of life that may be important to overall well-being," the responses were again uncorrelated ($r = 0.15$). This suggests that measures of SWB which incorporate queries about hearing loss will attune the respondent to the importance of hearing loss when responding to queries about their general SWB.

Several years ago, unaware of any existing survey that blended both hearing-related queries and general SWB items together, an internet-search was conducted for such measures for other disorders or treatments. This search led to a series of tools made available by www.FACIT.org. The Functional Assessment of Chronic Illness Therapy (FACIT) organization "…manages the distribution and information regarding administration, scoring and interpretation of a range of questionnaires that measure health-related quality of life for people with chronic illnesses." At the time of our search, one available measure seemed closest to our needs: Functional Assessment for Non-Life-Threatening Conditions (FANLTC). The FANLTC (Version 4) consisted of a total of 26 items, seven on physical well-being, seven covering social/family well-being, five dealing with emotional well-being, and seven concerned with functional well-being. The first seven items on physical well-being were deleted as they dealt with unlikely consequences of hearing loss including whether the respondent felt ill, had nausea, was bothered by treatment side effects, had pain, and so on. This scale on physical well-being reflects the roots of the FACIT organization's measures which began with the Functional Assessment of Cancer Therapy (FACT; Cella et al., 1993). As noted by Cella et al. (1993) regarding the FACT, for cancer patients undergoing treatment, the negative impacts of the treatment(s) on SWB often outweigh the impacts of the cancer being treated. For intervention with hearing aids, however, treatment side effects akin to those experienced during cancer therapy were not considered likely and the items pertaining to physical illness were dropped. In addition, one item in the social/family well-being section was deleted which inquired about the respondent's satisfaction with his or her sex life. This left a total of 18 items remaining. The other key change made to five of the remaining 18 items of the FANLTC involved changing all references to "my illness" to "my hearing loss." The response options and instructions remained the same as for the FANLTC. The response options for each item (points assigned) were: not at all (0); a little bit (1); somewhat (2); quite a bit (3); and very much (4). The final version that emerged from these analyses is referred to as the Subjective Well-Being of Older Adults with Hearing Loss (SWB-HL) and will be described in more detail below.

The SWB-HL was administered to a group of 173 older adults prior to being fit with amplification and again as an aided measure for 143 of the 173 who received hearing aids in this study. In addition to the SWB-HL, several other measures were obtained prior to being fit with amplification, ranging from general measures of depression, anxiety, and optimism to detailed assessment of the communication difficulties experienced by these older adults and their adjustment to these difficulties. These measures will be used here to assess the validity of the SWB-HL. In addition to examining the validity of the SWB-HL, analyses were performed to evaluate the reliability of the SWB-HL and its subscales. Classical test theory was used to assess reliability due, in part, to the sample size being insufficient for so-called modern psychometric analysis, such as item response theory and Rasch analysis, which have been applied to audiological self-report measures in recent years (Heffernan et al., 2019; Cassarly et al., 2020). Finally, as another assessment of the validity of the SWB-HL and its sensitivity to change in SWB following hearing-aid use, we compare the aided and unaided SWB-HL scores of the older adults fit in this study.

# MATERIALS AND METHODS

## Participants

A total of 173 older adults were recruited into the university clinic through advertisements in the newspaper, community centers, religious centers, and other facilities likely to be frequented by older adults in the Bloomington, Indiana community. The ads made it clear that the study was seeking those interested in hearing aids. In addition, some participants were recruited into the study from the clinic. That is, they came to the clinic on their own accord and once they completed the hearing evaluation, they were approached about participation in the study. As a result, this sample is a clinical convenience sample of older adults from the local community.

From 2004–2008, there were a total of 530 individuals who were screened for study eligibility. Of these, 162 (30.6%) were ineligible. Of the remaining 388 eligible individuals, 154 (39.7%) enrolled and purchased hearing aids with 143 completing the measures included in these analyses. They paid the full purchase price for the devices at the time of enrollment. Of the 234 eligible candidates who opted not to purchase hearing aids and enroll, 36 agreed to return to complete several of the pre-fit measures completed by the hearing-aid purchasers with 30 of these individuals completing all the measures included in these analyses. The total sample for these analyses was 173, including 143 individuals who were fitted with hearing aids and 30 who opted not to pursue hearing aids.

Of the 173 participants, 102 (59.0%) were male, 65 (37.6%) were female, and 6 (3.5%) did not indicate their gender. Participants ranged in age from 60 to 88 years with a mean age of 74.4 years ($SD = 7.2$ years). For 16%, formal education ended at high school, whereas 12.1% completed college, another 23.1% held a Masters' degree, and an additional 17.3% held a doctorate or medical degree. Regarding income, total annual income for the preceding tax year was reported to be >$45,000 annually by 55.6% with annual incomes of $5,000–15,000, $15,000–25,000, $25,000–35,000, and $35,000–45,000 reported by 7.6, 11.1, 13.5, and 12.3% of the participants, respectively. When queried regarding the duration of their hearing loss, the median response was 5.5 years with an interquartile range of 3–10 years. Regarding current or prior hearing aid use, 33 (19.1%) were currently using hearing aids and 47 (27.2%) reported ever having worn hearing aids with 14 (8.1%) no longer using them.

**Figure 1** provides the mean audiograms for the right and left ears of the 173 older adults. Standard deviations were 11–12 dB at 250 and 500 Hz in each ear and gradually increased with frequency peaking at 14–16 dB at and above 3000 Hz. The audiograms are consistent with those for older adults, showing a bilaterally symmetrical sloping sensorineural hearing loss. Otoscopy, bone-conduction thresholds, and immittance measures confirmed that the hearing loss was sensorineural in nature with no significant conductive components.

Informed consent was obtained from all participants and they were paid on a per-session basis for their participation. This study was approved by the Indiana University Bloomington Institutional Review Board.

## Measures Obtained

In addition to an audiological exam, several other measures pertinent to this study were obtained. These measures were included, in part, to evaluate the validity of the SWB-HL. The measures pertinent to this study could be divided into two categories: (1) psychological measures of affect or stress/anxiety, personality, and cognition; and (2) measures of communication performance and the participant's reaction to difficulties experienced. For the latter, we relied on the Communication Profile of the Hearing Impaired (CPHI; Demorest and Erdman, 1986, 1987). The CPHI represents one of the most comprehensive and rigorously evaluated self-report measures of communication difficulties and the impact of those difficulties on the person's well-being. The 163-item measure is reduced to 25 scales and, ultimately, to five factor scores.

There were three measures of the participant's affect included here. Measures of affect were included to validate the affect components of SWB assessed by the SWB-HL. The Life Orientation Test-Revised (LOT-R; Scheier et al., 1994) is a 10-item measure of optimism vs. pessimism. Three of the 10 items measure either optimism or pessimism with the other four serving as fillers. The participant is presented with a statement and then rates each on a four-point scale of agreement: strongly disagree, disagree, agree, or strongly agree with the three pessimism items reverse scored from the optimism items and the fillers unscored. Earlier work demonstrated the importance of optimism to perceived psychological and physical well-being (Scheier and Carver, 1992).

The Positive and Negative Affect Scale (PANAS), as noted in the introduction, is a brief self-report measure of affect and has been used frequently to capture the two affect components of the tripartite SWB model. The respondent is presented with a list of 20 words in a column and, for each, is asked to indicate whether he or she generally feels this way "on a regular basis" and selects among "very slightly or not at all," "a little," "moderately," "quite a bit" or "extremely" as responses, with points assigned from 1 to 5, respectively. Half of the 20 items convey positive affect, such as "excited," "strong," "enthusiastic," "proud," and "inspired," whereas the other have convey negative affect, such as "distressed," "upset," "scared," "ashamed," and "nervous." Two scores emerge: PANAS-positive and PANAS-negative. Each represents the total points for each set of 10 items, scores ranging from 10 to 50. For the positive scale, higher scores, and, for the negative scale, lower scores, reflect more positive affect.

The State-Trait Anxiety Inventory (STAI; Spielberger et al., 1970; Spielberger, 1983) is a self-report measure of feelings of

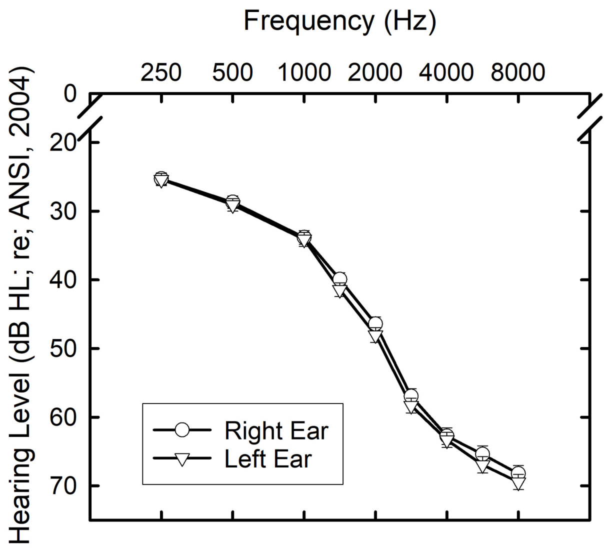


**FIGURE 1 |** Means and standard errors for the air-conduction thresholds for the right (circles) and left (triangles) ears of the 173 older adults who participated in this study.

anxiety, momentary or in current state and long-term or as a trait of the individual. There are forty items with the first 20 assessing situational or state anxiety, with a focus on how the respondent "feels right now, at this moment," and the last 20 measuring underlying trait anxiety, how he or she "generally feels." For all 40 items, four response choices are provided: "not at all," "somewhat," "moderately so," and "very much so" with points of 1, 2, 3, and 4, respectively. Within each scale, several items are reverse scored. Examples from the STAI-S include: "I feel content" and "I am worried." Examples from the STAI-T are: "I am happy" and "I lack self-confidence." Total scores, following reversal of some items, range from 20 to 80 for both the STAI-S and STAI-T with higher scores reflecting less anxiety and more positive affect.

In addition, given that other measures of SWB have been demonstrated to be correlated with personality (Steel et al., 2008) and cognitive function (Jones et al., 2003), both types of measures were included here. For personality, the Myers-Briggs Type Inventory (MBTI; Myers et al., 1998) was used and, for cognition, the full Wechsler Adult Intelligence Scale, 3rd Edition (WAIS-III; Wechsler, 1997) was used.

For most of these psychological measures and the CPHI, a Likert-type scale is used to collect participant responses and these responses have been treated like interval data since their inception with norms reported as total scores or mean scores. The assumption has been that the underlying latent constructs tapped by each scale are continuous and the responses represent interval-scale values along that underlying continuum. There has been considerable debate over the years as to whether such Likert-type ratings should be modeled as interval or ordinal data (e.g., Jamieson, 2004; Norman, 2010), but because each of these measures has considered the values to be interval in nature and the norms calculated under this assumption over many decades, the scores from these established scales were treated this way in these analyses as well. For the new SWB-HL scale proposed here, however, this assumption will be examined in greater detail below.

## Procedures

Following the initial case history, otoscopic examination, and audiological assessment, each participant completed a series of measures prior to being fit with amplification. These other measures were completed in separate sessions with the CPHI completed in one, full cognitive assessments with the WAIS-III in another, and the other psychological assessments completed together in a third session. All surveys were completed by the respondent in a pencil-and-paper format without examiner assistance.

After completion of these measures, as well as others not presented here, the participants were then fitted with hearing aids. The technology used varied among one of three options.

One group received four-channel wide-dynamic-range-compression (WDRC) circuits housed in full-concha ITE shells, half with directional microphones and half with omni-directional microphones. The other group received 6-channel open-fit mini-BTE devices with directional microphones. The directional microphones were a fixed super-cardioid configuration, and its function was verified using Verifit software and hardware (Etymonic Design, Dorchester, Canada). Verification of directional performance was obtained in the test box from both hearing aids. The frequency response of the hearing aids programmed for the directional microphone function was equalized to match the frequency response of the hearing aids when set to the omnidirectional-microphone mode. The same basic protocol was used to set and verify target gain for each participant in each group. First, based on audiological information obtained from each participant (air-conduction and bone-conduction hearing thresholds, as well as loudness discomfort levels), target 2-cm$^3$-coupler gain values were generated at octave intervals from 250 through 4000 Hz, as well as at 1500, 3000, and 6000 Hz. Hearing aids were adjusted in the 2-cm$^3$ coupler for a moderate level input (60–70 dB SPL, across studies) to match target in the coupler and were then fitted to the patient and verified using real-ear probe-tube microphone measurements with adjustments to better match targets performed as needed. The root-mean-square error between target and measured real-ear gain averaged across frequency (250–6000 Hz) was $\leq$ 5 dB.

The prescriptive procedure used to generate gain targets was NAL-NL1. With each group and technology, software from NAL was used to generate NAL-NL1 targets, rather than the manufacturer's version of that prescriptive protocol. Within a given group of participants, all were fitted bilaterally with identical make and models of hearing aids. In addition, participants paid the typical clinic price for the devices at the time of delivery and then were paid as research subjects for return visits during which they completed a variety of outcome measures.

After on-ear verification of real-ear gain, the participant was counseled about the use, function, and care of the hearing aids. Approximately 4–6 weeks post-fit, the participant returned to complete several outcome measures, with the focus here on the SWB-HL. Reports for several of the conventional outcomes have been published previously for subsets of the study sample reported here with a focus on differences in outcomes for various technologies (Humes et al., 2009, 2010).

# RESULTS AND DISCUSSION

## Preliminary Analysis of the SWB-HL Items

As noted, there were initially 18 items in the SWB-HL measure completed by the 173 participants. The responses made use of a 5-point Likert-scale rating for each of the 18 items. There is debate as to whether responses to such items should be modeled as ordinal data or interval data (e.g., Jamieson, 2004; Norman, 2010). Some have demonstrated that, except for extremely skewed distributions of responses, parametric data reduction and analyses of 5-point Likert ratings is appropriate (Flora and Curran, 2004; De Winter and Dodou, 2010). Most recently, Liddell and Kruschke (2018) examined the assumptions regarding the underlying distributions of Likert-type ratings in detail comparing the results for an interval-based metric model to those for an ordered-probit model. They noted that the item means and standard deviations can differ considerably under each model and argue in favor of the ordered-probit model when response distributions show such differences.

Liddell and Kruschke (2018) provided R code to analyze responses from 5-point Likert ratings under both models. This code was used here to generate means and standard deviations from each of the 18 items of the SWB-HL for the two models. Effect sizes were then used to compare those two sets of item means. Although there was a strong correlation between the means under the two model assumptions ($r = 0.92$), the Cohen's d effect sizes (Cohen, 1988) were small for three, medium for two, and large for one of the six items with significant differences in item means. The effect sizes for the difference between the metric and ordered-probit means were all less than small ($d < 0.2$) for the 12 items that were retained. In other words, for the 12 remaining items of the SWB-HL, assumptions about the underlying scale did not impact the item scores and parametric tests could therefore be applied with confidence in the validity of those analyses. In all cases, the six items eliminated had skewed distributions with ordered-probit means $\geq$ 3.7 (maximum = 4) and skewness values exceeding −1.2, consistent with an extremely skewed distribution associated with a ceiling effect. The 12 items that were retained following this analysis of response distributions were items 1, 2, 3, 4, 5, 8, 11, 13, 15, 16, 17, and 18.

## Factor Analysis of the SWB-HL

Following the paring down of the original 18 items to 12, factor analysis was performed on those 12 items. These analyses made use of the R "psych" package (Version 2.0.12; Revelle, 2020) to facilitate analysis with polychoric correlations. Given the preceding culling of items from the scale, use of parametric Pearson-r correlations for the factor analysis would have likely been acceptable (Flora and Curran, 2004). Nonetheless, non-parametric polychoric correlations optimized for the analysis of Likert-type responses were instead computed for these 12 items and the ensuing factor analysis was based on this matrix of polychoric correlations. The analysis of the SWB-HL began with an exploratory principal-axis factor analysis (Gorsuch, 1983) of the responses to the 12 survey items. Exploratory factor analysis was chosen because this was the first evaluation of the SWB-HL. Thus, no apriori assumptions were made about underlying scales, if any, with this to be determined by the results of the exploratory factor analysis. Initially, oblique rotation of factors was considered which allows for correlations among the factors that emerge. Three factors emerged and all resulting between-factor correlations were $< 0.35$. As a result of these weak inter-factor correlations, a second principal-axis factor analysis was performed, this time with orthogonal (varimax) rotation of factors resulting in three independent factors. A reasonable fit was obtained, accounting for 58% of the variance, but two items had low communalities (0.22, 0.33) indicating that neither was well-represented in the three-factor solution. These two items, item 11 and 16, were deleted. The remaining 10 items were analyzed again with the same principal-axis factor analysis with varimax rotation and a three-factor solution again emerged, accounting for 65% of the variance and all communalities $\geq$0.43. Each of the three rotated factors accounted for nearly equivalent amounts of variance. The root-mean-square of the residuals for this final factor solution for the SWB-HL was 0.05 with a Bayesian Information Criterion of 25.4, both reflecting a good fit of the solution to the data. Finally, a hierarchical factor analysis was performed using the *omega* function in the R psych package to determine whether all 10 items were represented by a general (g) underlying factor (Revelle, 2020). This would better validate the use of a total score for all 10 items as a measure representing this general factor common to all items.

**Figure 2** graphically depicts the results of the hierarchical factor analysis and **Table 1** provides the rotated factor loadings from the pattern matrix of the hierarchical solution for each of the 10 SWB-HL items. As can be seen, all 10 items are weighted both on the general factor (g) and one of the three domain-specific factors. Based on the factor loadings in **Table 1**, the three scale factors were interpreted as Life Satisfaction, Acceptance of Hearing Loss, and Social Support. The three factors accounted for 23, 21, and 21% of the variance, respectively, for the rotated solution.

These analyses suggest that the SWB-HL taps three separate aspects of SWB in older adults with hearing loss. The analyses also suggest that the SWB-HL Total score taps a general underlying construct common to all three factors.

## Reliability of the SWB-HL

The SWB-HL could be used to generate four scores based on the general factor, g, and each of the three domain-specific factors. Although the analyses of the SWB-HL support the existence of three domains captured by the 10-item scale, it is not recommended that such individual subscale scores be used due to the small number of items, 3–4, comprising each scale. Instead, it

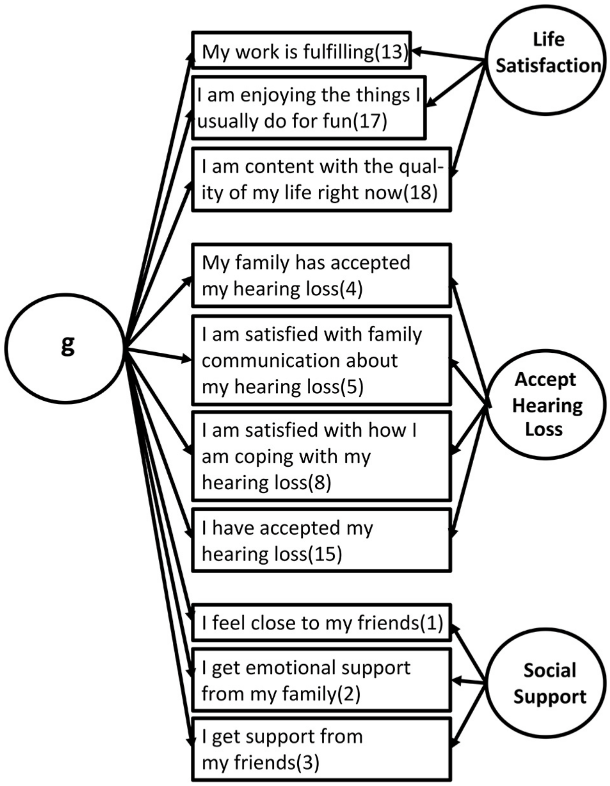


**FIGURE 2 |** Hierarchical factor structure for the 10 items of the SWB-HL scale showing the association of each item to the factors identified. The number in parentheses after each item is the item number in the original 18-item corpus.

is recommended that just the SWB-HL Total be used with the knowledge that it is tapping the constructs of Life Satisfaction, Acceptance of Hearing Loss, and Social Support. The full scale, 10 items, is itself brief and it is well-known that the reliability of any measure decreases as the number of items comprising the measure decrease. Moreover, the correlations of the SWB-HL Total score with the three subscale scores were $r = 0.71$, 0.67, and 0.84, all significant ($p < 0.01$), for the Life Satisfaction, Social Support, and Acceptance of Hearing Loss scales, respectively. This again reinforces that a single 10-item SWB-HL Total score will capture variations in each of these three domains.

With this in mind, reliability analyses included the computation of Cronbach's alpha (Cronbach, 1951) across all scale items, computation of the mean inter-item correlation ($r_{ii}$), and examination of the change in Cronbach's alpha following deletion of each item. The reliability analyses were completed using the *alpha* function of the R psych package and the polychoric correlation matrix for the 10 items. Cronbach's alpha for the SWB-HL Total score was 0.85 (95% confidence interval, 0.82–0.89) showing strong internal consistency of the measure. In general, the desired range for the mean inter-item correlation is $0.15 < r_{ii} < 0.50$ (Cronbach, 1984). The mean inter-item correlation for the 10 SWB-HL Total items ranged from 0.35 to 0.38, well within the desired range. Finally, Cronbach's alpha was recomputed several times after each item of the scale was deleted. This is a way to determine if the internal consistency of the scale would be increased by removing a particular item (Cronbach, 1984). For the SWB-HL Total score, Cronbach's alpha failed to increase above 0.85 after removal of any of the 10 items from the scale. Based on these analyses it is concluded that the 10-item SWB-HL Total score is reliable.

## Validity of the SWB-HL

In addition to completing the SWB-HL, as noted above, all participants also completed the full 163-item CPHI, several measures of affect, a personality scale, and a cognitive assessment. It is expected, if the SWB-HL is a valid measure of SWB, that there would be some associations between the SWB-HL and these other measures. This type of validity is generally referred to as criterion validity and is a common approach to validation of self-report measures (Cronbach, 1984). Basically, if the SWB-HL is a valid measure of SWB it would be expected to be related to measures found previously to be associated with SWB.

Linear multiple-regression analysis was used to examine the associations between SWB-HL Total scores and these other measures. First, however, the number of criterion variables was reduced through principal-components factor analysis (varimax rotation). It should be noted that, although several of these scales make use of Likert-type responses, factor analysis does not appear to be impacted by assumptions of ordinal or interval data except for extremely skewed ordinal distributions (Flora and Curran, 2004). When the 23 CPHI scale scores were analyzed, a good fit was obtained and six factors emerged, accounting for 76% of the variance and all communalities >0.62. Four of the six factors matched four of the five factors reported by Demorest and Erdman (1989) for a similar analysis of the CPHI, corresponding to factors of Personal Adjustment (PA), Communication Performance (CP), Communication Importance (CI), and Interactions with Others (Interax). The remaining factor from Demorest and Erdman (1989), Reaction to Hearing Loss, was split between two factors here: Communication Environment, need and physical characteristics (React1); and Communication Strategies, verbal and non-verbal (React2). Of the three datasets analyzed by Demorest and Erdman (1989), one of the three showed a similar 6-factor solution with a split of the Reaction factor along the same lines as observed here. It is concluded that the CPHI factors identified here are consistent with those identified originally by Demorest and Erdman (1989).

When the five measures of affect were subjected to a similar principal-components analysis a single factor emerged accounting for 58% of the variance and four of the five communalities were >0.49. Only the PANAS-positive score had a somewhat lower communality, 0.32, indicating that performance on it may not be quite as well-represented by the single factor score than the other four measures of affect. When the factor loadings for each of the five affect measures on the single factor were examined, negative weights were observed for the two positive measures, LOT-R and PANAS-positive, and positive weights for the three negative affect scores, PANAS-negative, STAI-state, STAI-trait. Thus, higher factors scores

**TABLE 1 |** Rotated factor loadings for the each of the 10 SWB-HL items (principal-axis factor analysis, Oblimin rotation) on the three resulting factors and the general factor, g.

| | g | AccHL | SocSupp | LifeSat |
|---|---|---|---|---|
| 1. I feel close to my friends. | 0.35 | | 0.77 | |
| 2. I get emotional support from my family. | 0.39 | | 0.57 | |
| 3. I get support from my friends. | 0.44 | | 0.77 | |
| 4. My family has accepted my hearing loss. | 0.50 | 0.64 | | |
| 5. I am satisfied with family communication about my hearing loss. | 0.58 | 0.74 | | |
| 8. I am satisfied with how I am coping with my hearing loss. | 0.49 | 0.38 | | |
| 13. My work (include work at home) is fulfilling. | 0.49 | | | 0.50 |
| 15. I have accepted my hearing loss. | 0.47 | 0.38 | | |
| 17. I am enjoying the things I usually do for fun. | 0.62 | | | 0.55 |
| 18. I am content with the quality of my life right now. | 0.62 | | | 0.61 |
| 6. I feel close to my partner (or my main support). | | | | |
| 7. I feel sad. | | | | |
| 9. I am losing hope in the fight against my hearing loss. | | | | |
| 10. I feel nervous. | | | | |
| 11. I worry that my hearing loss will get worse. | | | | |
| 12. I am able to work (include work at home). | | | | |
| 14. I am able to enjoy life. | | | | |
| 16. I am sleeping well. | | | | |

*Only factor loadings > 0.25 are shown for clarity. Items highlighted in gray at the bottom of the table were deleted from the original corpus of 18 items prior to this analysis and are shown for completeness (see text for details.). These factor loadings can be linked to the arrows in **Figure 2** to show the strength of association of each item to the various factors.*

reflect more negative affect. The affect factor score is referred to here as PC_Affect.

Finally, similarly good fits from principal-components factor analysis were obtained with four orthogonal factor scores emerging for the Myers-Briggs Type Inventory (MBTI) personality scale (Myers et al., 1998) and three orthogonal factor scores from the full 13-scale Wechsler Adult Intelligence Scale-3rd Edition (Wechsler, 1997: WAIS-III) cognitive measure. The four personality factors represented the dimensions of introversion/extraversion, perceiving/judging, thinking/feeling, and intuition/sensing at the core of the MBTI personality types. For the WAIS-III, the three factors represented the four major types of cognitive processing identified by the test, verbal comprehension (VC), working memory (WM), perceptual organization (PO), and processing speed (PS), with the latter two combined into a single factor.

A linear multiple-regression analysis was next conducted with the SWB-HL Total score as the dependent measure. The SWB-HL Total score was z-transformed prior to regression analysis. In addition to the orthogonal factor scores for the CPHI, Affect, MBTI, and WAIS-III, all of which have a mean of 0 and standard deviation of 1, age and better-ear four-frequency pure-tone average (PTA4 BtrE) were z-transformed and added as additional independent variables to each regression equation.

**Table 2** provides the results of the regression analyses for the SWB-HL Total score. Predictors with significant ($p < 0.05$) Beta coefficients in the standardized regression equation are highlighted in bold font and the associated zero-order, partial, and part correlations are italicized for those significant contributors. Five significant predictors emerged for the SWB-HL Total score with the solution accounting for 48% of the variance ($r^2$). In addition, collinearity among the predictors was not an issue with the Variance Inflation Factor (VIF) < 1.62 and the Condition Index < 2.7 for all predictors.

The linear multiple-regression analysis summarized in **Table 2** revealed significant associations between various SWB-HL scores and other criterion measures with the pattern reflecting expected associations. Of the five significant predictors in **Table 2**, two were captured by aspects of the CPHI; CPHI PA, a factor related to personal adjustment to hearing loss, and CPHI Interax, a factor representing communication interactions with others. It is not surprising that the CPHI PA scale emerged as a prominent predictor of SWB-HL as this measure includes assessment of self-acceptance, stress, anger, and withdrawal, among others, and would clearly impact one's self-reported SWB (Steel et al., 2008). Similarly, CPHI Interax would likely be tied closely to the social-support items of the SWB-HL scale and social support has a positive impact on SWB (Diener and Seligman, 2002; Siedlecki et al., 2014). Two other significant predictors identified were personality measures, the extraversion/introversion and thinking/feeling dimensions of the MBTI. Personality has long been known to influence general SWB, with consistent ties to extraversion and neuroticism (e.g., Steel et al., 2008; Strickhouser et al., 2017). The MBTI generally does not capture neuroticism (McCrae and Costa, 1989), but many neurotic personality characteristics are captured by several of the CPHI-PA scales as noted above. Finally, the affect factor score was the single-best predictor of SWB-HL Total scores. This reflects the long-time recognition of the strong association between affect and life-satisfaction measures of SWB (Diener, 1984; Diener et al., 1999)

**TABLE 2 |** Results of linear multiple-regression analyses for the SWB-HL Total.

| | Beta | *t* | *p* | Zero-order r | Partial r | Part r |
|---|---|---|---|---|---|---|
| (Constant) | | 103.234 | 0.000 | | | |
| Zscore: AGE | 0.051 | 0.709 | 0.479 | 0.053 | 0.057 | 0.041 |
| Zscore (PTA4 BtrE) | −0.115 | −1.565 | 0.120 | −0.217 | −0.124 | −0.090 |
| **PC_Affect** | **−0.342** | **−4.753** | **0.000** | *−0.548* | *−0.355* | *−0.275* |
| **PC_CPHI_PA** | **0.211** | **3.110** | **0.002** | *0.388* | *0.241* | *0.180* |
| PC_CPHI_CP | 0.039 | 0.578 | 0.564 | 0.116 | 0.046 | 0.033 |
| PC_CPHI_CI | 0.096 | 1.512 | 0.132 | 0.173 | 0.120 | 0.087 |
| **PC_CPHI_Interax** | **0.182** | **2.880** | **0.005** | *0.293* | *0.224* | *0.166* |
| PC_CPHI_React2 | 0.081 | 1.316 | 0.190 | 0.061 | 0.104 | 0.076 |
| PC_CPHI_React1 | 0.050 | 0.808 | 0.420 | −0.012 | 0.064 | 0.047 |
| PC MBTIsense_intuit | 0.064 | 0.908 | 0.365 | 0.123 | 0.072 | 0.052 |
| PC MBTIperceiv_judge | 0.097 | 1.626 | 0.106 | 0.185 | 0.129 | 0.094 |
| **PC MBTIthink_feel** | **0.169** | **2.751** | **0.007** | *0.177* | *0.214* | *0.159* |
| **PC MBTIintrov_extrav** | **0.123** | **1.981** | **0.049** | *0.235* | *0.156* | *0.114* |
| PC W3_VC | −0.075 | −1.045 | 0.298 | 0.086 | −0.083 | −0.060 |
| PC W3_PS_PO | −0.042 | −0.599 | 0.550 | −0.024 | −0.048 | −0.035 |
| PC W3_WM | 0.102 | 1.640 | 0.103 | 0.212 | 0.130 | 0.095 |

*Significant Beta regression coefficients are shown as p-values in bold and r values in italics. Dependent Variable: SWB-HL Total ($r^2$ = 0.48; [$F_{(16,157)}$ = 8.93], p < 0.001). PTA4 BtrE, pure-tone average for 500, 1000, 2000, and 4000 Hz in the better ear; PC, Principal Component; CPHI, Communication Profile of the Hearing Impaired; PA, Personal Adjustment; CP, Communication Performance; CI, Communication Importance; Interax, Interactions with others; React1 and React2 are two components regarding the individual's reaction to hearing impairment; MBTI, Myers-Briggs Type Inventory; W3, Wechsler Adult Intelligence Scale, 3rd Edition; VC, Verbal Comprehension; PS PO, Processing Speed/Perceptual Organization; WM, Working Memory.*

as well as the link between affect and hearing-loss-related quality of life (Preminger and Meeks, 2010).

Perhaps as interesting are the factors that proved not to be associated with SWB-Total performance. These include the severity of hearing loss, age, and cognition. Associations between cognitive function and general SWB have been observed previously (e.g., Jones et al., 2003; Siedlecki et al., 2020) but these associations may be mediated by several other factors (Yazdani and Siedlecki, 2020). Although the age range included here was restricted to older adults and no effect of age over this range of 60–88 years was observed here, SWB has been found to be relatively stable over the adult lifespan (Stone et al., 2010; Braun et al., 2017). Sensory loss, including hearing loss, has been associated with poorer SWB (Ciorba et al., 2012; Tseng et al., 2018) but no effects were observed here. A likely reason for the failure of hearing loss severity to emerge as a significant predictor in these analyses may be found in the limited range of hearing loss in this sample (**Figure 1**). This homogeneity of hearing loss severity may, in turn, have contributed to the inability of the CPHI Communication Performance (CP) and Communication Importance (CI) scales to predict SWB-HL Total scores. Regression analyses depend on sufficient variation in independent and dependent variables to identify significant predictors and the homogeneity of hearing loss and perceived communication performance among this sample may have impacted the results of those analyses. Nonetheless, except for the absence of the impact of hearing loss, those measures found to be significant predictors of SWB-HL Total scores, as well as those identified as not being predictive of performance, reveal a pattern consistent with the expectations for a general measure of SWB primed by probes of hearing loss.

## Aided and Unaided SWB-HL Scores

Another form of validation which focuses on the sensitivity of the instrument is to examine the scores before and after intervention (Ventry and Weinstein, 1983). As noted in Methods, 143 of the 173 participants were fitted with hearing aids and of these 143, 141 had complete pre-fit and post-fit SWB-HL data. The mean SWB-HL Total score was 31.32 (SD = 5.16) prior to the hearing-aid fit and 32.84 (SD = 5.07) after 4–6 weeks of hearing aid use. A paired-sample *t*-test showed this difference to be significant [$t_{(140)} = 4.29$, $p < 0.001$] with a medium effect size (Cohen's $d = 0.36$, 95% CI = 0.19–0.53). Significant improvements in SWB were observed following 4–6 weeks of hearing-aid use with the SWB-HL Total.

Perhaps the aided improvement in SWB-HL Total score is carried exclusively by the four items of the SWB-HL in the Acceptance of Hearing Loss domain. This pattern would be consistent with the prior literature reviewed in the Introduction suggesting that only domain-specific improvements are expected in SWB measures. Given that four of the 10 items in the SWB-HL Total score make queries about the impact of hearing loss perhaps aided improvements of these four items are sufficient to result in the observed improvement in SWB-HL Total score with amplification. To examine this, a 6-item SWB score was computed from the three Life Satisfaction and the three Social Support items. These non-HL SWB scores were then compared for the pre-fit and post-fit measurements. The mean SWB-HL Total "non-hearing loss" score was 19.45 (SD = 3.17) prior to the hearing-aid fit and 19.93 (SD = 3.28) after 4–6 weeks of hearing aid use. A paired-sample $t$-test showed this difference to be significant [$t_{(140)} = 2.33$, $p < 0.05$] but with a small effect size (Cohen's $d = 0.20$, 95% CI = 0.03–0.36). When just the four items tapping Acceptance of Hearing Loss were totaled for evaluation pre- and post-fit, the post-fit mean (12.91, SD = 2.57) was significantly greater [$t_{(140)} = 4.35$, $p < 0.001$] than the pre-fit mean (11.87, SD = 3.08) and a medium effect size was observed ($d = 0.37$). Given that the hearing aid intervention addresses the hearing loss of the respondent, and consistent with the prior literature on the demonstration of domain-specific improvements in SWB following hearing-aid use, the increased score for SWB-HL items tapping Acceptance of Hearing Loss is expected. More importantly, the increased score for the 6-item SWB-HL score calculated for the general SWB items reflects broader improvements in SWB from amplification in older adults with hearing loss. Again, such subscale scores of the SWB-HL are not being recommended here for use but as a means to demonstrate that hearing aid use impacted more than just the hearing-related items of the SWB-HL.

The Pearson-r correlation between pre-fit and post-fit SWB-HL Total scores was 0.66 and statistically significant ($p < 0.001$). This correlation, coupled with the increase in mean performance for the group, suggest that the general trend across individuals was for the use of hearing aids for this 4–6-week period to improve the measured SWB. **Figure 3** shows a scatterplot of the individual unaided and aided SWB-HL Total scores. The best-fitting linear regression equation is plotted in this figure as a dashed red line with 95% confidence intervals around this best-fitting line shown as blue solid lines. As can be seen, those with lower SWB pre-fit showed the largest improvements in SWB post-fit. Generally, if pre-fit SWB was good, SWB-HL Total > 30, then improvements in SWB with hearing aids were smaller and less frequent. Twenty-five percent of the 141 individuals showed an improvement of four points or more in the SWB-HL Total score which can range from 0 to 40.

As noted previously, of the 141 individuals with complete SWB-HL unaided and aided scores, 43 had worn hearing aids at some point in the past and 30 of those 43 were current hearing aid users when they enrolled in this study. As a result, assuming a positive impact of hearing aids on SWB, those with prior hearing aid use may have had higher pre-fit scores and would have shown less improvement in SWB over the 4–6-week course of this study. To determine whether inclusion of those with prior hearing aid experience impacted the analyses of pre- and post-fit SWB-HL Total scores, the paired-sample $t$-tests were again performed for the 98 who had never worn hearing aids and the 111 who were not currently using hearing aids. The pattern of significant differences described previously for the entire group of 141 was the same for the subgroups of 98 and 111 older adults with prior hearing aid experience. The same is true for the magnitude and significance of the correlations between the post-fit and pre-fit SWB-HL Total scores as well as the Cohen's d effect sizes for the aided improvements in SWB-HL Total.

Clearly, based on the data and correlation in **Figure 3**, the best predictor of *aided* SWB-HL scores is most likely the *unaided* SWB-HL score. To examine this further, the multiple linear-regression analysis conducted for the pre-fit SWB-HL Total scores was repeated for the post-fit scores but with the addition of the pre-fit SWB-HL and a hearing-aid expectations factor representing the four scales of the Expected Consequences of Hearing Aid Ownership (ECHO; Cox and Alexander, 2000). SWB-HL Total scores, pre-fit and post-fit, were converted to z-scores for these analyses. Conversion of several measures to z-scores matched the means and standard deviations of these z-transformed variables to those of the various factor scores used as predictors such that all variables in the analyses had means of 0 and standard deviations of 1.

With the z-transformed post-fit (aided) SWB-HL-Total score as the dependent measure, the best-fitting regression solution accounted for 50.8% of the variance but only had one significant predictor variable, the z-transformed pre-fit (unaided) SWB-HL Total score. Thus, these analyses with the inclusion of a broad range of potential predictors confirm the relationship between pre-fit and post-fit SWB-HL-Total scores illustrated previously in **Figure 3**. Aided SWB is largely determined by the pre-fit unaided SWB.

If the unaided SWB-HL Total score is not included among the independent measures and the regression analysis predicting aided SWB-HL Total score is repeated, only two significant predictors emerged: affect ($r = -0.42$) and ECHO ($r = 0.29$). This solution, however, only accounted for 34% of the variance compared to nearly 51% when unaided SWB-HL scores were included. The fact that neither affect or ECHO were significant predictors when the unaided SWB-HL total scores were included implies that the unaided SWB-HL Total score captures aspects of the individual's self-reported affect and hearing-aid expectations.

As demonstrated above, pre-fit, unaided SWB, as captured by the SWB-HL Total score, was largely determined by personality, affect, and the individual's adjustment to hearing loss, both in terms of the individual's affect and his or her interactions with others. Of course, personality, is a factor that would not be considered malleable by the audiologist to improve pre-fit and, consequently, post-fit SWB. The CPHI measures, on the other hand, can be shaped by the clinician through aural rehabilitation and counseling. The CPHI, in fact, was developed as an assessment tool to determine the focus of subsequent aural rehabilitation (Demorest and Erdman, 1986, 1987). Affect, both generally and as impacted by hearing loss, is also something that could be modified potentially through counseling. Whicker et al. (2020) recently noted associations between various psychological measures and SWB and encouraged audiologists to take a more active role in shaping the thoughts and emotions of patients regarding their hearing loss. The present findings suggest that doing so will enhance SWB, both unaided and aided, at least as measured by the SWB-HL.

In summary, the foregoing analyses of SWB-HL scale scores for unaided and aided listening conditions further validate this measure. When amplification was fit to these older adults, the SWB-HL Total score demonstrated sufficient sensitivity to support improved SWB after the 4–6-week hearing-aid trial. Regression analyses identified areas of focus for rehabilitation that may lead to enhanced SWB with amplification in older adults.

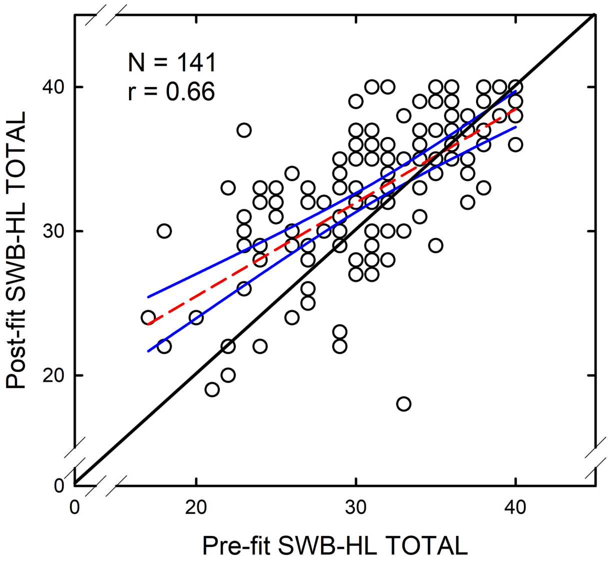


**FIGURE 3 |** Scatterplot of aided SWB-HL Total scores against unaided SWB-HL Total scores for 141 older adults fitted with hearing aids in this study. The correlation between these two scores was $r = 0.66$ ($p < 0.001$). Each circle represents an individual data point. The solid black line along the diagonal represents equality of the aided and unaided scores. The dashed red line is the best-fitting linear-regression equation and the blue lines represent 95% confidence intervals around that equation.

## CONCLUSIONS

The final 10-item version of the SWB-HL yielded a total score linked to the constructs of Life Satisfaction, Social Support, and Acceptance of Hearing Loss. This report documented the reliability and validity of the SWB-HL Total score. Generally, the greater the number of items comprising a test, the more reliable the result. As such, use of the SWB-HL Total score is recommended, although it is possible to get more specific information about SWB in older adults with hearing loss by examining the individual subscales. Regarding validity, the SWB-HL Total score was associated with criterion measures administered separately, especially the measures of Personal Adjustment from the CPHI, personality, and affect. Thus, the 10-item SWB-HL Total score captures both general and hearing-loss-specific components of SWB. Further demonstration of the validity of the 10-item SWB-HL Total was demonstrated through significant differences and medium effect sizes for comparisons before and after hearing-aid use. It is noteworthy that such differences in SWB were measurable following a 4–6-week period of hearing aid use. Regression analyses resulted in a potential model of the key determiners of SWB in older adults with hearing loss. This model requires further evaluation and validation.

A limitation of this study is the restricted nature of the study sample, both in terms of size and demographic characteristics (e.g., white, well-educated, and reasonable annual income). In addition, we noted the importance of item context in the Introduction, including the potential impact of priming. The final 10 items comprising the SWB-HL were extracted from the initial larger set of 18 items. It is unknown how the eight items that have been eliminated from the original corpus of 18 may have impacted the responses for the 10 items that remained. Further research is needed to overcome these limitations. In addition, these context effects should be kept in mind by developers of future questionnaires or surveys in the field.

## 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 Indiana University Bloomington IRB. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

LH performed all the data analyses described here and wrote the manuscript without assistance.

## FUNDING

This work was supported, in part, by a research grant from the National Institute on Aging, R01 AG008293, awarded to LH.

## ACKNOWLEDGMENTS

The author thanks Dana Kinney, Elizabeth Thompson, and several graduate students for their assistance in data collection as well as Melissa Su Ching Lee and Stephanie Dickinson for their help with data analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.640165/full#supplementary-material

## REFERENCES

Abrams, H. B., Chisolm, T. H., and McArdle, R. (2005). Health-related quality of life and hearing aids: a tutorial. *Trends Amplif.* 9, 99–109. doi: 10.1177/108471380500900302

Bertrand, M., and Mullainathan, S. (2001). Do people mean what they say? implications for subjective survey data. *Am. Econ. Rev.* 91, 67–108. doi: 10.1257/aer.91.2.67

Bess, F. H. (2000). The role of generic health-related quality of life measures in establishing audiological rehabilitation outcomes. *Ear Hear.* 21, 74S–79S. doi: 10.1097/00003446-200008001-00009

Bess, F. H., Lichenstein, M. J., and Logan, S. A. (1991). Making hearing impairment functionally relevant: linkage with hearing disability and handicap. *Acta. Otolaryngol.* 476, 226–231. doi: 10.3109/00016489109127283

Bess, F. H., Lichtenstein, M. J., Logan, S. A., Burger, M. C., and Nelson, E. (1989). Hearing impairment as a determinant of function in the elderly. *J. Am. Geriatr. Soc.* 37, 123–128. doi: 10.1111/j.1532-5415.1989.tb05870.x

Bowling, A., and Windsor, J. (2008). The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA). *J. Epidemiol. Community Health* 62, 81–85. doi: 10.1136/jech.2006.058214

Braun, T., Schmukle, S. C., and Kunzmann, U. (2017). Stability and change in subjective well-being: the role of performance-based and self-rated cognition. *Psychol. Aging* 32, 105–117. doi: 10.1037/pag0000153

Cassarly, C., Matthews, L. J., Simpson, A. N., and Dubno, J. R. (2020). The revised hearing handicap inventory and screening tool based on psychometric reevaluation of the hearing handicap inventories for the elderly and adults. *Ear Hear.* 41, 95–105. doi: 10.1097/AUD.0000000000000746

Cella, D. F., Tulsky, D. S., Gray, G., Saraflan, B., Linn, E., Bonomi, A., et al. (1993). The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J. Clin. Oncol.* 11, 570–579. doi: 10.1200/JCO.1993.11.3.570

Chia, E.-M., Wang, J. J., Rochtchina, E., Cumming, R. R., Newall, P., and Mitchell, P. (2007). Hearing impairment and health-related quality of life: the Blue Mountains Hearing Study. *Ear Hear.* 28, 187–195. doi: 10.1097/AUD.0b013e31803126b6

Chisolm, T. H., Abrams, H. B., McArdle, R., Wilson, R. H., and Doyle, P. J. (2005). The WHO-DAS II: psychometric properties in the measurement of functional health status in adults with acquired hearing loss. *Trends Amplif.* 9, 111–126. doi: 10.1177/108471380500900303

Chisolm, T. H., Johnson, C. E., Danhauer, J. L., Portz, L. J. P., Abrams, H., Lesner, S., et al. (2007). A systematic review of health-related quality of life and hearing aids: final report of the American Academy of Audiology task force on the health-related quality of life benefits of amplification in adults. *J. Am. Acad. Audiol.* 18, 151–183. doi: 10.3766/jaaa.18.2.7

Ciorba, A., Bianchini, C., Pelucchi, S., and Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clin. Interv. Aging* 7, 159–163. doi: 10.2147/CIA.S26059

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cox, R. M., and Alexander, G. C. (2000). Expectations about hearing aids and their relationship to fitting outcome. *J. Am. Acad. Audiol.* 11, 368–382.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555

Cronbach, L. J. (1984). *Essentials of Psychological Testing, 4th Edn*. Manhattan, NY: Harper and Row.

Dalton, D. S., Cruickshanks, K. J., Klein, B. E., Klein, R., Wiley,T. L., and Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *Gerontologist* 43, 661–668. doi: 10.1093/geront/43.5.661

Dawes, P., Cruickshanks, K. J., Fischer, M. E., Klein, B. E. K., Klein, R., and Nondahl, D. M. (2015). Hearing-aid use and long-term health outcomes: hearing handicap, mental health, social engagement, cognitive function, physical health, and mortality. *Int. J. Audiol.* 54, 838–844. doi: 10.3109/14992027.2015.1059503

De Winter, J. F. C., and Dodou, D. (2010). Five-point Likert items: t test versus Mann-Whitney-Wilcoxon. *Pract. Assess. Res. Eval.* 15, 1–12. doi: 10.7275/bj1p-ts64

Demorest, M. E., and Erdman, S. A. (1986). Scale composition and item analysis of the Communication Profile for the Hearing Impaired. *J. Speech Hear. Res.* 29, 515–535. doi: 10.1044/jshr.2904.535

Demorest, M. E., and Erdman, S. A. (1987). Development of the communication profile for the hearing impaired. *J. Speech Hear. Disord.* 52, 129–143. doi: 10.1044/jshd.5202.129

Demorest, M. E., and Erdman, S. A. (1989). Factor structure of the Communication Profile for the Hearing Impaired. *J. Speech Hear. Disord.* 54, 541–549. doi: 10.1044/jshd.5404.541

Diener, E. (1984). Subjective well-being. *Psychol. Bull.* 95, 542–575. doi: 10.1037/0033-2909.95.3.542

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13

Diener, E., and Seligman, M. E. P. (2002). Very happy people. *Psychol. Sci.* 13, 81–84. doi: 10.1111/1467-9280.00415

Diener, E., Suh, E. M., Lucas, R. E., and Smith, H. L. (1999). Subjective well-being: three decades of progress. *Psychol. Bull.* 125, 276–302. doi: 10.1037/0033-2909.125.2.276

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., et al. (2010). New well-being measures: short scales to assess flourishing and positive and negative feelings. *Soc. Indic. Res.* 97, 143–156. doi: 10.1007/s11205-009-9493-y

Flora, D. B., and Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Meth*ods 9:466. doi: 10.1037/1082-989X.9.4.466

Garbarski, D., Schaeffer, N. C., and Dykema, J. (2015). Effects of response option order and question order on self-rated health. *Qual. Life Res.* 24, 1443–1453. doi: 10.1007/s11136-014-0861-y

Gehlbach, H., and Barge, S. (2012). Anchoring and adjusting in questionnaire responses. *Basic Appl. Soc. Psychol.* 34, 417–433. doi: 10.1080/01973533.2012.711691

Gorsuch, R. (1983). *Factor Analysis, 2nd Edn*. Mahwah, NJ: Lawrence Erlbaum Associates.

Heffernan, E., Maidment, D. W., Barry, J. G., and Ferguson, M. (2019). Refinement and validation of the Social Participation Restrictions Questionnaire: an application of Rasch analysis and traditional psychometric analysis techniques. *Ear Hear.* 40, 328–339. doi: 10.1097/AUD.0000000000000618

Humes, L. E., Ahlstrom, J. B., Bratt, G. W., and Peek, B. F. (2009). Studies of hearing aid outcome measures in older adults: a comparison of technologies and an examination of individual differences. *Semin. Hear.* 30, 112–128. doi: 10.1055/s-0029-1215439

Humes, L. E., Kinney, D. L., and Thompson, E. (2010). "Comparison of benefits provided by various hearing aid technologies in older adults," in *Hearing Care for Adults 2009-The Challenge of Aging*, ed L. Hickson (Stafa: Phonak AG), 131–138.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Med. Educ.* 38, 1217–1218. doi: 10.1111/j.1365-2929.2004.02012.x

Jones, T. G., Rapport, L. J., Hanks, R. A., Lichtenberg, P. A., and Telmet, K. (2003). Cognitive and psychosocial predictors of subjective well-being in urban older adults. *Clin. Neuropsychol.* 17, 3–18. doi: 10.1076/clin.17.1.3.15626

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., and Redelmeier, D. A. (1993). When more pain is preferred to less: adding a better end. *Psychol. Sci.* 4, 401–405. doi: 10.1111/j.1467-9280.1993.tb00589.x

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., and Stone, A. (2004). Toward national well-being accounts. *Am. Econ. Rev.* 94, 429–434. doi: 10.1257/0002828041301713

Liddell, T. M., and Kruschke, J. K. (2018). Analyzing ordinal data with metric models: what could possibly go wrong? *J. Exp. Soc. Psychol.* 79, 328–348. doi: 10.1016/j.jesp.2018.08.009

Lucas, R. E. (2018). "Reevaluating the strengths and weaknesses of self-report measures of subjective well-being," in *Handbook of Well-Being*, eds E. Diener, S. Oishi, and L. Tay (Salt Lake City, UT: DEF Publishers).

McArdle, R., Chisolm, T. H., Abrams, H. B., Wilson, R. H., and Doyle, P. J. (2005). The WHO-DAS-II: measuring outcomes of hearing aid intervention for adults. *Trends Amplif.* 9, 127–143. doi: 10.1177/108471380500900304

McCrae, R. R., and Costa, P. T. Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. *J. Pers.* 57, 17–40. doi: 10.1111/j.1467-6494.1989.tb00759.x

Mulrow, C. D., Aguilar, C., Endicott, J. E., Tuley, M. R., Velez, R., Charlip, W. S., et al. (1990). Quality-of-life changes and hearing impairment: a randomized controlled trial. *Ann. Intern. Med.* 113, 188–194. doi: 10.7326/0003-4819-113-3-188

Myers, I. B., McCaulley, M. H., Quenk, N., and Hammer, A. (1998). *MBTI Handbook: A Guide to the Development and Use of the Myers-Briggs Type Indicator, 3rd Edn.* Palo Alto, CA: Consulting Psychologists Press.

Nordvik, O., Laugen Heggdal, P. O., Brannstrom, J., Hjermstad, M. J., Aatstad, A. K., and Aarstad, H. J. (2019). Quality of life in persons with hearing loss: a study of patients referred to an audiological service. *Int. J. Audiol.* 58, 696–703. doi: 10.1080/14992027.2019.1627010

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ. Theory Pract.* 15, 625–632. doi: 10.1007/s10459-010-9222-y

Preminger, J. E., and Meeks, S. (2010). The influence of mood on the perception of hearing-loss related quality of life in people with hearing loss and their significant others. *Int. J. Audiol.* 49, 263–271. doi: 10.3109/14992020903311396

Revelle, W. (2020). *Package "Psych." R Procedures for Psychological, Psychometric, and Personality Research. Version 2.0.12, 14 December 2020.* Evanston, IL: Northwestern University.

Scheier, M. F., and Carver, C. S. (1992). Effects of optimism on psychological and physical well-being: theoretical overview and empirical update. *Cogn. Ther. Res.* 16, 201–228. doi: 10.1007/BF01173489

Scheier, M. F., Carver, C. S., and Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a re-evaluation of the Life Orientation Test. *J. Pers. Soc. Psychol.* 67, 1063–1078. doi: 10.1037/0022-3514.67.6.1063

Schwarz, N., and Strack, F. (1991). "Context effects in attitude surveys: applying cognitive theory to social research," in *European Review of Social Psychology, Vol. 2*, eds W. Stroebe and M. Hewstone (Chichester: Wiley), 31–50.

Scollon, C. N., Kim-Prieto, C., and Diener, E. (2003). Experience sampling: promises and pitfalls, strengths and weaknesses. *J. Happiness Stud.* 4, 5–34. doi: 10.1023/A:1023605205115

Siedlecki, K. L., Salthouse, T. A., Oishi, S., and Jeswani, S. (2014). The relationship between social support and subjective well-being across age. *Soc. Indic. Res.* 117, 561–576. doi: 10.1007/s11205-013-0361-4

Siedlecki, K. L., Yazdani, N., Minahan, J., and Falzarano, F. (2020). Examining processing speed as a predictor of subjective well-being across age and time in the German Aging Survey. *Aging Neuropsychol. Cogn.* 27, 66–82. doi: 10.1080/13825585.2019.1585514

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Test Manual for the State Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Steel, P., Schmidt, J., and Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychol. Bull.* 134, 138–161. doi: 10.1037/0033-2909.134.1.138

Stone, A. A., Schwartz, J. E., Broderick, J. E., and Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9985–9990. doi: 10.1073/pnas.1003744107

Stone, A. A., Shiffman, S. S., and DeVries, M. W. (1999). "Ecological momentary assessment," in *Well-being: The Foundations of Hedonic Psychology*, eds D. Kahneman, E. Diener, and N. Schwarz (New York, NY: Russell Sage Foundation), 26–39.

Strickhouser, J. E., Zell, E., and Krizan, Z. (2017). Does personality predict health and well-being? a metasynthesis. *Health Psychol.* 36, 797–810. doi: 10.1037/hea0000475

Tseng, Y.-C., Liu, S. H.-Y., Lou, M.-F., and Huang, G.-S. (2018). Quality of life in older adults with sensory impairments: a systematic review. *Qual. Life Res.* 27, 1957–1971. doi: 10.1007/s11136-018-1799-2

United Nations (2020). *Resolution 75/131, United Nations Decade of Healthy Ageing (2021-2030). Adopted by the U.N. General Assembly on 14 December 2020*. New York, NY: United Nations.

Ventry, I. M., and Weinstein, B. E. (1983). Identification of elderly people with hearing problems. *ASHA* 25, 37–42.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale, 3rd Edn (WAIS-III)*. New York, NY: The Psychological Corporation.

Weinstein, B. E., Sirow, L. W., and Moser, S. (2016). Relating hearing aid use to social and emotional loneliness in older adults. *Am. J. Audiol.* 25, 54–61. doi: 10.1044/2015_AJA-15-0055

Whicker, J. J., Ong, C. W., Munoz, K., and Twohig, M. P. (2020). The relationship between psychological processes and indices of well-being among adults with hearing loss. *Am. J. Audiol.* 29, 728–737. doi: 10.1044/2020_AJA-20-00035

World Health Organization (2015). *World Report on Ageing and Health*. Geneva: World Health Organization.

World Health Organization (2021). *World Report on Hearing*. Geneva: World Health Organization.

Yazdani, N., and Siedlecki, K. L. (2020). Mediators of the relationship between cognition and subjective well-being. *J. Happiness Stud.* doi: 10.1007/s10902-021-00357-6. [Epub ahead of print].


**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 in Psychology

PERSPECTIVE
published: 05 November 2021
doi: 10.3389/fpsyg.2021.733060

# Clinical Trials and Outcome Measures in Adults With Hearing Loss

*Kevin J. Munro[1,2]*, William M. Whitmer[3,4] and Antje Heinrich[1]*

[1] *Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, United Kingdom,* [2] *Manchester Academic Health Science Centre, Manchester University Hospitals NNS Foundation Trust, Manchester, United Kingdom,* [3] *Hearing Sciences–Scottish Section, Division of Clinical Neuroscience, University of Nottingham, Glasgow, United Kingdom,* [4] *Institute of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom*

Clinical trials are designed to evaluate interventions that prevent, diagnose or treat a health condition and provide the evidence base for improving practice in health care. Many health professionals, including those working within or allied to hearing health, are expected to conduct or contribute to clinical trials. Recent systematic reviews of clinical trials reveal a dearth of high quality evidence in almost all areas of hearing health practice. By providing an overview of important steps and considerations concerning the design, analysis and conduct of trials, this article aims to give guidance to hearing health professionals about the key elements that define the quality of a trial. The article starts out by situating clinical trials within the greater scope of clinical evidence, then discusses the elements of a PICO-style research question. Subsequently, various methodological considerations are discussed including design, randomization, blinding, and outcome measures. Because the literature on outcome measures within hearing health is as confusing as it is voluminous, particular focus is given to discussing how hearing-related outcome measures affect clinical trials. This focus encompasses how the choice of measurement instrument(s) affects interpretation, how the accuracy of a measure can be estimated, how this affects the interpretation of results, and if differences are statistically, perceptually and/or clinically meaningful to the target population, people with hearing loss.

**Keywords: clinical trials, outcome measures, minimal important difference, interventions, hearing loss, hearing-related outcomes, clinically meaningful**

OPEN ACCESS

***Edited by:***
*Astrid van Wieringen, KU Leuven, Belgium*

***Reviewed by:***
*Paul Van de Heyning, University of Antwerp, Belgium*
*Philippe Fournier, Aix-Marseille Université, France*

****Correspondence:***
*Kevin J. Munro*
*kevin.munro@manchester.ac.uk*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *29 June 2021*
***Accepted:*** *20 October 2021*
***Published:*** *05 November 2021*

***Citation:***
*Munro KJ, Whitmer WM and Heinrich A (2021) Clinical Trials and Outcome Measures in Adults With Hearing Loss. Front. Psychol. 12:733060. doi: 10.3389/fpsyg.2021.733060*

## INTRODUCTION

Clinical trials are a type of research that study health interventions and evaluate their effects on human health outcomes (World Health Organisation, 2018). James Lind is credited for conducting the first clinical trial in humans (see for example, Collier, 2009). In 1747, Lind investigated different treatments for scurvy. He demonstrated that, in sailors living under the same conditions, it was only those who were provided with fruit (specifically, Vitamin C) that recovered. The purpose of the intervention in a clinical trial might be to prevent, diagnose or, in the case of Lind, treat a health condition. The conduct and quality of clinical trials is critical since they provide the evidence base for improving practice in health care. Many health professionals, including those working within or allied to hearing health, are expected to conduct or contribute to clinical trials.

Grading of Recommendations, Assessment, Development, and Evaluations (GRADE; Atkins et al., 2004) is a framework commonly used to assess quality of evidence based on study limitations, inconsistency, indirectness, imprecision, and publication bias, e.g., outcomes from non-randomized studies without blinding would be considered low. As with many areas of healthcare, systematic reviews in hearing science and audiology have highlighted a dearth of good quality clinical trials. Notable in this context are reviews published by the National Institute for Health and Care Excellence (NICE), a public body sponsored by the United Kingdom government that provides evidence to improve health and social care, and the Cochrane Database of Systematic Reviews (CDSR), the leading journal and database for systematic reviews in health care:

1. NICE published reviews as part of national guidelines (NG) on assessment and management of adult hearing loss (NG98; National Institute for Health and Care Excellence, 2018), and assessment and management of tinnitus (NG155; National Institute for Health and Care Excellence, 2020). Both guidelines include around 20 systematic reviews on areas of uncertainty or variation in clinical practice. 50–60% of the systematic reviews revealed no evidence on which to base clinical recommendations. The remaining 40–50% of the systematic reviews identified supporting evidence; however, the quality of the individual studies was mostly graded as low due to risk of bias (see later).
2. CDSR published a review of the effects of hearing aids in everyday life for people with mild to moderate hearing loss. This revealed benefits; however, the evidence was based on five studies, and their quality was graded as moderate (Ferguson et al., 2017).

The current article redresses the limited evidence base by providing an overview of the design, analysis and conduct of clinical trials. Judicious use of selected studies highlight potential methodological limitations as well as examples of good practice. The aim is to provide guidance to hearing health professionals about the key elements that define the quality of a trial. Detailed information is provided on outcome measures, and on how hearing-related outcome measures affect clinical trials: (i) how the choice of measurement instrument(s) affects interpretation, (ii) how the accuracy of a measure can be estimated, (iii) how this affects the interpretation of results, and (iv) if differences are statistically, perceptually and/or clinically meaningful to the target population, people with hearing loss.

## IDENTIFYING THE RESEARCH QUESTION

Before designing a clinical trial, an essential starting point is to craft a carefully worded research question. Evidence-based medicine provides an explicit framework for formulating research questions that can be used when: (i) designing clinical trials or (ii) searching the literature for studies to be included in a systematic review of the literature. The four components of the question are contained in the PICO mnemonic: Population (P), Intervention (I), Comparator (C), and Outcome (O).

An example of a research question in the PICO format would be, "What is the clinical- and cost-effectiveness [outcome] of monitoring and follow-up regimes [intervention] for adults offered NHS hearing aids for the first time [population], compared to usual care [comparator]. The same approach was used by NICE when preparing the clinical guidelines mentioned earlier.

## THE HIERARCHY OF EVIDENCE

Hierarchies of evidence, developed to aid the interpretation and evaluation of research findings, are a core principal of Evidence-Based Practice (EBP). They rank research according to its validity, and in particular, risk of bias. While many research study designs exist (e.g., cohort, case-controlled, cross-sectional and case series/reports), well conducted randomized controlled trials (RCT) are generally considered the gold standard because they provide the lowest risk of bias and, hence, the highest quality of evidence. The first step to building high-quality evidence for clinical practice should always be a recent well-conducted systematic review following a standardized reporting method such as the Preferred Reporting Items for Systematic Reviews (PRISMA).[1] An alternative design to a RCT is an observational study, so called because the researcher observes individuals without manipulation or intervention. These can be useful in instances where RCTs are not appropriate. For example, the effectiveness of parachutes has not been proven in a RCT where participants are randomized to parachute or placebo (Smith and Pell, 2003). In this example, the effect size would be very large because death and serious trauma is much more likely in the placebo group. However, when effect sizes are smaller (which applies to the vast majority of questions), confounds and bias may distort the effect size. In such cases, all efforts should be made to set up an RCT. To appreciate the potential disadvantages of observational designs compared to a RCT trial, consider the following study by Noble and Gatehouse (2006). They used an observational design to compare existing adult hearing aid users of bilateral or unilateral hearing aids. Their results showed that bilateral hearing aids offer advantages in demanding and dynamic listening situations that were not conferred by unilateral hearing aids. However, due to the design it is not possible to know if the natural selection of groups introduced a bias and led to a miss-estimation of the effect.

Systematic errors have the potential to result in the wrong conclusions about the effects of the intervention. The risk of systematic errors differs between designs and is more likely for observational designs than RCT. Two types of systematic errors are biases and confounds. An example of an experimental confound is age. If, for example, a higher proportion of older

[1] http://www.prisma-statement.org

people receive the intervention than the control, age-related differences, unrelated to the intervention, could affect the results. An example of bias is when researchers *or* participants expect the new intervention to generate a better outcome. For example, Dawes et al. (2011, 2013) examined the effect of participant expectation when comparing two hearing aids that were identical except one was labeled "new" and the other "conventional." Mean performance with the hearing aid labeled "new" was significantly higher on all outcome measures. These studies demonstrate that placebo effects can, and do, affect hearing aid trials. Initial preferences can dominate outcomes, as shown in hearing-aid RCTs investigating unilateral and bilateral fittings. For example, Cox et al. (2011) showed that 80% of participants could be predicted based on initial preference for one or two hearing aids. Additionally, Naylor et al. (2015) demonstrated that the outcome for the same technology was influenced by how involved the participant was in the fitting process. Therefore, measuring preferences and attitudes related to the intervention should be included to help control for such confounds in the analysis. Another set of biases are performance and detection biases when systematic differences exist between groups in terms of care and measurement of outcomes, which can be minimized through blinding. By reducing the risk of confounds and bias, any difference in outcome at the end of the trial can be more robustly attributed to the intervention.

Clinical trials in humans are commonly classified into four types or "phases," depending on their aim. Within a trial, there are typically four stages to its preparation and operation: pre-trial, trial set-up, during trial and end of trial. **Table 1** details the phases and gives examples of activities carried out at each stage of any clinical trial. Hackshaw (2009) provides a comprehensive overview of the design, conduct and analysis of trials, ideal for busy health professionals who read or undertake clinical research.

**TABLE 1 |** Phases (types) of clinical trials and examples of key activities in each clinical trial process.

| Phase | Explanation |
|---|---|
| One | An *exploratory* investigation of safety and the effects of dosage in a small number of healthy participants. |
| Two | A *preliminary* estimate of efficacy (i.e., the potential of the intervention to provide benefit), in a small number of participants with the specific health condition. |
| Three | A *definitive* trial of effectiveness, involving a relatively large number of participants who are randomized to the intervention(s) or control. |
| Four | *Monitors* side effects and how well the intervention works over a longer period and in a very large number of participants. |
| **Stage** | **Key activities** |
| Pre-trial | • Formulate the PICO research question<br>• Design trial |
| Trial set-up | • Protocol and ethics<br>• Operating procedures, including case report forms for collecting de-identified data<br>• Set-up site(s)<br>• Register trial |
| The trial | • Collect and store data<br>• Regularly review for protocol adherence |
| End of trial | • Lock database and undertake statistical analysis<br>• Identify and deal with missing data<br>• Disseminate trial findings |

## METHODOLOGICAL CONSIDERATIONS

In order to ensure that clinical trials are executed well, some key methodological issues need to be considered. These include design, randomization, and blinding; all three pose particular challenges to running a hearing-specific RCT.

### Design

A cardinal decision in every clinical trial is the choice of design. Fundamentally, the research team has the decision between two designs: a crossover design where participants receive all interventions in a randomized (or counter-balanced) order, or a parallel-group design where participants are randomized to a single intervention (**Figure 1**). An advantage of the crossover design is that it is a within-groups design: each participant acts as their own control, increasing statistical power. In hearing studies, where the emphasis is usually less on cure and more on benefit and quality of life (QoL), our preference is judicious use of crossover trials. Marriage et al. (2004) used a crossover design when comparing three prescriptions for hearing aid gain settings; there was, however, an issue in its crossover design: tolerance for greater gain increased over the course of the trial regardless of intervention order. Hearing-aid studies using a crossover design often do not include washout periods (Arlinger et al., 2008; cf. Cox et al., 2016) which may reduce carryover effects from one intervention to the next. For hearing training and support interventions, crossover randomization would confound the effect of the intervention with its order (i.e., outcomes following a training period would not be expected to be equivalent to outcomes preceding a training period), hence parallel designs have been used (e.g., Meijerink et al., 2020). A parallel group design contains more natural variation, making it harder to know whether any variation in results is due to the intervention or differences between the participants in the groups. Humes et al., 2017 used a parallel design to study the efficacy of generically fit hearing aids vs. individually fit and placebo devices, randomly assigning participants to one of the three arms. The population to be tested also needs to be considered; in interventions with hearing-aid users, for example, halo effects may lead to greater effects for new compared to experienced users (Ivory et al., 2009).

### Randomization

In an RCT, a sample of participants from the population of interest are randomly allocated to receive the experimental or control/comparator intervention (the latter may be "usual care" or a placebo). The purpose of randomization is to reduce systematic differences in the characteristics of participants allocated to each group. In the case of Lind's scurvy trial, the population of interest (sailors with scurvy) were randomly allocated to receive interventions including seawater, nutmeg and garlic and fruit. In hearing science, within-group crossover

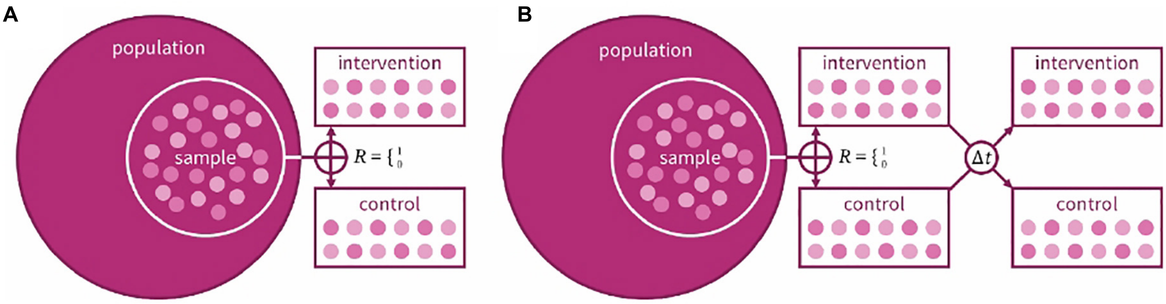


**FIGURE 1 |** Schematic example of a **(A)** parallel and **(B)** crossover design for a two-arm clinical trial. In both designs, a sample is taken from the target population, randomly allocated to intervention or control group. In the crossover design, the same participants are then given the other intervention after an interim washout period (At).

designs are much more common. The type of randomization can be critical to allocation and analyses of the trial (Lachin et al., 1988). For hearing-related RCTs the sample size is usually <200, so simple randomization could lead to imbalanced group sizes. When using multiple clinics or outcomes with known covariates, both common in hearing trials, stratified randomization is necessary to insure reasonably balanced allocation across sites and/or covariates. For example, Humes et al. (2017) first stratified participants by unaided speech-in-noise performance, an expected covariate with their outcomes, before then allocating each stratum randomly to a different arm. A newer randomization technique, called merged block (van der Pas, 2019), combines block and simple randomization while avoiding the biases of both, and could be well-suited to hearing RCTs.

## Blinding

Blinding is a critical methodological feature of RCTs that reduces the risks of confounds and biases. Ideally, blinding should extend to everyone associated with the trial including clinicians, data collectors and data analysts. Clinical trials are described as double-blinded if both the researcher and participant are unaware of treatment allocation. A single-blinded study usually means the participant is unaware which treatment has been allocated. Blinding is more difficult to incorporate in trials of medical devices and surgical interventions than trials of medical therapies, which usually include placebo medications. Cox et al. (2016) investigated the effect of basic vs. premium hearing-aid features on subjective outcomes in a single-blinded study with no statistically significant difference between feature levels. In theory, this could have been a double-blinded study if the researcher responsible for data collection and analysis was also blinded from the hearing aid prescription and fitting. For many studies involving standard hearing aids, the devices need to be individually fit, potentially unblinding the audiologist. The audiologist would then need to be outwith the research team and blind to the aims of the study. In the Cochrane systematic review evaluating the effects of hearing aids for mild-to-moderate hearing loss in adults (Ferguson et al., 2017), the risk of performance and detection bias was rated as high because blinding was inadequate or absent. More recently, there have been attempts to maintain blinding. The use of placebo hearing aids allows blinding if they are visibly identical to active hearing aids. Studies by Adrait et al. (2017) and Humes et al. (2017) both used placebo hearing aids that provided minimal gain. These studies demonstrate that it is possible to blind participants and outcome assessors in hearing aid trials where the amplification characteristics can be concealed. Also, in a double-blinded RCT investigating the effectiveness of sound therapy in people with a reduced audiometric dynamic range, Formby et al. (2015) used conventional and placebo-controlled sound generators where the output of the placebo decayed to silence after 1 h of use in the ear.

# HEALTH-RELATED OUTCOME MEASURES

The most important question of any clinical trial is whether the trial's intervention was successful. The question is answered by means of primary and secondary outcome measures. Primary outcome measures capture the most evident or most important changes connected to the intervention (Vetter and Mascha, 2017). Secondary outcome measures assess aspects of the intervention in finer detail, for example, in order to understand mechanisms of change.

Once it is clear what the main expected change is, the vital question is how to capture this change. Outcome measurements can be objective (physiological or behavioral) or subjective, generalized or specific and clinician- or patient-reported. Other important considerations are the period being measured, and the measures' generalizability, reliability, and relevance.

## Objective Versus Subjective Outcomes

Some changes are only measurable by one type of outcome. One example is satisfaction, which can only be assessed as a subjective measure. However, subjective measures always need to be treated with caution. Satisfaction is a good example as the aforementioned study by Humes et al. (2017) found relatively good satisfaction with a placebo hearing aid.

For other outcomes, both objective and subjective measures exist. The combination of different instruments, such as objective and subjective measures of change in hearing ability, often will provide greater sensitivity and interpretability than a single measure. Further, using multiple measures will help counteract any dependence a single outcome has on participants or practitioners when blinding is an issue (e.g., the intervention difference cannot be concealed). However, the more outcome measures included in a trial, the greater the risk that results do not concur and potentially lead to opposing interpretations. One example is hearing aid use, which can differ between patients' self-reported use and their devices' data logging. For example, Solheim and Hickson (2017) reported a mean of 8.4 and 6.1 h for self-report and data logging, respectively. A possible alternative to measuring hearing aid use by data logging is to measure persistence through requests for supply of batteries (Zobay et al., 2021). Future measures may also be able to tap into usefulness – the desired outcome for which use and persistence are surrogates.

## Outcome Measurement Period

Deciding on the time point of assessment is particularly difficult, as it needs to include considerations of the temporal nature of the intervention. For hearing-aid trials, there may be an auditory acclimatization period before achieving full objective benefit (Dawes and Munro, 2017), whereas initial subjective benefit may decline over time (Humes et al., 2002). In addition, care must be taken to monitor the environments during the measurement period (e.g., via data logging) to ensure it is homogenous (Humes et al., 2018). For other studies, including training studies, the main interest might be in the time course and longevity of change. In the case of the latter, it needs to be carefully considered whether change is best assessed immediately after the intervention, or 6 weeks, 6 months or a year later. Wisely chosen test intervals may, for example, show whether training effects persist or weaken after the end of regular training (Henshaw et al., 2021).

## Generalizability of Outcomes

The question of generalizability reflects the tension between choosing standardized tools that are validated but have limited specificity to a particular health condition versus tools that are specific to a health condition but possibly newly created or modified, and often insufficiently validated. One example are QoL measures. As shown by Heinrich et al. (2015), a standardized generic QoL questionnaire such as the EQ-5D may not show any correlation with speech-in-noise performance, while a hearing-specific extension, the HR EQ-5D, does, but has not been appropriately standardized and validated. In the interest of building a body of evidence that can support CDSR and healthcare-system decisions (e.g., NICE) to improve clinical practice, some standardization and validation of outcomes measures will be essential. The Health Utility Index (HUI3) may provide a compromise as it is a standardized tool that has shown some sensitivity to hearing-aid provision (Barton et al., 2004).

A number of initiatives have been set up to understand what measurement instruments are being used within a field, how accurate, reliable and valid they are for what they aim to assess and how a core minimum outcome set could look like. Initiatives such as COMET (Core Outcome Measures in Effectiveness Trials)[2] bring together research groups interested in the development and application of agreed standardized sets of outcomes that should be measured and reported as minimum core sets in all clinical trials of a specific condition. One hearing-aid related outcome measure that was developed in a consortium resembling (but prior to) COMET is the seven-item International Outcome Inventory for Hearing Aids (IOI-HA; Cox et al., 2000).

## Outcome Reliability

If the validity and reliability of an outcome measure are in doubt, any interpretation of the results may suffer. COSMIN (Consensus-based Standards for the selection of health Measurement INstruments)[3] is an expert-led initiative that developed standards for the evaluation of health-status measures. Any outcome measure included in a trial should conform to their standards. A critical aspect of an outcome measure's methodological quality is its retest reliability. There are various ways of calculating reliability estimates (see Heinrich and Knight, 2020 for a discussion). The broader point, however, is that the retest reliability for many hearing outcome measures is rather poor, leading to "non-trivial" minimum/critical differences required to show an effect of an intervention (Weinstein et al., 1986; Cox and Rivera, 1992). Retest reliability and critical differences are also rather poor for standard speech-in-noise tests (Heinrich and Knight, 2020), making it a challenge to use them as outcome measures for a hearing RCT in which small effects may be expected.

## Relevance of Outcomes

Statistical significance is only one aspect of change. Equally important is that changes are perceptually noticeable and clinically relevant. Often it is possible to show that a change is statistically significant, particularly on a group level, but not perceptually noticeable or meaningful for an individual (e.g., improvement in signal-to-noise ratio that was not perceived by the participants; McShefferty et al., 2015, 2016), hence may lack relevance for the patient. Relevance at the clinic level can be achieved from comparing results against a (minimal) clinically important difference [(M)CID], a stakeholder-defined threshold of the proportional alleviation of a dysfunction or reduction in its prevalence. As hearing-loss interventions are compensatory, not restorative, (M)CIDs can seem ill suited to measuring clinically important differences, though it is possible, as demonstrated by Skarżyński et al. (2018) for tinnitus improvement after middle ear surgery. By first defining the threshold for a successful intervention, abetted by using validated measures that have a no-change midpoint, it is possible to report the percentage in alleviation for a particular hearing problem.

# REPRODUCIBILITY

The reproducibility of research is key to scientific advancement. It means that comparable results are obtained by methodologically

[2] https://www.comet-initiative.org/

[3] https://www.cosmin.nl/

closely matched but independent studies. Many fields, including biomedical science, suffer from a reproducibility crisis (de Vries et al., 2018) led by poor research practices and a well-established bias in scientific journals to preferentially publish novel and statistically significant findings which support the experimental hypothesis (Fanelli, 2012; Open Science Collaboration, 2015). Reproducibility can be increased in a number of ways, many of them applicable to clinical trials research. First, it is important to ensure that every phase of the research cycle is as transparent and open as possible, so that readers can fully evaluate the work. This research practice is referred to as "open science" (Kathawalla et al., 2021) and often contains the following three components: pre-registration, open data and open materials (Svirsky, 2019). Pre-registration makes information available in the public domain about the design and conduct of an intended study *before* collection of data (Munro and Prendergast, 2019). Open data and materials refers to depositing the datasets and test materials from the trial in the public domain. In addition to adhering to open science principles, the robustness of results are further bolstered by conducting collaborative multi-laboratory studies to understand the conditions for and boundaries of replication (Heinrich and Knight, 2020).

## CONCLUSION

There is a dearth of high quality evidence to support much of our existing clinical practice. This can be addressed by clinical trials but only if the conduct is rigorous and the quality is high. Good quality clinical trials have a research question based on PICO guidelines, follow best practice on methodological issues such as design, randomizing treatments and full blinding (participants and assessors) and choose optimal outcomes to assess the research questions in the correct timeframe and with reliability and validity. The importance of transparency and open science practices cannot be over-estimated.

## 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

KM proposed the topic. All authors contributed equally to draft and revised the manuscript and approved the final submission.

## FUNDING

KM and AH are supported by the NIHR Manchester Biomedical Research Centre (NIHR BRC-20007). WW was supported by the Medical Research Council (grant number MR/S003576/1) and the Chief Scientist Office of the Scottish Government.

## REFERENCES

Adrait, A., Perrot, X., Nguyen, M. F., Gueugon, M., Petitot, C., Collet, L., et al. (2017). Do hearing aids influence behavioural and psychological symptoms in hearing impaired Alzheimer's disease? *J. Alzheimers Dis.* 58, 109–121. doi: 10.3233/JAD-160792

Arlinger, S., Gatehouse, S., Kiessling, J., Naylor, G., Verschuure, H., and Wouters, J. (2008). The design of a project to assess bilateral versus unilateral hearing aid fitting. *Trends Amplif.* 12, 137–144. doi: 10.1177/1084713808316171

Atkins, D., Best, D., Briss, P. A., Eccles, M., Falck-Ytter, Y., Flottorp, S., et al. (2004). Grading quality of evidence and strength of recommendations. *Br. Med J* 328:1490. doi: 10.1136/bmj.328.7454.1490

Barton, G. R., Bankart, J., Davis, A. C., and Summerfield, Q. A. (2004). Comparing utility scores before and after hearing-aid provision : results according to the EQ-5D, HUI3 and SF-6D. *Appl. Health Econ. Health Policy* 3, 103–105. doi: 10.2165/00148365-200403020-00006

Collier, R. (2009). Legumes, lemons and streptomycin: a short history of the clinical trial. *Can. Med. Assoc. J.* 180, 23–24. doi: 10.1503/cmaj.081879

Cox, R. M., Hyde, M., Gatehouse, S., Noble, W., Dillon, H., Bentler, R., et al. (2000). Optimal outcome measures, research priorities and international cooperation. *Ear Hear.* 21, 106S–115S. doi: 10.1097/00003446-200008001-00014

Cox, R. M., Johnson, J. A., and Xu, J. (2016). Impact of hearing aid technology on outcomes in daily life I: the patients' perspective. *Ear Hear.* 37, e224–e237. doi: 10.1097/AUD.0000000000000277

Cox, R. M., and Rivera, I. M. (1992). Predictability and reliability of hearing aid benefit measured using the PHAB. *J. Am. Acad. Audiol.* 3, 242–254.

Cox, R. M., Schwartz, K. S., Noe, C. M., and Alexander, G. C. (2011). Preference for one or two hearing aids among adult patients. *Ear Hear.* 32, 181–197. doi: 10.1097/AUD.0b013e3181f8bf6c

Dawes, P., Hopkins, R., and Munro, K. J. (2013). Placebo effects in hearing aid trials are reliable. *Int. J. Audiol.* 52, 472–477.

Dawes, P., and Munro, K. J. (2017). Auditory distraction and acclimatization to hearing aids. *Ear Hear.* 38, 174–183. doi: 10.1097/AUD.0000000000000366

Dawes, P., Powell, S., and Munro, K. J. (2011). The placebo effect and the influence of participant expectation on outcome of hearing aid trials. *Ear. Hear.* 32, 767–774. doi: 10.1097/AUD.0b013e3182251a0e

de Vries, Y. A., Roest, A. M., de Jonge, P., Cuijpers, P., Munafò, M. R., and Bastiaansen, J. A. (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychol. Med.* 48, 2453–2455. doi: 10.1017/S0033291718001873

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientmetrics* 90, 891–904. doi: 10.1007/s11192-011-0494-7

Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-Jones, M., Barker, F., and Hoare, D. J. (2017). Hearing aids for mild to moderate hearing loss in adults. *Cochrane Database Syst. Rev.* 9:CD012023. doi: 10.1002/14651858.CD012023.pub2

Formby, C., Hawley, M. L., LaGuinn, P. S., Gold, S., Payne, J., Brooks, R., et al. (2015). A sound therapy-based intervention to expand the auditory dynamic range for loudness among persons with sensorineural hearing losses: a randomized placebo-controlled clinical trial. *Sem. Hear.* 36, 77–110. doi: 10.1055/s-0035-1546958

Hackshaw, A. (2009). *A Concise Guide to Clinical Trials*. London: Blackwell Publishing. doi: 10.1002/9781444311723

Heinrich, A., Henshaw, H., and Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Front. Psychol.* 6:782. doi: 10.3389/fpsyg.2015.00782

Heinrich, A., and Knight, S. (2020). Reproducibility in cognitive hearing research: theoretical considerations and their practical application in multi-lab studies. *Front. Psychol.* 11:1590. doi: 10.3389/fpsyg.2020.01590

Henshaw, H., Heinrich, A., Tittle, A., and Ferguson, M. A. (2021). Cogmed training does not generalise to real-world benefits for adult hearing aid users: Results of a blinded, active-controlled randomised trial. *Ear Hear.* doi: 10.1097/AUD.0000000000001096 [Epub ahead of print].

Humes, L. E., Rogers, S., Main, A. K., and Kinney, D. L. (2018). The acoustic environments in which older adults wear their hearing aids: insights from datalogging sound environment classification. *Am. J. Audiol.* 27, 594–603. doi: 10.1044/2018_AJA-18-0061

Humes, L. E., Rogers, S., Quigley, T. M., Main, A. K., Kinney, D. L., and Herring, C. (2017). The effects of service-delivery model and purchase price on hearing aid outcomes in older adults: a randomized double-blind placebo-controlled clinical trial. *Am. J. Audiol.* 26, 53–79. doi: 10.1044/2017_AJA-16-0111

Humes, L. E., Wilson, D. L., Barlow, N. N., and Garner, C. (2002). Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *J. Speech Lang. Hear. Res.* 45, 772–782. doi: 10.1044/1092-4388(2002/062)

Ivory, P. J., Hendricks, B. L., Van Vliet, D., Beyer, C. M., and Abrams, H. B. (2009). Short-term hearing aid benefit in a large group. *Trends Amplif.* 13, 260–280. doi: 10.1177/1084713809354902

Kathawalla, U. K., Silverstein, P., and Syed, M. (2021). Easing into Open Science: a guide for graduate students and their advisors. *Collabra Psychol.* 7:18684. doi: 10.1525/collabra.18684

Lachin, J. M., Matts, J. P., and Wei, L. J. (1988). Randomization in clinical trials: conclusions and recommendations. *Control Clin. Trials* 9, 365–374. doi: 10.1016/0197-2456(88)90049-9

Marriage, J., Moore, B. C. J., and Alcantara, J. I. (2004). Comparison of three procedures for initial fitting of compression hearing aids. III. Inexperienced versus experienced users. *Int. J. Audiol.* 43, 198–210. doi: 10.1080/14992020400050028

McShefferty, D., Whitmer, W. M., and Akeroyd, M. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends Hear.* 19:2331216515572316. doi: 10.1177/2331216515572316

McShefferty, D., Whitmer, W. M., and Akeroyd, M. (2016). The just-meaningful difference in speech-to-noise ratio. *Trends Hear.* 20:2331216515626570. doi: 10.1177/2331216515626570

Meijerink, J. F. J., Pronk, M., Lissenberg-Witte, B. I., Jansen, V., and Kramer, S. E. (2020). Effectiveness of a Web-Based SUpport PRogram (SUPR) for hearing aid users aged 50+: two-arm, cluster randomized controlled trial. *J. Med. Internet Res.* 22:e17927. doi: 10.2196/preprints.17927

Munro, K. J., and Prendergast, G. (2019). Editorial. Encouraging pre-registration of research studies. *Int. J. Audiol.* 58, 123–124. doi: 10.1080/14992027.2019.1574405

National Institute for Health and Care Excellence (2018). *Hearing Loss in Adults: Assessment and Management (NICE National Guideline 98)*. Available online at: https://www.nice.org.uk/guidance/ng98 (accessed January 13, 2021)

National Institute for Health and Care Excellence (2020). *Tinnitus Assessment and Management (NICE National Guideline 155)*. Available online at https://www.nice.org.uk/guidance/ng155 (accessed January 13, 2021)

Naylor, G., Öberg, M., Wänström, G., and Lunner, T. (2015). Exploring the effects of the narrative embodied in the hearing aid fitting process on treatment outcomes. *Ear Hear.* 36, 517–526. doi: 10.1097/AUD.0000000000000157

Noble, W., and Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ). *Int. J. Audiol* 45, 172–181. doi: 10.1080/14992020500376933

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Sci. Transl. Med.* 349:aac4716. doi: 10.1126/science.aac4716

Skarżyński, H., Gos, E., Dziendziel, B., Raj-Koziak, D., Włodarczyk, E. A., and Skarżyński, P. H. (2018). Clinically important change in tinnitus sensation after stapedotomy. *Health Qual. Life Outcomes* 16:208. doi: 10.1186/s12955-018-1037-1

Smith, G. C. S., and Pell, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *Br. Med. J.* 327:1459. doi: 10.1136/bmj.327.7429.1459

Solheim, J., and Hickson, L. (2017). Hearing aid use in the elderly as measured by datalogging and self-report. *Int. J. Audiol.* 56, 472–479. doi: 10.1080/14992027.2017.1303201

Svirsky, M. A. (2019). Editorial. Preregistration and open science practices in hearing science and audiology: the time has come. *Ear Hear.* 41, 1–2. doi: 10.1097/AUD.0000000000000817

van der Pas, S. L. (2019). Merged block randomisation: a novel randomisation procedure for small clinical trials. *Clin. Trials* 16, 246–252. doi: 10.1177/1740774519827957

Vetter, T. R., and Mascha, E. J. (2017). Defining the primary outcomes and justifying secondary outcomes of a study: usually, the fewer, the better. *Anesth. Analg.* 125, 678–681. doi: 10.1213/ANE.0000000000002224

Weinstein, B. E., Spitzer, J. B., and Ventry, I. M. (1986). Test-retest reliability of the Hearing Handicap Inventory for the Elderly. *Ear. Hear.* 7, 295–299. doi: 10.1097/00003446-198610000-00002

World Health Organisation (2018). *International Standards for Clinical Trial Registries v3.0*. Geneva: World Health Organisation.

Zobay, O., Dillard, L. K., Naylor, G., and Saunders, G. H. (2021). A measure of long-term hearing aid use persistence based on battery reordering data. *Ear Hear.* 42, 1441–1444. doi: 10.1097/AUD.0000000000001032

**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.

frontiers in Neuroscience

METHODS
published: 24 November 2021
doi: 10.3389/fnins.2021.718416

# Best Practices in the Development, Translation, and Cultural Adaptation of Patient-Reported Outcome Measures for Adults With Hearing Impairment: Lessons From the Cochlear Implant Quality of Life Instruments

*Ariane Laplante-Lévesque*[1,2]*, *Judy R. Dubno*[3], *Isabelle Mosnier*[4,5], *Evelyne Ferrary*[4,5] *and Theodore R. McRackan*[3]

[1] Department of Clinical Evidence Cochlear Implants, Oticon Medical A/S, Smørum, Denmark, [2] Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden, [3] Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, Charleston, SC, United States, [4] Hearing Institute, Institut Pasteur/Université de Paris/Inserm, Paris, France, [5] Unité Fonctionnelle Implants Auditifs, ORL, GH Pitié-Salpêtrière, AP-HP Sorbonne Université, Paris, France

OPEN ACCESS

**Edited by:**
Astrid van Wieringen,
KU Leuven, Belgium

**Reviewed by:**
Zahra Jafari,
University of Lethbridge, Canada
Yonghee Oh,
University of Florida, United States

***Correspondence:**
Ariane Laplante-Lévesque
arll@oticonmedical.com

**Specialty section:**
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

**Received:** 31 May 2021
**Accepted:** 02 November 2021
**Published:** 24 November 2021

**Citation:**
Laplante-Lévesque A, Dubno JR, Mosnier I, Ferrary E and McRackan TR (2021) Best Practices in the Development, Translation, and Cultural Adaptation of Patient-Reported Outcome Measures for Adults With Hearing Impairment: Lessons From the Cochlear Implant Quality of Life Instruments. Front. Neurosci. 15:718416. doi: 10.3389/fnins.2021.718416

This manuscript summarizes available evidence-based best practices in the development, translation, and cultural adaptation of one type of outcome measure for adults with hearing impairment, patient-reported outcome measures (PROMs). It presents the development of the Cochlear Implant Quality of Life (CIQOL) instruments and the ongoing translation and cultural adaptation of the CIQOL-35 Profile from English to French as case studies and discusses useful lessons for selecting, developing, translating, culturally adapting, and using PROMs. Relevant best practice guides are introduced, described and their steps are illustrated with examples. Future trends in hearing-related PROMs, including computerized adaptive testing, patient-reported experience measures (PREMs), economic evaluation and allocation of scarce resources, and PROMs in low-resource settings, are discussed. The manuscript concludes on the lessons that can be learned from implementation science for the successful and sustainable integration of PROMs in clinical practice.

**Keywords: patient-reported outcome measure (PROM), questionnaire, PROM development, PROM translation, PROM cultural adaptation, quality of life, hearing impairment, cochlear implant**

## INTRODUCTION

This manuscript, part of the Research Topic "Outcome Measures to Assess the Benefit of Interventions for Adults with Hearing Loss: From Research to Clinical Application," summarizes already available evidence-based best practices in the development, translation, and cultural adaptation of patient-reported outcome measures (PROMs) for adults with hearing impairment. It presents the development of the Cochlear Implant Quality of Life (CIQOL) instruments and the ongoing translation and cultural adaptation of the CIQOL-35 Profile from English to French as illustrative case studies for those interested in selecting, developing, translating, culturally adapting, and using PROMs.

## Hearing Impairment and Patient-Reported Outcome Measures

Hearing loss is the most prevalent sensory disorder and the third most common cause of Years Lived with Disability (YLDs) after low back pain and migraine, in the 2019 Global Burden of Disease (GBD) study, a systematic overview of the prevalence of 369 diseases and injuries (Haile et al., 2021). The World Hearing Organization urges for multi-disciplinary hearing health care action including prevention and rehabilitation. The GBD study defines hearing loss as a pure-tone average of audiometric thresholds at 0.5, 1, 2, and 4 kHz $\geq$ 20 dB HL in the better ear. This definition focuses only on hearing detection/acuity and, therefore, does not consider functional abilities, self-reported hearing difficulties, or their impact on quality of life. Globally, more than 1.5 billion people, or 20% of the population, experience some degree of hearing loss (World Health Organization [WHO], 2021b).

Hearing evaluation is mostly performed through pure-tone thresholds measurements, speech recognition tests, and other standard diagnostic assessments designed to differentiate conductive and sensorineural hearing loss, and less-so on PROMs (Granberg et al., 2014; Hill-Feltham et al., 2020). A systematic review reported that PROMs represent only 9% of the total hearing outcome measures ($n$ = 837), whereas pure-tone thresholds measurements and speech recognition tests accounted for 65 and 20%, respectively (Hill-Feltham et al., 2020).

Quality of life refers to an "individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" (World Health Organization [WHO], 2021a). Health-related quality of life focuses on the aspects of quality of life most relevant for health, i.e., the physical, mental and social well-being (Guyatt et al., 1993). To evaluate the impact of hearing loss on quality of life or the benefit of hearing-related interventions on quality of life, numerous PROMs have been developed, including hearing-specific instruments, hearing-aid-specific instruments, and cochlear-implant-specific (CI) instruments (Andries et al., 2021). Health-related quality of life measures, such as the Euro-QoL (EQ-5D), the Health Utilities Index (HUI3), and the SF-36, are often used to measure health utility, but are weakly associated with hearing-specific PROMs due to the lack of items that are related to everyday functional communication and social interaction (McRackan et al., 2019a, 2021).

In contrast to standard audiometric test batteries, a direct input assessment from patients of improvement of their health and quality of life following CI is recommended to evaluate the positive and multidimensional impact of hearing rehabilitation. PROMs/quality of life measures are increasingly being regarded as quality indicators. For example, in England the National Health Service has a National PROMs Program under which it coordinates the national collection of PROMs for four elective surgical procedures (National Health Service [NHS], 2021). In the United States, the Meaningful Measures framework identifies priority areas that promote quality healthcare for Medicare and Medicaid patients and include functional outcomes and patient experiences of care feature (Centers for Medicare and Medicaid Services [CMMS], 2021). Moreover, PROMs/quality of life measures are now mandatory in some countries for reimbursement of medical devices and requested to identify most personalized care pathways (Patrick et al., 2007; Artières-Sterkers et al., 2020; Fraysse et al., 2020). For example, the U.S. Food and Drug Administration positions appropriate PROMs as central to clinical studies that evaluate the effectiveness of new medical products including hearing devices (Patrick et al., 2007).

## Currently Available Patient-Reported Outcome Measures

A synthesis of available PROMs identified that they target the following three domains: auditory (listening, communicating, and speaking), social (relationships, isolation, social life, occupational, and interventions), and self (effort and fatigue, emotions, identity, and stigma; Vas et al., 2017). However, limited evidence is available to support the unidimensionality of these domains. For example, recent re-evaluation of the Hearing Handicap Inventory for the Elderly did not show the social and emotional domains to be independent (Cassarly et al., 2020). Indeed, some of the legacy measures sometimes have unknown or unsound psychometric properties, including face/content validity. It is common for researchers and clinicians to have developed PROM items without relying on expert panels of patients, focusing instead on input from clinicians. This means the domains/items included may not cover the issues most important to the patient population. It is recommended to include qualitative research methods and literature reviews in the development of outcome measures, as detailed in the next sections. Another limitation of some legacy PROMs is that they are not always efficient, with some domains including more items than necessary, which creates burden for the patient, the clinician and the researcher, and reduces the likelihood that the PROM will be used in routine clinical practice and clinical research.

## Development of Patient-Reported Outcome Measures–Cochlear Implant Quality of Life

Modern development standards for PROMs aim to create efficient, precise, and responsive instruments that represent the values most important to the population of interest. The Patient-Reported Outcomes Measurement Information System (PROMIS) and COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) have established standards that aim to improve the quality of PROMs used to measure clinical and research outcomes (Mokkink et al., 2010; PROMIS, 2013). While differences exist, both support the use of a mixed methods research design and agree on an overall structure. We illustrate this process through a case study, the development of the CIQOL-35 Profile instrument and CIQOL-10 Global measure.

### Systematic Literature Review

A systematic literature review is necessary for step one in the PROM development process in order to develop a comprehensive understanding of the previous work done in the area of interest. In addition, the previous items and concepts included in legacy instruments can help form the protocols for future focus groups or key informant interviews. As a part of the CIQOL development process, two systematic reviews and meta-analyses identified 21 studies that used pre- and post-implantation PROMs to monitor adult CI outcomes (McRackan et al., 2018a,b). Overall, this identified a clear improvement in QOL using both hearing- and CI-specific instruments and general-health QOL instruments. However, hearing- and CI-specific instruments demonstrated substantially greater improvements in QOL than general health instruments (McRackan et al., 2018a,b, 2019a). In addition, this analysis found negligible to low positive correlations between speech recognition scores (words in quiet, sentences in quiet, and sentences in noise) and patient self-reported functional ability (**Table 1**). This relationship was maintained even when comparing communication domains in QOL instruments to speech recognition outcomes. These results are consistent with the assumption that how patients with CIs communicate in their everyday functioning is more complex than revealed by speech recognition tasks routinely used in clinical care, which further supports the use of PROMs as part of a test battery to comprehensively assess CI outcomes.

### Focus Groups and Cognitive Interviews

The next step in the PROM development process is to conduct focus groups or key informant interviews to ensure the themes and topics that affect the population interest, in this case adult patients with CIs, are included in the items in the PROM. This qualitative analysis is critical as it provides the face and content validity of the instrument. For the CIQOL, adult patients with CIs with a wide range of speech recognition outcomes took part in focus groups (McRackan et al., 2017). The 23 patients were stratified into 3 focus groups based on communication abilities with their CI as measured by word scores on the consonant–vowel nucleus–consonant (CNC) test in quiet presented at 60 dB SPL (group 1: <36%; group 2: 36–66%; group 3: >66%). Analysis of the focus group transcripts identified seven themes: communication, emotion, entertainment, environmental sounds, independence, listening effort, and social. Individual focus group participant statements related to these themes were then developed into items. This generated a 101-item pool, which served as a potential source of items to include in subsequent instruments. Audiologists, physicians and hearing science researchers then carefully reviewed the items and also ensured the items were at or below a 6th grade level reading level, using the Lexile Analyzer.[1] Item clarity was then confirmed through cognitive interviews with 20 adult patients with CIs who were not involved in the focus groups (McRackan et al., 2017). These interviews confirmed that the items were easy to read and understand, unambiguous, and culturally appropriate. No items required revision based on the cognitive interviews.

**TABLE 1** | Meta-analysis of correlations between speech recognition scores and patient self-reported functional ability.

| Speech recognition scores | *r* | 95% confidence intervals |
|---|---|---|
| **Cochlear implant-specific quality of life** | | |
| Word recognition in quiet | 0.21 | 0.12 to 0.30 |
| Sentence recognition in quiet | 0.24 | 0.08 to 0.39 |
| Sentence recognition in noise | 0.26 | −0.08 to 0.54 |
| **Hearing-specific quality of life** | | |
| Word recognition in quiet | 0.28 | 0.14 to 0.37 |
| Sentence recognition in quiet | 0.20 | 0.07 to 0.33 |
| Sentence recognition in noise | NA | NA |
| **Health-related quality of life** | | |
| Word recognition in quiet | 0.33 | 0.19 to 0.46 |
| Sentence recognition in quiet | 0.34 | 0.18 to 0.48 |
| Sentence recognition in noise | 0.32 | 0.19 to 0.44 |

### Psychometric Testing to Develop the Cochlear Implant Quality of Life Item Bank

One of the most recent significant changes to PROM development is the increased use of item response theory (IRT). IRT is the core of modern psychometric analyses used to develop PROMs and has several advantages over classical test theory (CTT), which was the previous standard. First, CTT is grounded on observed and true scores, which focuses on the measurement of an underlying trait—referred to as person ability or person measure. Therefore, CTT-derived instruments are sample-dependent as subjects will have higher true scores on easier tests and lower true scores on more difficult tests. In contrast, IRT-developed instruments remain sample and test independent (Prieto et al., 2003).

Second, whilst CTT focuses on test-level psychometrics, IRT focuses on item-level psychometrics. IRT analyses concentrate on each individual item and determine its measurement characteristics and utility for inclusion in subsequent instruments. IRT analyses not only evaluate the ceiling and floor effects for each item, but also identify fit to the hierarchical model, matches individual item difficulty level to person ability level, and ensures that the items cover the full ability range of the patient population. Application of IRT analyses to the item pool results in the final item bank, which serves as the source for items to be used for subsequent PROMs (including short form, profile, and computerized adaptive testing (CAT) instruments, which will be discussed in a later section). With the psychometric properties established for each item, researchers can select items for each instrument based on their highest discrimination across the ability range and best match between item difficulty and patient ability. This results in optimized instruments with maximized capacity to differentiate individuals across a greater range of the latent trait—termed precision (Rose et al., 2008).

The third advantage is related to the stricter assumptions that must be met before IRT analysis is performed compared to CTT (Reeve et al., 2007). This includes (1) unidimensionality—items only contribute to one domain construct, (2) local independence—responses to each item are unrelated to responses

[1] https://hub.lexile.com/analyzer

to other items, and (3) item fit—the items must fit the IRT measurement model. Confirmatory factor analysis (CFA) is used to confirm unidimensionality and local independence. For item bank development, items are eliminated if they do not significantly contribute to the unidimensional construct captured by the other items in a domain, or if responses to an item are dependent upon responses to other items in the pool. In addition, item fit to the IRT model, such as infit and outfit, are examined to ensure that the included items sufficiently measure the construct of interest for individuals at ability levels close to and far from the item difficulty.

For the CIQOL item bank, the item pool of 101 items organized into 7 domain constructs was completed by 371 adult patients with CIs from all regions of the United States (McRackan et al., 2019c). By completing the psychometric analyses described earlier, one domain construct was found to lack unidimensionality (i.e., independence) and was removed from the item pool. All other domains were found to represent unidimensional constructs. In addition, several items were removed for being locally dependent on other items ($n$ = 3) and misfitting the IRT model ($n$ = 6). This resulted in the final item bank item which consisted of 81 items in 6 domains (communication, emotion, entertainment, environmental sounds, listening effort, and social).

### Development of the Cochlear Implant Quality of Life-35 Profile Instrument and Cochlear Implant Quality of Life-10 Global Measure

The item-level psychometric analyses results were then used to guide the development of the subsequent instruments (McRackan et al., 2019b). Here, items are selected for each domain that represent the full ability continuum (based on item difficulty) and have the greatest capacity to discriminate individual patient ability. Additional IRT analyses can then be performed to ensure that the items selected fit each domain's IRT model. The CIQOL-35 Profile was developed using this method and assesses outcomes represented in the 6 domains (McRackan et al., 2019b). A single factor CFA was then performed on the CIQOL-35 Profile to ensure it was psychometrically sound to use as a source for items in a global measure (CIQOL-10 Global). After this was confirmed, the CIQOL-10 Global was created based on the above parameters. This measure provides an overall assessment of CI-related functional outcomes but does not provide domain-specific information. Importantly, all items for the global measure are included in the profile instrument so a global score can be easily calculated from the CIQOL-35 Profile.

### Final Validation of the Cochlear Implant Quality of Life-35 Profile and Cochlear Implant Quality of Life-10 Global

After the creation of the instrument, final validation typically includes comparison of psychometric properties of the newly developed PROMs to legacy instruments. Available guidelines are less concrete regarding the analyses needed for this final stage. In general, there are three main components to this comparison. First, construct validity determines whether each purported domain represents a unidimensional concept. This includes analysis of all domains, subdomains, and total scores. Second, convergent validity evaluates the degree to which scores from an instrument are associated with conceptually similar measures. This can range from correlation with physiological findings when available or legacy PROMs. Third, reliability determines the consistency of PROM scores across time.

To accomplish this, results from the CIQOL instruments were compared to results from the Nijmegen Cochlear Implant Questionnaire (NCIQ) and HUI3 in 334 adult patients with CIs who were not involved in previous development stages (McRackan et al., 2021). The results demonstrated that all CIQOL domains as well as the global measure had strong construct validity, strong convergent validity, and strong to very strong reliability. In contrast, 8 of the 10 NCIQ domains/subdomains as well as the NCIQ total score demonstrated poor construct validity. The remaining NCIQ subdomains (basic sound performance and activity limitation) demonstrated strong psychometric properties and test–retest reliability. Interestingly, HUI3 reliability was moderate to weak in adult patients with CIs with the weakest reliability in the hearing dimension. This is likely related to the use of "hearing aid" in several items, which may confuse patients with CIs.

The final product of this process are two instruments that represent the values of adult patients with CIs and are more psychometrically sound and comprehensive than previously developed PROMs. The CIQOL-35 Profile and CIQOL-10 Global are available for use in clinical and research settings and are free to download at http://education.musc.edu/CIQOL. The CIQOL instruments have been downloaded by over 210 CI centers and are undergoing translation and cross-cultural adaptation in 8 languages.

## Hearing Related Patient-Reported Outcome Measures in French Language

Most of the world's population does not speak English, yet exchanging information beyond and across linguistic communities is crucial. PROMs developed and validated using rigorous procedures as described earlier should then be carefully translated, culturally adapted, and validated to other linguistic and cultural groups.

Currently available hearing PROMs in French language include the Abbreviated Profile of Hearing Aid Benefit (APHAB), the Client Oriented Scale of Improvement (COSI), the Hearing Implant Sound Quality Index (HISQUI), the NCIQ, and the Speech, Spatial and Qualities of Hearing Scale (SSQ). All the above-mentioned PROMs were published in English and their French translation process is undocumented. To the best knowledge of the authors, the NCIQ, which was developed in Dutch and published in English but unfortunately without a description of the translation process, is the only PROM designed specifically for patients with CIs available in French and its use in clinical practice is complex due to its length (60 items). In contrast, the Evaluation of the Impact of Hearing Loss in Adults (ERSA) was developed and validated in French and is relevant for hearing aid and adult patients with CIs. It has good reliability, validity, and sensitivity to change, but it is difficult to compare scores against other PROMs as it has yet to be translated to other languages (Ambert-Dahan et al., 2018).

The following sections describe the process of translation, cultural adaptation, and validation through a case study, the translation of the CIQOL-35 Profile instrument from English, its source language, to French. Throughout the manuscript, for clarity of expression, the term translation is used to refer to the iterative process of both translation and cross-cultural adaptation.

## MATERIALS, EQUIPMENT AND METHODS

The translation of PROM items, response choices, and instructions should be obtained through an iterative process of forward and back-translation by qualified translators and bilingual content experts, bilingual expert review, and testing with the patient group. A hearing-related PROM translation good practice guide recommending six steps can be used to guide this iterative process (Hall et al., 2018b).

Materials include the source-language PROM (i.e., the CIQOL-35), good practice guides (PROMIS, 2013; Hall et al., 2018b), and a location to archive all steps and related documentation. The "reconciliation report" provided by Hall et al. (2018b) as supplemental file 3 is especially helpful in documenting the translation process.

The following section describes the six steps of Hall et al. (2018b) and illustrates how they guide activities in the case study of the translation of the CIQOL-35 instrument from English to French. In this PROM translation project, Steps 1–4 are completed and Steps 5–6 are yet to be completed. **Figure 1** summarizes the steps completed so far; these are described below.

### Step 1. Preparation

#### Summary of Best Practices

According to Hall et al. (2018b), this step sets the scene for the translation and includes checking whether a translation of the instrument already exists and gaining approval from the source-language PROM copyright holders for the translation. Source-language PROM developers should be invited to be involved as their input is important to clarify the original intent of the PROM instructions, items, and response options. The translation project should have clear aims and intended audience and the main concepts that underpin the PROM should be defined. Finally, template documents for documenting the translation should be prepared.

#### Case Study: Translation of the Cochlear Implant Quality of Life-35 Instrument From English to French

No French version of the CIQOL-35 instrument existed, as it is a recent PROM. The translation project was instigated and took form through a close collaboration between hearing clinicians and researchers who are native French speakers and the CIQOL-35 developers. The copyright holder, the Medical University of South Carolina Foundation for Research Development, provided written approval for the translation project. Aims and intended audience were discussed and agreed upon. The primary audience was similar to the source-language CIQOL-35 instrument, e.g., adults with hearing loss regardless of hearing device status. Although the CIQOL suite of instruments is designed for adult patients with CIs, it should also be valid for pre-CI measures, i.e., for adults with hearing loss before they receive one or two CIs. Efforts were made to reach all people of adult age irrespective of literacy and to create a translation that could be administered as pen-and-paper as well as electronic.

The linguistic and cultural profile of the intended audience includes French-speaking people living in France or elsewhere. French is spoken by 300 million people globally, of which 59.3% live in Africa, 33.4% in Europe, 7.0% in the Americas, and 0.3% in Asia and Oceania (Organisation Internationale de la Francophonie, 2019). The 300 million French speakers spread across the globe do not use the French language in a uniform manner. In the different geographical areas, the French language has evolved into different dialects, i.e., varieties of French that are mostly mutually intelligible, especially if they are close on the dialect continuum. This is the same phenomenon that distinguishes, for example, the English or Spanish spoken in different parts of the world. Dialects do not respect country borders and dialects can be multiple in the same country. Dialects vary in their vocabulary, grammar, and pronunciation, the latter important for example for speech perception test stimuli but not for PROMs to be administered in written form. Given the presence of dialects, two approaches to PROM translation can be taken. A specific and localized approach produces as many translations as the number of dialects. A universal approach favors a "standard" version of the language and avoids regionalisms (i.e., vocabulary words, grammatical structures, or expressions favored by speakers in a particular geographic area). For this translation, a universal approach was adopted. Universal translation runs the risk of using terms or grammatical structures that are not immediately recognized by some speakers, or that require more cognitive effort to be understood. To support a universal approach to translation, involving people familiar with different dialects, referring to linguistic resources that recognize standard and colloquial usage of terms by region (such as Joseph Wright's English Dialect Dictionary or the Real Academia Española's Spanish dictionary), and testing on speakers of different dialects is recommended. These considerations are also relevant during the development of a PROM and for which linguistic communities it is intended.

To prepare for the translation, a list of resources was created, which included the concept definitions used in the source-language CIQOL-35 instrument development, further concept definitions, as well as examples of hearing-related written documents available in both English and French obtained from the World Health Organization, Hear-it.org, and Oticon Medical A/S. These resources served two purposes. First, they presented background information about hearing, hearing impairment, its consequences, and its treatment. Second, they provided a range of examples of English-French translations of relevant terms. The "reconciliation report" provided by Hall et al. (2018b) as supplemental file 3 was adapted to the purposes of this translation

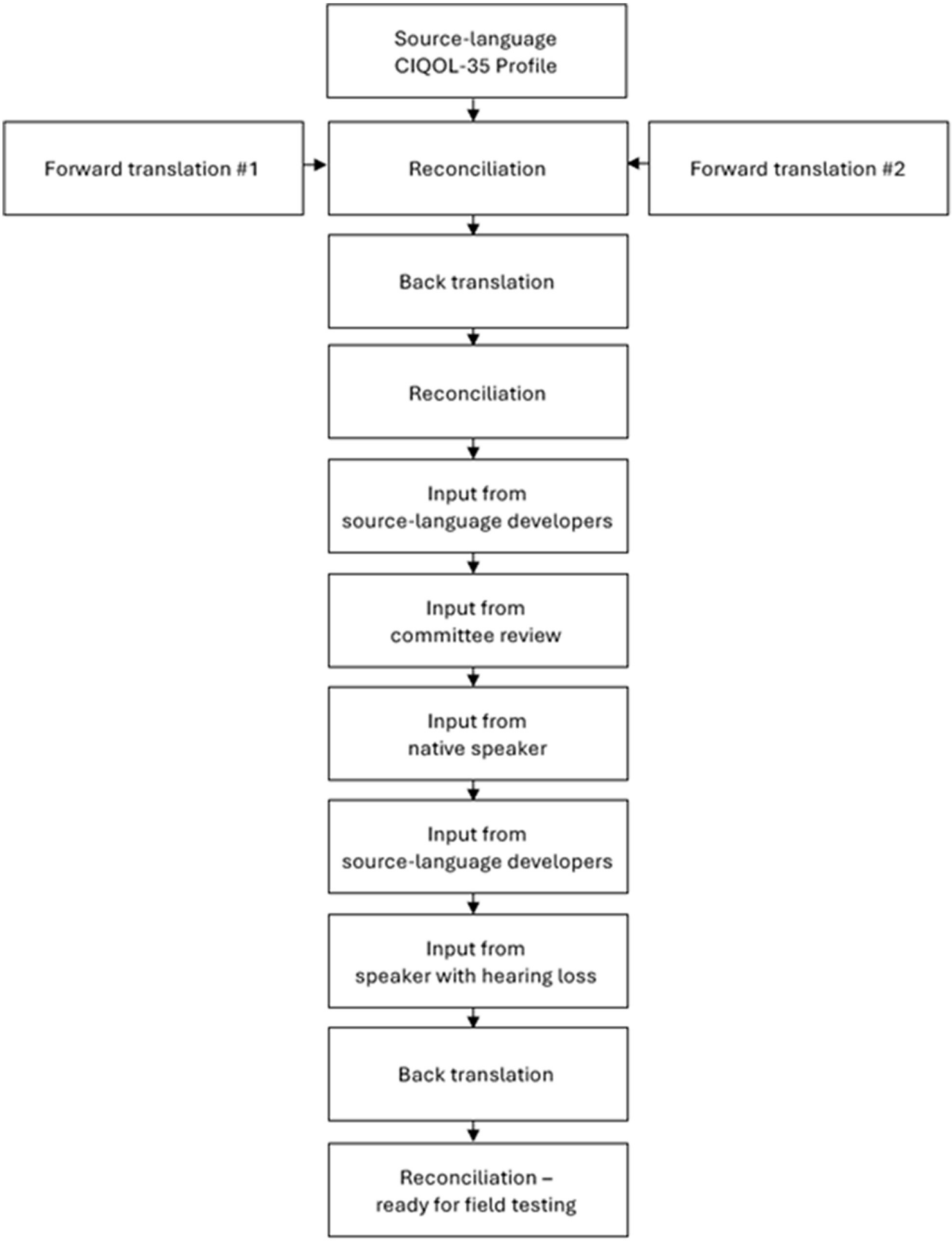


**FIGURE 1 |** Translation and cross-cultural adaptation process of the CIQOL-35 Profile from English to French.

project, including copying all CIQOL-35 instructions, items, and response options as separate spreadsheet rows and copying all supporting information and links into a separate spreadsheet within the same document, for easy access by all people involved in the translation. **Table 2** lists the people involved in this translation project.

## Step 2. Forward Translation

### Summary of Best Practices

According to Hall et al. (2018b), this step includes the translation of the PROM from the source language to the target language. For this, translators whose first language is the target language, and ideally, have the same dialect and reside or have lived experience of the region/culture of the intended audience should be recruited. It is recommended that at least two translators are involved, one translator that is a professional translator with training/certification in linguistics and another translator that is a healthcare professional with experience working with adult patients with CIs. The translators should be introduced to the PROM to be translated, the health condition and related concepts, as well as the concept definitions in Step 1 described earlier. They should also be instructed on the translation and adaptation task and this should be done in a uniform fashion for all translators. Each translator should work independently to produce a forward translation of the PROM instructions, items, and response options. The reconciliation of the forward translations by another person involves creating from the multiple forward translations a single forward translation.

### Case Study: Translation of the Cochlear Implant Quality of Life-35 Instrument From English to French

The two translators recruited were both French native-speaking and had a high command of English. One was a certified linguist with experience in translation of hearing-related documents and the other held a clinical support position for Oticon A/S that includes regular contacts with patients with CIs. They received the same instructions and background information together with the spreadsheet described earlier in the previous step. They were instructed to maintain conceptual, item and semantic equivalence and that it was more important to preserve meaning than to provide a literal translation. Everyday non-technical language was to be used and a "universal translation" approach avoiding regionalisms was prioritized. As they translated each section of the PROM (i.e., each spreadsheet row), they were instructed to rate how difficult they found each translation (from 0 extremely easy to 10 extremely hard). These ratings were useful in the reconciliation step, which involves comparing and contrasting the different translations. A bilingual hearing clinician/researcher completed this task, using the same spreadsheet described earlier. The first step involved highlighting sections where the translations differed. As stated by Hall et al. (2018b), dedicated effort was spent on the sections that the translators rated as relatively more difficult to translate compared to other sections. The person completing the reconciliation documented the reasoning behind reconciliation decisions. This step resulted in one single forward translation of the CIQOL-35.

## Step 3. Back Translation

### Summary of Best Practices

According to Hall et al. (2018b), this step involves the translation of the PROM from the target language back to the source language for comparison with the source-language PROM. The person conducting the back translation should be naive to the source language PROM. The assumption is that if the translation and adaptation process is carefully done, any differences between the source-language PROM and the back translation would reflect cultural adaptation. For this, at least one translator should be recruited, ideally a professional translator with training/certification in linguistics. The back translation is then carefully compared with the source PROM and equivalence is classified from perfect (A) to null (D) equivalence, in both choice of words and semantics conveyed, and this is recorded in the reconciliation report.

### Case Study: Translation of the Cochlear Implant Quality of Life-35 Instrument From English to French

One certified linguist with experience in translation of hearing-related documents was recruited. This translator received the same instructions, background information and spreadsheet described in Step 2 earlier, except that the source-language CIQOL-35 instrument was not shown in the spreadsheet. The same person who completed the reconciliation task in Step 2 then compared the source-language CIQOL-35 instrument with its back translation. Sections of the back translation that

**TABLE 2 |** People involved in this translation project along with their roles.

| Person | Role in this translation project |
|---|---|
| Translation lead | Project management, resource management, procedure documentation/archiving, reconciliation of the translation, oversight of the field testing |
| Source-language PROM developers | Provision of concept definitions, consulting on queries arising during translation |
| Linguist #1 | Forward translation including difficulty rating and participation in committee review |
| Native speaker health professional #1 | Forward translation including difficulty rating and participation in committee review |
| Linguist #2 | Back translation and participation in committee review |
| Native speaker health professional #2 | Participation in committee review, also field testing investigator |
| Native speaker health professional #3 | Participation in committee review, also field testing investigator |
| Native speaker reviewer #1 | Revision after committee review |
| Native speaker reviewer #2 | Revision after committee review |
| Linguist #3 | Back translation after committee review |

were discrepant to the source were documented in the same spreadsheet using color coding for easy identification of sections that were problematic and/or required review.

## Step 4. Committee Review

### Summary of Best Practices

According to Hall et al. (2018b), this step recommends appointing a multi-disciplinary committee that includes linguistic and healthcare expertise to review the translation steps including the forward and back-translations and review and solve the problematic sections identified in Step 3.

### Case Study: Translation of the Cochlear Implant Quality of Life-35 Instrument From English to French

The committee review included the three people who translated the PROM in Steps 2–3, the person who completed the reconciliation of the two forward translations and the comparison of the source-language instrument with the back translation, and two bilingual hearing care professionals. All committee members were provided the spreadsheet documenting all steps before the review. During the review meeting, all problematic sections identified in Step 3 as well as any other sections that committee members deemed relevant to discuss were reviewed. As much as possible, consensus was sought on preferable translation. Discussions and reasons underlying translation choices were documented in the spreadsheet. The optimal translation and cultural adaptation of English expressions and figures of speech that do not have direct equivalents in French language/culture such as “crowded environments,” “to socialize,” “social situations,” or “to feel inadequate” generated the most discussion, to ensure semantics were preserved as much as possible whilst creating a culturally appropriate translation. Overall, the translation was at times too literal and benefited from a deeper translation and slight cultural adaptation. During the committee review, any questions raised regarding the original intent of the CIQOL-35 items were noted.

Because significant improvements were suggested to the translation, four additional steps not mentioned in Hall et al. (2018b) were taken. First, a native French speaker naive to the CIQOL-35 reviewed the latest French translation to ensure natural language. As a result, minor changes such as in the choice of prepositions and adverbs were implemented. Second, questions regarding the original intent of the CIQOL-35 were raised with its developers and the translation was slightly revised accordingly. Third, as some questions were raised about the appropriateness of the translation for the intended audience, a native French speaker with hearing loss naive to the CIQOL-35 reviewed the latest French translation to ensure appropriateness. As a result, one minor change to one item was made. Fourth, because significant changes were made since the previous back translation, another back translation was performed on the latest version of the translation. This back translation was completed by a professional translator not involved in previous steps. The same process of comparison and reconciliation described earlier was completed to result in a French translation of the CIQOL-35 ready for field testing (i.e., validation). Whilst it could be argued that the extra steps taken are evidence of suboptimal translation practices, it is believed that they reflect an attention to detail that PROM translation deserves. **Table 3** is an excerpt of the reconciliation report: it presents the 26 columns documenting the different steps of the translation for one of the CIQOL-35 items. This reconciliation report was adapted from Hall et al. (2018b) supplemental file 3.

## Step 5. Field Testing

### Summary of Best Practices

According to Hall et al. (2018b), this step involves testing the translated PROM on a small group of people drawn from the intended audience. The aim of the field test is to ensure the intended audience understands the translation and finds it acceptable. The field test also aims to ensure that the translation is equivalent to the source-language PROM. Qualitative and/or quantitative methods can be used to reach these aims. Two types of equivalence are typically sought: equivalence of meaning, also called semantic or conceptual equivalence, obtained through careful translation process and qualitative field test, and equivalence of measurement, obtained through careful development and quantitative field test, with CTT (internal consistency) or with IRT (differential item functioning; Petersen et al., 2003; Eremenco et al., 2005), as described earlier. Hall et al. (2018b) state that field testing “is important before proceeding to a wider evaluation of its psychometric properties or before using the translation in real clinical research” (p. 171).

### Case Study: Translation of the Cochlear Implant Quality of Life-35 Instrument From English to French

It is planned to conduct field testing in two parts. The first part will involve cognitive interviews, a qualitative research method, in a purposely selected and small sample of French-speaking adult patients with CIs. Data collection in close collaboration with two French-speaking CI centers located on two continents (Africa and Europe) have been initiated. This geographical diversity will help determine whether the French-language CIQOL-35 can be used in different French-speaking communities around the world. An interview guide in French has been prepared to facilitate the cognitive interviews. The interview guide queries the respondents on their understanding of the items and the response options and of the cognitive processes engaged when mapping their experiences to the response options. The interview guide also asks whether any item is unsuitable or offensive in the culture of the person completing the PROM. The goal is for the PROM to be a valid representation of the lived experiences of the patients whilst not being cognitively taxing and being acceptable. The responses to the cognitive interview questions will be noted by the interviewers and will be analyzed using content analysis (Graneheim and Lundman, 2004). The second part of the field testing will involve pilot testing in a randomly selected sample of French-speaking adult patients with CIs, without the same emphasis on geographical diversity. Their scores will be summarized using descriptive statistics.

**TABLE 3 |** Reconciliation report and how it was adapted from Hall et al. (2018b) supplemental file 3 for the purposes of the present translation project.

| **Step** (Hall et al., 2018b) | **Column in reconciliation report** | **Title**<br>**Description** | **Example** |
|---|---|---|---|
| 2. Forward translation | 1 | **Descriptor of CIQOL section**<br>Identifier of the section of the SL CIQOL | Item 3 |
| | 2 | **SL English CIQOL (United States 2019)**<br>Section of the SL CIQOL | If I am interested, I will join family or friends for a social event |
| | 3 | **FT #1 French**<br>Made by native speaker professional translator; With red font used to flag discrepancies identified during reconciliation (7–8 below) | Je n'hésite pas à participer à des réunions d'amis ou à des réunions de famille si j'en ai envie. |
| | 4 | **Scoring of FT difficulty by professional translator**<br>[0–10], where 0 is extremely easy and 10 is extremely hard | 7 |
| | 5 | **FT #2 French**<br>Made by native speaker health professional; With red font used to flag discrepancies identified during reconciliation (7–8 below) | Si un événement social m'intéresse, j'y participe avec ma famille ou mes amis |
| | 6 | **Scoring of FT difficulty by health professional**<br>[0–10], where 0 is extremely easy and 10 is extremely hard (same as 4 above) | 4 |
| | 7 | **Reconciliation of FT**<br>Combination of the two independent FTs (3 and 5 above), with dedicated effort spent on the sections rated as relatively more difficult to translate compared to other sections (4 and 6 above) | Si un événement social avec ma famille ou mes amis m'intéresse, j'y participe. |
| | 8 | **Reconciliation Reasoning**<br>Where red font flagged discrepancies between the two FTs, reasons for selecting one translation over another | "N'hésite pas" and "j'ai envie" are less neutral in meaning |
| 3. Back translation | 9 | **BT**<br>Made by native speaker professional translator naive to the SL CIQOL | If I am interested in a family gathering or social event, I participate. |
| | 10 | **SL-BT discrepancy classification**<br>A: Perfect semantic equivalence and good literal and semantic parallels<br>B: Satisfactory semantic equivalence, but have used one or two different words<br>C: Preserves the meaning of the SL, but without satisfactory semantic equivalence<br>D: No agreement | B |
| | 11 | **Reconciliation Reasoning**<br>Reasons for adjusting the FT based on the input of the BT, if relevant | Form different but meaning mostly preserved |
| | 12 | **Updated FT**<br>With sections in red requiring input from SL developers and/or committee review | Si un événement social m'intéresse, j'y participe avec ma famille ou mes amis. |
| | 13 | **Questions for the SL developers and their comments**<br>To clarify original meaning of SL CIQOL sections, with comments identified by SL initials | Clarify meaning of SL item to identify best translation<br>TRM: SL meaning mostly preserved in BT<br>JRD: Slight difference, SL meaning is "going/joining with family and friends to a social event" |
| | 14 | **Updated FT**<br>After input from SL developers, before committee review, with sections in red requiring input from committee review | Si un événement social m'intéresse, j'y participe avec ma famille ou mes amis. |
| 4. Committee Review | 15 | **Questions for committee review**<br>Divided into background and question<br>With reference to 14 above for context | Background: Let's check the FT against the SL given the BT-SL discrepancies<br>Question: Ok with "Si un événement social m'intéresse, j'y participe avec ma famille ou mes amis."? |
| | 16 | **Comments from committee review**<br>With comments identified by committee review participant initials | ACB: Si un événement social m'intéresse, j'y rejoins ma famille ou mes amis.<br>KJ/EF: Supports FT offered in item 5 above<br>EF: Je n'hésite pas à participer. si j'en ai envie. |

*(Continued)*

**TABLE 3 |** (Continued)

| **Step** (Hall et al., 2018b) | **Column in reconciliation report** | **Title** **Description** | **Example** |
|---|---|---|---|
| x. Additional steps not mentioned in Hall et al. (2018b) | 17 | **Review from native French speaker naive to the SL CIQOL** To ensure natural language | N/A: no comment provided on this item |
| | 18 | **Updated FT** On the basis of 16–17 above | Je n'hésite pas à participer à un événement social avec des amis ou de la famille si j'en ai envie. |
| | 19 | **Questions for the SL developers and their comments** To clarify original meaning of SL CIQOL sections, with comments identified by SL initials | N/A: no comment provided on this item |
| | 20 | **Updated FT** After input from SL developers, | Je n'hésite pas à participer à un événement social avec des amis ou de la famille si j'en ai envie. |
| | 21 | **Review from native French speaker with hearing impairment naive to the SL CIQOL** To ensure appropriateness | N/A: no comment provided on this item |
| | 22 | **Updated FT** On the basis of 21 above | Je n'hésite pas à participer à un événement social avec des amis ou de la famille si j'en ai envie. |
| | 23 | **BT** Performed by professional translator naive to the translation process so far | I do not hesitate to participate in a social event with friends or family if I want to. |
| | 24 | **SL-BT discrepancy classification** A: Perfect semantic equivalence and good literal and semantic parallels B: Satisfactory semantic equivalence, but have used one or two different words C: Preserves the meaning of the SL, but without satisfactory semantic equivalence D: No agreement (same as 10 above) | B |
| | 25 | **Reconciliation Reasoning** Reasons for adjusting the FT based on the input of the BT, if relevant | Wording slightly different but meaning fully preserved |
| | 26 | **Updated FT** FT ready for field test | Je n'hésite pas à participer à un événement social avec des amis ou de la famille si j'en ai envie. |

*We also added the date and person responsible for each step/column in the reconciliation report. BT, back translation; FT, forward translation; SL, source language.*

## Step 6. Review and Translation Finalization

### Summary of Best Practices

According to Hall et al. (2018b), this final step includes the review of the Step 5 results and their incorporation into the final translation, archiving, and dissemination.

### Case Study: Translation of the Cochlear Implant Quality of Life-35 Instrument From English to French

It is planned to incorporate results of both quantitative and qualitative field testing activities in the final version of the French-language CIQOL-35. It is hoped that the final translation of the PROM will be widely shared free of charge on the Medical University of South Carolina website and disseminated to researchers, clinicians, hearing intervention program developers, and any other people interested in the measurement of CI-related quality of life in adults.

# ANTICIPATED RESULTS AND DISCUSSION

It is hoped that this careful translation and cultural adaptation of the CIQOL-35 instrument will lead to a psychometrically sound PROM for French-speaking populations. Ideally, French-speaking adults will find the instrument a relevant and suitable tool to capture the extent to which hearing impairment and hearing interventions impact their CI-related quality of life. Such PROMs are instrumental to quality of care monitoring and improvement.

The translation and cultural adaptation of the CIQOL instruments to eight other languages (Arabic, Danish, German, Hebrew, Malay, Mandarin, Spanish, and Turkish) has been initiated. Cultural adaptation is facilitated by translators living in the location where the translated PROM will be used and by qualitative validation (e.g., with cognitive interviews) to ensure the items are appropriate, understandable, relevant and respectful (i.e., not offensive). In some of the regions in which these languages are spoken, there are no validated speech recognition word/sentence lists. Thus, these PROMs will be heavily relied on to monitor treatment outcomes. The extent of relationships between PROMs needs to continue to be investigated and reported. A related movement is the development of core outcome sets. A core outcome set is an agreed minimum battery of outcome measures to be included in clinical trials. Clinicians and researchers are free to add additional outcome measures, but adherence to a core outcome set ensures that some outcome measures are consistently collected and reported. Core sets have the advantage to allow for cross-trial comparisons and data pooling in meta-analyses. An iterative and multisectoral project led to the World Health Organization's International Classification of Functioning, Disability and Health Core Sets for hearing loss (ICF Research Branch, 2017). Only a few of the Core Set categories can be measured through physiological and behavioral tests routinely used in clinical and research practice: many more can be evaluated with PROMs. This core set is currently being validated in the population of interest (Karlsson et al., 2021). Core sets are starting to emerge regarding different sub-populations of people with hearing disorders (Hall et al., 2018a; Hill-Feltham et al., 2020; Katiri et al., 2020).

Overall, our experience has taught us that the process of PROM development, translation, and cross-cultural adaptation requires significant time and resources. Therefore, it is best to consider this type of work as a stand-alone project well ahead of time rather than as a quick endeavor when the need for a translation of a PROM becomes apparent.

## Future Trends in Hearing-Related Patient-Reported Outcome Measures

The promises that technologies offer in improving healthcare delivery have been described for decades. The COVID-19 pandemic, and its physical distancing imperative, has accelerated trends toward hybrid hearing healthcare services, combining traditional face-to-face as well as remote care modalities, such as telehealth. Care modalities can be chosen based on patient and context needs. This trend is relevant to PROMs as remote data collection is time efficient and allows the measurement to take place at a time and place that is convenient for the patient, rendering the measure a better reflection of true everyday functioning. Ecological momentary assessment is interesting in a world where ubiquitous technology is increasingly following us, quantifying, tracking, and even pre-empting our behaviors and thoughts. The convergence of PROMs with hearing device usage, acoustic environments, and health and wellbeing data provides a holistic view of a patient's level of functioning against the context and environment in which they evolve (Timmer et al., 2018). Cloud-based programming of hearing devices also calls for PROMs that are easier and closer to the patient so that replacing face-to-face appointments with remote care does not have to compromise opportunities for outcome measures. Method of administration as well as timing in the course of care can impact on completion and scores. For example, evidence shows that people with hearing impairment may complete PROMs differently when administration is done with pen and paper vs. online (Thorén et al., 2012).

### Computerized Adaptive Testing

A promising mode of PROM administration is CAT, where an algorithm selects, based on item difficulty and patient responses to previous items, an individualized set of items from a bank of IRT-calibrated items. Items are presented until a predefined measurement precision is reached, or a pre-set maximum number of presented items is reached. CAT increases measurement precision without increasing administration time, thus reducing burden for patients, clinicians, and researchers. Often CATs can provide a similar degree of precision as the full item bank, with completion of far fewer items (Choi et al., 2010; Fries et al., 2014; Pilkonis et al., 2014), and are easily adapted to smartphone or tablet administration. CATs have been developed for each of the CIQOL domains (CIQOL-CAT). Final validation and reliability testing is pending.

### Patient-Reported Experience Measures

Whilst this manuscript focuses on PROMs, some authors differentiate those from patient-reported experience measures (PREMs) (Kingsley and Patel, 2017). PREMs gather the patient's perspectives and views of their care experience. Whilst PROMs measure care outcomes, PREMs measure how the patient experienced the care process, for example in terms of communication skills, patient-centeredness, and timeliness. PREMs lead to information central to improve care; they are currently underused within hearing care.

### Economic Evaluation and Allocation of Scarce Resources

In an era of accountability in healthcare, economic evaluations are increasingly needed to inform the careful allocation of scarce resources. These compare the benefits and costs of several treatment options and use health state values, or utilities, representing people's preferences for a given health state. The cost per quality-adjusted life year (QALY) or disability-adjusted life year (DALY) and the comparison of incremental cost-effectiveness ratios to thresholds inform the evidence-based prioritization of interventions across health conditions. A recent systematic review identified 117 published economic analyses of hearing healthcare across the continuum of care from prevention and screening to CI and hearing aid provision (Borre et al., 2021). Of those, 62% measured health outcomes in QALYs and 12% in DALYs.

The measurement and valuation of the benefits of medical devices have challenges. First, costs are easier to measure and value than benefits (Thum et al., 2020). Second, generic PROMs are not suitable given limitations. The impact of CI on quality of life in older adults has been measured with the HUI2 and HUI3 (Andries et al., 2021), however EQ-5D lacks construct validity for hearing and HUI3 exhibits ceiling effects, uses "hearing aid" in the item, and measures hearing ability through speech reception only; generic PROMs underestimate the impact of hearing intervention such as CIs (McRackan et al., 2019a). Therefore, it is imperative that PROMs allow for the suitable valuation of hearing intervention benefits.

### Patient-Reported Outcome Measures in Low-Resource Settings

Patient-reported outcome measures are also of interest in low-resource settings, such as low- and middle-income countries. Their advantages include administration that does not require trained professionals and allows rapid assessment, which makes them interesting as part of monitoring and evaluation of both clinical as well as public health initiatives (Kaspar et al., 2021).

Also, PROMs, unlike other forms of outcome measures, do not rely on specialized equipment that requires frequent calibration.

## Clinical Applications of Patient-Reported Outcome Measures

The clinical application of PROMs should be both the start and the end point in PROM development. There is a misconception that if a PROM is carefully developed and translated, it will automatically, or almost magically, be implemented when ready. Careful knowledge translation and implementation science are mandatory for sustainable changes in practice. PROMs improve communication and counseling between professionals and patients (as well as inter-professional communication) regarding the health condition and its impact on quality of life (Greenhalgh et al., 2018). Still, they are underused. Learnings from implementation science can support the successful usage of PROMs to address common implementation barriers such as PROMs inadequately integrated in electronic health record systems, uncertainty about how or why to use PROMs to improve patient care, and clinical workflows that are not conducive, for example due to time pressures (Stover et al., 2020). The successful implementation of PROMs in routine clinical care requires that organizations invest time and resources in the early stages of "designing" the processes for using PROMs (i.e., planning not just which PROMs to use and how to administer them, but also how the data would be used for clinical purposes) and "preparing" an organization and its staff (i.e., getting an organization and its staff ready to use PROMs, particularly persuading clinicians of the validity and value of PROMs, delivering training, and developing electronic systems; Foster et al., 2018).

Selecting the best PROM is paramount. A systematic review concluded that eight criteria should inform PROM selection: appropriateness (match between PROM and specific purpose including research questions if relevant), reliability (reproducibility and internal consistency), validity (whether the PROM measures what it intends to), responsiveness (PROM sensitivity to changes of importance to patients), precision (number and accuracy of distinctions the PROM make), interpretability (how meaningful the PROM scores are), acceptability (how acceptable patients find PROM completion), and feasibility (extent of effort, burden and disruption to staff and clinical care arising from using the PROM; Fitzpatrick et al., 1998). The target patient group, the treatment, and the outcome of the PROM should match the clinical needs. Prior use in groups of similar patients is particularly helpful, and a pilot of the PROM implementation questionnaire can help identify any potential barriers. A short and relevant instrument that is future proof is more likely to be sustainably implemented. An easy, license-free online access to the latest version of any PROM is also conducive to implementation. Of course, if the PROM has been translated, the quality of this process should be documented. The timing and the method of administration should also be carefully considered (Bernstein et al., 2020).

Furthermore, the hearing community needs to reach consensus on the most important outcome domains and the core set of measures to assess these domains in a consistent and therefore comparable fashion in clinical research, clinical trials, and in the monitoring of the impact of hearing health policies. Current minimum reporting standards for adult CI do not include PROMs (Adunka et al., 2018). The hearing community can learn from other fields where core sets of measures of treatment effects include PROMs. Medical device regulators worldwide are also increasingly asking for the systematic collection and reporting of PROMs during the entire product lifecycle.

## Summary and Conclusion

This case study centered around the CIQOL-35 Profile instrument measuring functional abilities and quality of life in adults with hearing impairment, showed that the development of PROMs should be driven by real-world needs. The development of PROMs must start with clinical need and must ensure active involvement of important stakeholders at all stages. This is the case of the CIQOL suite of instruments, which benefited from a systematic literature review and focus groups with patients, who are experts in lived experiences of hearing impairment. The mixed methods used in the development of the CIQOL suite of instruments enhance and expand their potential applications.

This paper concludes with four suggestions for people embarking on similar endeavors:

1. Think clinical applications first, in terms of populations, concepts to be measured, etc.
2. Start with performing a literature review and inviting patient perspectives, because there is no need to reinvent the wheel and because patients are the experts into the lived experiences of a health condition.
3. Adhere to published standards of both development as well as translation and cultural adaptation, because there exists a large body of psychometric science to draw from.
4. Aim for PROMs that will stand the test of time, for example in terms of content and modes of administration.

# 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/s.

# AUTHOR CONTRIBUTIONS

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

## FUNDING

Some of the work presented was funded by the NIH grant # K23 DC016911 awarded to TM. Oticon Medical A/S covers the expenses related to the translation, cross-cultural adaptation, and validation of the CIQOL-35 Profile from English to French.

## ACKNOWLEDGMENTS

The authors thank Marine Ardoint from Oticon Medical A/S and the translators from Translate Plus for their input to the translation work. The authors acknowledge the support of Sarah Chouikh from Oticon Medical A/S in the planning of the validation study.

## REFERENCES

Adunka, O. F., Gantz, B. J., Dunn, C., Gurgel, R. K., and Buchman, C. A. (2018). Minimum reporting standards for adult cochlear implantation. *Otolaryngol. Head Neck Surg.* 159, 215–219. doi: 10.1177/0194599818764329

Ambert-Dahan, E., Laouénan, C., Lebredonchel, M., Borel, S., Carillo, C., Bouccara, D., et al. (2018). Evaluation of the impact of hearing loss in adults: validation of a quality of life questionnaire. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 135, 25–31. doi: 10.1016/j.anorl.2017.09.003

Andries, E., Gilles, A., Topsakal, V., Vanderveken, O., Van de Heyning, P., Van Rompaey, V., et al. (2021). The impact of cochlear implantation on health-related quality of life in older adults, measured with the Health Utilities Index Mark 2 and Mark 3. *Eur. Arch. Oto Rhino Laryngol.* 26, 61–75. doi: 10.1007/s00405-021-06727-3

Artières-Sterkers, F., Mondain, M., Aubry, K., Bordure, P., Bozorg-Grayeli, A., Deguine, O., et al. (2020). The French national cochlear implant registry (EPIIC): results, quality of life, questionnaires, academic and professional life. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 137, S57–S63. doi: 10.1016/j.anorl.2020.07.013

Bernstein, D. N., McIntyre, A. W., and Baumhauer, J. F. (2020). Effect of assessment administration method and timing on patient-reported outcome measures completion and scores: overview and recommendations. *Musculoskeletal Care* 18, 535–40. doi: 10.1002/msc.1476

Borre, E. D., Diab, M. M., Ayer, A., Zhang, G., Emmett, S. D., Tucci, D. L., et al. (2021). Evidence gaps in economic analyses of hearing healthcare: a systematic review. *EClinicalMedicine* 35:100872. doi: 10.1016/j.eclinm.2021.100872

Cassarly, C., Matthews, L. J., Simpson, A. N., and Dubno, J. R. (2020). The revised hearing handicap inventory and screening tool based on psychometric reevaluation of the hearing handicap inventories for the elderly and adults. *Ear Hear.* 41, 95–105. doi: 10.1097/AUD.0000000000000746

Centers for Medicare and Medicaid Services [CMMS] (2021). *Meaningful Measures Hub*. Available online at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/MMF/General-info-Sub-Page (accessed May 3, 2021).

Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., and Cella, D. (2010). Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual. Life Res.* 19, 125–136. doi: 10.1007/s11136-009-9560-5

Eremenco, S. L., Cella, D., and Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval. Health Prof.* 28, 212–232. doi: 10.1177/0163278705275342

Fitzpatrick, R., Davey, C., Buxton, M. J., and Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technol. Assess.* 2:14.

Foster, A., Croot, L., Brazier, J., Harris, J., and O'Cathain, A. (2018). The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J. Patient-Reported Outcomes* 2:46. doi: 10.1186/s41687-018-0072-3

Fraysse, B., Van, H. T., Collignon, C., and Ruidavets, B. (2020). The French cochlear implant registry (EPIIC): place and role of registries in post-registration studies of medical devices. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 137, S1–S3. doi: 10.1016/j.anorl.2020.07.008

Fries, J. F., Witter, J., Rose, M., Cella, D., Khanna, D., and Morgan-DeWitt, E. (2014). Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J. Rheumatol.* 41, 153–158. doi: 10.3899/jrheum.130813

Granberg, S., Dahlström, J., Möller, C., Kähäri, K., and Danermark, B. (2014). The ICF core sets for hearing loss–researcher perspective. Part I: systematic review of outcome measures identified in audiological research. *Int. J. Audiol.* 53, 65–76. doi: 10.3109/14992027.2013.851799

Graneheim, U. H., and Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ. Today* 24, 105–112. doi: 10.1016/j.nedt.2003.10.001

Greenhalgh, J., Gooding, K., Gibbons, E., Dalkin, S., Wright, J., Valderas, J., et al. (2018). How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. *J. Patient Reported Outcomes* 2:42. doi: 10.1186/s41687-018-0061-6

Guyatt, G. H., Feeny, D. H., and Patrick, D. L. (1993). Measuring health-related quality of life. *Ann. Intern. Med.* 118, 622–629. doi: 10.7326/0003-4819-118-8-199304150-00009

Haile, L. M., Kamenov, K., Briant, P. S., Orji, A. U., Steinmetz, J. D., Abdoli, A., et al. (2021). Hearing loss prevalence and years lived with disability, 1990–2019: findings from the global burden of disease study 2019. *Lancet* 397, 996–1009. doi: 10.1016/S0140-6736(21)00516-X

Hall, D. A., Smith, H., Hibbert, A., Colley, V., Haider, H. F., Horobin, A., et al. (2018a). The COMiT'ID study: developing core outcome domains sets for clinical trials of sound-, psychology-, and pharmacology-based interventions for chronic subjective tinnitus in adults. *Trends Hear.* 22:2331216518814384. doi: 10.1177/2331216518814384

Hall, D. A., Zaragoza Domingo, S., Hamdache, L. Z., Manchaiah, V., Thammaiah, S., Evans, C., et al. (2018b). A good practice guide for translating and adapting hearing-related questionnaires for different languages and cultures. *Int. J. Audiol.* 57, 161–175. doi: 10.1080/14992027.2017.1393565

Hill-Feltham, P. R., Johansson, M. L., Hodgetts, W. E., Ostevik, A. V., McKinnon, B. J., Monksfield, P., et al. (2020). Hearing outcome measures for conductive and mixed hearing loss treatment in adults: a scoping review. *Int. J. Audiol.* 60, 239–245. doi: 10.1080/14992027.2020.1820087

ICF Research Branch (2017). *ICF Core Set for Hearing Loss*. Available online at: http://www.icf-research-branch.org/icf-core-sets-projects2/other-health-conditions/icf-core-set-for-hearing-loss (accessed May 3, 2021).

Karlsson, E., Mäki-Torkko, E., Widén, S., Gustafsson, J., Manchaiah, V., Mahomed-Asmail, F., et al. (2021). Validation of the Brief International Classification of Functioning, Disability and Health (ICF) core set for hearing loss: an international multicentre study. *Int. J. Audiol.* 60, 412–420. doi: 10.1080/14992027.2020.1846088

Kaspar, A., Pifeleti, S., and Driscoll, C. (2021). The need for translation and cultural adaptation of audiology questionnaires to enable the development of hearing healthcare policies in the Pacific Islands: a Samoan perspective. *Arch. Pub. Health* 79:80. doi: 10.1186/s13690-021-00606-3

Katiri, R., Hall, D. A., Buggy, N., Hogan, N., Horobin, A., Van De Heyning, P., et al. (2020). Core Rehabilitation Outcome Set for Single Sided Deafness (CROSSSD) study: protocol for an international consensus on outcome measures for single sided deafness interventions using a modified Delphi survey. *Trials* 21:238. doi: 10.1186/s13063-020-4094-9

Kingsley, C., and Patel, S. (2017). Patient-reported outcome measures and patient-reported experience measures. *Br. J. Anaesth Educ.* 17, 137–144. doi: 10.1093/bjaed/mkw060

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., and Nguyen, S. A. (2018a). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope* 128, 982–990. doi: 10.1002/lary.26738

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Velozo, C. A., et al. (2018b). Meta-analysis of cochlear implantation outcomes evaluated with general health-related patient-reported outcome measures. *Otol. Neurotol.* 39, 29–36. doi: 10.1097/MAO.0000000000001620

McRackan, T. R., Fabie, J. E., Bhenswala, P. N., Nguyen, S. A., and Dubno, J. R. (2019a). General health quality of life instruments underestimate the impact of bilateral cochlear implantation. *Otol. Neurotol.* 40, 745–753. doi: 10.1097/MAO.0000000000002225

McRackan, T. R., Hand, B. N., Cochlear Implant Quality of Life Development Consortium, Velozo, C. A., and Dubno, J. R. (2019b). Cochlear implant quality of life (CIQOL): development of a profile instrument (CIQOL-35 profile) and a global measure (CIQOL-10 Global). *J. Speech Lang Hear Res.* 62, 3554–3563. doi: 10.1044/2019_JSLHR-H-19-0142

McRackan, T. R., Hand, B. N., Velozo, C. A., and Dubno, J. R., Cochlear Implant Quality of Life Consortium (2019c). Development of the Cochlear Implant Quality of Life (CIQOL) item bank. *Ear Hear.* 40, 1016–1024. doi: 10.1097/aud.0000000000000684

McRackan, T. R., Hand, B. N., Velozo, C. A., Dubno, J. R., and Cochlear Implant Quality of Life Consortium. (2021). Validity and reliability of the Cochlear Implant Quality of Life (CIQOL)-35 profile and CIQOL-10 Global instruments in comparison to legacy instruments. *Ear Hear.* 42, 896–908. doi: 10.1097/AUD.0000000000001022

McRackan, T. R., Velozo, C. A., Holcomb, M. A., Camposeo, E. L., Hatch, J. L., Meyer, T. A., et al. (2017). Use of adult patient focus groups to develop the initial item bank for a cochlear implant quality-of-life instrument. *JAMA Otolaryngol. Head Neck Surg.* 143, 975–982. doi: 10.1001/jamaoto.2017.1182

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual. Life Res.* 19, 539–549. doi: 10.1007/s11136-010-9606-8

National Health Service [NHS] (2021). *Patient Reported Outcome Measures (PROMs)*. Available online at: http://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms (accessed May 3, 2021).

Organisation Internationale de la Francophonie (2019). *La langue française dans le monde ("The French language in the world*"). Available online at: http://www.francophonie.org/sites/default/files/2021-04/LFDM-20Edition-2019-La-langue-fran%C3%A7aise-dans-le-monde.pdf (accessed May 3, 2021).

Patrick, D. L., Burke, L. B., Powers, J. H., Scott, J. A., Rock, E. P., Dawisha, S., et al. (2007). Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 10, S125–S137. doi: 10.1111/j.1524-4733.2007.00275.x

Petersen, M. A., Groenvold, M., Bjorner, J. B., Aaronson, N., Conroy, T., Cull, A., et al. (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Qual. Life Res.* 12, 373–385. doi: 10.1023/a:1023488915557

Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Maihoefer, C. C., and Lawrence, S. M. (2014). Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *J. Psychiatr Res.* 56, 112–119. doi: 10.1016/j.jpsychires.2014.05.010

Prieto, L., Alonso, J., and Lamarca, R. (2003). Classical Test Theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual. Life Outcomes* 1:27. doi: 10.1186/1477-7525-1-27

PROMIS (2013). *PROMIS: Instrument Development and Validation Scientific Standards. Version 2.0.* Available online at: http://www.mcgill.ca/can-pro-network/files/can-pro-network/promisstandards_vers2.0_final.pdf (accessed May 3, 2021).

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med. Care* 45, S22–S31. doi: 10.1097/01.mlr.0000250483.85507.04

Rose, M., Bjorner, J. B., Becker, J., Fries, J. F., and Ware, J. E. (2008). Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J. Clin. Epidemiol.* 61, 17–33. doi: 10.1016/j.jclinepi.2006.06.025

Stover, A. M., Haverman, L., van Oers, H. A., Greenhalgh, J., and Potter, C. M. (2020). Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. *Qual. Life Res.* 30, 3015–3033. doi: 10.1007/s11136-020-02564-9

Thorén, E. S., Andersson, G., and Lunner, T. (2012). The use of research questionnaires with hearing impaired adults: online vs. paper-and-pencil administration. *BMC Ear Nose Throat Disord.* 12:12. doi: 10.1186/1472-6815-12-12

Thum, C., Lenarz, T., Lesinski-Schiedat, A., and Fleßa, S. (2020). Lifetime cost of unilateral cochlear implants in adults: a monte carlo simulation. *Eur J. Health Econ.* 21, 963–975. doi: 10.1007/s10198-020-01188-7

Timmer, B. H., Hickson, L., and Launer, S. (2018). The use of ecological momentary assessment in hearing research and future clinical applications. *Hear. Res.* 369, 24–28. doi: 10.1016/j.heares.2018.06.012

Vas, V., Akeroyd, M. A., and Hall, D. A. (2017). A data-driven synthesis of research evidence for domains of hearing loss, as reported by adults with hearing loss and their communication partners. *Trends Hear.* 21:2331216517734088. doi: 10.1177/2331216517734088

World Health Organization [WHO] (2021a). *WHOQOL: Measuring Quality of Life*. Available online at: http://www.who.int/tools/whoqol (accessed May 3, 2021).

World Health Organization [WHO] (2021b). *World Report on Hearing*. Available online at: http://www.who.int/publications/i/item/world-report-on-hearing (accessed May 3, 2021).

**Conflict of Interest:** AL-L is employed by Oticon Medical A/S, part of Demant A/S, a hearing healthcare company. TM is an educational consultant with Stryker Corporation and on the medical advisory board of Envoy Medical.

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

The authors declare that this study received funding from Oticon Medical (ALL) and NIH/NIDCD K23 DC016911 (TRM). The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 30 November 2021
doi: 10.3389/fnins.2021.773427

# Home-Based Speech Perception Monitoring for Clinical Use With Cochlear Implant Users

*Astrid van Wieringen*, Sara Magits, Tom Francart and Jan Wouters*

*Experimental ORL, Department of Neurosciences, KU Leuven, Leuven, Belgium*

Speech-perception testing is essential for monitoring outcomes with a hearing aid or cochlear implant (CI). However, clinical care is time-consuming and often challenging with an increasing number of clients. A potential approach to alleviating some clinical care and possibly making room for other outcome measures is to employ technologies that assess performance in the home environment. In this study, we investigate 3 different speech perception indices in the same 40 CI users: phoneme identification (vowels and consonants), digits in noise (DiN) and sentence recognition in noise (SiN). The first two tasks were implemented on a tablet and performed multiple times by each client in their home environment, while the sentence task was administered at the clinic. Speech perception outcomes in the same forty CI users showed that DiN assessed at home can serve as an alternative to SiN assessed at the clinic. DiN scores are in line with the SiN ones by 3–4 dB improvement and are useful to monitor performance at regular intervals and to detect changes in auditory performance. Phoneme identification in quiet also explains a significant part of speech perception in noise, and provides additional information on the detectability and discriminability of speech cues. The added benefit of the phoneme identification task, which also proved to be easy to administer at home, is the information transmission analysis in addition to the summary score. Performance changes for the different indices can be interpreted by comparing against measurement error and help to target personalized rehabilitation. Altogether, home-based speech testing is reliable and proves powerful to complement care in the clinic for CI users.

**Keywords: speech understanding in noise, digits in noise, phoneme identification in quiet, CI users, home testing**

OPEN ACCESS

***Edited by:***
*Monita Chatterjee, Boys Town, United States*

***Reviewed by:***
*Tobias Goehring, University of Cambridge, United Kingdom*
*Mary M. Flaherty, University of Illinois at Urbana-Champaign, United States*

****Correspondence:***
*Astrid van Wieringen*
*astrid.vanwieringen@kuleuven.be*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *09 September 2021*
***Accepted:*** *28 October 2021*
***Published:*** *30 November 2021*

***Citation:***
*van Wieringen A, Magits S, Francart T and Wouters J (2021) Home-Based Speech Perception Monitoring for Clinical Use With Cochlear Implant Users. Front. Neurosci. 15:773427. doi: 10.3389/fnins.2021.773427*

## INTRODUCTION

Speech perception assessment is a cornerstone of audiological rehabilitation (Boothroyd, 1994). It is usually assessed in the clinic with meaningful words and sentences in quiet and (sometimes) in noise. These scores reflect large variability in performance for persons with hearing aids (HA) and cochlear implants (CI), especially in noise (e.g., Gifford et al., 2008, 2015; Zeitler et al., 2008; Meister et al., 2015; Ricketts et al., 2019), due to differences in patient demographics, as well as technical, linguistic, and cognitive factors (Rählmann et al., 2018; James et al., 2019; de Graaff et al., 2020; Zhao et al., 2020). The multidisciplinary nature of audiological rehabilitation requires a wide range of performance measures to capture bottom-up and top-down neurocognitive skills (Moberly et al., 2016; Rählmann et al., 2018; Skidmore et al., 2020; Tamati et al., 2020;

Völter et al., 2020; Zhan et al., 2020; Biever et al., 2021; Lundberg et al., 2021). However, clinical time is scarce. Demand for care has increased over the past years due to the expansion of candidacy criteria for cochlear implantation, advancements in technology, and improved surgical techniques (van der Straaten et al., 2020; Perkins et al., 2021). A potential approach to alleviating some of the work on clinical care and possibly making room for other outcome measures is to employ technologies that assess performance in the home environment. An increasing number of people are using their smartphones or tablets for healthcare assessment, and home-based testing could be used to monitor potential changes in hearing performance and provide guidance for audiological rehabilitation. Such an approach may be good for the clinic (reduced workload/more testing) and enhance the user's self-efficacy.

Over the past decade, different audiological service deliveries via telepractice have been explored (Swanepoel and Hall, 2010; Muñoz et al., 2021). Several applications for hearing screening have demonstrated the feasibility and reliability of telehealth (Smits et al., 2013; Louw et al., 2017). For experienced HA users, face-to-face and remote programming of hearing aids give similar speech perception results (Venail et al., 2021). Also, CI programming levels are similar when done remotely compared to the face-to-face method in the clinic, not only with adults (Ramos et al., 2009; Wesarg et al., 2010; Hughes et al., 2012; Eikelboom et al., 2014) but also with children using visual reinforcement audiometry (Hughes et al., 2018). Additionally, speech recognition of CI users can be assessed at home (de Graaff et al., 2018, 2019), although presentation mode requires some attention. With direct-connect from the computer to the CI processor, different physiological and basic perceptual measures yielded similar scores whether assessed in person or remotely (Goehring et al., 2012; Hughes et al., 2012). However, speech perception scores of CI users were significantly poorer in an office/conference room simulating remote testing than in the in-person condition in the sound booth at the clinic, presumably because of the higher background noise level and longer reverberation times at the remote sites. To overcome the adverse effects of background noise and reverberation, speech sounds can be delivered via direct audio input (DAI), bypassing the microphones. While testing with DAI has proved to be a valid alternative to standard sound-booth testing (de Graaff et al., 2016, 2019; Cullington and Aidi, 2017; Sevier et al., 2019), wireless streaming from a device to the sound processor has also become possible and can be used for testing in the home environment.

Using remote tools may also lead to increased confidence to manage one's hearing and identify problems quicker instead of waiting for a scheduled appointment at the clinic. A randomized control trial using a well-validated generic measure of patient activation showed that CI users who received remote care for device adjustment and assessment demonstrated greater user activation after 6 months than those who received the clinic-based care pathway (Cullington et al., 2018). A custom-made satisfaction questionnaire revealed that patients and clinicians were generally positive about remote care tools and wanted to continue. They liked the idea that tests can be used any time, that they receive instant feedback on progress, and that less staff is needed. These findings related to audiological rehabilitation align with a systematic review analysis and meta-analysis showing that self-management support interventions can reduce health service utilization without compromising patient health outcomes (Panagioti et al., 2014).

Not all outcome measures are suitable for remote self-testing. In the clinic, speech understanding is usually assessed with an open-set response format. The client responds verbally to the presented word or sentence, and the clinician notes down the responses. Home-based testing requires a closed-set response format, where the client chooses from a pre-defined set of alternatives unless auto-correction is applied with open-set testing (e.g., Francart et al., 2009). Another prerequisite for home-based testing is that the materials can be used repeatedly. Meaningful words and sentences cannot be used repeatedly unless an infinitive number of alternatives can be generated, such as with the Matrix sentences (Kollmeier et al., 2015) or the Coordinate Response Measure (Bolia et al., 2000). Digits and phonemes can be used repeatedly.

The digit triplet test also called the digits in noise test (DiN), is increasingly used for hearing assessment. It was initially developed for hearing screening (Smits et al., 2013; for a review, we refer to Van den Borre et al., 2021), but with persons with a cochlear implant, it is also used as an alternative for the sentence in noise (SiN) task (Kaandorp et al., 2016; Cullington and Aidi, 2017; Zhang et al., 2019). Using an adaptive procedure, the speech reception threshold is determined for digits presented in speech-weighted noise. Even persons with limited language ability are familiar with digits and can use a keypad. Long before this paradigm was developed, it was clear that an extensive range of hearing abilities can be mapped with numbers (van Wieringen and Wouters, 2008). The DiN paradigm can be used repeatedly since learning of the content is less likely to occur.

Phoneme identification, or the nonsense syllable test, is also assessed with an n-alternative closed-set response format. The summary scores (percentage correct) reflect how well a listener perceives the spectral and temporal properties of vowels and consonants (e.g., Gordon-Salant, 1985; Dorman et al., 1990; Tyler and Moore, 1992; van Wieringen and Wouters, 1999; Välimaa et al., 2002a,b; Munson et al., 2003; Nie et al., 2006; Shannon et al., 2011; Rødvik et al., 2018). Phoneme identification is not often assessed in the clinic, although responses are very insightful, as they can yield both a summary score and detailed analysis of confused speech features by means of information transmission analyses (Miller and Nicely, 1955). Phonemes are characterized by distinctive acoustic features that produce differences in voicing, manner, place of articulation, etc., Per phoneme, the transmission of different speech features is determined. The relative information transmitted is the ratio of the transmitted information calculated from the confusion matrix to the maximal possible information transferred by the stimuli and features under test. The more phonemes share distinctive features, the more likely they are confused perceptually (Miller and Nicely, 1955). The results of the information transmission analysis can guide the rehabilitation process (e.g., optimize the fitting of the device). Nonsense syllable tests also have the advantage that learning effects in multiple experiments with the same stimuli are minimal compared with tests using real-word stimuli (Dubno and Dirks, 1982).

In summary, clinical care is time-consuming and often challenging with an increasing number of clients. Speech-perception testing is essential for monitoring outcomes with a HA or CI and should encompass various measures to gain insight into variability in performance. Some of these could be done at home to complement assessment in the clinic. The study aimed to investigate performance on three different speech perception tasks, i.e., sentence identification in noise (SiN), digits in noise (DiN), and phoneme identification in quiet, in the same CI recipients during 16 weeks. We expect the digit scores to be associated with the sentence scores, and we anticipate that the vowel and consonant errors will provide additional insight into individual performance patterns. Additionally, we investigate the reliability of these indices in the home-based setting and potential differences between response scores determined at the beginning and at the end of the trial.

# METHODOLOGY

## Participants, Outcome Measures and Procedure

Forty CI users, 26 with Cochlear device, 14 with AB device, performed the phoneme and DiN tasks at home. Their median age was 64.3 years [IQR 10.4, min 28 yrs, max 75 yrs], median experience with their CI 2.1 years [IQR 4.2 yrs, range 0.1–15.9 yrs]. Thirty-six out of forty CI users had progressive hearing loss. Twenty-seven participants wore a hearing aid contralaterally (CI-HA), eight persons had one CI, three persons bilateral CIs, and two persons 1 CI and residual hearing. The participants' average pure tone average (PTA4, average of 500, 1,000, 2,000, 4,000 Hz), determined in free field at the clinic with their CI only, was 26.4 dB HL (SD 5.3). All participants presented with a postlingually acquired profound hearing impairment, and they communicated through spoken language in their daily life. The median period of education was 12.5 years [IQR 3.3].

These participants participated in a more extensive study dealing with the efficacy of a personalized listening training program LUISTER compared to a non-personalized one (Magits et al., under revision). In that study, participants were asked to practice segmental and suprasegmental speech tasks five times per week for 15 to 20 min on a tablet at home. The efficacy of the two training tasks was based on the SiN scores (pre- versus post-training) assessed at the clinic. Once a week, before practicing with a training program, the participants were asked to complete a DiN test twice and either a vowel or a consonant phoneme identification task (in quiet) at home. At home, the stimuli were streamed via Bluetooth and a streaming device to one CI. The participant chose which CI if they had two. Speech understanding in noise (SiN, pre-and post-training) was assessed at the clinic, via streaming. Three conditions were tested: (1) SiN presented via streaming to one CI (same as DiN and phoneme in quiet, "SiN streaming"), (2) in sound field to the CI only ("SiN CI-SF"), and (3) in sound field as in daily life (with CI and HA if applicable, "SiN daily settings"). The same CI devices were used at home and at the clinic. Logged data were automatically transferred to a repository hosted on the server of the research group via a restricted one-way communication from tablet to server.

Participants provided written informed consent, and the Ethics Committee approved the study of the University Hospitals Leuven (approval no. B322201731501). Participants were paid for the testing sessions but not compensated for the practicing sessions at home. The study protocol is registered on ClinicalTrials.gov (I.D. = NCT04063748).

## Outcome Measures

### Speech Understanding in Noise (SiN)

Sentence understanding in stationary speech-weighted noise (SiN) was assessed with the LIST speech materials (van Wieringen and Wouters, 2008). An adaptive method was used to determine the speech reception threshold (SRT), the signal-to-noise ratio at which 50% of the sentences are repeated correctly. Each sentence contains two to three keywords. The level of the sentences was held fixed at 65 dB SPL, the level of the noise was varied. The level of the noise for the first sentence varied until all keywords were repeated correctly. For each subsequent sentence, the level of the noise was increased or decreased in steps of 2 dB until ten sentences had been presented (Plomp and Mimpen, 1979). The SRT was the average of the last five presented signal-to-noise ratios and the signal-to-noise ratio of the imaginary 11th sentence, with lower SRT values indicating better performance.

Participants completed two lists for each of the three conditions before and at the end of the 16-week trial. A third list was completed if the two lists differed by more than 2 dB, and the average was taken. In the sound field room at the clinic speech sounds were played using APEX (Francart et al., 2017) from a tablet via a streaming device to the CI or a computer via an external sound card to the loudspeaker at 65 dB SPL. The median duration for SiN testing ranges from 2.2 min [0.5 min] to 2.4 min [0.7 min] per list of 10 sentences, hence 6–8 min in total.

### Digits in Noise (DiN)

Participants identified 17 digit triplets in stationary speech-weighted noise on the touch screen of the tablet. The development and validation of the Flemish DiN (female speaker) are described by Jansen et al. (2013). The level of the speech was fixed at 65 dB A, and the first triplet was presented at + 4 dB signal-to-noise ratio. An adaptive procedure using triplet and digit scoring and an adaptive step size converged to a threshold in noise (Denys et al., 2019). One DiN trial takes about 2.3 min [0.4 min].

### Phoneme Identification in Quiet

Both vowel and consonant identification in quiet were assessed separately. The vowel identification task consisted of 10 Dutch/Flemish vowels presented in p-t context: /oe, oo, i, I, o, u, e, ee, aa, a/. The consonant identification test consisted of 12 consonants presented in/a/context: /p, t, b, d, m, n, s, f, ch, z, v, w/. Stimuli were produced by a female speaker (van Wieringen and Wouters, 1999). Each phoneme was routed ten times from the tablet to the streaming device in random order ($n$ = 100 for vowel, $n$ = 120 for consonant). Testing was self-paced. No training nor feedback was provided. Vowel identification (100 items) takes 6.0 min [2.2] and consonant identification (120 items] takes 9.4 min [2.8 min].

Responses were cast into stimulus-response confusion matrices. Information transmission (Miller and Nicely, 1955) was determined of three speech features for the Dutch vowels: Duration, First formant frequency (F1), and Second formant frequency (F2). Classification of the vowels into these categories is the same as documented in van Wieringen and Wouters (1999, Table 3). Seven features distinguish consonants: presence/absence of voicing (voicing), perception of release burst (plos), perception of relatively high or low amplitude envelope (envel), place of articulation (place), perception of frication (fric), manner of articulation (manner), and perception of nasal cues (nasal). The classification of the consonants into these categories follows van Wieringen and Wouters (1999, Table 5).

## Procedure

### Tablet and Calibration

Testing was done with a 7.0″ Samsung Galaxy Tab A tablet and a streaming device, the phone clip or minimic for the Cochlear device ($n$ = 26) and compilots for the AB device ($n$ = 14). The output level for the speech tasks was calibrated with a personal audio cable, and the overall intensity level was set to 65 dBA. During the initial visit at the clinic participants were shown how to connect their streaming device and to run the tasks. A blue light indicated that the streaming device was connected. Participants also received manuals with clear instructions or could contact the clinician via email if needed. They were allowed to adjust the volume settings of their streaming devices but nobody reported having done this. At the end of the 16 weeks participants were asked to rate the usability of the tablet using the System Usability Scale (from 0 to 100), developed by Brooke (1996). The average SUS score was 90.5 (SD 10.4), the median is 95 (IQR 5).

### Number of Trials

All participants performed the DiN test twice sequentially and completed either a vowel identification task or a consonant identification task each week during the 16 weeks. This resulted in 1269 DiN trials, 307 vowel identification trials, and 326 consonant identification trials. The average number of trials per person was 31.7 (SD 4.1) for the DiN 7.7 (SD 1.0) for vowel identification and 8.2 (SD 1.2) for consonant identification, respectively. Since 2 (out of 40) participants performed the vowel and consonants tasks only five times, the averaged values of DiN and phoneme identification are based on the last five trials (=weeks) per participant. SiN is based on one value (average of 2–3 lists of sentences), determined in the first week and one value determined in the last week.

## Statistics

Statistical analyses were performed using IBM SPSS Statistics for Windows version 27 (2020). Data were tested for normality and homogeneity of variance. The Shapiro-Wilk showed that the DiN data distribution did not significantly differ from normal, $W(40) = 0.946$, $p = 0.057$, but the SiN data did $W(40) = 0.907$, $p = 0.003$. Vowel identification scores were normally distributed: $W(40) = 0.978$, $p = 0.628$, as well as consonant identification scores: $W(40) = 0.984$, $p = 0.845$. The pure tone average (PTA) data were also normally distributed, $W(40) = 0.967$, $p = 0.296$, but not "years of CI use," $W(40) = 0.836$, $p < 0.001$. Since the SiN data were not normally distributed we opted for median and interquartile ranges when presenting SiN with other performance measures. The non-parametric Spearman's Rho was used to determine the strength of an association between SiN and other variables, while Pearson correlation (r) was used for the normally distributed performance measures. Linear regression analyses were performed to study the relationship between different performance measures and to determine how much the different predictors explain the response. Potential differences in performance between the start and end of the 16-week trial were analyzed with the non-parametric Friedman test of differences among repeated measures, followed by a Wilcoxon signed rank test for paired comparisons.

# RESULTS

## SiN and DiN

**Figure 1** illustrates SiN (streaming) and DiN (A) and percentage vowel and consonant identification in quiet (B) for each of the 40 participants separately. Participants are ranked according to increasing (poorer) SiN scores determined at the end of the 16-week trial. These scores range from −5.2 to + 11.6 dB SNR. Generally, DiN scores are in line with the SiN ones by 3–4 dB improvement. The median SRT of the last five trials for DiN is −3.8 dB SNR [IQR, 5.0), and for SiN streamed to the device −0.3 dB SNR [IQR, 4.5). The difference between the DiN and the SiN in this study, about 3.5 dB, is also in line with the difference between the norm values of the SiN for normal hearing young persons (−7.8 dB SNR, van Wieringen and Wouters, 2008) and the norm values of the DiN (−11.7 dB SNR, Jansen et al., 2014).

Spearman's rho indicates a statistically significant relationship between the SRTs of SiN and DiN ($rs$ [40] = 0.767, $p < 0.001$). Linear regression analyses showed that DiN significantly predicts SiN, thereby explaining 74% of the variance, $F(1,38) = 621.34$, $p < 0.0001$. The model for SiN is $y = 4.52 + (1.098 * \text{score})$ with a narrow 95% confidence interval to predict SiN from the DiN score [3.5–5.4].

## Phoneme Identification in Quiet

The bottom panel (**Figure 1B**) illustrates phoneme identification in quiet for each of the participants. All participants performed well above chance (10% for vowels and 8.3% for consonants), but a wide range of performance is observed. Median vowel identification is 70.0% [IQR 17.8], median consonant identification is 64.4% [IQR 24.1]. Vowel and consonant perception in quiet are highly correlated [$r(40) = 0.678$, $p < 0.001$], the difference between the two measures is in the same order of magnitude for most participants.

Spearman's rho indicated a significant negative relationship between SiN assessed at the clinic and vowel identification in quiet assessed at home ($rs$ [40] = −0.611, $p < 0.001$), and

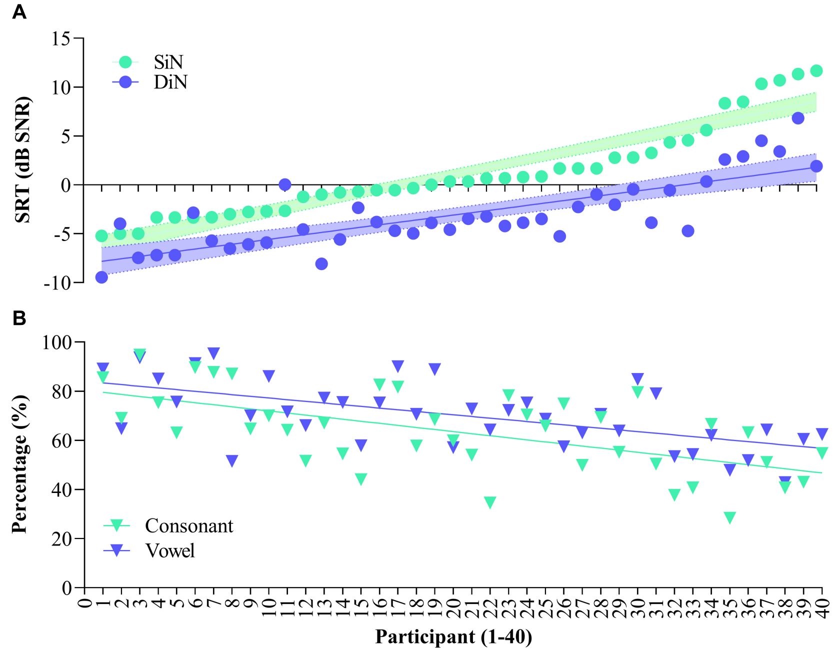


**FIGURE 1 |** Speech reception thresholds for Sentence in noise (SiN) and Digits in noise (DiN) for the 40 participants **(A)** and concomitant vowel and consonant scores [percentage correct, **(B)**]. Data are ranked according to SiN. DiN and phoneme recognition data are based on the average of the last 5 trials.

a significant negative relationship between SiN and consonant identification in quiet ($rs$ [40] = −0.587, $p < 0.001$). In other words, the more negative (better) sentence identification in noise, the higher the vowel and consonant recognition in quiet. Vowel and consonant recognition significantly predict SiN, with the linear regression model explaining 41% of the variance ($p < 0.001$). Semi partial correlations, which explain the unique contribution of each predictor variable, are 28% for vowel identification, $p = 0.031$, and 26% for consonant identification, $p = 0.043$.

As with SiN, a significant negative relationship was observed between DiN and vowel identification: $r$ (40) = −0.537, $p < 0.001$, and between DiN and consonant identification: $r$ (40) = −0.520, $p = 0.001$.

Vowel and consonant identification also significantly predict DiN, albeit somewhat less than SiN: the model explains 30% variance ($p < 0.001$). Semi partial correlations show that vowels predict 25% and consonants predict 21% of DiN.

## Perception of Vowel Features

**Figure 2A** illustrates the average percentages for each of the three vowel features for each 40 participants, together with the percentage correct score (blue stars). Participants are ranked according to increasing (poorer) SiN, as in **Figure 1**. Averaged over participants, "duration" is 50.4% (SD (19.0%), "F1" is 57.1% (SD 19%) and "F2" is 72.5% (SD 21%). F2 predicts 65% of the variance for vowel identification and was significant $F(1,305) = 574.0$, $p < 0.0001$. The linear regression analysis was performed on the individual data. Due to multicollinearity, the features "duration" and "F1" were dropped from the model ($r > 0.7$).

Information transmission analyses show that CI users with similar percent correct recognition scores can make different errors. Compare, for instance, vowel recognition of participants 15, 20, and 26 in **Figure 2A**. While percentage correct scores are similar (57%), the distributions of errors are different: participant 26 perceives the high-frequency spectral information much better (F2 cue, 76%) than participant 15 (40%) or participant 20 (57%). Likewise, participant 30 discriminates long and short vowels much better (80%) than participant 31 (53%) despite similar percentage correct scores (80%). Compare the data of participants 2 & 25, 21 & 23, and others who have similar percentage correct scores but perceive the different speech features differently.

## Perception of Consonant Features

**Figure 2B** illustrates the average percentages of the seven features per participant, together with the percentage correct scores ranked from low to high (blue crosses). Again, the order of participants is according to **Figure 1** (SiN). Perception of voicing (AVG 61.6%, SD 27.1%), and place of articulation (AVG 47.0%, SD 19.1%) remain difficult, but the perception of plosives (AVG 79.2%, SD 23.4%), the coding of temporal envelope cues (envelop, AVG 88.2%, SD 16.6%), manner of articulation (AVG = 87.1%, SD 13.8%), fricatives 86.2% (SD 18.6%), and nasals (AVG 84.8%, SD 20%) are generally good. As with vowel identification, the feature transmission analyses of the consonant confusions provide additional information on differential performance.

The perception of the seven features can vary widely for a similar percentage correct score: compare, for instance, the data of participants 14 and 21 who both have similar

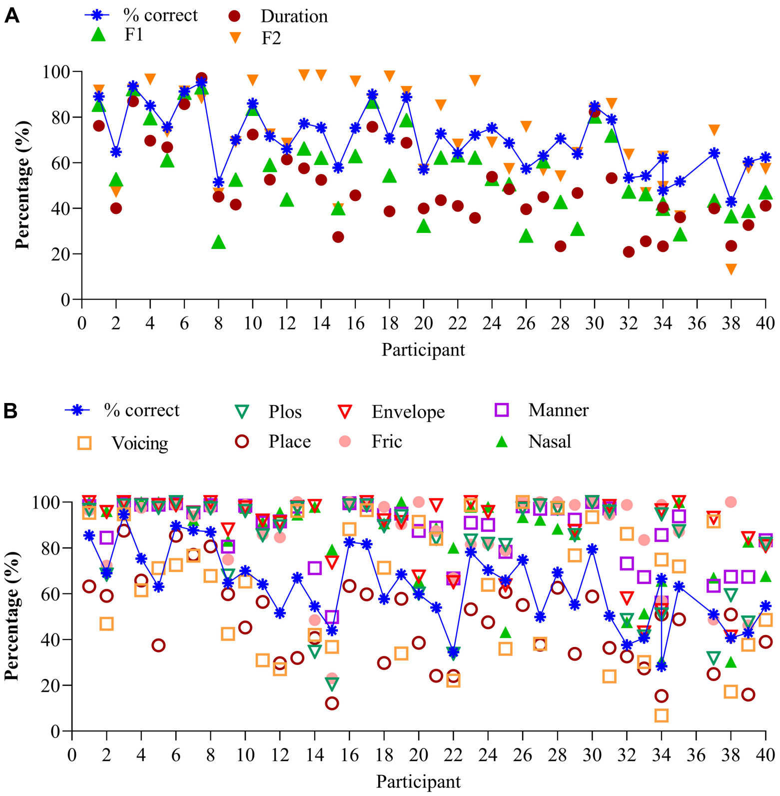


**FIGURE 2 | (A)** Vowel identification for the 40 participants, together with the speech features. Data are ranked according to increasing (poorer) SiN scores. **(B)** Consonant identification for the 40 participants, together with the speech features. Data are ranked according to increasing (poorer) SiN scores.

recognition scores (54%). However, participant 21 mainly has difficulty perceiving place of articulation and perceives the other cues very well (>80%). In contrast, participant 14 has difficulty perceiving the correct place of articulation, perception of the burst, and frication (all below 50%). Compare also data of participants 6, 7, & 8, 9 & 11 to name a few. These participants yield the same percentage correct score, yet a different distribution of errors. Analysis of the errors can guide the mapping of the device and the rehabilitation process.

Of the seven consonantal speech features, "voicing" and "frication" significantly predict 53% of the variance of consonant identification in quiet, $F(2,323) = 186.1$, $p < 0.0001$ ($n = 326$). Both features contribute uniquely to predicting consonant recognition in quiet ($r = 0.433$ for "frication" and $r = 0.3$ for "voicing." Due to multicollinearity ($r > 0.7$) the features "plos," or perception of release burst ($r = 0.77$), envelope ($r = 0.72$), place of articulation ($r = 0.85$), manner of articulation ($r = 0.8$), and nasal ($r = 0.72$) were removed from the model.

## Longitudinal Analyses and Measurement Error

A primary reason to use the DiN test is the limited content learning and thus the possibility to use the paradigm repeatedly (Smits et al., 2013). During the 16-week training trial (Magits et al., under revision), participants performed the DiN at home each week. **Figure 3** illustrates the speech reception thresholds (SRTs) as a function of time for each participant separately (ranked in the same order as in previous figures). Each data point is the average of 2 SRTs administered consecutively at the beginning of each week. For the sake of clarity each figure illustrates the data of 5 participants. Most participants yield low SRTs that vary minimally with time, especially those with very good SiN scores (participants 1–10). Test-retest reliability was determined for each participant by taking the standard deviations of the differences between the two consecutive scores, divided by square 2. This procedure outbalances a procedural learning effect (Smits and Houtgast, 2005). Averaged over all participants, the measurement error is 2.0 dB (SD 0.9). Individual measurement errors are indicated next to the participant number in **Figure 3** and range from 0.8 dB (participant 3) to 4.7 dB (participant 32). Generally, the values are larger for the participants with poorer DiN (and SiN), cf participants 30–40. These higher and more variable measurement errors were not related to the age of the participants ($r$ (40) = −0.595, $p$ = 0.09).

## Changes in Phoneme Identification in Quiet With Time

Closed-set phoneme identification in quiet also offers the possibility to monitor changes with time. Recall that the

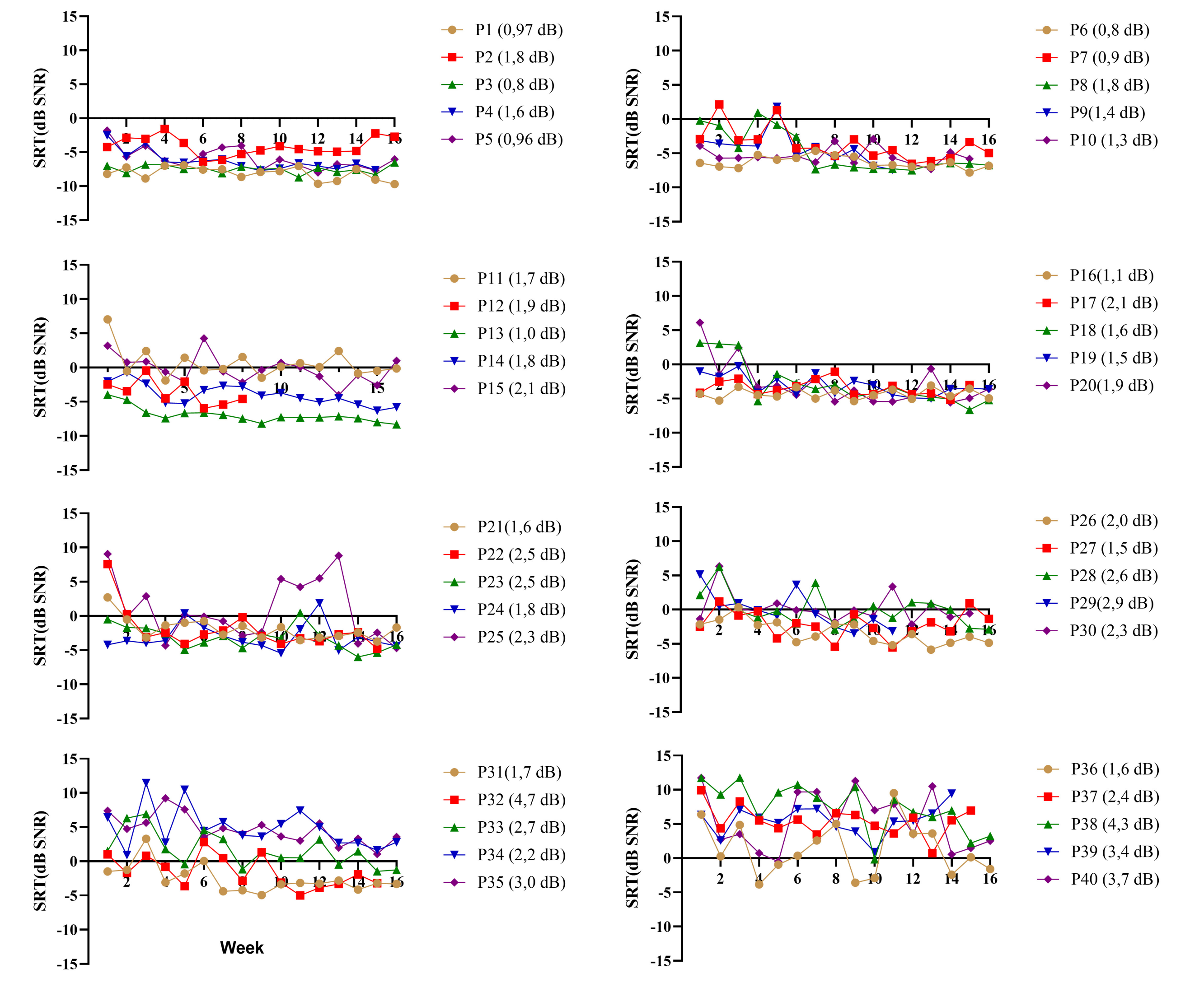


**FIGURE 3 |** DiN thresholds as a function of week (*n* = 16) for each participant (p*) separately. Each data point is based on two estimates taken consecutively. Individual measurement errors are indicated next to the participant number and range from 0.8 dB (participant 3) to 4.7 dB (participant 32). Participants (P1-P40) are ranked according to increasing SiN.

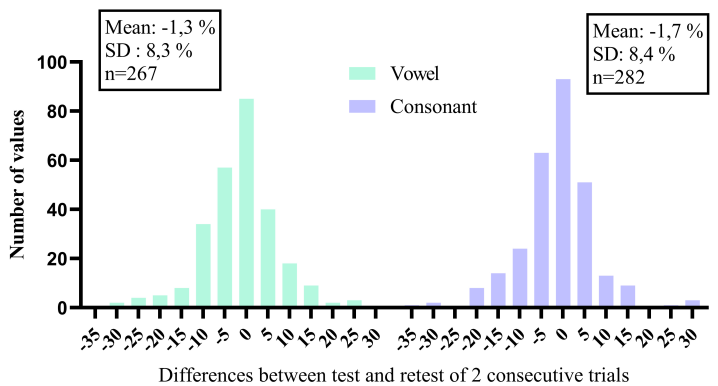


**FIGURE 4 |** Histograms of the differences in percentage correct of consecutive trials, for vowels and consonants separately.

vowel and consonant tasks were performed every other week during the 16-week trial. Summary scores and speech features of individual trials are illustrated for each participant under **Supplementary Material**. Vowel and consonant data of the 40 participants are presented in the same order as before. Many participants, especially the lowest ranked ones, show little improvement with time because they already perceive the different features very well. However, others show improvements with time, possibly because of practicing the listening training tasks or due to the weekly testing regime of DiN and phoneme identification to know whether a change in summary score is meaningful, histograms of the differences between consecutive scores were constructed for vowels and consonants separately. These are illustrated in **Figure 4**. The standard deviation of the distribution can be used as a guideline to make changes to the mapping of the device or to the audiological management of the client. For the current data, changes smaller than 10% may be meaningful, but changes larger than 10% certainly are.

## First Versus Last Measurement

We also compared potential performance differences between the first and the last measurement. For this, we compared 1 SiN value (streaming mode), the average of 2 DiN scores, and one vowel and one consonant identification score assessed at the beginning of the trials with the same outcomes assessed at the end. Phoneme data were transformed to RAU scores (Studebaker, 1985) for the statistical analyses. **Figures 5A,B** illustrate the difference in performance between the beginning and end of the period. Wilcoxon signed-rank tests for paired comparisons revealed that speech perception scores were significantly lower (better) after 16 weeks than before for all outcome measures: for SiN $z = -2.88$, $p = 0.004$, for DiN, $z = -4.7$, $p < 0.0001$, for vowel identification in quiet $z = -4.5$, $p < 0.0001$ and consonant identification in quiet $z = -5.0$, $p < 0.0001$. Median and IQR scores are presented in **Table 1**.

For the sake of comparison, **Table 1** also lists the median SiN thresholds for speech stimuli presented to only the CI (CI-SF) and in the daily settings condition (with CI and HA if applicable). These three different SiN outcomes do not differ statistically from each other at the beginning of the trial. However, at the end of the trial, the non-parametric Friedman test of differences among repeated measures rendered a Chi-square value of 12.1, which was significant ($p < 0.005$). A Wilcoxon signed-rank tests subsequently showed that the SiN "daily settings" was significantly better than the SiN in streaming mode ($x = -2.574$, $p = 0.010$) and SiN CI-SF ($X = -3.968$, $p < 0.0001$).

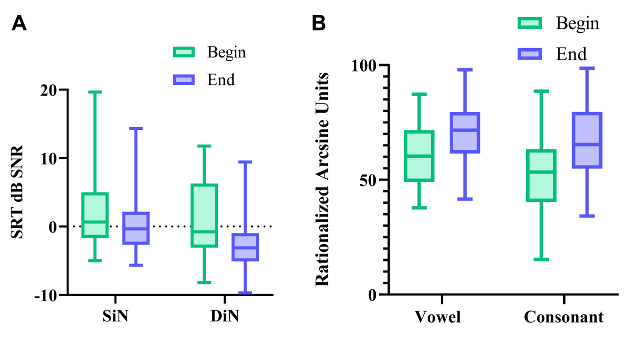


**FIGURE 5 | (A)** median SRT in noise (and IQR) for SiN, DiN, and **(B)** median percentage vowel and consonant identification in quiet (and IQR) at the beginning and end of the 16-week trial.

# DISCUSSION

## Sentence and Digits Understanding in Noise

Understanding speech in noise is the most common complaint of persons with hearing impairment, and several indices can be used to document listening difficulties and guide hearing rehabilitation. The present study reports three different indices for speech perception in the same 40 CI users, of which two are administered at home. Where possible, we present individual results instead of group mean average to better understand individual differences in speech recognition outcomes (which, in turn, enables personalized rehabilitation).

SiN performance of contemporary CI users is excellent in the current study, even when the sentences are only streamed to the implanted side. Candidacy criteria for cochlear implantation have changed considerably over the past years, and several CI users have residual hearing (Snel-Bongers et al., 2018). Nevertheless, variability in performance is large, and to understand the source of variability, it is important to look at individual performance and do this from different

**TABLE 1 |** Median SRT in noise and IQR, min and max for SiN in streaming mode, daily settings, and sound-field, only 1 CI (CI-SF) for the first and last session separately ($n = 40$).

| | Median | IQR | Min | Max |
|---|---|---|---|---|
| **First week** | | | | |
| SiN streaming (dB SNR) | 0.7 | 6.7 | −5.0 | 19.7 |
| SiN daily settings (dB SNR) | 0.3 | 3.8 | −3.7 | 20.0 |
| SiN CI-SF (dB SNR) | 1.5 | 4.9 | −3.7 | 20.0 |
| DiN (dB SNR) | 0.02 | 9.9 | −8.8 | 37.0 |
| Vowel% | 61 | 23.5 | 37.0 | 86.0 |
| Consonant% | 55 | 24.6 | 15.8 | 88.7 |
| **Last week** | | | | |
| SiN streaming (dB SNR) | −0.3 | 4.5 | −5.7 | 14.3 |
| SiN daily settings (dB SNR) | −1.2 | 4.0 | −7.0 | 9.3 |
| SiN CI-SF (dB SNR) | 0.7 | 5.8 | −5.7 | 20.0 |
| DiN (dB SNR) | −3.2 | 4.4 | −10.0 | 10.9 |
| Vowel% | 72.5 | 17.5 | 41.0 | 93.0 |
| Consonant% | 66.3 | 24.6 | 33.3 | 93.3 |

*Median SRT in noise and IQR, min and max for DTT streaming, the first two sessions (n = 80) or the last two sessions (n = 80) separately. Median vowel percentage (and IQR) of the first vowel identification task and the first consonant identification task (week 1 and 2) as well as for the last two weeks of the trials (week 15 and 16).*

perspectives. Word and sentence identification remain important measures to evaluate an intervention (e.g., Zhang and Coelho, 2018; Kelsall et al., 2021). While open-set word and sentence understanding lack full external generalizability to speech perception in daily life, they are most closely related to capturing some of the real-world listening difficulties. These measures involve phonological, lexical, grammatical skills, and semantic/contextual knowledge (Heald and Nusbaum, 2014), especially when administered using an open-set response format (Clopper et al., 2006). In an open-set task, listeners compare the stimulus to all possible candidate words in lexical memory.

In contrast, in closed set tests, the listeners need to make only a limited number of comparisons among the response alternatives. An advantage of SiN above word identification is the steeper slope of the performance intensity function of the former. The slope measures how rapidly performance changes with a change in level or signal to noise ratio (Leek, 2001).

However, SiN cannot be assessed too often (due to learning and limited test materials), while DiN can be used repeatedly and without a clinician. The high correlation between DiN and SiN is in line with the results of Smits et al. (2013) for persons with normal hearing and Kaandorp et al. (2015), Kaandorp et al., 2016, 2017 Cullington and Aidi (2017), and Zhang et al. (2019) for persons with cochlear implants, thereby indicating that the two measures share some common mechanisms. The difference between the two is in the order of 4 dB SNR in the current study. DiN may even be more sensitive than SiN to capture changes in auditory performance, which can be done for each ear separately. The large dataset of the current study did reveal individual differences in performance which do not necessarily change with time. For some participants, especially those struggling most with SiN, performance varied substantially. Here, measurement error of subsequent trials can be used as a guideline for potential changes in performance.

## Phoneme Identification in Quiet

Phoneme identification in quiet sheds additional light on variability in speech perception. While sentence and word recognition in quiet often yield ceiling scores, phoneme scores provide specific information on the perception of speech cues in the absence of context cues. Phoneme perception has a long history in research (Miller and Nicely, 1955) but is not often used as a standard metric in the clinic. At least two arguments plead in favor of incorporating phoneme identification in clinical care. First, it is essential to know how vowel and consonant identification in quiet relate to performance on tests in noise. Our study shows that vowel and consonant recognition in quiet contribute (uniquely) to SiN (and to DiN) and yields additional information on the audibility and discriminability of speech cues. At the clinic, phoneme recognition is often assessed via meaningful words (phoneme score of a word recognition test). However, nonsense syllables are preferred over meaningful words, because context can affect the recognition of phoneme scores in the latter (Donaldson and Kreft, 2006). With a nonsense syllable task, each phoneme can be presented an equal number of times. The task takes only a few minutes and can easily be done remotely. In the future, phoneme perception in noise will also be considered.

Second, phoneme identification can also be used as a diagnostic tool. While percentage correct is a summary score of phoneme perception, the information transmission analyses reveal which spectral and temporal cues are most challenging for the recipient. This information can help optimize the mapping and provide targeted rehabilitation. For instance, a low score for duration discrimination in vowels could guide the clinician to provide tasks to improve discriminability between short and long vowels. A low score on “frication” or “voicing” could guide the clinician to optimize the mapping of high- and low-frequency cues, respectively. The value of this metric was recognized several years ago (van Wieringen and Wouters, 2000) but seemed cumbersome to implement in clinical care. With novel, cost-effective technologies, the benefit of assessing phoneme perception at regular intervals can be reconsidered. Note that sufficient data scores are required to draw conclusions from the information transmission analyses. The maximum-likelihood estimate for information transfer is biased to overestimate its true value when the number of stimulus presentations is small (Sagi and Svirsky, 2008).

## Procedural Learning

Learning either the content or the procedure of a test could improve performance when presented repeatedly. During the 16-weeks, the participants also practiced training modules (Magits et al., under revision). Comparison of SiN pre- versus post-training showed a significant improvement in speech understanding in noise for both the personalized LUISTER and the non-personalized listening training programs. Since the same sentences were never presented twice to the same participant, the observed differences are more likely to result from practicing than repeated testing. However, it is difficult to determine whether the observed improvements for DiN and phoneme identification result from the content of the listening training (perceptual learning) or procedural learning (repeated listening to a task). All perceptual experiments involve some procedural learning, such as getting acquainted with a voice, the characteristics of the speech material, etc., (Nygaard and Pisoni, 1998; Yund and Woods, 2010). A procedural learning effect is larger for a closed-set than for open-set response format, but Smits et al. (2013) report that procedural learning with DiN is accomplished after 1 trial with normal-hearing persons. Nevertheless, de Graaff et al. (2019) report that DiN data of CI users reveal improvements in speech recognition over time, without a clear relation to fitting appointments with an audiologist. These improvements could result from procedural learning or improved perception of speech perception in general.

## Remote Care

Rehabilitation following cochlear implantation is demanding and requires several visits to the clinic to fine-tune the device. With the growing number of clients, improved technology, and public health concerns surrounding the COVID-19 pandemic, remote testing has sparked a lot of interest to complement care at the clinic. The shared responsibility between professional and client may also empower clients to take action if needed. Home-based testing has the potential to change and improve the hearing care pathway. It would not only lead to a reduction in the required number of visits and thus reduction in cost— and time savings for both clinics and patients, it would even improve the quality and richness of data obtained during audiological rehabilitation. The importance of speech in noise testing cannot be overestimated, but note that it entails more than the perception of the auditory signal in noise which can be captured with a DiN task. When the acoustical signal is difficult to perceive, as in noisy conditions, speech understanding places more demands on linguistic knowledge and executive functioning (Mattys et al., 2012; Moberly et al., 2019; Zhan et al., 2020). Remote monitoring of speech-in-noise performance possibly makes room to assess neurocognitive abilities that differentially explain speech in noise performance, which may lead to a personalized holistic management of hearing impairment.

For remote testing to be successful, the obtained data should be clinically valid and accurate, and clients should feel confident handling the device. In our study, all participants felt comfortable doing tests remotely because the professional had provided sufficient information prior to the trial and was online available to address any concerns or technical problems. Data collection with wireless streaming was reliable as repeated testing yielded similar results in the same CI user.

During the COVID-19 pandemic, face-to-face care was brought to a halt, and interest in tele-audiology surged out of necessity. While a recent survey reports that audiologists are generally positive about teleaudiology, infrastructure and training should not be underestimated, and hybrid care remains necessary (Saunders and Roughley, 2021). Also, more research is needed to examine reimbursement and cost-effectiveness of remote services (Bush et al., 2016). Such factors may represent a barrier to the practical delivery of telemedicine services, and these topics represent areas for further research. Technical advances in connectivity now allow for wireless streaming capabilities for current CI systems. Wireless streaming provides good quality audio and is less susceptible to noise or signal processing introduced by the connection cable. Only one calibration is needed for a given digital communication set. In the current study, only the implanted side was assessed at home, but stereo streaming is possible. It remains important to evaluate the whole hearing pathway in the sound field too.

## CONCLUSION

Speech perception assessment in the same forty CI users showed that DiN assessed at home is a powerful alternative to SiN in the clinic to monitor performance at regular intervals and detect changes in auditory performance. Phoneme identification in quiet also explains a significant part of speech perception in noise and provides additional information on the detectability and discriminability of speech cues. DiN and phoneme identification in quiet can be assessed reliably at home in a limited amount of time. Home-based testing with wireless streaming can be complementary to testing in the clinic. Embracing these technologies could reduce the cost, serve clients who would otherwise not have access to clinical services, and open the door to holistic hearing care.

## 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.

## ETHICS STATEMENT

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

## AUTHOR CONTRIBUTIONS

AW, SM, TF, and JW contributed to the conceptualization of this research project. AW wrote the manuscript. JW, TF, and SM provided critical revision and feedback. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by a TBM-FWO grant from the Research Foundation-Flanders (grant number T002216N). The authors acknowledge the financial support from the Legacy Ghislaine Heylen (Tyberghein).

## ACKNOWLEDGMENTS

We thank all our participants for their time and effort and Linus Demeyere for technical developments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.773427/full#supplementary-material

## REFERENCES


Biever, A., Amurao, C., and Mears, M. (2021). Considerations for a revised adult cochlear implant candidacy evaluation protocol. *Otol. Neurotol.* 42, 159–164. doi: 10.1097/MAO.0000000000002966

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.* 107, 1065–1066. doi: 10.1121/1.428288

Boothroyd, A. (1994). Speech perception. *Curr. Opin. Otolaryngol. Head Neck Surg.* 2, 186–194.

Brooke, J. (1996). "SUS: a 'quick and dirty' usability scale," in *Usability Evaluation in Industry*, eds P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland (London: Taylor & Francis), 189–194.

Bush, M. L., Thompson, R., Irungu, C., and Ayugi, J. (2016). The role of telemedicine in auditory rehabilitation: a systematic review. *Otol. Neurotol.* 37, 1466–1474.

Clopper, C. G., Pisoni, D. B., and Tierney, A. T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *J. Am. Acad. Audiol.* 17, 331–349. doi: 10.3766/jaaa.17.5.4

Cullington, H., Kitterick, P., Weal, M., and Margol-Gromada, M. (2018). Feasibility of personalized remote long-term follow-up of people with cochlear implants: a randomized controlled trial, Randomized Controlled Trial. *BMJ Open* 8:e019640.

Cullington, H. E., and Aidi, T. (2017). Is the digit triplet test an effective and acceptable way to assess speech recognition in adults using cochlear implants in a home environment? *Cochlear Implants Int.* 18, 97–105. doi: 10.1080/14670100.2016.1273435

de Graaff, F., Huysmans, E., Merkus, P., Theo Goverts, S., and Smits, C. (2018). Assessment of speech recognition abilities in quiet and in noise: a comparison between self-administered home testing and testing in the clinic for adult cochlear implant users. *Int. J. Audiol.* 57, 872–880. doi: 10.1080/14992027.2018.1506168

de Graaff, F., Huysmans, E., Philips, B., Merkus, P., Goverts, S. T., Kramer, S. E., et al. (2019). Our experience with home self-assessment of speech recognition in the care pathway of 10 newly implanted adult cochlear implant users. *Clin. Otolaryngol.* 44, 446–451. doi: 10.1111/coa.13307

de Graaff, F., Huysmans, E., Qazi, O. U., Vanpoucke, F. J., Merkus, P., Goverts, S. T., et al. (2016). The development of remote speech recognition tests for adult cochlear implant users: the effect of presentation mode of the noise and a reliable method to deliver sound in home environments. *Audiol. Neurootol.* 21(Suppl. 1), 48–54. doi: 10.1159/000448355

de Graaff, F., Lissenberg-Witte, B. I., Kaandorp, M. W., Merkus, P., Goverts, S. T., Kramer, S. E., et al. (2020). Relationship between speech recognition in quiet and noise and fitting parameters, impedances and ECAP thresholds in adult cochlear implant users. *Ear Hear.* 41, 935–947. doi: 10.1097/AUD.0000000000000814

Denys, S., Hofmann, M., van Wieringen, A., and Wouters, J. (2019). Improving the efficiency of the digit triplet test using digit scoring with variable adaptive step sizes. *Int. J. Audiol.* 58, 670–677. doi: 10.1080/14992027.2019.1622042

Donaldson, G. S., and Kreft, H. A. (2006). Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users. *Ear Hear.* 27, 658–677. doi: 10.1097/01.aud.0000240543.31567.54

Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., and Parkin, J. (1990). Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant. *J. Acoust. Soc. Am.* 88, 2074–2079.

Dubno, J. R., and Dirks, D. D. (1982). Evaluation of hearing-impaired listeners using a Nonsense-Syllable Test. I. Test reliability. *J. Speech Hear. Res.* 25, 135–141. doi: 10.1044/jshr.2501.135

Eikelboom, R. H., Jayakody, D. M., Swanepoel, D. W., Chang, S., and Atlas, M. D. (2014). Validation of remote mapping of cochlear implants. *J. Telemed. Telecare* 20, 171–177. doi: 10.1177/1357633X14529234

Francart, T., Hofmann, M., Vanthornhout, J., Van Deun, L., van Wieringen, A., and Wouters, J. (2017). APEX/SPIN: a free test platform to measure speech intelligibility. *Int. J. Audiol.* 56, 137–143. doi: 10.1080/14992027.2016.1247215

Francart, T., Moonen, M., and Wouters, J. (2009). Automatic testing of speech recognition. *Int. J. Audiol.* 48, 80–90.

Gifford, R. H., Driscoll, C. L., Davis, T. J., Fiebig, P., Micco, A., and Dorman, M. F. (2015). A within-subject comparison of bimodal hearing, bilateral cochlear implantation, and bilateral cochlear implantation with bilateral hearing preservation: high-performing patients. *Otol. Neurotol.* 36, 1331–1337. doi: 10.1097/MAO.0000000000000804

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (2008). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol. Neurootol.* 13, 193–205. doi: 10.1159/000113510

Goehring, J. L., Hughes, M. L., Baudhuin, J. L., Valente, D. L., McCreery, R. W., Diaz, G. R., et al. (2012). The effect of technology and testing environment on speech perception using telehealth with cochlear implant recipients. *J. Speech Lang. Hear. Res.* 55, 1373–1386. doi: 10.1044/1092-4388(2012/11-0358)

Gordon-Salant, S. (1985). Phoneme feature perception in noise by normal-hearing and hearing-impaired subjects. *J. Speech Hear. Res.* 28, 87–95. doi: 10.1044/jshr.2801.87

Heald, S. L., and Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Front. Syst. Neurosci.* 8:35. doi: 10.3389/fnsys.2014.00035

Hughes, M. L., Goehring, J. L., Baudhuin, J. L., Diaz, G. R., Sanford, T., Harpster, R., et al. (2012). Use of telehealth for research and clinical measures in cochlear implant recipients: a validation study. *J. Speech Lang Hear. Res.* 55, 1112–1127. doi: 10.1044/1092-4388(2011/11-0237)

Hughes, M. L., Sevier, J. D., and Choi, S. (2018). Techniques for remotely programming children with cochlear implants using pediatric audiological methods via telepractice. *Am. J. Audiol.* 27, 385–390. doi: 10.1044/2018_AJA-IMIA3-18-0002

*IBM SPSS Statistics for Windows, version 27*. (2020). Armonk, NY: IBM Corp.

James, C. J., Karoui, C., Laborde, M. L., Lepage, B., Molinier, C. E., Tartayre, M., et al. (2019). Early sentence recognition in adult cochlear implant users. *Ear Hear.* 40, 905–917. doi: 10.1097/aud.0000000000000670

Jansen, S., Luts, H., Dejonckere, P., van Wieringen, A., and Wouters, J. (2013). Efficient hearing screening in noise-exposed listeners using the digit triplet test. *Ear Hear.* 34, 773–778. doi: 10.1097/AUD.0b013e318297920b

Jansen, S., Luts, H., Dejonckere, P., van Wieringen, A., and Wouters, J. (2014). Exploring the sensitivity of speech-in-noise tests for noise-induced hearing loss. *Int. J. Audiol.* 53, 199–205. doi: 10.3109/14992027.2013.849361

Kaandorp, M. W., De Groot, A. M., Festen, J. M., Smits, C., and Goverts, S. T. (2016). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in Noise. *Int. J. Audiol.* 55, 157–167. doi: 10.3109/14992027.2015.1104735

Kaandorp, M. W., Smits, C., Merkus, P., Festen, J. M., and Goverts, S. T. (2017). Lexical-access ability and cognitive predictors of speech recognition in noise in adult cochlear implant users. *Trends Hear.* 21:2331216517743887. doi: 10.1177/2331216517743887

Kaandorp, M. W., Smits, C., Merkus, P., Goverts, S. T., and Festen, J. M. (2015). We are assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *Int. J. Audiol.* 54, 48–57. doi: 10.3109/14992027.2014.945623

Kelsall, D., Lupo, J., and Biever, A. (2021). Longitudinal outcomes of cochlear implantation and bimodal hearing in a large group of adults: a multicenter clinical study. *Am. J. Otolaryngol.* 42:102773. doi: 10.1016/j.amjoto.2020.102773

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., et al. (2015). The multilingual matrix test: principles, applications, and comparison across languages: a review. *Int. J. Audiol.* 54(Suppl. 2), 3–16. doi: 10.3109/14992027.2015.1020971

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Percept Psychophys.* 63, 1279–1292.

Louw, C., Swanepoel, W., Eikelboom, R. H., and Myburgh, H. C. (2017). Smartphone-based hearing screening at primary health care clinics. *Ear Hear.* 38, e93–e100. doi: 10.1097/AUD.0000000000000378

Lundberg, E. M. H., Strong, D., Anderson, M., Kaizer, A. M., and Gubbels, S. (2021). Do patients benefit from a cochlear implant when they qualify only in the presence of background noise? *Otol. Neurotol.* 42, 251–259. doi: 10.1097/MAO.0000000000002878

Magits, S., Boon, E., De Meyere, L., Dierckx, A., Vermaete, E., Francart, T., et al. (under revision). Computerized listening training improves speech understanding in noise in many CI users. *Ear Hear.*

Mattys, S. V., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006

Meister, H., Rählmann, S., Walger, M., Margolf-Hackl, S., and Kießling, J. (2015). 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* 10, 435–443. doi: 10.2147/CIA.S77096

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338–352.

Moberly, A. C., Bates, C., Harris, M. S., and Pisoni, D. B. (2016). The enigma of poor performance by adults with cochlear implants. *Otol. Neurotol.* 37, 1522–1528. doi: 10.1097/MAO.0000000000001211

Moberly, A. C., Doerfer, K., and Harris, M. S. (2019). Does cochlear implantation improve cognitive function? *Laryngoscope* 129, 2208–2209. doi: 10.1002/lary.28140

Muñoz, K., Nagaraj, N. K., and Nichols, N. (2021). Applied tele-audiology research in clinical practice during the past decade: a scoping review. *Int. J. Audiol.* 60(Suppl. 1), S4–S12. doi: 10.1080/14992027.2020.1817994

Munson, B., Donaldson, G. S., Allen, S. L., Collison, E. A., and Nelson, D. A. (2003). Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. *J. Acoust. Soc. Am.* 113, 925–935. doi: 10.1121/1.1536630

Nie, K., Barco, A., and Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear Hear.* 27, 208–217.

Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376. doi: 10.3758/bf03206860

Panagioti, M., Richardson, G., Small, N., Murray, E., Rogers, A., Kennedy, A., et al. (2014). Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. *BMC Health Serv. Res.* 14:356. doi: 10.1186/1472-6963-14-356

Perkins, E., Dietrich, M. S., Manzoor, N., O'Malley, M., Bennett, M., Rivas, A., et al. (2021). Further evidence for the expansion of adult cochlear implant candidacy criteria. *Otol. Neurotol.* 42, 815–823. doi: 10.1097/mao.0000000000003068

Plomp, R., and Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology* 18, 43–52. doi: 10.3109/00206097909072618

Rählmann, S., Meis, M., Schulte, M., Kießling, J., Walger, M., and Meister, H. (2018). Assessment of hearing aid algorithms using a master hearing aid: the influence of hearing aid experience on the relationship between speech recognition and cognitive capacity. *Int. J. Audiol.* 57(Suppl. 3), S105–S111. doi: 10.1080/14992027.2017.1319079

Ramos, A., Rodriguez, C., Martinez-Beneyto, P., Perez, D., Gault, A., Falcon, J. C., et al. (2009). Use of telemedicine in the remote programming of cochlear implants. *Acta Otolaryngol.* 129, 533–540. doi: 10.1080/00016480802294369

Ricketts, T. A., Picou, E. M., Shehorn, J., and Dittberner, A. B. (2019). Degree of hearing loss affects bilateral hearing aid benefits in ecologically relevant laboratory conditions. *J. Speech Lang. Hear. Res.* 62, 3834–3850. doi: 10.1044/2019_JSLHR-H-19-0013

Rødvik, A. K., von Koss Torkildsen, J., Wie, O. B., Storaker, M. A., and Silvola, J. T. (2018). Consonant and vowel identification in cochlear implant users measured by nonsense words: a systematic review and meta-analysis. *J. Speech Lang Hear. Res.* 61, 1023–1050. doi: 10.1044/2018_JSLHR-H-16-0463

Sagi, E., and Svirsky, M. A. (2008). Information transfer analysis: a first look at estimation bias. *J. Acoust. Soc. Am.* 123, 2848–2857. doi: 10.1121/1.2897914

Saunders, G. H., and Roughley, A. (2021). Audiology in the time of COVID-19: practices and opinions of audiologists in the UK. *Int. J. Audiol.* 60, 255–262. doi: 10.1080/14992027.2020.1814432

Sevier, J. D., Choi, S., and Hughes, M. L. (2019). Use of direct-connect for remote speech-perception testing in cochlear implants. *Ear Hear.* 40, 1162–1173. doi: 10.1097/AUD.0000000000000693

Shannon, R. V., Cruz, R. J., and Galvin, J. J. 3rd. (2011). Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise. *Audiol. Neurootol.* 16, 113–123. doi: 10.1159/000315115

Skidmore, J. A., Vasil, K. J., He, S., and Moberly, A. C. (2020). Explaining speech recognition and quality of life outcomes in adult cochlear implant users: complementary contributions of demographic, sensory, and cognitive factors. *Otol. Neurotol.* 41, e795–e803. doi: 10.1097/MAO.0000000000002682

Smits, C., and Houtgast, T. (2005). Results from the Dutch speech-in-noise screening test by telephone. *Ear Hear.* 26, 89–95. doi: 10.1097/00003446-200502000-00008

Smits, C., Theo Goverts, S., and Festen, J. M. (2013). The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J. Acoust. Soc. Am.* 133, 1693–1706. doi: 10.1121/1.4789933

Snel-Bongers, J., Netten, A. P., Boermans, P. B. M., Rotteveel, L. J. C., Briaire, J. J., and Frijns, J. H. M. (2018). Evidence-based inclusion criteria for cochlear implantation in patients with postlingual deafness. *Ear Hear.* 39, 1008–1014. doi: 10.1097/AUD.0000000000000568

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J. Speech Hear. Res.* 28, 455–462. doi: 10.1044/jshr.2803.455

Swanepoel, D. W., and Hall, J. W. 3rd. (2010). A systematic review of telehealth applications in audiology. *Telemed. J. E Health* 16, 181–200. doi: 10.1089/tmj.2009.0111

Tamati, T. N., Ray, C., Vasil, K. J., Pisoni, D. B., and Moberly, A. C. (2020). High- and low-performing adult cochlear implant users on high-variability sentence recognition: differences in auditory spectral resolution and neurocognitive functioning. *J. Am. Acad. Audiol.* 31, 324–335. doi: 10.3766/jaaa.18106

Tyler, R. S., and Moore, B. C. (1992). Consonant recognition by some of the better cochlear-implant patients. *J. Acoust. Soc. Am.* 92, 3068–3077. doi: 10.1121/1.404203

Välimaa, T. T., Määttä, T. K., Löppönen, H. J., and Sorri, M. J. (2002a). Phoneme recognition and confusions with multichannel cochlear implants: consonants. *J. Speech Lang Hear. Res.* 45, 1055–1069. doi: 10.1044/1092-4388(2002/085)

Välimaa, T. T., Määttä, T. K., Löppönen, H. J., and Sorri, M. J. (2002b). Phoneme recognition and confusions with multichannel cochlear implants: vowels. *J. Speech Lang Hear. Res.* 45, 1039–1054. doi: 10.1044/1092-4388(2002/084)

Van den Borre, E., Denys, S., van Wieringen, A., and Wouters, J. (2021). The digit triplet: a scoping review. *Int. J. Audiol. [Online ahead of print]* 1–18. doi: 10.1080/14992027.2021.1902579

van der Straaten, T. F. K., Briaire, J. J., Vickers, D., Boermans, P. P. B. M., and Frijns, J. H. M. (2020). Selection criteria for cochlear implantation in the United Kingdom and flanders: toward a less restrictive standard. *Ear Hear.* 42, 68–75. doi: 10.1097/AUD.0000000000000901

van Wieringen, A., and Wouters, J. (1999). Natural vowel and consonant recognition by Laura cochlear implantees. *Ear Hear.* 20, 89–103. doi: 10.1097/00003446-199904000-00001

van Wieringen, A., and Wouters, J. (2000). "Assessing progress in speech perception with the LAURA cochlear implant device," in *Cochlear Implants*, eds S. B. Waltzman and N. L. Cohen (Stuttgart: Thieme), 355–356.

van Wieringen, A., and Wouters, J. (2008). LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *Int. J. Audiol.* 47, 348–355. doi: 10.1080/14992020801895144

Venail, F., Picot, M. C., Marin, G., Falinower, S., Samson, J., Cizeron, G., et al. (2021). Speech perception, real-ear measurements and self-perceived hearing impairment after remote and face-to-face programming of hearing aids: a randomized single-blind agreement study. *J. Telemed. Telecare* 27, 409–423. doi: 10.1177/1357633X19883543

Völter, C., Oberländer, K., Carroll, R., Dazert, S., Lentz, B., Martin, R., et al. (2020). Non-auditory functions in low-performing adult cochlear implant users. *Otol. Neurotol.* 42, e543–e551.

Wesarg, T., Wasowski, A., Skarzynski, H., Ramos, A., Falcon Gonzalez, J. C., Kyriafinis, G., et al. (2010). Remote fitting in Nucleus cochlear implant recipients. *Acta Otolaryngol.* 130, 1379–1388. doi: 10.3109/00016489.2010.492480

Yund, E. W., and Woods, D. L. (2010). Content and procedural learning in repeated sentence tests of speech perception. *Ear Hear.* 31, 769–778.

Zeitler, D. M., Kessler, M. A., Terushkin, V., Roland, T. J. Jr., Svirsky, M. A., Lalwani, A. K., et al. (2008). Speech perception benefits of sequential bilateral cochlear implantation in children and adults: a retrospective analysis. *Otol. Neurotol.* 29, 314–325. doi: 10.1097/mao.0b013e3181662cb5

Zhan, K. Y., Lewis, J. H., Vasil, K. J., Tamati, T. N., Harris, M. S., Pisoni, D. B., et al. (2020). Cognitive functions in adults receiving cochlear implants: predictors of speech recognition and changes after implantation. *Otol. Neurotol.* 41, e322–e329. doi: 10.1097/MAO.0000000000002544

Zhang, E., and Coelho, D. H. (2018). Beyond sentence recognition in quiet for older adults: implications for cochlear implant candidacy. *Otol. Neurotol.* 39, 979–986. doi: 10.1097/MAO.0000000000001885

Zhang, F., Underwood, G., McGuire, K., Liang, C., Moore, D. R., and Fu, Q. J. (2019). Frequency change detection and speech perception in cochlear implant users. *Hear. Res.* 379, 12–20. doi: 10.1016/j.heares.2019.04.007

Zhao, E. E., Dornhoffer, J. R., Loftus, C., Nguyen, S. A., Meyer, T. A., Dubno, J. R., et al. (2020). Association of patient-related factors with adult cochlear implant speech recognition outcomes: a meta-analysis. *JAMA Otolaryngol. Head Neck Surg.* 146, 613–620. doi: 10.1001/jamaoto.2020.0662

**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.

frontiers in Neuroscience

SYSTEMATIC REVIEW
published: 10 January 2022
doi: 10.3389/fnins.2021.788899

# Perception of Environmental Sounds in Cochlear Implant Users: A Systematic Review

*Valeriy Shafiro*[1]*, *Nathan Luzum*[2], *Aaron C. Moberly*[3] and *Michael S. Harris*[4]

[1] Department of Communication Disorders and Sciences, Rush University Medical Center, Chicago, IL, United States, [2] Medical College of Wisconsin, Milwaukee, WI, United States, [3] Department of Otolaryngology–Head and Neck Surgery, The Ohio State University Wexner Medical Center, Columbus, OH, United States, [4] Department of Otolaryngology and Communication Sciences, Medical College of Wisconsin, Milwaukee, WI, United States

**Objectives:** Improved perception of environmental sounds (PES) is one of the primary benefits of cochlear implantation (CI). However, past research contains mixed findings on PES ability in contemporary CI users, which at times contrast with anecdotal clinical reports. The present review examined extant PES research to provide an evidence basis for clinical counseling, identify knowledge gaps, and suggest directions for future work in this area of CI outcome assessment.

**Methods:** Six electronic databases were searched using medical subject headings (MeSH) and keywords broadly identified to reference CI and environmental sounds. Records published between 2000 and 2021 were screened by two independent reviewers in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement to identify studies that met the inclusion criteria. Data were subsequently extracted and evaluated according to synthesis without-meta-analysis (SWiM) guidelines.

**Results:** Nineteen studies met the inclusion criteria. Most examined PES in post-lingually implanted adults, with one study focused on pre/perilingual adults. Environmental sound identification (ESI) in quiet using open- or closed-set response format was most commonly used in PES assessment, included in all selected studies. ESI accuracy in CI children (3 studies) and adults (16 studies), was highly variable but generally mediocre (means range: 31–87%). Only two studies evaluated ESI performance prospectively before and after CI, while most studies were cross-sectional. Overall, CI performance was consistently lower than that of normal-hearing peers. No significant differences in identification accuracy were reported between CI candidates and CI users. Environmental sound identification correlated in CI users with measures of speech perception, music and spectro-temporal processing.

**Conclusion:** The findings of this systematic review indicate considerable limitations in the current knowledge of PES in contemporary CI users, especially in pre/perilingual late-implanted adults and children. Although no overall improvement in PES following implantation was found, large individual variability and existing methodological limitations

OPEN ACCESS

**Edited by:**
*Astrid van Wieringen, KU Leuven, Belgium*

**Reviewed by:**
*Justin Aronoff, University of Illinois at Urbana-Champaign, United States*
*Paul van de Heyning, University of Antwerp, Belgium*

***Correspondence:**
*Valeriy Shafiro*
*valeriy_shafiro@rush.edu*

**Specialty section:**
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

**Received:** *03 October 2021*
**Accepted:** *07 December 2021*
**Published:** *10 January 2022*

**Citation:**
*Shafiro V, Luzum N, Moberly AC and Harris MS (2022) Perception of Environmental Sounds in Cochlear Implant Users: A Systematic Review. Front. Neurosci. 15:788899. doi: 10.3389/fnins.2021.788899*

in PES assessment may potentially obscure potential CI benefits for PES. Further research in this ecologically relevant area of assessment is needed to establish a stronger evidence basis, identify CI users with significant deficits, and improve CI users' safety and satisfaction through targeted PES rehabilitation.

**Keywords: cochlear implant, systematic review, perception of environmental sounds, hearing loss, auditory assessment**

## INTRODUCTION

Improved perception of environmental sounds (PES) is considered a major benefit of cochlear implantation (Duchesne et al., 2017; McRackan et al., 2017, 2019). Environmental sounds can be defined as non-speech, non-musical sounds in the listener's surroundings that convey information about places, objects, and actions. These sounds can help listeners navigate their surroundings, warn of potential dangers, and provide a sense of aesthetic satisfaction. From avoiding a road collision, answering a doorbell, to enjoying birdsongs or waves crashing on the shore, environmental sounds provide a sense of connection to the environments and enhance awareness of it (Ramsdell, 1978). Outside of the early years of cochlear implant (CI) development and clinical use, however, there has been relatively little research attention to PES in CI users (Tyler and Kelsay, 1990), even as implantation criteria have expanded over time. Relevant findings from early studies with profoundly deaf individuals using first generation CIs with a single or several electrodes may not accurately represent PES performance of more recently implanted individuals. Although qualified CI candidates who are considering implantation are often counseled about increased access to environmental sounds, without a clear evidence basis PES in contemporary CI users remains largely a presumed benefit. To address the knowledge gap in this area of CI outcomes assessment, the present review provides a systematic evaluation of the extant published research on CI users' ability to perceive environmental sounds.

Cochlear implants are the treatment of choice for a growing number of people afflicted with sensorineural hearing loss beyond the therapeutic capabilities of acoustic amplification with hearing aids (NIDCD, 2021). Although CIs were initially approved by the Food and Drug Administration (FDA) for adults with profound hearing loss in both ears (Sladen et al., 2017), today's CI candidates may include adults and children who still retain usable, and sometimes normal hearing in at least one ear (e.g., Benchetrit et al., 2021). Patients with greater overall hearing abilities prior to implantation may expect more from their implants afterwards. In addition to improved speech perception, CI users often expect better perception of music and environmental sounds. Recognizing the importance of research in this area, the National Institutes of Health (NIH) Consensus Development Panel on CIs in Adults & Children highlighted "nonspeech benefits of implantation," such as PES, as a vital future direction for CI research more than two and a half decades ago (NIH Consensus Conference, 1995). Since that time, however, PES in CI users has remained minimally assessed, and the development of new processing strategies and most common outcome measures of auditory performance in CI users have continued to focus primarily on the speech perception and, to a lesser extent, spectro-temporal processing and music perception (McRackan et al., 2017, 2019; Shekar et al., 2021).

In daily life, PES is central to independence and safety of CI users (Bond et al., 2009; Debruyne et al., 2017; Hamel et al., 2020). Both CI candidates and CI users specifically identify PES as an important contributor to quality of life (QOL) (Tyler and Kelsay, 1990; McRackan et al., 2017, 2019). It has been proposed that PES may explain significant improvements in CI-specific QOL in patients who do not demonstrate proportional speech perception gains with CIs (Capretta and Moberly, 2016; Zaidman-Zait et al., 2017; Moberly et al., 2018; Vasil et al., 2020). Distinct from speech and musical sounds, environmental sounds comprise acoustic byproducts of mechanical interactions of sound-producing objects, such as sounds of machinery or nature, or they can be learned, arbitrary associations between a specific sound and its meaning, such as warning signals and alarms (Shafiro et al., 2020). Outside of the laboratory, environmental sounds tend to occur in the presence of other sounds populating a given auditory scene, and their perception can be affected by both energetic and informational masking (Gygi and Shafiro, 2013; Shafiro et al., 2016). Nevertheless, past research indicates that healthy normal-hearing listeners can readily identify a wide variety of common environmental sounds and can infer detailed information about their sources (Carello et al., 1998; Pastore et al., 2008; Lemaitre and Heller, 2013). Much less is known, however, about PES in contemporary CI users.

This systematic review was designed and conducted to identify and examine published studies of PES in CI users in order to synthesize relevant empirical evidence and appraise existing methods of PES assessment. The review's primary objectives were to (a) ascertain the ability of CI users to perceive environmental sounds and (b) to determine whether PES improves following implantation. To our knowledge, no systematic review in this area has been previously conducted. Given the clinical importance of PES for CI users and limited research in this area of assessment, the inclusion criteria were set broadly to capture as much pertinent research as possible across patient populations, implant models, and assessment methods. To make review findings relevant to contemporary CI users, only studies that provided a quantitative assessment of PES in CI users published in the 21st century were included.

## METHODS

The goal of this systematic review was to provide a broad assessment of CI users' abilities to perceive environmental sounds. In addition, the following specific questions were addressed:

(1) Does current evidence indicate an improvement in perception of environmental sounds (PES) following CI?
(2) Does the degree of improvement in PES following CI differ between CI populations (pre-lingual and post-lingual children and adults)?
(3) What are predictors of PES improvement in CI users?
(4) What assessment methods have been used to evaluate PES in CI users?

The protocol for this systematic review was registered with PROSPERO (International Prospective Register of Systematic Reviews, University of York, https://www.crd.york.ac.uk/prospero, Protocol number CRD42021248601).

## Search Strategy

The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) Extension for Scoping Reviews (PRISMA-ScR) checklist (Peters et al., 2015; Tricco et al., 2018) was used as the reporting guide for this review (**Figure 1**). A comprehensive literature search was developed by a medical librarian and reviewed using the Peer Review of Electronic Search Strategies (PRESS) guidelines (McGowan et al., 2016). Searches were conducted in February and March, 2021 in MEDLINE (Ovid), Scopus, Web of Science, Cochrane Library, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and ComDisDome. Searches were limited to articles from 2000 to 2021. The search strategies were created using medical subject headings (MeSH) and keywords combined with database-specific advanced search techniques. MeSH terms and keywords were broadly identified to reference CIs and environmental sounds. The full search strategy is further detailed in **Supplementary Material**. A total of 2,598 results from the literature searches were saved and imported into Covidence (www.covidence.org), a web application for managing systematic reviews. After 1,247 duplicate entries were removed, the remaining 1,351 were screened by two independent reviewers to determine eligibility for this review. The first phase of screening was a title/abstract review, and potentially relevant articles were moved to the second phase of screening for the full text of the publications. The screening was conducted in Covidence. All conflicts were resolved with group consensus.

## Eligibility Criteria

Study selection was based on Population, Intervention, Control, and Outcomes (PICOS) guidelines (Tacconelli, 2009), summarized in **Table 1**. Studies were selected if they contained quantitative assessment of PES in CI users of any age, etiology, or duration of hearing loss, all language abilities (pre-lingual/perilingual or post-lingual), with any CI model or hearing modality (unilateral CI, bilateral CI, bimodal). Studies were excluded from the review if they were published prior to the year 2000, were based on single-channel CIs, or assessed PES based solely on anecdotal reports or expert opinions.

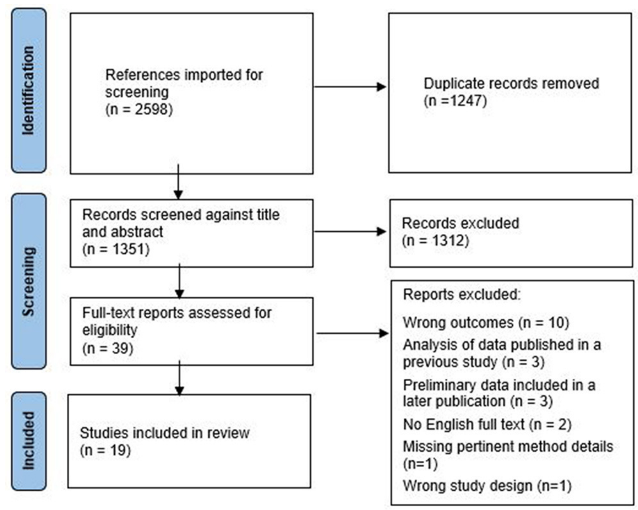


**FIGURE 1 |** PRISMA flow diagram.

**TABLE 1 |** PICOS (Patient, Intervention, Control, Outcome, and Study Design).

| | |
|---|---|
| P | Adults or children users of cochlear implants |
| I | Cochlear implantation |
| C | A control group of normal hearing or hearing impaired peers with or without hearing loss, pre-to-post-implantation comparisons |
| O | Quantitative assessments of environmental sound perception, associations with speech and auditory processing measures |
| S | Observational studies: cross-sectional, pre- and post-implantation repeated measures. |

## Data Extraction

Information from the full texts of selected studies that met the inclusion criteria was extracted. This information included study design and methods (**Table 2**), study sample size and subject characteristics (**Table 3**), type of PES assessment used, and task characteristics (**Table 4**), as well as correlations to other auditory performance outcome measures (**Table 2**).

## Quality Evaluation and Risk of Bias

The risk of bias for each study was assessed using (a) The Oxford Centre for Evidence-Based Medicine (CEBM) Levels of Evidence guidelines (Oxford Centre for Evidence-Based Medicine, 2009) and (b) the NIH study quality assessment tools (NIH, 2021). These quality evaluation tools provide additional criteria for specific study designs which complement the level of evidence metric and can be used to assign a quality rating of Good, Fair, or Poor.

## Data Analysis

The extracted data were evaluated following synthesis without meta-analysis (SWiM) guidelines (Campbell et al., 2020). SWiM guidelines are consistent with and further expand

**TABLE 2 |** Characteristics of the studies included in the systematic review.

| Author and year | LoE and quality | Study design | Subjects | | Assessment | | Correlations |
|---|---|---|---|---|---|---|---|
| | | | CI | Cntrl | PES tests | Non-PES auditory tests | CI PES and other auditory tests |
| Harris et al. (2021)** | 2b Good | Pre-post | Adult | Self | ID | AzBio; SMRT | Speech ♦, Spectro-temporal ♦ |
| Shafiro et al. (2020) | 2b Good | Cross-sectional | Adult | ONH | ID and Serial recall | 15 additional tests of speech, music and psychoacoustic spectro-temporal processing | Speech ■ ♦, Music ♦, Spectro-temporal ●■ |
| McMahon et al. (2018) | 2b Good | Cross-sectional | Adult | OHI | ID | AzBio | NR |
| Strelnikov et al. (2018) | 2b Good | Cross-sectional | Adult | – | ID and Ctgrs. | Disyllabic words; Sentences in noise | Speech ♦ |
| Chang et al. (2017) | 2b Good | Cross-sectional | Adult | YNH | ID | Vowels, Consonant | Speech ♦ |
| Zhang et al. (2016) | 2b* Fair | Cross-sectional | Adult | NH | ID | NR | NR |
| Shafiro et al. (2016) | 2b Good | Cross-sectional | Adult | YNH MON MOI | ID and Serial recall | BKB-SIN | Speech ♦ |
| Shafiro et al. (2015) | 2b Good | Cross-sectional | Adult | – | ID | CNC, SPIN-R | Speech ■, ♦ |
| Heo et al. (2013) | 2b Good | Cross-sectional | Adult | – | ID Locl. | NR | NR |
| Shafiro et al. (2011) | 2b Good | Cross-sectional | Adult | – | ID | CNC, HINT | Speech ♦ |
| Lee and Kim (2011) | 2b* Fair | Cross-sectional | Adult | HA | ID | Monosyllabic words | Speech ♦ |
| Looi and Arnephy (2010) | 2b Good | Cross-sectional Pre-post | Adult | NH Self | ID | Speech perception (specific test not described) | NR |
| Inverso and Limb (2010) | 2b Good | Cross-sectional | Adult | – | ID Ctgrs. | CNC-Words, CNC-Phonemes, HINT-Quiet, HINT-Noise | Speech ● ■♦ |
| Kaga and Akamasu (2009) | 4 Poor | Cross-sectional | Adult | CD AN | ID | NR | NR |
| Reed and Delhorne (2005) | 4 Poor | Cross-sectional | Adult | – | ID | NU-6 | NR |
| Peasgood et al. (2003) | 2b Fair | Cross-sectional | Adult (non-traditional candidates) | – | ID | Speech pattern perception, CUNY sentences | Speech ●♦ |
| Berland et al. (2019) | 2b Good | Cross-sectional | Children | NH | ID Ctgrs. | NR | NR |
| Liu et al. (2013) | 2b Good | Cross-sectional | Children | – | ID | PPVT-R vocabulary test | Speech ■ |
| Kim and Lee (2012) | 2b* Good | Cross-sectional | Children | NH, HA | ID | Word and sentence recognition (specific tests not described) | Speech ♦ |

*LoE, levels of evidence; Cntrl., controls; PES, Perception of environmental sounds; NR, not reported; ID, identification; Ctgris., Categorization; Locl., Localization; NH, normal hearing; ONH, older normal hearing; YNH, younger normal hearing; MON, middle/older aged normal hearing; MOI, middle/older aged impaired; HA, hearing aid users; CD, cortical deafness (auditory agnosia); AN, auditory neuropathy; Correlation magnitude symbols: ● = low $r < 0.3$, ■ = medium $r = 0.3 \leqslant 0.49$, large $> 0.5$ ♦; * = not published in English; ** = correlation symbols reflect synchronous results for a 12-month time point for 11 subjects, to be comparable with other studies in the table.*

PRISMA methodology (www.prisma-statement.org) to provide formal guidance for the synthesis of quantitative studies for which meta-analysis cannot be completed. This type of analysis was deemed most appropriate given the large methodological variation in previous PES studies including differences in assessment methods, study design, and populations.

# RESULTS

## Study Characteristics

The 19 selected studies were published between 2003 and 2021. Nine studies were conducted in the United States, four in Korea, two in France, and one at each of the following: the United Kingdom, New Zealand, Taiwan, and China. Sixteen

**TABLE 3 |** Characteristics of cochlear implant participants across the studies reviewed.

| Author and year | N | Age (years) mean (range) | CI experience (years) mean (range) | Language history | Modalities tested |
|---|---|---|---|---|---|
| Harris et al. (2021) | 20 | 67 (49–82) | pre-CI, 0.5 and 1 | Post-lingual | Bimodal |
| Shafiro et al. (2020) | 40 | 61 (24–84) | 6 (1–29) | Post-lingual | 15 bimodal, 17 unilateral, 8 bilateral |
| McMahon et al. (2018) | 39 | 68 (50–83) | 7 (1.5–34) | Post-lingual | 12 bilateral, 14 bimodal, 13 unilateral |
| Strelnikov et al. (2018) | 17 | 60 (46–74) | 0. | NR | NR |
| | 15 | 45 (23–67) | 0.8 | | |
| | 16 | 56 (41–71) | 5 | | |
| Chang et al. (2017) | 10 | 45 (19–65) | 3.5 (1–4.5) | Post-lingual | Unilateral |
| Zhang et al. (2016) | 9 | 31 (18–45) | 5.1 (0.5–13) | NR | NR |
| Shafiro et al. (2016) | 8 | 54 (25–68) | 3.6 (1.3–9) | Post-lingual | Unilateral |
| Shafiro et al. (2015) | 14 | 63 (51–87) | 5 (1–8) | Post-lingual | Unilateral |
| Heo et al. (2013) | 14 | 51 (35–66) | 1.2 (0.6–2.6) | Post-lingual | Bimodal, Unilateral |
| Shafiro et al. (2011) | 17 | 58 (40–80) | 3.2 (1-7) | Post-lingual | Unilateral |
| Lee and Kim (2011) | 9 | 35 (24–69) | 3 | 4 post-lingual, 5 pre-lingual | Unilateral |
| Looi and Arnephy (2010) | 10 | 58 (29–77) | 2.3 (0.8–4.8) | Post-lingual | Unilateral |
| | 4 | 55 (43–66) | pre-CI and 0.25 | Post-lingual | Unilateral |
| Inverso and Limb (2010) | 22 | 59 (39–75) | At least 1 year | Post-lingual | NR |
| Kaga and Akamasu (2009) | 17 | 50 (14–75) | NR | Post-lingual | NR |
| Reed and Delhorne (2005) | 11* | 42 (29–67) | 6.9 (1-12) | 10 post-lingual and 1 pre-lingual | Unilateral |
| Peasgood et al. (2003) | 10 | 31 (15–52) | 3.4 (0.8–6.3) | Pre-lingual | NR |
| Berland et al. (2019) | 24 | 9 (6–11) | 6.3 (0.8–7.6) | Pre-lingual and Early Implanted | Unilateral |
| Liu et al. (2013) | 21 | 5 (3–6) | 1.6 | Pre-lingual | NR |
| | 26 | 8 (6–10) | 2.9 | Pre-lingual | |
| Kim and Lee (2012) | 22 | 12 (7–15) | 5.7 | Pre-lingual | Unilateral |

*For the columns "Age" and "CI experience" the average and range are provided in years; N, number of participants; NR, not reported; *, only 7 of 11 participants completed all testing.*

studies were published in English, one in Chinese, and two in Korean. Pertinent details for the three studies not published in English were obtained through translation specifically for this review or provided in personal communications by the study authors. All studies were published in peer-reviewed journals, except (Kaga and Akamasu, 2009), which was published as a book chapter.

Sixteen of the selected studies examined PES in adult CI users and three studies examined PES in children with CIs with congenital or early onset hearing loss (Kim and Lee, 2012; Liu et al., 2013; Berland et al., 2019). Most of the adult CI studies focused on post-lingually implanted adults, with only one of the adult studies (Peasgood et al., 2003) focused exclusively on pre/perilingual CI users. Two additional studies also included pre/perilingual CI users: one had about an equal number of post- and pre/perilingual adults (Lee and Kim, 2011) and the other included only one pre-lingual participant (Reed and Delhorne, 2005).

Study quality, assessed with the NIH-NHLBI study quality assessment tool, was judged "Poor" for two studies, "Fair" for six studies, and "Good" for 11 studies. The 17 studies rated as "Good" or "Fair" were classified as 2b on the Oxford Level of Evidence scale, while the studies of "Poor" quality were classified as 4 on this scale (Oxford Centre for Evidence-Based Medicine, 2009).

## Study Designs

The majority of studies used a cross-sectional design. Two studies utilized a longitudinal pre-to-post implantation paradigm (Looi and Arnephy, 2010; Harris et al., 2021). One of the two longitudinal studies, however, had a small sample of only four participants in its pre-to-post-implantation arm (Looi and Arnephy, 2010). In addition to participants serving as their own control, at least one control group was included in 10 studies, while the remaining studies referenced prior research using the same assessment instruments or otherwise deemed their stimuli to have high or near-ceiling accuracy for healthy individuals with normal hearing. When a control group was used, control listener populations were quite variable across the studies and included, for pediatric studies, children with normal hearing or with a hearing impairment and, for adult studies, adults who were young normal hearing, older normal hearing, older hearing impaired, or had other comorbid conditions affecting auditory processing (auditory neuropathy or cortical deafness).

## CI Participants

A total of 395 CI users (302 adults and 93 children) were evaluated across the selected studies. Sample sizes of CI participants varied considerably (range 8–48) with an average of 20 participants per study. In most studies, participants were experienced CI users, having various hearing loss etiologies and

**TABLE 4 |** Environmental sound assessment tasks and results.

| Author and year | Task | Response options | Stimuli | Group | PES result |
|---|---|---|---|---|---|
| Harris et al. (2021) | Identification | 25 names | 25 sounds (1 token each) | CI (post-test–6 months) | 65% (SD = 14.3) |
| | | | | CI (post-test–12 months) | 69.1% (SD = 15.7) |
| | | | | HI-CIC (pretest) | 64% (SD = 14.1) |
| Shafiro et al. (2020) | Identification | 15 names | 24 sounds (1 token each) | CI | 74% (SD = 16.8) |
| | | | | ONH | 95% (SD = 5) |
| | Identification and Serial recall | 3, 4, or 5 names (based on the number sounds in a given sequence) | 24 sounds—sequences of 3, 4, or 5 sounds | CI | 59% (SD = 23) |
| | | | | ONH | 70% (SD = 18) |
| McMahon et al. (2018) | Identification | 25 names | 25 sounds (1 token each) | CI | 59% (SD = 14.3) |
| | | | | HI-CIC | 55% (SD = 26.4) |
| Strelnikov et al. (2018) | Identification | Open/3 categories | 16 sounds (1 token each, included music) | CI (new users) | 33% (SD = 30) |
| | | | | CI (intermediate users) | 35% (SD = 29) |
| | | | | CI (experienced users) | 30% (SD = 17) |
| | Categorization | Free sorting | | CI (new users) | 43% |
| | | | | CI (intermediate users) | 55% |
| | | | | CI (experienced users) | 60% |
| Chang et al. (2017) | Identification | 9 names | 9 sounds (1 token each) | CI | 78.9% (SD = 20.6) |
| | | | | YNH | 98.9% (SD = 3.5) |
| Zhang et al. (2016) | Identification | 16 names | 67 sounds | CI | 63.18% |
| | | | | NH | 96.16% |
| Shafiro et al. (2016) | Identification (percent correct sound name regardless of order accuracy) | 25 names | 20 sounds—sequences of 5 | CI | 69% (SD = 25) |
| | | | | YNH | 78% (SD = 4.4) |
| | | | | MON | 73% (SD = 11.8) |
| | | | | MOI | 73% (SD = 13.9) |
| | Serial Recall (percent correct sound names placed in correct order) | | | CI | 45% (SD = 20.1) |
| | | | | YNH | 65% (SD = 8.2) |
| | | | | MON | 44% (SD = 18.2) |
| | | | | MOI | 44% (SD = 20.3) |
| | Serial recall (percent entire sequences corrects) | | | CI | 14% (SD = 16.9) |
| | | | | YNH | 43% (SD = 11.1) |
| | | | | MON | 14% (SD = 13.4) |
| | | | | MOI | 14% (SD = 18.6) |
| Shafiro et al. (2015) | Identification | 60 names | 40 sounds (4 tokens each) | CI | 47% (SD = 14.9) |
| Heo et al. (2013) | Identification | Open | 40 sounds (4 tokens each) | CI (bimodal) | 36% (SD = 10.3) |
| | | | | CI (unilateral) | 29% (SD = 11.9) |
| | Localization | 8 speakers | | CI (bimodal) | 75% (SD = 7.4) |
| | | | | CI (unilateral) | 63% (SD = 5.0) |
| Shafiro et al. (2011) | Identification | 60 names | 40 sounds (4 tokens each) | CI | 45% (SD = 16.2) |
| Lee and Kim (2011) | Identification | 10 names | 40 sounds (2 tokens each) | CI | 33% (SD = 17.9) |
| | | | | HI-HA | 40% (SD = 19.2) |
| Looi and Arnephy (2010) | Identification | 45 names | 45 sounds (2 tokens each) | CI (experienced) | 59% (SD = 11.5) |
| | | | | NH | 93% (SD = 4.3) |
| | | | | CIC | 40% (SD = 14.3) |
| | | | | CI (3 month post-test) | 57% (SD = 21.4) |
| Inverso and Limb (2010) | Identification | Open set | 40 sounds (50 total tokens) | CI | 48% (SD = 13.5) |
| | Categorization | 5 names | | | 71% (SD = 13.5) |
| Kaga and Akamasu (2009) | Identification | Open set | 24 sounds | CI | 42% |
| | | | | CD | 8% |
| | | | | AN | 50% |

*(Continued)*

**TABLE 4 |** Continued

| Author and year | Task | Response options | Stimuli | Group | PES result |
|---|---|---|---|---|---|
| | | 4 images | | CI | 88% |
| | | | | CD | 46% |
| | | | | AN | 92% |
| Reed and Delhorne (2005) | Identification | 10 names | 40 sounds (3 tokens each) | CI | 79% (SD = 15.5) |
| Peasgood et al. (2003) | Identification | Open set | 20 sounds (1 token each) | CI | 41% (SD = 13.7) |
| Berland et al. (2019) | Identification | Open set | 18 sounds (1 token each) —includes musical, vocal, and environmental sounds | CI | 35% |
| Liu et al. (2013) | Identification | 4 images | 30 sounds (single token) | CI (younger group) | 61% (SD = 23.8) |
| | | | | CI (older group) | 73% (SD = 20.5) |
| Kim and Lee (2012)* | Identification | 10 images | 40 sounds (4 tokens) | CI | 31.67% |
| | | | | NH | 96.5% |
| | | | | HA | 30.7% |

*CI, cochlear implant; CIC, cochlear implant candidates; NH, normal hearing; ONH, old normal hearing; YNH, young normal hearing; MON, middle/older aged normal hearing; MOI, middle/older aged impaired; HA, hearing aid; CD, cortical deafness (auditory agnosia); AN, auditory neuropathy; *the reported PES score is the average of two similar scores obtained with 5 dB SNR using background noise recorded before the class and after the class.*

at least a year of CI experience. All were implanted with devices approved for implantation at the study site prior to participation. The average duration of CI experience in cross-sectional studies that reported this value for adults was 4.2 years and for children 4.1 years. Harris et al. (2021) examined CI users at both 6 and 12 months after implantation and three other studies did not report an average duration of CI experience (Kaga and Akamasu, 2009; Inverso and Limb, 2010; Zhang et al., 2016). Performance of CI users with <1 year of CI experience was specifically examined in three studies (Looi and Arnephy, 2010; Strelnikov et al., 2018; Harris et al., 2021). Several studies also included some participants that had 3–12 months of CI experience, although the average duration of CI experience for participants in those studies was considerably longer than 1 year (Peasgood et al., 2003; Looi and Arnephy, 2010; Heo et al., 2013; Liu et al., 2013; Zhang et al., 2016; Berland et al., 2019).

## Assessment Tasks

The most common type of assessment, included in all 19 studies, was environmental sound identification (ESI), in which CI users heard a single environmental sound, and, in two studies, also sequences of several environmental sounds in series. The participants were asked to either provide their own name for the sound they heard in an open-set response format or to select the most appropriate name for the sound from a closed-set of names. When such closed-set response formats were used, there was further variability across studies in the number of response options provided for naming the sound stimuli, ranging from 4 to 60 response options. One study (Reed and Delhorne, 2005) additionally constrained response options by including names of the settings in which sounds could be heard (e.g., "Kitchen," "Office"). The two studies that used sequences of several environmental sounds to examine CI users' ability to name the sound (Shafiro et al., 2016, 2020), also examined the ability to recall the specific order in which the sounds were presented, thus placing a greater demand on auditory working memory. In addition to sound identification, three studies also examined categorization of environmental sounds either by providing participants with specific category names (Inverso and Limb, 2010) or asking participants to group sounds into categories of their choice in a free sorting task (Strelnikov et al., 2018; Berland et al., 2019). A single study also examined localization of environmental sounds, in addition to identification (Heo et al., 2013).

## Stimuli and Procedures

In all studies, environmental sound stimuli were sourced from publicly available audio recording libraries or online databases, and sometimes included recordings made specifically for the study. The stimuli in most tests tended to be broadly sampled from different categories of meaningful environmental sounds, including sounds of nature, urban environments, machinery, household, alarms and warnings, animal and human non-speech vocalizations or bodily sounds. In some studies (e.g., Inverso and Limb, 2010; Strelnikov et al., 2018), stimuli also included sounds of musical instruments and samples of human speech for judgments of indexical properties.

The number of stimuli in a single test varied between nine (Chang et al., 2017) and 160 (Shafiro et al., 2011), with several studies using multiple sound tokens of the same type of sound (e.g., four different "dog barking" sounds). The maximum number of different types of sounds in one test was 67 (Zhang et al., 2016). In most studies, the test stimuli were presented to participants only once in a single session. In some studies, stimuli were presented more than once for different tasks, for example, first for free sorting of sounds into groups and then for identification (Strelnikov et al., 2018), or when stimuli were modified by different lowpass and highpass filters (Chang et al., 2017). In Zhang et al. (2016), participants could replay the sound up to three times, and in Berland et al. (2019), there was no limit to the number of times the participants could replay the sounds. In two other studies that used sequences of individual environmental sounds on each trial (Shafiro et al., 2016, 2020), participants were first tested on individual sounds and then heard each sound twice but in two different sound sequences.

Most participants were tested at the study sites in a sound booth or a quiet room, with a loudspeaker positioned one-meter away from the participant, presenting stimuli at either a comfortable or a set presentation level (65–70 dB SPL). In one study, a subgroup of participants was tested at home with their preferred audio settings following a calibration with multitalker babble, during which sound levels could be adjusted (Shafiro et al., 2020). Nine studies tested participants with a unilateral CI alone, one tested all participants bimodally, with a CI and a hearing aid, three studies included participants in some mix of three modalities: unilateral, bimodal, bilateral, and six studies did not specify listening modality during testing.

Only one study (Kim and Lee, 2012) examined ESI in the presence of background noise, using a fixed 5 dB signal-to-noise ratio (SNR) and two types of classroom noises recorded either before the class begins or during the break period. In all other studies, environmental sounds were always presented in quiet.

## Accuracy

Identification accuracy scores for isolated environmental sounds differed considerably across the 19 studies. For post-lingual adults, identification accuracy ranged between ∼33% correct (Strelnikov et al., 2018) and 87.5% correct (Kaga and Akamasu, 2009). For children with CIs, two studies reported sound identification accuracy of 31.6 and 35.3% (Kim and Lee, 2012; Berland et al., 2019), while a third study reported 67.6% accuracy (Liu et al., 2013). The single study, which focused specifically on pre/perilingual late-implanted adults, reported identification accuracy of 40.5% (Peasgood et al., 2003).

To an extent, such wide variation in ESI accuracy appears to be related to response format. For instance, in one study, when the same environmental sound stimuli that produced 87.5% correct in a 4-alternative forced choice (4AFC) response format were presented to the same CI users in an open set that required them to name each sound, accuracy decreased to 41.7% correct (Kaga and Akamasu, 2009). The general relationship between identification accuracy and response set size is further illustrated in **Figure 2**. Excluding the five studies with open set responses, there is a negative Spearman Rho correlation of −0.39 ($p > 0.05$). However, if the open set studies are conservatively assigned the value above 60 response options (since all five open set studies reported accuracy which was close to or below that of studies with 60 response options), the rank order correlation magnitude increases to Rho = −0.64 ($p < 0.01$). The two outlier studies in **Figure 2** with lower accuracy scores obtained on tests with a relatively small number of response options (i.e., 10; Lee and Kim, 2011; Kim and Lee, 2012) included early deafened participants, some of whom relied primarily on sign language prior to implantation.

Considering the wide variation in ESI scores of CI users, further comparisons with ESI accuracy in control groups provides a useful context for evaluating CI performance. It is notable that when normal hearing controls were included (Looi and Arnephy, 2010; Zhang et al., 2016; Chang et al., 2017; Shafiro et al., 2020), their ESI for isolated sounds was quite high (>90% correct) regardless of the number of response options. The authors of the studies that did not include normal hearing control groups similarly claimed that the stimuli used in the studies were selected to be highly identifiable by normal hearing listeners, as confirmed through pilot testing or in prior research. Furthermore, when control groups comprised individuals with hearing loss (Kaga and Akamasu, 2009; Looi and Arnephy, 2010; Lee and Kim, 2011; Kim and Lee, 2012; Shafiro et al., 2016; McMahon et al., 2018; Harris et al., 2021), study findings did not reveal significant differences in environmental sound identification accuracy between hearing impaired adults and CI users. The only exception was Kaga and Akamasu (2009), who found a better ESI performance in CI users compared to individuals with cortical deafness (auditory agnosia), while CI users performed similarly to individuals with auditory neuropathy included in the same study. Although not all of the control participants with hearing loss could be considered CI candidates and some may have had milder hearing loss, the lack of significant differences in any of these seven studies is concerning since it indicates no overall ESI improvement following implantation. Notably, in three of the above studies where controls were known to meet CI candidacy criteria (Looi and Arnephy, 2010; McMahon et al., 2018; Harris et al., 2021), likewise, no significant differences in ESI between CI candidates and CI users were observed.

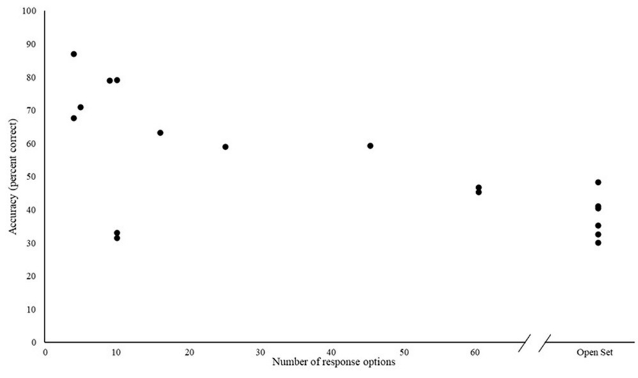


**FIGURE 2 |** Identification accuracy and the number of response options in a test.

Higher accuracy among CI users was obtained when participants were asked to categorize environmental sounds rather than to identify each sound individually. In one study of experienced adult CI users (Inverso and Limb, 2010), environmental sounds identified individually with 48.3% accuracy in open set naming were categorized with 71.1% accuracy when participants were offered to choose from five category names for each sound. As with individual sounds, however, the increase in accuracy scores could be also related to the reduction in the number of response options. Nevertheless, in another study where three predefined categories were applied to sound groupings created by CI users themselves in a free sorting task using 16 individual sounds (Strelnikov et al., 2018), categorization accuracy varied between 43%, in the first 3 months of CI use, to 60% for patients with more than 12 months of CI experience. In contrast, identification accuracy for the same sounds individually ranged from 30 to 35% across the three CI experience groups.

Accuracy scores were also affected when more than one environmental sound was presented on a single trial. For example, in Shafiro et al. (2016), when asked to identify five sounds presented together in a specific order by selecting sound names from a set of 25 response options, whole sequence identification accuracy was 14%, for placing correct sound names in the correct presentation order. In a follow up study, when Shafiro et al. (2020) modified the number of sounds in each sequence to match the number of sounds in the stimulus sequence, overall accuracy rose to 59%.

Another factor that seems to have influenced CI users' environmental sound identification scores was listening modality. However, only a couple of studies reported scores based on the CI listening modality. McMahon et al. (2018) found bimodal CI listeners performed similarly to bilateral CI listeners (64.6 and 63.7%, respectively), while both groups significantly outperformed unilateral listeners (51.4%) on environmental sound identification. These findings were more recently confirmed by Nyirjesy et al. (2020), who expanded the participant pool from the McMahon et al. (2018) study from 39 experienced CI users to 50. Similarly, McMahon et al. (2018) reported that bimodal and bilateral CI users achieved scores of 65.8 and 63.7%, respectively, outperforming unilateral CI users who scored 55.4% correct on the same 25-alternative forced choice test. Heo et al. (2013) reported smaller sound identification differences of 35.5 and 29.5% correct for bimodal and unilateral adult CI users, respectively. It is possible that the smaller differences might have resulted from score range compression due to the overall lower identification scores. In the same study, somewhat larger modality differences were also observed for CI participants for localizing environmental sounds in space with accuracy scores of 74.9% for bimodal and 63.2% for bilateral CI users. Thus, overall, it appears that bimodal and bilateral CI users have some advantage in ESI compared to unilateral CI users.

## Correlations With Speech Perception and Other Auditory Abilities

Correlation analyses of ESI scores with speech and other measures of auditory function were performed in all but three studies (Reed and Delhorne, 2005; Kaga and Akamasu, 2009; Zhang et al., 2016). Note that Reed and Delhorne (2005) did not perform correlation analysis but rather observed that "[p]erformance on the environmental-sound identification test was roughly related to [Northwestern University-6] NU-6 word recognition ability." In this study, those who scored higher than 34% correct on monosyllabic NU-6 words scored higher on environmental sound identification. When conducted, correlation analyses were based on test scores collected synchronously around the same time period, except in one study (Harris et al., 2021), which also examined the associations between pre-CI and post-CI performance for environmental sounds, speech, and spectral-temporal processing test scores. Because the studies tended to have relatively small sample sizes with a large intra-subject variance typical of CI listeners and used several scoring metrics, the foregoing discussion will focus on correlation magnitudes (Cohen, 1988) that may help to reveal converging patterns across studies.

Correlations between ESI and various measures of speech perception abilities were performed in 13 studies (**Table 2**). In the majority of these studies, speech materials were presented in quiet, and in five studies were also presented in noise. Across the 13 studies in which associations between environmental sound identification and speech perception in quiet and/or in noise were examined, all 13 studies reported correlations of moderate-to-large magnitude (i.e., $r > 0.3$) with two of the 13 studies also reporting small correlation magnitudes for additional measures of speech perception (i.e., $r < 0.3$) (Peasgood et al., 2003; Inverso and Limb, 2010). In one study, ESI was also examined in relation to indexical properties of speech, gender, and emotion identification, reporting moderate to large effect sizes for each (Shafiro et al., 2020). In the same study, Shafiro et al. also reported correlations between PES and music perception (i.e., musical instrument and genre identification) with large effect sizes for both. The associations between ESI and spectro-temporal processing abilities were examined in three studies (Shafiro et al., 2011, 2020; Harris et al., 2021). Across the three studies, correlation magnitudes were distributed between small and large depending on the type of test and also, for Harris et al. (2021), across time-points of analysis relative to the time of implantation.

# DISCUSSION

This systematic review examined published studies of PES in CI users, a perceptual ability which is generally considered to be highly valuable in daily living and an important benefit of implantation. Only studies published since the year 2000 were included to reflect performance of contemporary CI users with multichannel devices. The search strategy and inclusion criteria for the present review were broadly set to allow for the maximal inclusion of any published quantitative assessment of environmental sound perception regardless of participant age, hearing loss etiology, implant type or language and communication background.

The majority of the 19 studies that met the inclusion criteria focused on post-lingually implanted adults. One study focused on pre/perilingual adults and three focused on children with CIs. The most common assessment method used in all studies was ESI, although several studies also included categorization, localization and serial recall. Study results, based primarily on ESI, consistently indicate (1) marked deficits in CI users in comparison to normal-hearing peers, regardless of participant age and language learning background, (2) lack of evidence indicating an overall improvement in ESI following implantation,

(3) similar performance across different CI populations, (4) a tendency for bimodal and bilateral CI users to outperform unilateral CI users, and (5) mostly moderate-to-high correlations of ESI with other auditory abilities, including speech and music perception and spectro-temporal processing. This review also highlighted significant limitations in the breadth and depth of research in this area of CI outcomes assessment. Given the recognized ecological importance of PES, the present findings underscore the need for further investigation.

The limited knowledge regarding PES in contemporary CI users is concerning because both the eligibility criteria for implantation and implant technology have changed considerably over the decades (Varadarajan et al., 2021). These changes make extrapolation from earlier studies problematic. Unlike CI patients implanted in the 1980s and 1990s, the majority of whom were profoundly deaf in both ears, today's CI candidates often include adults and children who still retain usable, and in case of single-sided deafness, normal hearing in one ear (Benchetrit et al., 2021). Patients with greater overall hearing abilities prior to implantation may expect more from their hearing after implantation, and post-implantation PES scores that indicated an improvement in the past may no longer be sufficiently high. Nevertheless, evidence from studies of speech perception in contemporary CI users consistently indicate an overall improvement in speech recognition performance following implantation, particularly in quiet and, to a lesser extent in noise (Zwolan et al., 2014; Kelsall et al., 2020; Harris et al., 2021). In contrast, the small number of studies that have investigated PES in contemporary CI users do not indicate a comparable improvement in PES following implantation. Although the results are limited by the reliance on ESI as the primary PES assessment method, they reveal a generally mediocre performance and a large variability in CI users' performance, even for environmental sounds presented in quiet.

Overall, the present findings contrast with commonly held clinical views and anecdotal reports that environmental sound perception improves following implantation. Reasons for this apparent contradiction may reflect the large variability in PES performance levels of individual CI users, limitation in assessment methods and the general lack of clinical and research attention to PES as a post-CI assessment area. It is possible that following implantation, CI patients who can successfully recognize new or previously inaudible sounds are more likely to share their positive experience than those who have no or marginal changes in PES. That is, the lack of awareness in environmental sound recognition may be less readily apparent to the CI user compared to difficulties in recognition of speech, which tend to be overt and obvious – oral language users are usually well-aware when their speech perception is disrupted and they are not able to understand the words of another talker. However, CI candidates who had limited access to environmental sounds prior to implantation, often for extended periods of time, may not realize that they still cannot recognize many common environmental sounds unless they are specifically asked about it or formally tested.

The apparent discrepancy between research findings and anecdotal clinical experiences with respect to PES in CIs may also result from limitations in the assessment methods used to examine PES. The most common type of assessment administered across the studies was identification of isolated environmental sounds presented in quiet. There was also a large variation across studies in the rigor of stimulus development and selection, the number of the stimuli and the number of response options. The wide range of identification accuracy scores from different tests can give a skewed sense, especially since the number of response options used in closed set identification may influence the result (**Figure 2**) and because certain environmental sounds, such as those with strong temporal patterning, may be inherently more identifiable to CI users' than others (Reed and Delhorne, 2005; Shafiro, 2008a,b). Thus, without rigorous sampling, some stimulus sets used in ESI tests may contain inherently more or less identifiable sounds, biasing the overall outcome.

Furthermore, in everyday ecological encounters environmental sounds are rarely heard in isolation and tend to be accompanied by some contextual cues. Listeners are usually aware of the environment they are in and can leverage situational context and information from other sensory modalities to optimize PES. However, only one study (Kim and Lee, 2012) has examined environmental sound identification in the presence of background noise, while two studies have assessed the effect of context in sequences of environmental sounds distinguished by their semantic coherence with each other (Shafiro et al., 2016, 2020). Although environmental sound identification does not appear to improve following implantation, CIs may still positively contribute to environmental sound awareness, for example by informing the listener that something is happening in the environment, which may in turn lead to more accurate source identification when supplemented by visual or other contextual cues.

In natural settings, outside of the laboratory, it is also quite common for environmental sounds to be in motion, rather than stationary (e.g., a car driving by). However, perception of motion in environmental sounds was not investigated in any of the 19 studies included in the present review. Only one of these studies by Heo et al. (2013) investigated environmental sound localization. However, in Heo et al. environmental sounds were presented from one of eight stationary locations evenly distributed around the listener. A more recent study (Bahadori et al., 2021), published after the current literature search was completed, investigated judgments of distance of moving sound objects for two environmental sounds distinguished by their emotional content – either negative (car wreck) or positive (applause). The judgments of distance were modulated by the emotional content of sounds for 30 normal hearing adults, but not for 10 unilateral CI participants. On the other hand, the authors found a generally comparable ability to localize sounds in space for the CI users and normal hearing controls. Therefore, it is conceivable that following implantation, listeners with CIs may develop improved awareness of objects and events, be more likely to broadly categorize sounds and more accurately perceive the nature of interacting objects and materials, even as their identification accuracy for specific environmental sounds does not improve. With the exception of several studies that examined environmental sound categorization (Inverso and Limb, 2010; Strelnikov et al., 2018; Berland et al., 2019), other potential CI benefits for PES might not be reflected in the existing body of research.

The present review has further revealed that PES assessment is particularly lacking for two CI populations: pre/perilingual late-implanted adult CI users and children. Only one study (Peasgood et al., 2003), published nearly two decades ago, focused exclusively on PES in pre/perilingual adults. The lack of attention to this CI population is surprising given that environmental sound awareness is often one of the main reasons pre/perilingually deafened adults elect to undergo implantation, despite limited expectations about speech perception. Peasgood's et al. (2003) findings are reassuring since pre/perilingual adults in that study demonstrated environmental sound identification scores comparable to those obtained in post-lingual CI users. However, it is worth noting that six of the 10 participants were exclusively aural language users and eight were continuous hearing aid users from their first hearing loss diagnosis through CI surgery. Thus, it remains unclear how much the study findings are applicable to pre/perilingual adults with lesser oral language experience or even more limited access to sound. More research specifically focusing on more recently implanted pre/perilingual adults is needed to estimate their performance and inform pre-CI counseling in this population.

Similarly, only three studies that met this systematic review's inclusion criteria examined PES in children (Kim and Lee, 2012; Liu et al., 2013; Berland et al., 2019). All three studies demonstrated substantial deficits in ESI and categorization for the pediatric population. Two studies demonstrated a low performance of ∼30–35%, on average, while the higher score of 61–73% in CI children in the remaining study was obtained in a 4AFC format, while the CI results were still lower than normal hearing peers of the same chronological age. Partly, the low number of PES reports for children may reflect the greater difficulty of administering quantitative tests in this population, combined with the paucity of available tests. It is important to note that several studies that reported on PES in children were not included in the present review because they provided only anecdotal reports and clinician impressions or used rating scales that were not specific to environmental sounds. Although it is possible that, similar to adults, the limited available quantitative assessments do not capture all PES benefits of CI in children, PES remains an area of concern in this population and may benefit from more targeted intervention (Liu et al., 2013).

Surprisingly, only two studies have prospectively examined PES (using ESI) comparing pre- and post-CI performance in post-lingually implanted adults. However, one of these studies (Looi and Arnephy, 2010) tested only four participants in the pre-post-study arm, while the other (Harris et al., 2021) had a larger but still relatively small sample (20 participants at 6-months post-CI and 11 participants at 12-months). Furthermore, all participants in Harris et al. (2021) were bimodal CI users, which may have also affected their performance (Nyirjesy et al., 2020). Neither study found a significant overall improvement in ESI scores compared to pre-CI performance. Both studies, however, reported considerable individual variation in performance. Thus, one goal for future research in PES among CI users is to determine factors that may distinguish patients for whom PES improves from those for whom it does not. Another important goal for future research is to broaden the range of assessment methods used to evaluate potential PES benefits. An evaluation of other ecologically relevant aspects of PES in addition to ESI, and the role of attention, memory and other cognitive abilities, can lead a fuller understanding of potential PES benefits for CI users and indicate areas of strength and weakness. These may include awareness and recognition of events and objects in naturalistic auditory scenes, ability to recognize action and material properties of sound sources, integration of contextual cues provided by vision and/or other sensory modalities, judgments of location, distance and motion of common environmental sounds, perception of emotional aspects and the ability to recognize specific safety-relevant sounds.

The present systematic review considered research studies published between 2000 and 2021 that quantitatively examined environmental sound perception in CI users. Despite the generally recognized importance of environmental sound perception for individual safety, quality of life, and well-being (McRackan et al., 2017, 2019; Vasil et al., 2020), research in this area of assessment of CI users' performance appears to be significantly lacking. The 19 reviewed studies revealed generally mediocre levels of environmental sound identification and an apparent lack of improvement in group performance following implantation relative to pre-CI baseline. A wide variation in PES ability among CI users was also observed. Importantly, sounds that pose perceptual difficulty for CI users are distributed quite broadly in terms of their acoustic and semantic properties (Inverso and Limb, 2010; Shafiro et al., 2011, 2020; McMahon et al., 2018 ), and identification of sounds relevant to individual safety is not significantly different from that of non-safety relevant sounds (Hamel et al., 2020; Luzum et al., 2021). On the other hand, PES assessment methods used in the reviewed studies may not have captured some important aspects of environmental sound perception relevant to daily living of CI users. The lack of widely used validated tests that tap into different aspects of environmental sound perception may thus be a major contributing factor to the limited knowledge in this area of CI performance. Thus, strong conclusions about CI users' PES abilities seem premature. A comprehensive assessment of environmental sound perception in the post-implantation follow up can help to identify CI users with PES deficits and serve as an important step toward developing effective rehabilitation.

## 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/s.

## AUTHOR CONTRIBUTIONS

VS, MH, AM, and NL designed the study, discussed the results, and wrote the paper. NL, VS, and MH performed the paper search, paper selection, and data extraction. All authors have read and approved the publication of the final manuscript.

## FUNDING

This study was supported by: (1) the National Institutes of Health, National Institute on Deafness and Other Communication Disorders Career Development Award 5K23DC015539-02 and the American Otological Society Clinician-Scientist Award to AM; (2) the Triological Society Research Career Development Award and the National Institutes of Health (NIH)/National Deafness and Other Communication Disorders (NIDCD) R21 (1R21DC018871-01A1) to MH.

## ACKNOWLEDGMENTS

We are grateful to Rita Sieracki and Jennifer Westrick for their expert librarian assistance, to Drs. Ivy Hsiao, Jae Hee Lee, Son-A Chang, and Jong Ho Won for providing English translation and pertinent details of relevant studies.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.788899/full#supplementary-material

## REFERENCES


Bahadori, M., Barumerli, R., Geronazzo, M., and Cesari, P. (2021). Action planning and affective states within the auditory peripersonal space in normal hearing and cochlear-implanted listeners. *Neuropsychologia* 155:107790. doi: 10.1016/j.neuropsychologia.2021.107790

Benchetrit, L., Ronner, E. A., Anne, S., and Cohen, M. S. (2021). Cochlear implantation in children with single-sided deafness: a Systematic review and meta-analysis. *JAMA Otolaryngol. Head Neck Surg*. 147, 58–69. doi: 10.1001/jamaoto.2020.3852

Berland, A., Collett, E., Gaillard, P., Guidetti, M., Strelnikov, K., Cochard, N., et al. (2019). Categorization of everyday sounds by cochlear implanted children. *Sci. Rep.* 9:3532. doi: 10.1038/s41598-019-39991-9

Bond, M., Mealing, S., Anderson, R., Elston, J., Weiner, G., Taylor, R. S., et al. (2009). The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model. *Health Technol. Assess*. 13, 1–330. doi: 10.3310/hta13440

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., et al. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 368:l6890. doi: 10.1136/bmj.l6890

Capretta, N. R., and Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *Laryngoscope* 126, 699–706. doi: 10.1002/lary.25525

Carello, C., Anderson, K. L., and Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychol. Sci*. 9, 211–214. doi: 10.1111/1467-9280.00040

Chang, S. A., Won, J. H., Kim, H., Oh, S. H., Tyler, R. S., and Cho, C. H. (2017). Frequency-limiting effects on speech and environmental sound identification for cochlear implant and normal hearing listeners. *J. Audiol. Otol*. 22, 28–38. doi: 10.7874/jao.2017.00178

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates: Hillsdale, NJ.

Debruyne, J., Janssen, M., and Brokx, J. (2017). Late cochlear implantation in early-deafened adults: a detailed analysis of auditory and self-perceived benefits. *Audiol. Neurotol*. 22, 364–376. doi: 10.1159/000488023

Duchesne, L., Millette, I., Bhérer, M., and Gobeil, S. (2017). Auditory performance and subjective benefits in adults with congenital or prelinguistic deafness who receive cochlear implants during adulthood. *Cochlear Implants Int*. 18, 143–152. doi: 10.1080/14670100.2017.1290925

Gygi, B., and Shafiro, V. (2013). Auditory and cognitive effects of aging on perception of environmental sounds in natural auditory scenes. *J. Speech Lang. Hear. Res*. 56, 1595–1612. doi: 10.1044/1092-4388

Hamel, B. L., Vasil, K., Shafiro, V., Moberly, A. C., and Harris, M. S. (2020). Safety-relevant environmental sound identification in cochlear implant candidates and users. *Laryngoscope* 130, 1547–1551. doi: 10.1002/lary.28285

Harris, M. S., Moberly, A. C., Hamel, B. L., Vasil, K., Runge, C. L., Riggs, W. J., et al. (2021). A longitudinal comparison of environmental sound recognition in adults with hearing aids before and after cochlear implantation. *J. Speech Lang. Hear. Res*. 64, 1040–1052. doi: 10.1044/2020_JSLHR-20-00400

Heo, J. H., Lee, J. H., and Lee, W. S. (2013). Bimodal benefits on objective and subjective outcomes for adult cochlear implant users. *Korean J. Audiol*. 17, 65–73. doi: 10.7874/kja.2013.17.2.65

Inverso, Y., and Limb, C. J. (2010). Cochlear implant-mediated perception of nonlinguistic sounds. *Ear Hear*. 31, 505–514. doi: 10.1097/AUD.0b013e3181d99a52

Kaga, K., and Akamasu, Y. (2009). "Environmental sound perception in patients with cochlear implants compared with that in patients with auditory nerve diseases (auditory neuropathy) and cortical deafness," in *Neuropathies of the Auditory and Vestibular Eighth Cranial Nerves*. eds. K. Kaga and S. Arnold (New York, NY: Springer), 53–59. doi: 10.1007/978-4-431-09433-3

Kelsall, D., Lupo, J., and Biever, A. (2020). Longitudinal outcomes of cochlear implantation and bimodal hearing in a large group of adults: a multicenter clinical study. *Am. J. Otolaryngol*. 42:102773. doi: 10.1016/j.amjoto.2020.102773

Kim, Y. J., and Lee, H. J. (2012). A study of the speech and environmental sound recognition in the classroom noise for school-aged children with hearing Loss. *Audiology* 8, 101–110. doi: 10.21848/audiol.2012.8.1.101

Lee, H. J., and Kim, Y. J. (2011). Comparison of word and environmental sound recognition by cochlear implant and hearing aid users. *Audiology* 7, 28–39. doi: 10.21848/audiol.2011.7.1.28

Lemaitre, G., and Heller, L. (2013). Evidence for a basic level in a taxonomy of everyday action sounds. *Exp. Brain Res*. 226, 253–264. doi: 10.1007/s00221-013-3430-7

Liu, S. Y., Liu, T. C., Teng, Y. L., Lee, L. A., Lai, T. J., and Wu, C. M. (2013). Environmental sounds recognition in children with cochlear implants. *PLoS ONE* 8:e66100. doi: 10.1371/journal.pone.0066100

Looi, V., and Arnephy, J. (2010). Environmental sound perception of cochlear implant users. *Cochlear Implants Int*. 11, 203–227. doi: 10.1002/cii.428

Luzum, N., Hamel, B. Harris, M.S., and Shafiro, V. (2021). "Recognition of safety-relevant environmental sounds by cochlear implant users," in *The Fifth International Meeting on Internet & Audiology, Virtual*. Available Online at: https://www.internetaudiology.com/2021/

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., and Lefebvre, C. (2016). PRESS peer review of electronic search strategies: 2015 guideline statement. *J. Clin. Epidemiol*. 75, 40–46. doi: 10.1016/j.jclinepi.2016.01.021

McMahon, K. R., Moberly, A. C., Shafiro, V., and Harris, M. S. (2018). environmental sound awareness in experienced cochlear implant users and cochlear implant candidates. *Otol. Neurotol*. 39, e964–e971. doi: 10.1097/MAO.0000000000002006

McRackan, T. R., Hand, B. N., Velozo, C. A., and Dubno, J. R. (2019). Development of the cochlear implant quality of life item bank. *Ear Hear*. 40, 1016–1024. doi: 10.1097/AUD.0000000000000684

McRackan, T. R., Velozo, C. A., Holcomb, M. A., Camposeo, E. L., Hatch, J. L., Meyer, T. A., et al. (2017). Use of adult patient focus groups to develop the initial item bank for a cochlear implant quality-of-life instrument. *JAMA Otolaryngol. Head Neck Surg*. 143, 975–982. doi: 10.1001/jamaoto.2017.1182

Moberly, A. C., Harris, M. S., Boyce, L., Vasil, K., Wucinich, T., Pisoni, D. B., et al. (2018). Relating quality of life to outcomes and predictors in adult cochlear implant users: are we measuring the right things? *Laryngoscope* 128, 959–966. doi: 10.1002/lary.26791

NIDCD. (2021). *Quick Statistics About Hearing*. Available online at: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing (accessed September 21, 2021).

NIH (2021). *Study Quality Assessment Tools*. Available online at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools (accessed September 22, 2021).

NIH Consensus Conference (1995). Cochlear implants in adults and children *JAMA* 274:24.

Nyirjesy, S., Rodman, C., Tamati, T. N., and Moberly, A. C. (2020). Are there real-world benefits to bimodal listening? *Otol. Neurotol*. 41, e1111–e1117. doi: 10.1097/MAO.0000000000002767

Oxford Centre for Evidence-Based Medicine (2009). *Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009)*. Available online at: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 (accessed September 22, 2021).

Pastore, R. E., Flint, J. D., Gaston, J. R., and Solomon, M. J. (2008). Auditory event perception: the source-perception loop for posture in human gait. *Percept Psychophys*. 70, 13–29. doi: 10.3758/pp.70.1.13

Peasgood, A., Brookes, N., and Graham, J. (2003). Performance and benefit as outcome measures following cochlear implantation in non-traditional adult candidates: a pilot study. *Cochlear Implants Int*. 4, 171–190. doi: 10.1179/cim.2003.4.4.171

Peters, J. P. M., Hooft, L., Grolman, W., and Stegeman, I. (2015). Reporting quality of systematic reviews and meta-analyses of otorhinolaryngologic articles based on the PRISMA statement. *PLoS ONE* 10:e0136540. doi: 10.1371/journal.pone.0136540

Ramsdell, D. A. (1978). "The psychology of the hard-of-hearing and the deafened adults," in *Hearing and Deafness*, eds H. Davis and S.R. Silverman (New York, NY: Holt, Rinehart and Winston), 499–510.

Reed, C. M., and Delhorne, L. A. (2005). Reception of environmental sounds through cochlear implants. *Ear Hear*. 26, 48–61. doi: 10.1097/00003446-200502000-00005

Shafiro, V. (2008a). Identification of environmental sounds with varying spectral resolution. *Ear Hear*. 29, 401–420. doi: 10.1097/AUD.0b013e31816a0cf1

Shafiro, V. (2008b). Development of a large-item environmental sound test and the effects of short-term training with spectrally-degraded stimuli. *Ear Hear*. 29, 775–790. doi: 10.1097/AUD.0b013e31817e08ea

Shafiro, V., Gygi, B., Cheng, M. Y., Vachhani, J., and Mulvey, M. (2011). Perception of environmental sounds by experienced cochlear implant patients. *Ear Hear*. 32, 511–523. doi: 10.1097/AUD.0b013e3182064a87

Shafiro, V., Hebb, M., Walker, C., Oh, J., Hsiao, Y., Brown, K., et al. (2020). Development of the basic auditory skills evaluation battery for online testing of cochlear implant listeners. *Am. J. Audiol*. 29, 577–590. doi: 10.1044/2020_AJA-19-00083

Shafiro, V., Sheft, S., Kuvadia, S., and Gygi, B. (2015). Environmental sound training in cochlear implant users. *J. Speech Lang. Hear Res*. 58, 509–519. doi: 10.1044/2015_JSLHR-H-14-0312

Shafiro, V., Sheft, S., Norris, M., Spanos, G., Radasevich, K., Formsma, P., et al. (2016). Toward a nonspeech test of auditory cognition: semantic context effects in environmental sound identification in adults of varying age and hearing abilities. *PLoS ONE* 11:e0167030. doi: 10.1371/journal.pone.0167030

Shekar, R. C. M. C., Belitz, C., and Hanson, J. H. L. (2021). "Development of CNN-based cochlear implant and normal hearing sound recognition models using natural and auralized environmental audio," in *2021 IEEE Spoken Language Technology Workshop (SLT)*, Shenzhen, China. doi: 10.1109/SLT48900.2021.9383550

Sladen, D. P., Gifford, R. H., Haynes, D., Kelsall, D., Benson, A., Lewis, K., et al. (2017). Evaluation of a revised indication for determining adult cochlear implant candidacy. *Laryngoscope* 127, 2368–2374. doi: 10.1002/lary.26513

Strelnikov, K., Collett, E., Gaillard, P., Truy, E., Déguine, O., Marx, M., et al. (2018). Categorisation of natural sounds at different stages of auditory recovery in cochlear implant adult deaf patients. *Hear. Res*. 367, 82–194. doi: 10.1016/j.heares.2018.06.006

Tacconelli, E. (2009). Systematic reviews: CRD's guidance for undertaking reviews in health care. *Lancet Infect. Dis*. 10:226. doi: 10.1016/S1473-3099(10)70065-7

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med*. 169, 467–473. doi: 10.7326/M18-0850

Tyler, R. S., and Kelsay, D. (1990). Advantages and disadvantages reported by some of the better cochlear-implant patients. *Am. J. Otol*. 11, 282–289.

Varadarajan, V. V., Sydlowski, S. A., Li, M. M., Anne, S., and Adunka, O. F. (2021). Evolving criteria for adult and pediatric cochlear implantation. *Ear Nose Throat J*. 100, 31–37. doi: 10.1177/0145561320947258

Vasil, K. J., Lewis, J., Tamati, T., Ray, C., and Moberly, A. C. (2020). How does quality of life relate to auditory abilities? A subitem analysis of the Nijmegen Cochlear Implant Questionnaire. *J. Am. Acad. Audiol*. 31, 292–301. doi: 10.3766/jaaa.19047

Zaidman-Zait, A., Curle, D., Jamieson, J. R., Chia, R., and Kozak, F. K. (2017). Health-related quality of life among young children with cochlear implants and developmental disabilities. *Ear Hear*. 38, 399–408. doi: 10.1097/AUD.0000000000000410

Zhang, X., Qinglin, M., Guofang, T., Nenheng, Z., and Xia, L. (2016). Environmental sound recognition with cochlear implants. *Acta Acustica* 41, 891–896.

Zwolan, T. A., Henion, K., Segel, P., and Runge, C. (2014). The role of age on cochlear implant performance, use, and health utility: a multicenter clinical trial. *Otol. Neurotol*. 35, 1560–1568. doi: 10.1097/MAO.0000000000000583

**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 in Neuroscience

ORIGINAL RESEARCH
published: 26 January 2022
doi: 10.3389/fnins.2021.715660

# Speech Perception in Bilateral Hearing Aid Users With Different Grades of Asymmetric Hearing Loss

**Ulrich Hoppe[1]*, Anne Hast[1] and Thomas Hocke[2]**

[1] Department of Audiology, ENT-Clinic, University of Erlangen-Nürnberg, Erlangen, Germany, [2] Cochlear Deutschland GmbH & Co. KG, Hanover, Germany

Hearing loss is associated with decreased speech perception as well as with changes in the auditory pathway. The effects of those changes on binaural speech perception with hearing aids are not yet fully understood. To provide further evidence on the functional changes of the auditory pathway, several speech perception tests (unilateral and bilateral, aided and unaided, in quiet, and in noise) were conducted in a population of 370 bilateral hearing aid users covering the entire range of the World Health Organization's most recent classification of hearing loss. To characterize the effects of asymmetric hearing thresholds, a generalized linear model was used for regression analysis. The model revealed a detrimental effect of the poorer ears' thresholds on both the unaided and the aided unilateral word recognition scores that were attained by the better ear. Moreover, aided binaural word recognition (in quiet and in noise) was affected to a degree that cannot be explained on the sole basis of bilateral summation. Thus, this study provides evidence that there is reorganization and altered functioning of the afferent and efferent auditory pathways due to asymmetric hearing loss. Consequently, more attention should be paid to provision with a hearing aid as early as possible, and separately for each ear.

**Keywords: asymmetric hearing loss, efferent auditory system, afferent auditory system, auditory deprivation, hearing aids, speech recognition model**

OPEN ACCESS

***Edited by:***
*Melanie Ferguson, Curtin University, Australia*

***Reviewed by:***
*Pekcan Ungan, The Science Academy Society of Turkey, Turkey*
*Fernando R. Nodal, University of Oxford, United Kingdom*

***Correspondence:***
*Ulrich Hoppe*
*Ulrich.hoppe@uk-erlangen.de*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *27 May 2021*
***Accepted:*** *07 December 2021*
***Published:*** *26 January 2022*

***Citation:***
*Hoppe U, Hast A and Hocke T (2022) Speech Perception in Bilateral Hearing Aid Users With Different Grades of Asymmetric Hearing Loss. Front. Neurosci. 15:715660. doi: 10.3389/fnins.2021.715660*

## INTRODUCTION

Hearing loss is associated with a number of negative effects (Chia et al., 2007) and represents the fifth largest burden of disability (Vos et al., 2015). Additionally, according to the latest World Health Organization world report on hearing, hearing loss is the third largest cause of years with disability, and unaddressed hearing loss is estimated to impose a global cost of more than US $980 billion annually (WHO, 2021). For most people with chronic hearing loss, hearing aids (HAs) are the primary therapeutic option. HAs provide amplification and therefore better speech understanding in quiet and in noise. However, the degree of benefit varies substantially, and little is known about the actual causes of this variability. In consequence, the prevalence of HA use is rather low. The reported overall prevalence of HA use among adults varies between 9.7% (Sawyer et al., 2019) and 30% (Anovum, 2018; WHO, 2021). The actual use of HAs tends to increase with higher age, greater degree of hearing loss, the presence of comorbidities, and self-perceived limitations of hearing in everyday situations (Sawyer et al., 2019). Earlier classification of hearing loss by the WHO was based upon the pure-tone average (PTA) of the better-hearing ear throughout. Consequently, there was no recommendation for treatment of the worse ear. Recently, the WHO refined their classification and included an additional class for unilateral hearing loss. Additionally, the comment is made that "unilateral hearing loss can pose a significant challenge for an individual at any level of asymmetry. It therefore requires suitable attention and intervention based on the difficulty experienced by the person" (WHO, 2021). For several reasons (Lin and Reed, 2021), the most commonly used reference for hearing loss is the PTA. However, this measure certainly fails to reflect the full impact of hearing loss (e.g., Plomp, 1978). Therefore, standardized speech perception tests should complement pure-tone audiometry as an indispensable measure for individuals with hearing loss.

For many patients with hearing loss, its etiology is unknown. Though genetics play an important role, hearing decline generally starts in adult age and progresses over the years, either in steps or smoothly. Typically, hearing loss affects both ears in a similar way, and large asymmetries are rare. Asymmetric hearing loss has been estimated to affect 8.5–13.3% of the general population (Chia et al., 2007). The causes of asymmetric hearing loss are usually the same as for hearing loss in general; these include aging (age-related hearing loss), noise (noise-induced hearing loss), metabolic causes, genetic causes (genetic hearing loss), ototoxic drugs, viral infection, Ménière's disease, and injuries to the head or the ear. However, some of these causes are more closely associated with symmetric hearing loss and others with asymmetric hearing loss.

In summary, for both symmetric and asymmetric hearing loss, causal therapies are often not available and bilateral HAs are recommended in order to obtain best hearing outcomes. Unfortunately, for hearing loss that progresses over the years, it is quite usual for the better ear to be provided with an HA later than the worse ear. This assumption is supported by monaural and binaural HA adoption rates in Germany of about 29 and 71%, respectively (Anovum, 2018). Hence, the better ear may remain understimulated, and this could lead to detrimental effects for hearing.

Recently, Kurioka et al. (2021) investigated speech perception in twenty-eight participants with asymmetric hearing loss. In particular, they measured word recognition scores at the highest just tolerable level ($WRS_{max}$) for the participants' ears separately. They found that the worse ears exhibited significantly reduced $WRS_{max}$ when compared with ears of persons with symmetric hearing loss for given (equal) pure-tone hearing thresholds. They concluded that decreased auditory utilization of the worse-hearing ear may impair speech discrimination ability, and they identified a need for special rehabilitation. Their findings strengthen the deprivation hypothesis (Silman et al., 1984; Glick and Sharma, 2020). Silman et al. found poorer speech discrimination in the unfitted ear compared with the fitted ear. They postulated an auditory deprivation effect, indicating that reduced auditory input can induce adverse auditory plasticity through the central auditory pathway.

The aim of this study was to investigate speech-recognition scores with HAs at the conversation level, $WRS_{65}$(HA), with reference to the most recent WHO classification. Another established reference measure for HAs and other technical interventions, $WRS_{max}$ (Hoppe et al., 2014, 2016; Maier et al., 2018; McRackan et al., 2018; Franks and Jacob, 2019), was assessed. Furthermore, we measured the unaided speech perception threshold in quiet and word recognition scores in noise, and we also investigated the relationship between these routine clinical measures and the grade of asymmetry of hearing loss.

# MATERIALS AND METHODS

## Ethics Approval Statement

This study was carried out in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board at the University of Erlangen (No. 162_17Bc). All participants provided written informed consent before participation in the study.

## Patients

In this retrospective study, more than 2,000 HA examinations were screened; they were performed between August 2012 and September 2017 in the Erlangen ENT Clinic. Bilateral HA users with at least 3 months of HA experience, German as mother tongue, and a minimum age of eighteen were included. The exclusion criteria were a mean air–bone gap at 0.5, 1, 2, and 4 kHz of more than 5 dB and any technical defects of the patients' HAs. Prior to measurements, otoscopy was performed and, if needed, cerumen was removed. The test results of 370 bilateral HA users (182 men, 188 women) aged 21–98 years (mean, 62.8 years; standard deviation, 16.2 years) were eligible for assessment.

## Measurements

### Pure-Tone Air-Conduction

Thresholds were measured for frequencies between 0.125 and 8 kHz, and bone-conduction thresholds between 0.25 and 6 kHz, by using a standard clinical audiometer (AT900/AT1000 Auritec, Hamburg, Germany) with appropriate headphones (DT48, Beyerdynamic, Heilbronn, Germany). For each patient and ear, the pure-tone threshold was summarized by averaging the thresholds found at 0.5, 1, 2, and 4 kHz; these thresholds are referred to hereinafter as PTAs.

### Speech Audiometry With Headphones

Speech recognition was assessed by using the Freiburg number test and the Freiburg word test (Hahlbrock, 1957). Both were conducted with monaural presentation using headphones. Multisyllabic (two-digit) numbers were used to measure the speech-recognition threshold (SRT) in quiet, i.e., the sound pressure level (SPL) that corresponds to a recognition score of 50%. Roughly, this level corresponds to the pure-tone loss at 500 Hz + 20 dB (Braun et al., 2012). For individuals with normal hearing, the SRT is at 18 dB SPL (Brinkmann, 1974).

This relationship is an established measure in German speech audiometry to check the consistency of audiometric findings. The Freiburg monosyllable test was used to measure speech-recognition scores at higher levels and in particular the maximum word recognition score ($WRS_{max}$): Starting with 65 dB SPL, the presentation level was increased in increments of 5–15 dB until 100% speech intelligibility was attained, unless the sound level became intolerable for the user or the audiometer limit of 120 dB SPL was reached. The uncomfortable level corresponds to speech presentation at the lowest SPL that is no longer tolerated. For analysis, we used the variable "better ear $WRS_{max}$," defined as $WRS_{max}$ for the ear with the better PTA.

### Speech Audiometry in Free Sound Field

Additionally, the Freiburg monosyllable test was used to assess aided speech recognition in free sound field. Word recognition scores in quiet were determined with HA for the left and right ear separately, $WRS_{65}$ (HA). For the monaural measurements, the contralateral side was adequately blocked with earplugs. Additionally, binaural measurements for $WRS_{65}$ (HA) were performed in quiet and with masking noise at a signal-to-noise ratio of +5 dB.

Before performance of the speech perception measurements, HAs were checked by visual inspection and dynamic elicitation of acoustic feedback by shifting the earmolds, removing HA, and cupping the HA in hand. In addition, qualified personnel (HA acousticians) checked whether the type and model of the HA provided, and the amplification, were appropriate for the individual's hearing loss. Amplification was checked by real-ear measurements (Aurical, Natus, Münster, Germany).

## Data Analysis

MATLAB software version R2019b (MathWorks, Natick MA, United States) was used for all calculations and figures. A generalized linear regression model (GLM) was applied to the data. For speech recognition scores, model data for sigmoid regression were calculated according to equation 1. For speech recognition threshold, a linear fit was derived by using equation 2.

$$Score\,[\%] = \frac{100}{1 + e^{-(\beta_0 + \beta_1 \cdot PTA + \beta_2 \cdot Asymmetry)}} \qquad (1)$$

$$SRT\,[dB] = \beta_0 + \beta_1 \cdot BEA + \beta_2 \cdot Asymmetry \qquad (2)$$

BEA (better ear average) refers to the better-ear four-frequency average of the pure-tone thresholds. The asymmetry refers to the difference between BEA and the poorer-ear four-frequency average.

Any effects of hearing thresholds and asymmetric hearing loss on speech perception measures were considered significant if the $p$-value was below 0.05.

# RESULTS

The results of the speech perception measurements are shown in **Figures 1–4**. **Figure 1** shows the unaided speech perception scores in quiet, while **Figures 2**, **3** show the aided speech perception scores in quiet. **Figure 4** refers to the aided speech perception in noise. The boxplots result from the grouping of hearing loss according to the WHO grade. In cases where there was a significant effect of asymmetric hearing loss on unilateral or bilateral scores, the results of the regression model are shown as examples for (i) symmetric hearing, (ii) for asymmetry of 15 dB, and (iii) for asymmetry of 30 dB. The characteristics of the study population are summarized in **Table 1**: age, WHO classification, and PTA asymmetry. Asymmetric hearing was similar across WHO grades up to Grade 6. Mean age was approximately the

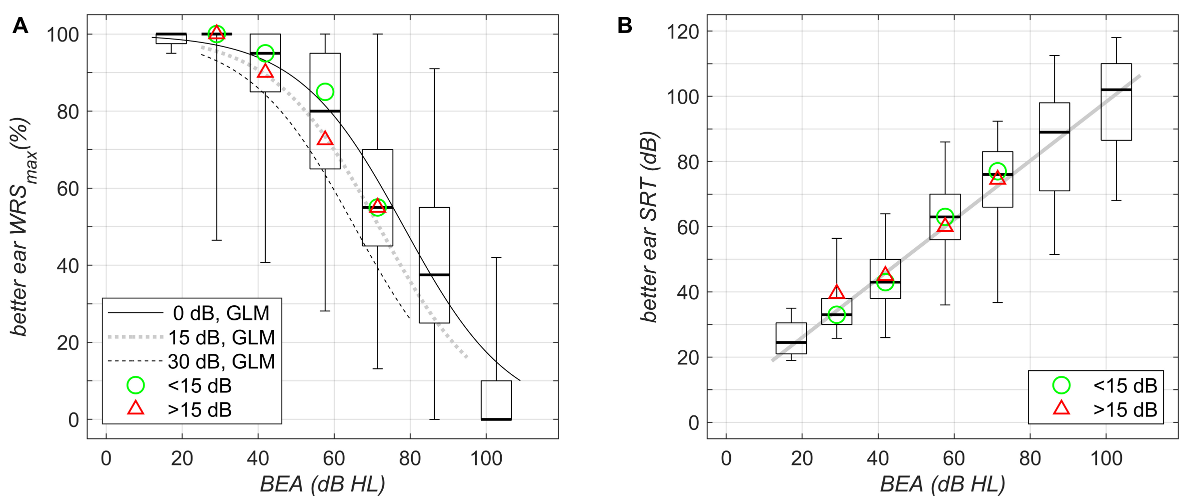


**FIGURE 1 |** Unilateral unaided better-ear speech-recognition scores grouped by WHO classification of hearing loss (boxplots; left to right, Grades 0–6) and plotted against better-ear PTA. **(A)** Maximum word recognition score, $WRS_{max}$, for the different WHO-grade groups as measured by the Freiburg monosyllable test. The boxplots summarize the $WRS_{max}$ results; boxes show the medians and 1st and 3rd quartiles, while whiskers denote the 2.5 and 97.5 percentiles. The lines show the results of the fitted GLM: solid line, calculation performed for symmetric PTA (interaural difference 0 dB); dotted line, asymmetric PTA with interaural difference 15 dB; dashed line, asymmetric PTA with interaural difference 30 dB. **(B)** Speech-recognition score (SRT) in quiet as measured by the Freiburg multisyllable test. The solid line shows the result of linear regression; boxplots as in A summarize the SRT results. The green circle and the red triangle in **(A,B)** denote the median of the data with corresponding WHO grade for asymmetries below and above 15 dB, respectively.

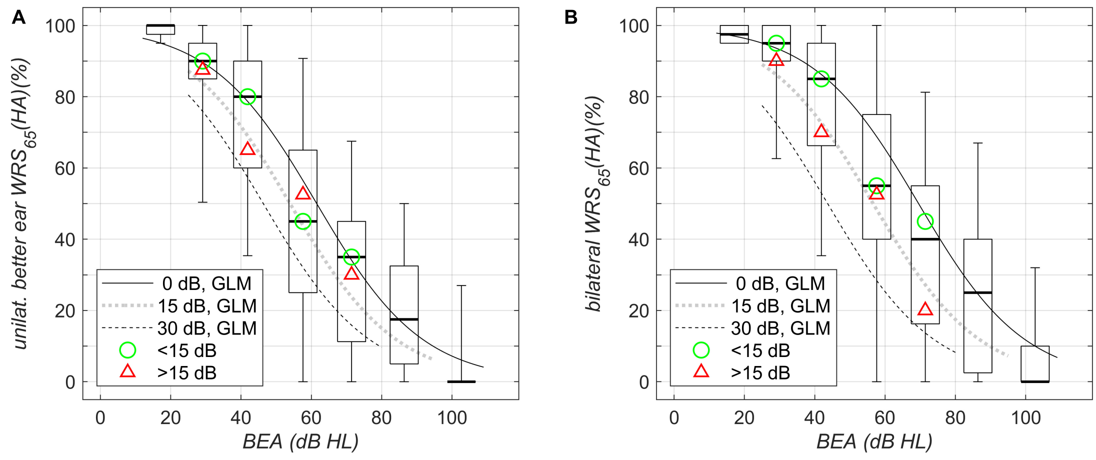


**FIGURE 2 |** Aided speech-recognition scores in quiet grouped by WHO classification of hearing loss and plotted against better-ear PTA (BEA). **(A)** Unilateral WRS $(HA)_{65}$. **(B)** Bilateral WRS $(HA)_{65}$. Boxplots and solid/dotted/dashed lines as in **Figure 1A**. The green circle and the red triangle in **(A,B)** denote the median of the data with corresponding WHO grade for asymmetries below and above 15 dB, respectively.

same in each WHO-grade group (**Table 1**), and no correlation was seen between age and the degree of hearing asymmetry ($r_{Spearman} = 0.07$, $p = 0.18$). For WHO Grade 6, the mean PTA difference was smaller because of audiometer limits; thresholds beyond the audiometer limits were set to the limit values.

## Unaided Speech Perception in Quiet

**Figure 1A** shows boxplots for $WRS_{max}$ in dependence upon the BEA. BEAs were grouped according to WHO Grades 0–6. The scores show the largest variability of around 90 percentage points (pp) for WHO Grades 4 and 5. The model regression revealed a decrease in $WRS_{max}$ of up to 20 pp; the greatest decrease was seen for an asymmetric hearing loss of 30-dB interaural PTA difference at a better-ear PTA of around 60-dB hearing loss (HL).

**Figure 1B** shows a corresponding analysis of SRT for multisyllables; here we found a large, continuously increasing variability with an increasing degree of hearing loss. The model regression did not reveal any significant effect of asymmetry on the better-ear SRT.

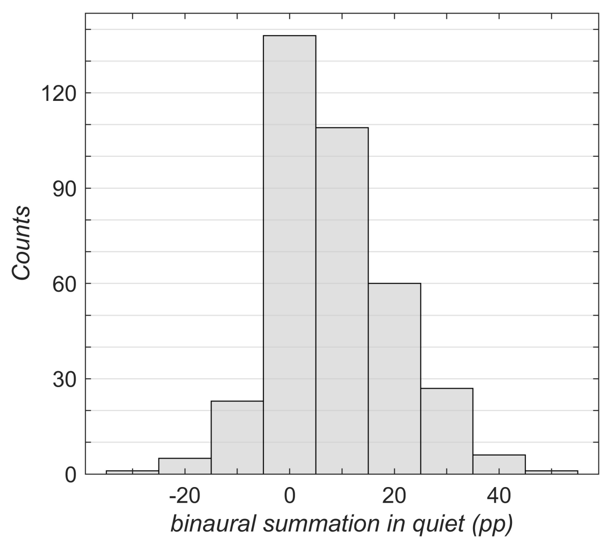


**FIGURE 3 |** Overview of the study population broken down by binaural summation. The binaural summation was calculated as the difference in percentage points (pp) between the unilateral aided better-ear score (**Figure 2A**) and the bilateral aided score (**Figure 2B**). A negative value corresponds to binaural interference in which the bilateral score was poorer than the unilateral better-ear score.

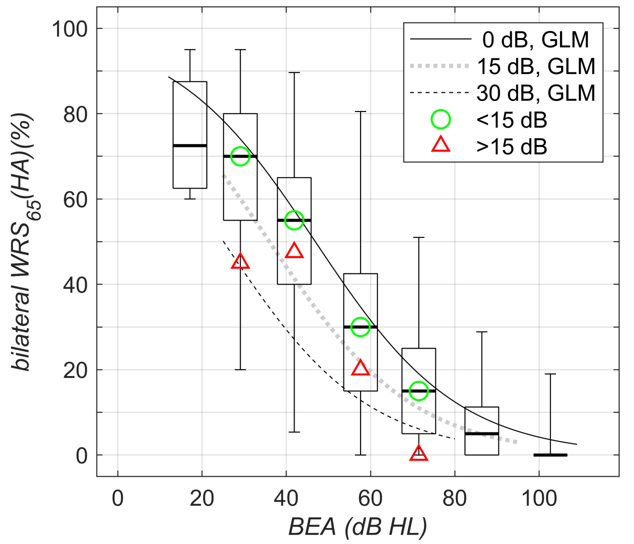


**FIGURE 4 |** Bilateral aided speech-recognition scores in noise plotted against WHO classification of hearing loss. Boxplots and solid/dotted/dashed lines as in **Figure 1A**. The green circle and the red triangle denote the median of the data with corresponding WHO grade for asymmetries below and above 15 dB, respectively.

## Aided Speech Perception in Quiet

**Figure 2A** shows boxplots for aided word recognition score at a presentation level of 65 dB SPL, $WRS_{65}$(HA), plotted against BEA. The scores show the largest variability, around 90 pp, for WHO Grade 3. For asymmetric hearing loss, with BEA around 60 dB HL and poorer-ear PTA around 90 dB HL, the model regression revealed a decrease of up to 24 pp in aided unilateral score. For the bilateral score (**Figure 2B**), the largest variability was again observed for WHO Grade 3. For symmetric hearing and PTAs of 60 dB HL, an improved score was found, compared with unilateral scores (53%), by about 12 pp and up to 65%. For asymmetric hearing loss with an interaural difference of 30 dB, there was already a decreased unilateral baseline performance (29%). Additionally, the model results not only revealed the absence of any binaural summation effect but also showed a slight binaural interference of 4 pp (25%). Therefore, the overall disadvantage for HA users with a better-ear PTA of 60 dB HL and an asymmetry of 30 dB adds up to 40 pp. With respect to asymmetry, the break-even for a binaural summation effect in quiet with BEA of 60 dB HL was found to be around 20 dB: for those cases, the two models yielded a unilateral score equal to the binaural score. Where the asymmetry was larger, binaural interference was dominant, while for binaural summation, a smaller asymmetry was found to be a precondition.

**Figure 3** shows an overview of the binaural summation effect across all grades of hearing loss. A considerable part of the population (37%) has no significant binaural summation (i.e., below 5 pp). More than one-half (55%) of the population was assigned to positive categories, exhibiting a binaural summation effect. Less than one-tenth (8%) of the patients exhibited binaural interference of more than 5 pp.

## Aided Speech Perception in Noise

For bilateral speech perception in noise, **Figure 4**, the largest variability was found for WHO Grades 2 and 3. Owing to the test characteristic (with saturation and floor effects), for a BEA of 60 dB HL, the detrimental effect of asymmetric hearing was found to be up to 20 pp, while for a BEA of 40 dB HL, the detrimental effect of asymmetric hearing was found to be up to 31 pp. Both decrements are for asymmetric hearing of 30-dB side difference.

## Generalized Linear Regression Model

**Table 2** summarizes the results for the GLM parameters. Parameters for sigmoid regression were calculated according to equation 1. For the linear fit according to equation 2, the GLM yielded a non-significant $\beta_2$. Hence, we simplified equation 2 to equation 3:

$$SRT\left[dB\right] = \beta_0 + \beta_1 \cdot PTA \quad (3)$$

In summary, the suprathreshold measures, unilateral $WRS_{max}$, unilateral/bilateral $WRS_{65}$ (HA) in quiet, and bilateral $WRS_{65}$ (HA) in noise, depend on asymmetry. This was not found for the near threshold measure of the unilateral better-ear SRT.

# DISCUSSION

Hearing outcomes for a large group of bilateral HA users were investigated within the context of routine clinical measurements. The population covered the degrees of hearing loss from WHO Grades 0–6. Outcome measures with and without HAs in quiet and in noise were found to depend significantly on the degree of asymmetry. For bilateral conditions, this is a well-known finding (Vannson et al., 2015; Jerger et al., 2017). However, the present study showed that even the unilateral scores in quiet on the better side are negatively affected by the degree of hearing loss on the contralateral (worse) side.

For the clinically relevant measures of the unilateral maximum word recognition score and the unilateral score with HAs, our results were similar to those of earlier studies (Hoppe et al., 2014). Unilaterally aided speech perception scores above 50%

**TABLE 1 |** Patient characteristics.

| WHO grade (PTA [dB]) | Number of patients | Mean age [years] | PTA difference [dB] | | | |
|---|---|---|---|---|---|---|
| | | | Mean ± SD | No. of participants with PTA difference | | |
| | | | | 0–10 dB | 10–20 dB | > 20 dB |
| 0 (< 20) | 4 (1%) | 63 ± 9 | 8 ± 5 | 2 | 2 | 0 |
| 1 (20–< 35) | 51 (14%) | 65 ± 10 | 8 ± 8 | 38 | 12 | 1 |
| 2 (35–< 50) | 103 (28%) | 65 ± 14 | 7 ± 6 | 68 | 31 | 4 |
| 3 (50–< 65) | 97 (26%) | 66 ± 16 | 7 ± 7 | 72 | 20 | 5 |
| 4 (65–< 80) | 55 (15%) | 61 ± 17 | 10 ± 8 | 35 | 15 | 5 |
| 5 (80–< 95) | 32 (8.5%) | 55 ± 17 | 8 ± 7 | 23 | 5 | 4 |
| 6 (≥ 95) | 28 (7.5%) | 49 ± 18 | 3 ± 3 | 25 | 3 | 0 |
| All | 370 (100%) | 63 ± 16 | 8 ± 7 | 263 (71%) | 88 (24%) | 19 (5%) |

**TABLE 2 |** Parameters of the generalized linear regression models.

| | | Parameter | Estimate | Standard error | *t*-statistic | *p* | [β] | |
|---|---|---|---|---|---|---|---|---|
| Unaided scores in quiet | $WRS_{max}$ (**Figure 1A**) | $\beta_0$ | 5.60 | 0.13 | 43.4 | <0.0001 | | |
| | | $\beta_1$ | –0.0714 | 0.0018 | –40.2 | <0.0001 | 1/dB | $PTA_{betterear}$ |
| | | $\beta_2$ | –0.0308 | 0.0044 | –7.0 | <0.0001 | 1/dB | Asymmetry |
| | | 7,400 observations, 7,397 error degrees of freedom, $\chi^2$-statistic vs. constant model: $2.5{\cdot}10^3$, $p < 0.0001$ | | | | | | |
| | SRT (**Figure 1B**) | $\beta_0$ | 7.97 | 1.803 | 4.4 | <0.0001 | | |
| | | $\beta_1$ | 0.903 | 0.0307 | 29.4 | <0.0001 | 1/dB | $PTA_{betterear}$ |
| | | 355 observations, 353 error degrees of freedom, F-statistic vs. constant model: 864, $p < 0.0001$ | | | | | | |
| Aided scores in quiet | Unilateral $WRS_{65}$ (HA) (**Figure 2A**) | $\beta_0$ | 4.09 | 0.10 | 39.5 | <0.0001 | | |
| | | $\beta_1$ | –0.0664 | 0.0017 | –39.7 | <0.0001 | 1/dB | $PTA_{betterear}$ |
| | | $\beta_2$ | –0.0337 | 0.0040 | –8.4 | <0.0001 | 1/dB | Asymmetry |
| | | 7,400 observations, 7,397 error degrees of freedom, $\chi^2$-statistic vs. constant model: $2.5{\cdot}10^3$, $p < 0.0001$ | | | | | | |
| | Bilateral $WRS_{65}$ (HA) (**Figure 2B**) | $\beta_0$ | 4.61 | 0.11 | 42.2 | <0.0001 | | |
| | | $\beta_1$ | –0.0663 | 0.0016 | –40.2 | <0.0001 | 1/dB | $PTA_{betterear}$ |
| | | $\beta_2$ | –0.0572 | 0.0041 | –14.0 | <0.0001 | 1/dB | Asymmetry |
| | | 7,400 observations, 7,397 error degrees of freedom, $\chi^2$-statistic vs. constant model: $2.6{\cdot}10^3$, $p < 0.0001$ | | | | | | |
| Aided scores in noise | Bilateral $WRS_{65}$(HA) (**Figure 3**) | $\beta_0$ | 2.76 | 0.10 | 28.7 | <0.0001 | | |
| | | $\beta_1$ | –0.0589 | 0.00172 | –34.2 | <0.0001 | 1/dB | $PTA_{betterear}$ |
| | | $\beta_2$ | –0.0428 | 0.00427 | –10.0 | <0.0001 | 1/dB | Asymmetry |
| | | 7,400 observations, 7,397 error degrees of freedom, $\chi^2$-statistic vs. constant model: $1.8{\cdot}10^3$, $p < 0.0001$ | | | | | | |

are typically found for hearing loss below 60 dB. All of the above studies referred to unilateral scores. Kronlachner et al. (2018) found a minimal effect of cognition on the success of HA provision. Additionally, in their population of 40 HA users, they found a deterioration of $WRS_{max}$ with age. Müller et al. (2016) investigated age effects in elderly HA users and found effects for both measures [$WRS_{max}$ and $WRS_{65}$ (HA)] of the order of 10–20 pp. Regrettably, neither of these studies considered contralateral hearing or included binaural measurements.

## Effects of Asymmetric Hearing Thresholds on Better-Ear Speech Perception

For our study population, there was no correlation between age and degree of hearing asymmetry. Otherwise, the detrimental effect of age on speech perception would have been superimposed upon, or even have masked, the effect of asymmetric hearing. Most remarkably, even for unilateral $WRS_{max}$, there were effects of the order of 20 pp. The PTA range in which this effect was the greatest is obviously determined by the ceiling effects of the speech material used and the presentation levels applied. $WRS_{max}$ is typically measured near the discomfort level, while SRT is measured at a low level; the different presentation levels are probably the root cause of the different findings; as for SRTs in quiet, the asymmetry did not show significant effects in our study population.

For the purpose of simplification, the impact of hearing loss can be attributed in terms of functionality to two different components (Plomp, 1978; Plomp and Mimpen, 1979): (i) The attenuation component simply describes the effect of weakened sound perception due to the sensorineural component of hearing loss. This component should be easily compensated for by acoustic amplification. (ii) The distortion component refers to the loss of dynamic, frequency dependence of hearing loss, and the loss of temporal processing.

Complementary to this functional description of the impact of hearing loss, there is pathophysiological classification for hearing loss, which was originally applied to different types of presbyacusis. Schuknecht (1964) and Johnsson and Hawkins (1972) proposed for presbyacusis the terms sensory, metabolic, mechanical, vascular, and neural; for the latter, we prefer the term "central," reserving "neural presbyacusis" for hearing loss due to the degeneration of the cochlear nerve. Within this classification, sensory presbyacusis is equivalent to the attenuation component. All the other types are summarized by the distortion component. Commonly, age-related hearing loss (presbyacusis) means hearing loss in the elderly. However, the term does not refer exclusively to aging of the auditory pathway; it can be interpreted as having a much broader meaning. Consequently, it includes the cumulative, genetically determined effects of aging, and this may also include possible damage to the auditory system caused by environmental noise (Johnsson and Hawkins, 1972; Schacht and Hawkins, 2005). One may therefore apply the above classification to the findings in our population, i.e., HA users with sensorineural hearing loss.

One possible cause for the observed detrimental effect of asymmetric PTA on unilateral better-ear scores could be the deprivation of the contralateral ear (Kurioka et al., 2021). Our data suggest that this deprivation may have an effect not only on the afferent auditory pathway but also on the efferent system. For the $WRS_{max}$ at the near-uncomfortable presentation level, we found an effect, but not for the SRT, which is measured at near-threshold levels. The effects of the efferent system are believed to begin significantly above the hearing threshold if they are to be measurable. Hence, it is reasonable to consider that $WRS_{max}$ is influenced by asymmetry while SRT is not. Therefore, one may at least partially assign the lower scores for the asymmetric PTA cases to a compromised efferent system in those patients. Unfortunately, the objective assessment of such effects by otoacoustic emission in individuals with hearing impairment presents a Gordian knot. Following the discovery of otoacoustic emissions (Gold, 1948; Kemp, 1978), their measurement was soon found useful for the objective assessment of effects that can be assigned to the efferent auditory pathway (Guinan, 2018; Lopez-Poveda, 2018; Fuchs and Lauer, 2019; Lauer et al., 2021). In patients with significant hearing loss, this approach is not possible owing to the lack of measurable otoacoustic emissions. Even though ultimate evidence is still lacking, we hypothesize that the missing efferent mechanisms result in deteriorated speech recognition. Following the functional description by Plomp (1978), these effects of impairment can be attributed to an increased distortion component of hearing loss.

The GLM revealed detrimental effects of asymmetric hearing, namely, the effect of the poorer-ear PTA on speech perception by the better ear. However, such a model does not permit a more detailed analysis. It remains unclear whether those effects can be attributed exclusively to interaural asymmetry. For higher degrees of hearing loss in the better ear, the poorer-ear PTA is subject to "numerical" saturation effects due to the limits of the audiometer. It is reasonable to assume that higher degrees of hearing loss have a detrimental effect on the efferent system as well.

## Bilateral Speech Perception

The binaural summation effect might be regarded as a result of the loudness increase from one to two ears for unimodal bilateral and symmetric listening (Christen, 1980; Rawool and Parrill, 2018). It is well known that binaural summation does not occur in all HA users (Arkebauer et al., 1971; Allen et al., 2000). Arkebauer et al. (1971) found a detrimental interaction between ears exhibiting bilateral asymmetric hearing loss, later referred to a binaural interference. They already highlighted the observation that in "cases with bilateral hearing loss, candidacy for binaural amplification should be determined from each ear independently, and the combined effect of both aids." Our study design did not allow for determination of an effect of age on binaural interference. A corresponding matching of HA users was not possible. The model revealed a detrimental effect of asymmetry on binaural summation in quiet and noise. Trivially, asymmetry is equivalent to poorer PTA on the poorer ear and is therefore less surprising than a detrimental effect of asymmetry on unilateral better-ear scores. However, as the example of HA users with better-ear PTA around 60 dB impressively illustrates (**Figure 2**), the two disadvantages add up. For asymmetric hearing loss, both symptoms (the already decreased perception on the better ear and the missing binaural summation, or even binaural interference) might be caused by afferent and efferent deprivation.

## Limitations of the Study

The retrospective approach of this study certainly needs confirmation from prospective studies. Even though the number of HA users included was relatively high, the large variability in age, in hearing loss, and in experience with HAs prevents a more sophisticated evaluation. The GLM is *per se* an average-based model. The average model output was based on many different patients with highly variable progress of hearing loss, in some cases differing very strongly between the ears of the same study participant. Attempts at in-depth analysis, especially if retrospective, can easily result in an overfit of a model if it includes too many parameters. Even though in our study the GLM yielded significant effects of asymmetric PTA, it has to be stressed that these findings are preliminary owing to the retrospective study design.

According to clinical routine at our institution, HA users undergo unilateral assessment of speech perception on each side in quiet. Speech in noise is assessed bilaterally only. Altogether, each HA user routinely undergoes six different speech tests. Consequently, one cannot exclude the possibility that in some patients fatigue effects may have played a role, and thus increased the variability in the data. Probably the most important shortcoming of this study with respect to the hypothesized root causes of the effect of asymmetry, namely deprivation of afferent and efferent pathways, is the lack of detailed knowledge about the progress of individuals' hearing loss. In a first attempt, it seems to be obvious to assign larger detrimental effects of asymmetry to a longer period of asymmetry. However, although this may fit in better with what we know so far, one cannot exclude the possibility that, after asymmetry has set in, reorganization of the auditory pathway may help in overcoming such detrimental effects. Recent findings in unilaterally deaf patients indicate such reorganization within as little as 1 year (Müller et al., 2017). The present study is a snapshot of a typical clinical population of HA users, and as such, it does not reveal deeper insights into the time course and the direction of detrimental effects in patients with and without asymmetric hearing loss. However, in view of the large effects of asymmetric PTA on speech perception that we have observed, further and deeper investigations of auditory deprivation effects are needed, particularly with reference to the efferent innervation of the better ear.

## Clinical Consequences for the Treatment of Asymmetric Hearing Loss

The results of this study suggest strongly that early treatment of hearing loss may be beneficial, even if the hearing loss is asymmetric and the prescription of an HA for both ears may not be considered urgently needed by the patient. Reimbursement criteria should reflect the detrimental effects of asymmetric hearing loss such that "in cases with bilateral hearing loss, candidacy for binaural amplification should be determined from each ear independently, and the combined effect of both aids" (Arkebauer et al., 1971).

## CONCLUSION

In a population of hearing-aid users, including symmetric and asymmetric PTA, the asymmetry exerts a detrimental effect on both unaided and aided word recognition by the better ear. Also, the binaural speech perception with HAs worsens with increasing asymmetry. This decrease exceeds the limits of a missing binaural summation by far. More attention has to be paid to provision of an HA as early as possible. There is an evident need for more research on the short- and long-term effects of asymmetric hearing on the afferent and efferent auditory pathways in individuals with hearing impairment.

## 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 Ethik-Kommission der Friedrich-Alexander-Universität Erlangen-Nürnberg. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

UH: conception and design, acquisition and interpretation of data, and final approval. AH: revision and final approval. TH: drafting and design, figure preparation, and final approval. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

We would like to thank Peter Woolley for his careful revision of language and careful proofreading.

## REFERENCES

Allen, R. L., Schwab, B. M., Cranford, J. L., and Carpenter, M. D. (2000). Investigation of binaural interference in normal-hearing and hearing-impaired adults. *J. Am. Acad. Audiol.* 11, 494–500.

Anovum (2018). *Anovum EuroTrak Germany*. Available online at: https://www.ehima.com/wp-content/uploads/2018/06/EuroTrak_2018_GERMANY.pdf. Download 19-05-2021. (accessed January 11, 2022)

Arkebauer, H. J., Mencher, G. T., and McCall, C. (1971). Modification of speech discrimination in patients with binaural asymmetrical hearing loss. *J. Speech Hear. Disord.* 36, 208–212. doi: 10.1044/jshd.3602.208

Braun, T., Wimmer, M., and Hempel, J. M. (2012). [Two formulas for exact calculation of hearing loss for numbers]. *HNO* 60, 814–816. doi: 10.1007/s00106-011-2474-x

Brinkmann, K. (1974). Die neuaufnahme der „wörter für gehörprüfung mit sprache". *Z. Hörgeräteakust* 13, 12–40.

Chia, E. M., Wang, J. J., Rochtchina, E., Cumming, R. R., Newall, P., and Mitchell, P. (2007). Hearing impairment and health-related quality of life: the blue mountains hearing study. *Ear Hear.* 28, 187–195. doi: 10.1097/AUD.0b013e31803126b6

Christen, R. (1980). Binaural summation at the most comfortable loudness level (MCL). *Aust. J. Audiol.* 2, 92–98.

Franks, Z. G., and Jacob, A. (2019). The speech perception gap in cochlear implant patients. *Cochlear Implants Int.* 20, 176–181. doi: 10.1080/14670100.2019.1582165

Fuchs, P. A., and Lauer, A. M. (2019). Efferent inhibition of the Cochlea. *Cold Spring Harb. Perspect. Med.* 9:a033530. doi: 10.1101/cshperspect.a033530

Glick, H. A., and Sharma, A. (2020). Cortical neuroplasticity and cognitive function in early-stage, mild-moderate hearing loss: evidence of neurocognitive benefit from hearing aid use. *Front. Neurosci.* 14:93. doi: 10.3389/fnins.2020.00093

Gold, T. (1948). Hearing. The physiological basis of action in the cochlea. *Proc. R. Soc. B* 135, 492–498.

Guinan, J. J. Jr. (2018). Olivocochlear efferents: their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses. *Hear. Res.* 362, 38–47. doi: 10.1016/j.heares.2017.12.012

Hahlbrock, K. H. (1957). *Sprachaudiometrie: Grundlagen und Praktische Anwendung Einer Sprachaudiometrie für das Deutsche Sprachgebiet*. Stuttgart: Thieme.

Hoppe, U., Hast, A., and Hocke, T. (2014). Speech perception with hearing aids in comparison to pure-tone hearing loss. *HNO* 62, 443–448.

Hoppe, U., Hocke, T., Müller, A., and Hast, A. (2016). Speech perception and information-carrying capacity for hearing aid users of different ages. *Audiol. Neurootol.* 21 (Suppl. 1), 16–20. doi: 10.1159/000448349

Jerger, J., Silman, S., Silverman, C., and Emmer, M. (2017). Binaural interference: quo vadis? *J. Am. Acad. Audiol.* 28, 266–270.

Johnsson, L. G., and Hawkins, J. E. Jr. (1972). Sensory and neural degeneration with aging, as seen in microdissections of the human inner ear. *Ann. Otol. Rhinol. Laryngol.* 81, 179–193. doi: 10.1177/000348947208100203

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* 64:1386e1391.

Kronlachner, M., Baumann, U., Stover, T., and Weissgerber, T. (2018). [Investigation of the quality of hearing aid provision in seniors considering cognitive functions]. *Laryngorhinootologie* 97, 852–859. doi: 10.1055/a-0671-2295

Kurioka, T., Sano, H., Furuki, S., and Yamashita, T. (2021). Speech discrimination impairment of the worse-hearing ear in asymmetric hearing loss. *Int. J. Audiol.* 60, 54–59. doi: 10.1080/14992027.2020.1795282

Lauer, A. M., Jimenez, S. V., and Delano, P. H. (2021). Olivocochlear efferent effects on perception and behavior. *Hear. Res.* 108207. doi: 10.1016/j.heares.2021.108207 [Epub ahead of print].

Lin, F. R., and Reed, N. S. (2021). The pure-tone average as a universal metric-knowing your hearing. *JAMA Otolaryngol. Head Neck Surg.* 147, 230–231. doi: 10.1001/jamaoto.2020.4862

Lopez-Poveda, E. A. (2018). Olivocochlear efferents in animals and humans: from anatomy to clinical relevance. *Front. Neurol.* 9:197. doi: 10.3389/fneur.2018.00197

Maier, H., Baumann, U., Baumgartner, W. D., Beutner, D., Caversaccio, M. D., Keintzel, T., et al. (2018). Minimal reporting standards for active middle ear hearing implants. *Audiol. Neurootol.* 23, 105–115. doi: 10.1159/000490878

McRackan, T. R., Fabie, J. E., Burton, J. A., Munawar, S., Holcomb, M. A., and Dubno, J. R. (2018). Earphone and aided word recognition differences in cochlear implant candidates. *Otol. Neurotol.* 39, e543–e549. doi: 10.1097/mao.0000000000001873

Müller, A., Hocke, T., Hessel, H., and Mir-Salim, P. (2017). [Results of speech audiometry testing in case of transcranial CROS depend on duration of deafness]. *Laryngorhinootologie* 96, 293–298. doi: 10.1055/s-0042-111079

Müller, A., Hocke, T., Hoppe, U., and Mir-Salim, P. (2016). [The age effect in evaluation of hearing aid benefits by speech audiometry]. *HNO* 64, 143–148. doi: 10.1007/s00106-015-0115-5

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J. Acoust. Soc. Am.* 63, 533–549.

Plomp, R., and Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *J. Acoust. Soc. Am.* 66, 1333–1342.

Rawool, V. W., and Parrill, M. (2018). Objective evaluation of binaural summation through acoustic reflex measures. *Int. J. Audiol.* 57, 493–501.

Sawyer, C. S., Armitage, C. J., Munro, K. J., Singh, G., and Dawes, P. D. (2019). Correlates of hearing aid use in UK adults: self-reported hearing difficulties, social participation, living situation, health, and demographics. *Ear Hear.* 40, 1061–1068. doi: 10.1097/aud.0000000000000695

Schacht, J., and Hawkins, J. E. (2005). Scetches of otohistory. Part 9: Presby[a]cusis. *Audiol. Neurotol.* 10, 243–247. doi: 10.1159/000086524

Schuknecht, H. F. (1964). Further observations on the pathology of presbycusis. *Arch. Otolaryngol.* 80, 369–382. doi: 10.1001/archotol.1964.00750040381003

Silman, S., Gelfand, S. A., and Silverman, C. A. (1984). Late-onset auditory deprivation: effects of monaural versus binaural hearing aids. *J. Acoust. Soc. Am.* 76, 1357–1362. doi: 10.1121/1.391451

Vannson, N., James, C., Fraysse, B., Strelnikov, K., Barone, P., Deguine, O., et al. (2015). Quality of life and auditory performance in adults with asymmetric hearing loss. *Audiol. Neurootol.* 20(Suppl. 1), 38–43. doi: 10.1159/000380746

Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 386, 743–800. doi: 10.1016/s0140-6736(15)60692-4

WHO (2021). *World Report on Hearing*. Geneva: World Health Organization.

**Conflict of Interest:** TH was employed by Cochlear GmbH & Co. KG, Germany.

The remaining 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 in Neuroscience

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/fnins.2021.725412

Check for updates

# Speech Recognition and Listening Effort in Cochlear Implant Recipients and Normal-Hearing Listeners

*Khaled H. A. Abdel-Latif*† *and Hartmut Meister**†

*Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Cologne, Germany*

The outcome of cochlear implantation is typically assessed by speech recognition tests in quiet and in noise. Many cochlear implant recipients reveal satisfactory speech recognition especially in quiet situations. However, since cochlear implants provide only limited spectro-temporal cues the effort associated with understanding speech might be increased. In this respect, measures of listening effort could give important extra information regarding the outcome of cochlear implantation. In order to shed light on this topic and to gain knowledge for clinical applications we compared speech recognition and listening effort in cochlear implants (CI) recipients and age-matched normal-hearing listeners while considering potential influential factors, such as cognitive abilities. Importantly, we estimated speech recognition functions for both listener groups and compared listening effort at similar performance level. Therefore, a subjective listening effort test (adaptive scaling, "ACALES") as well as an objective test (dual-task paradigm) were applied and compared. Regarding speech recognition CI users needed about 4 dB better signal-to-noise ratio to reach the same performance level of 50% as NH listeners and even 5 dB better SNR to reach 80% speech recognition revealing shallower psychometric functions in the CI listeners. However, when targeting a fixed speech intelligibility of 50 and 80%, respectively, CI users and normal hearing listeners did not differ significantly in terms of listening effort. This applied for both the subjective and the objective estimation. Outcome for subjective and objective listening effort was not correlated with each other nor with age or cognitive abilities of the listeners. This study did not give evidence that CI users and NH listeners differ in terms of listening effort – at least when the same performance level is considered. In contrast, both listener groups showed large inter-individual differences in effort determined with the subjective scaling and the objective dual-task. Potential clinical implications of how to assess listening effort as an outcome measure for hearing rehabilitation are discussed.

**Keywords: listening effort, speech recognition, effort scaling, dual-task, cognition, working memory**

OPEN ACCESS

***Edited by:***
*Todd Andrew Ricketts,*
*Vanderbilt University, United States*

***Reviewed by:***
*Erin Margaret Picou,*
*Vanderbilt University Medical Center, United States*
*Erol J. Ozmeral,*
*University of South Florida, United States*

****Correspondence:***
*Hartmut Meister*
*hartmut.meister@uni-koeln.de*

*†These authors have contributed equally to this work*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *15 June 2021*
***Accepted:*** *23 December 2021*
***Published:*** *10 February 2022*

***Citation:***
*Abdel-Latif KHA and Meister H (2022) Speech Recognition and Listening Effort in Cochlear Implant Recipients and Normal-Hearing Listeners. Front. Neurosci. 15:725412. doi: 10.3389/fnins.2021.725412*

## INTRODUCTION

Cochlear implants (CI) have been established as the treatment of severe to profound hearing loss in both children and adults with hearing impairment. CIs aim at restoring hearing by means of electrical stimulation of the auditory nerve. In comparison to healthy hearing, sounds transmitted *via* CIs are largely limited especially in terms of spectro-temporal cues. Despite these limitations

CIs allow open speech understanding in many patients at least in favorable surroundings (Clark, 2015).

Clinically, the functional outcome of cochlear implantation is determined by a number of measurements. In this respect, speech audiometry plays an outstanding role since it directly addresses verbal communication. Speech audiometry is typically assessed both in quiet and against background noise considering different speech materials such as phonemes, single words or sentences giving comprehensive information on speech recognition abilities (Boisvert et al., 2020).

While numerous outcome measures are established, the challenges listeners face in everyday communication are not fully addressed by common audiometric tests, since speech understanding in detrimental acoustic situations (e.g., in presence of people talking nearby, environmental sounds, or reverberation) relies not only on peripheral hearing. Amongst others, different cognitive abilities might play a role. In two meta-analyses, the role of working memory capacity (WMC) has been highlighted for listeners with healthy hearing or mild to moderate hearing loss (Akeroyd, 2008; Dryden et al., 2017). Additionally, processing speed and subdomains of executive mechanisms such as inhibitory control may play a role (Dryden et al., 2017). Less is known about the influence of cognitive factors on CI-mediated speech recognition. However, recent work has shown associations of speech recognition in CI users and in NH listeners presented with spectrally degraded (i.e., noise-vocoded) speech with WMC (Kaandorp et al., 2017), non-verbal reasoning (Mattingly et al., 2018; Moberly et al., 2018), inhibition control (Zhan et al., 2020) and processing speed as well as executive functions (Rosemann et al., 2017; Völter et al., 2021).

The role of cognition for understanding speech in adverse situations is advocated by the Ease of Language Understanding (ELU)-model (Rönnberg et al., 2013). This model postulates that understanding speech is an implicit, automated and seemingly effortless process as long as the input signal is clear. Any distortions (noise, signal processing, hearing loss) are detrimental to this process consequently activating an explicit processing putting strain on cognitive resources (i.e., working memory). Due to the generally limited capacity (Kahneman, 1973) this constitutes a cognitive load that makes performing a specific task effortful. The ELU model posits that the degree of explicit processing needed for speech understanding is positively related to effort (Rönnberg et al., 2019). Hence, it may be assumed that adverse conditions yield higher listening effort (LE) than favorable conditions despite a listener may exhibit reasonable speech recognition in both situations.

Though there is no uniform definition (McGarrigle et al., 2014) the concept of LE is increasingly common in hearing research. A number of publications define this term in the sense of the attention and cognitive resources required to understand speech (Hick and Tharpe, 2002; Fraser et al., 2010; Picou et al., 2011). The FUEL-model ("Framework for Understanding Effortful Listening," Pichora-Fuller et al., 2016) sets a somewhat broader focus and defines listening effort as "the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task that involves listening." Moreover, it proposes that LE depends on factors such as input-related demands (noise, signal processing, hearing loss), cognitive factors, and motivation, making it a complex multifactorial construct. According to this concept, two individuals can exhibit similar speech recognition but may differ tremendously in the effort accomplished to achieve this performance. Amongst others this might be due to differences in their cognitive abilities, as described above. For instance, Desjardins and Doherty (2013) showed that listening effort was significantly negatively correlated with working memory capacity (WMC) and processing speed. Similarly, Stenbäck et al. (2021) found a negative relation between subjectively assessed listening effort and WMC, in line with the view that larger cognitive capacity is associated with less effort. However, it should be noted that such an association was not found in all studies (cf. Rönnberg et al., 2014).

Due to the relevance of effort to daily-life communication (cf. Nachtegaal et al., 2009) and the fact that it may be related to individual factors not necessarily captured by audiometry it is reasonable to assume that determining LE could give important extra information to clinical diagnostics. In recent years there has been much research devoted to assess LE but no "gold standard" or consensus of clinical measurement has been established. Basically, subjective and objective measurements can be applied. Besides questionnaires (cf. Hughes et al., 2019) subjective measurements include rating scales (Rennies et al., 2014; Krueger et al., 2017). Mostly Likert-scales with verbal categorization ranging from "no effort" to "extreme effort" are used. Rating is typically quantified by presenting speech in the presence of a background masker with different signal-to-noise ratios (SNRs). The SNR may be adjusted adaptively in order to cover a wide range of subjectively perceived effort ("ACALES," Krueger et al., 2017).

Objective measurements include physiological tests and behavioral performance measures. The former consider methods such as electroencephalography, pupillometry, assessment of heart rate variability, or skin conductance (e.g., Bernarding et al., 2013; Holube et al., 2016; Mackersie and Calderon-Moultrie, 2016; Winn et al., 2018) and reflect the mental load associated with listening in adverse conditions. Behavioral measures of LE are based on the fact that cognitive capacity is limited (Kahneman, 1973) and that understanding speech in detrimental situations results in fewer resources available for other tasks, in line with both the ELU- and the FUEL-model. From this rationale, listening effort can be objectively measured *via* the dual-task paradigm (Gagné et al., 2017). In this paradigm, listeners perform a primary speech recognition task simultaneously with a secondary task. In comparison to performing the tasks alone (i.e., single-task) it is assumed that the depletion of resources due to demanding listening shows up in a decline in the secondary task when keeping speech recognition stable. While the primary task typically involves presenting words or sentences in noise, a large number of secondary tasks have been proposed, both within the same modality as the primary task (i.e., auditory) as well as a different modality (i.e., tactile, visual). Moreover, secondary tasks differ largely in terms of their complexity, a factor that might affect the sensitivity of the measurements (Picou and Ricketts, 2014). Frequently, reaction times are captured for the secondary task assuming that the depletion of cognitive capacity associated with effortful listening slows down processing speed. Using these different methods it has been well established that adverse acoustic conditions, typically reflected by decreased signal to noise ratio (SNR), increase both subjectively and objectively assessed listening effort.

In the framework of clinical studies such measures of LE have also been used to assess specific signal processing strategies in cochlear implants (e.g., Stronks et al., 2020) or to compare the effort of CI recipients and NH listeners. For instance, Perreau et al. (2017) applied subjective ratings and a dual-task paradigm while modifying the SNR of the speech presented. Compared to the CI users they found a larger reduction of LE in the NH listeners when the SNR was improved suggesting that effort is different in these two groups. A meta-analysis by Ohlenforst et al. (2017) revealed that hearing-impaired persons show larger LE than normal-hearing subjects, but clear evidence was only given for electroencephalographic measures. However, Alhanbali et al. (2017) applied a subjective effort assessment scale based on six questions and also showed that hearing-impaired subjects, including groups of hearing aid and CI users, revealed significantly higher perceived effort than a control group of normal-hearing listeners. Similarly, Hughes et al. (2018) stated that hearing impaired individuals may need to invest more effort to participate successfully in everyday listening situations despite provision of hearing aids (HAs) and cochlear implants (CIs). Thus, at least during daily verbal communication hearing impaired listeners may show additional demands, even when provided with appropriate rehabilitative technologies. In terms of CIs the rationale is that the limitations in spectro-temporal processing yield extra demands that cannot readily be compensated for. Limited transmission of acoustic details in combination with adverse environments calls for cognitive compensation of speech perception constraints (Başkent et al., 2016). In line with this, pupillometry data by Winn et al. (2015) showed an impact of auditory spectral resolution beyond speech recognition when normal-hearing listeners were subjected to noise-vocoded speech aiming at simulating the spectro-temporal limits of cochlear implants. In contrast, it has also been shown in adolescent CI and NH listeners that both groups show similar effort once performance has been balanced (Hughes and Galvin, 2013). Thus, it remains unclear if and under what circumstances hearing impairment and CI-mediated listening yield increased effort.

In the present study, we compared listening effort in experienced CI recipients and age-matched NH listeners while considering potential influential factors, such as cognitive abilities. Based on the outcome of this comparison we discuss implications for the use as a clinical outcome measure. To this end two measurements of listening effort previously applied in clinical studies, a subjective scaling procedure as well as an objective test (dual-task paradigm), were applied and compared. Importantly, we estimated speech recognition functions for both listener groups and contrasted listening effort at similar performance levels. We hypothesized that listening effort is higher for CI users than NH listeners due to the degraded signal conveyed by the CI and that individual cognitive abilities of the participants mediate listening effort.

# MATERIALS AND METHODS

## Participants

Two groups ($n$ = 14 each) of cochlear implant users with at least 2 years of CI experience and age-matched NH listeners were recruited for participation in this study. The CI recipients used different devices and all except three were fitted bilaterally. Detailed information is given in **Table 1**. The NH listeners had pure tone thresholds $\leq$ 25 dB HL across all frequencies of 125 to 4,000 Hz and were chosen to match the age of the CI users as closely as possible. The NH group involved 11 female and 3 male listeners. The maximum age difference between each CI-NH pair was 3 years. Thus, both groups did not differ regarding their age (61.9 $\pm$ 12.4 years for CI and 62.4 $\pm$ 12.6 years for NH). All participants were native German speakers and had normal or corrected-to-normal vision. Prior to the experiment they were given detailed information about the study and informed consent was obtained. Participants were reimbursed with € 10,-/h. The study protocol was approved by the local ethics committee.

## Cognitive Tests

As described in the introduction several cognitive functions are potentially related to recognizing speech in adverse conditions as well as the associated listening effort. From the variety of these functions we selected three that are suited for clinical assessment based on appropriate neuropsychological tests.

Working memory capacity (WMC) was assessed by the German version of the Reading Span Test (RST; Carroll et al., 2015). This test presents sentences in blocks of 2 to 6 stimuli on a computer screen. The task is to read each sentence aloud and to judge immediately after presentation whether the sentence is meaningful or not. At the end of each block, the participant is asked to recall the first or last word of the sentences. The percentage of correctly recalled words across all trials is determined and taken as an indicator of WMC.

Furthermore, processing speed and executive functions were assessed by the Trail Making Test (TMT; Reitan, 1958). The TMT consists of two subsets: In TMT-A the participants are asked to connect digits shown on a sheet of paper in ascending numerical order. In TMT-B the participants are required to alternate between digits and letters in ascending order. In both tests the time to complete the task is assessed. TMT-A and TMT-B are thought to give an indication of different cognitive abilities (Sanchez-Cubillo et al., 2009). Specifically, TMT-A is associated with processing speed and TMT-B is assumed to reflect executive control and cognitive flexibility.

## Speech Recognition in Noise

The Oldenburg sentence test (OLSA, Wagener et al., 1999) was used for assessing speech recognition in noise. This test is frequently applied in clinical routine in Germany. The OLSA is a matrix test presenting sentences composed of five words (name – verb – numeral – adjective – object) and ten possible

**TABLE 1** | Characteristics of the cochlear implant recipients.

| ID | Gender | Age (years) | Fitting | Hearing loss right ear since (years) | Hearing loss left ear since (years) | Experience right CI (years) | Experience left CI (years) | Word recognition score (%) | Right CI type | Left CI type |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | m | 47 | bilateral | 22 | 22 | 4 | 4 | 90 | Cochlear® N6 | Cochlear® N6 |
| 2 | m | 67 | bilateral | childhood | childhood | 14 | 19 | 45 | Advanced Bionics, Auria (SAS) | Advanced Bionics, Auria (HiRes-P) |
| 3 | f | 74 | bilateral | 40 | 35 | 15 | 8 | 85 | Advanced Bionics, Naída CI Q90 | Advanced Bionics, Naída CI Q90 |
| 4 | m | 83 | unilateral | childhood | childhood | – | 16 | 70 | – | MED-EL, Sonnet |
| 5 | f | 68 | unilateral | na | na | 10 | – | 55 | MED-EL, Opus2 | – |
| 6 | f | 59 | bilateral | 41 | 41 | 6 | 4 | 80 | MED-EL, Opus2 | MED-EL, Opus2 |
| 7 | m | 71 | bilateral | childhood | childhood | 2 | 11 | 90 | MED-EL, Sonnet | MED-EL, Opus2 |
| 8 | f | 57 | bilateral | 32 | 32 | 7 | 3 | 75 | Cochlear®, CP810 | Cochlear®, CP810 |
| 9 | m | 60 | unilateral | 18 | 18 | 16 | – | 60 | Cochlear®, CP910 | – |
| 10 | f | 61 | bilateral | 41 | 41 | 19 | 8 | 85 | Advanced Bionics, Harmony | Advanced Bionics, Harmony |
| 11 | f | 52 | bilateral | 47 | 47 | 5 | 6 | 90 | MED-EL, Opus2 | MED-EL, Opus2 |
| 12 | f | 78 | bilateral | na | na | 11 | 4 | 55 | MED-EL, Sonnet | MED-EL, Sonnet |
| 13 | m | 39 | bilateral | 18 | 18 | 3 | 3 | 90 | Cochlear®, CP910 | Cochlear®, CP910 |
| 14 | f | 51 | bilateral | 26 | 26 | 5 | 5 | 90 | MED-EL, Opus2 | MED-EL, Synchrony |

alternatives for each word position. Sentences are syntactically correct but semantically unpredictable thus allowing repeated testing. The male voice of the OLSA was used. The masker was a test-specific stationary noise ("olnoise") generated by multiple random superpositions of the sentences of the OLSA corpus. These stimuli were used for examining speech recognition as well as for the subjective and objective assessment of listening effort.

An important aspect of the study was to estimate the speech recognition function of the listeners. To this end the 50% speech recognition threshold (SRT50) as well as the slope of the recognition function were assessed concurrently following the procedure suggested by Brand and Kollmeier (2002). This procedure adaptively tracks correct response probabilities of 19 and 81% in an interleaved fashion during one test list of 30 trials. Initial step-width for varying the SNR is 1.5 dB and reduced after each reversal yielding a final step-width of 0.25 dB to stabilize presentation levels near the targets. The SNRs presented after five reversals of the adaptive procedures were averaged to determine the two targets. Based on the estimates of 19 and 81% intelligibility the SRT50 and the slope are determined. The noise was fixed at 65 dB SPL and the speech level was varied depending on the subject's responses, who were asked to repeat back as many words as possible. The stimuli were routed from a PC to an audiometer (Siemens Unity) and sent to a free-field loudspeaker (Events Electronics, Australia) placed at a distance of 1.2 m from the listener's head located at 0°. In order to test reliability and to improve accuracy of the psychometric function this measurement was performed twice using test lists of 30 sentences each. Based on the individual threshold and the slope derived from the measurements a logistic function

$$y = \frac{100}{1 + e^{-\frac{(x - SRT50)}{s}}} \quad (1)$$

was fitted, with SRT50 as the SNR associated with 50% intelligibility, s as the slope at 50% intelligibility, x as the level in dB SNR, and y as the percentage of words correctly understood.

This function was used to estimate the SNR associated with 80% intelligibility that was applied for assessing objective listening effort in the dual-task paradigm.

## Objective Listening Effort

Listening effort was measured with a dual-task paradigm, consisting of a listening task (primary task) and a visual reaction time task (secondary task). This behavioral paradigm determines performance and thus assesses effort objectively. The primary task was to recognize speech at a performance level of 80%. Choosing this level represented a situation where performance was relatively high but still demanding and followed the recommendation to avoid unfavorable SNRs with dual-task paradigms in order to prevent cognitive overload (Wu et al., 2016). Since it was difficult to target exactly 80% for each listener a range of ±8% was allowed. This range of maximum 16% was not expected to have a significant influence on listening effort, in line with the psychometric functions of dual-task paradigms given in Wu et al. (2016). If this criterion was not met the SNR was readjusted and the measurement was repeated until the desired range was reached. This was necessary in seven cases.

The secondary task was a visual reaction time task. We chose a simple task in order to maximize the possibility that the primary task was unaffected. A white fixation cross (visual angle = 5.2°) was shown on a black background *via* a computer screen (ELO TouchSystems) placed about 65 cm in front of the subject. The cross briefly disappeared at arbitrary points in time during the presentation of half of the sentences of a test list at random intervals. The task of the participants was to react as fast as possible by pressing the left mouse button.

The dual-task paradigm was administered using a custom made computer program, implemented using the *Presentation* software (Neurobehavioral Systems Inc., Berkeley, CA, United States). Sentences of the OLSA masked by olnoise were sent *via* an external sound-card (Hammerfall DSP Multiface II) to the loudspeaker as described with the speech recognition procedure.

The primary and secondary tasks were measured separately *via* single-task, as well as in a combined fashion *via* dual-task. The single-task measurements served as baselines. Here, the participants were asked to concentrate on the task at hand (speech recognition or visual reaction) and to ignore the other task (visual reaction or speech recognition). In the dual-task instructions were given to the participants to optimize performance in the primary task (speech recognition) but also to perform the secondary task as accurately and fast as possible (cf. Gagné et al., 2017). In each condition test lists of 40 sentences were presented. Because in the secondary task only half of the stimuli were randomly associated with the fixation cross disappearing, twenty reaction time scores were recorded across a test list. Since reaction times typically show a non-normal distribution a median score was calculated across a test list for each participant.

In order to derive a measure of listening effort, proportional dual-task costs (pDTC%) indicating the load on the secondary task (Fraser et al., 2010) was calculated by the formula

$$pDTC\% = 100 * (\text{Secondary (dual task)} - \text{Secondary (single task)})/\text{Secondary (single task)} \quad (2)$$

Likewise, proportional dual-task costs can be calculated for the primary task. However, as intended and shown below, the primary task was not critically affected by combining both tasks.

## Subjective Listening Effort

Listening effort was measured subjectively with the "Adaptive Categorical Listening Effort Scaling" (ACALES, Krueger et al., 2017). Similar to the speech recognition test this method presents sentences of the OLSA masked by olnoise at various SNRs. Again, stimuli were sent *via* an external sound-card (Hammerfall DSP Mulitface II) to the loudspeaker as described above. With each SNR two sentences were presented allowing a reasonable amount of time to listen to the stimuli. After each presentation the listeners were asked to answer the question "How much effort does it require for you to follow the speaker?" (German: "Wie anstrengend ist es für Sie, dem Sprecher zu folgen?"). LE is assessed on a categorical scale showing the labels "no effort," "very little effort," "little effort," "moderate effort," "considerable effort," "very much effort," "extreme effort," displayed on a touch screen (ELO TouchSystems). These labels corresponded to 1, 3, 5, 7, 9, 11, and 13 effort scale categorical units (ESCU), respectively. There were six unlabelled intermediate steps and an additional category ("only noise") that allowed for a response when no speech was perceived. The ESCU-values were not shown to the subjects.

The adaptive procedure consists of three phases (details in Krueger et al., 2017). In the first phase the boundaries for "no effort" and "extreme effort" are searched by varying the SNR by a step-width of 3 dB. These boundaries are used for the second phase that presents five intermediate SNRs to estimate the five categories "very little effort," "little effort," "moderate effort," "considerable effort," and "very much effort." By linear interpolation of these data the SNRs for "no" and "extreme effort" are re-estimated and SNRs for the five intermediate categories are re-calculated and presented to the listeners in a third phase. Based on these presentations LE estimates were determined by linear regression for each listener.

## Procedures

After giving informed consent the participants first completed the cognitive tests beginning with the TMT and followed by the RST. Speech recognition testing and listening effort experiments were run in a sound treated booth (l:4 × w:3 × h:2 m). Speech recognition in noise was preceded with a training phase presenting two tests lists of 20 sentences each in quiet in order to familiarize the participants with the OLSA-material. After that, subjective listening effort was assessed. Prior to the measurement a short training by presenting 20 stimuli at different SNRs was performed in order to familiarize the participants with the method and the rating scale. Finally, the dual-task paradigm was performed in order to assess listening effort objectively. Again, prior to conducting the actual experiment a training phase familiarized the subjects with the tasks and the stimuli provided. Testing was accomplished in a single visit lasting approximately 3 h, including several individual breaks.

## Statistical Analyses

Kolmogorov-Smirnov-Tests and visual inspection of Q-Q-plots revealed that the data were mostly normally distributed. In that case, repeated measures analyses of variance (rmANOVA) were performed. If the assumption of sphericity was violated, Greenhouse-Geisser corrections were used. The association of listening effort outcome and cognitive tests was assessed by correlation analysis. In the case of non-normally distributed data non-parametric tests were used as documented in the results section. IBM SPSS v. 25 was used for all calculations.

# RESULTS

## Speech Recognition in Noise

Individual speech recognition functions were estimated based on the procedure described above. Test and retest were highly correlated (Pearson's coefficients $r_p = 0.95$ for SRT50, $r_p = 0.83$ for slope, both $p < 0.001$) and thus outcome was averaged across the two measurements. Hence, estimates of the functions were based on 60 sentences in total.

**Figure 1** shows the individual functions of both listener groups. As expected, speech recognition was clearly better for the NH than the CI listeners. A rmANOVA on SNR with target speech recognition (50%, 80%) as within-subjects variable and listener group (CI, NH) as between-subjects variable

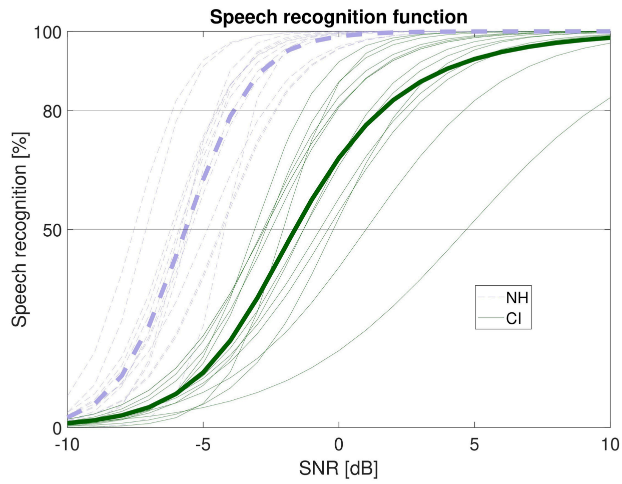


**FIGURE 1 |** Estimated speech recognition functions for the CI recipients (green) and the NH listeners (violet). Thin lines show individual functions, bold lines show the group mean. SNR = signal-to-noise ratio.

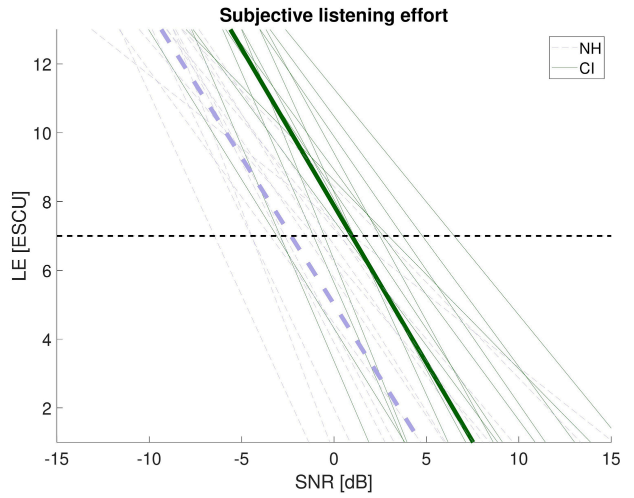


**FIGURE 2 |** Listening effort (LE) assessed by ACALES as a function of the signal-to-noise ratio (SNR) in CI recipients (green) and NH listeners (violet). Bold lines represent the mean. ESCU = Effort Scaling Category Units. The dashed horizontal line shows the value of 7 ESCU ("moderate effort").

revealed a significant main effect of target speech recognition ($F_{1,26}$ = 338.96, $p < 0.001$, $\eta_p{}^2 = 0.93$), a significant main effect of group ($F_{1,26}$ = 49.52, $p < 0.001$, $\eta_p{}^2 = 0.66$) and a speech recognition by group interaction ($F_{1,26}$ = 23.65, $p < 0.001$, $\eta_p{}^2 = 0.48$). The mean SNR associated with 50% recognition was −5.6 ± 0.9 dB SNR in the NH listeners and −1.2 ± 2.0 dB SNR in the CI listeners. The estimation of 80% speech recognition revealed a SNR of −4.1 ± 1.1 dB SNR in the NH listeners and +1.4 ± 2.7 dB SNR in the CI users. Follow-up of the significant interaction revealed that the difference in SNR between 50 and 80% target speech recognition was significantly larger in the CI listeners than in the NH listeners ($t_{1,26}$ = 4.86, $p < 0.001$). This shows that the slope of the function was typically steeper in NH than CI listeners.

Further analyses revealed that SRT50 and slope were significantly correlated in the CI recipients ($r_p = -0.71$, $p = 0.005$) but not in the NH listeners ($r_p = -0.26$, $p = 0.372$) which might be attributed to the relatively low variability in speech recognition in the latter group. However, for the CI users it could be approximated that the slope changed by about 1% per dB/SRT, which might be helpful for estimating speech recognition at different SNRs.

## Subjective Listening Effort – ACALES

For each participant listening effort outcome was fitted by a simple linear regression function which is suitable when using a stationary test-specific masker (i.e., olnoise, see Krueger et al., 2017). **Figure 2** shows the results for both listener groups in dependence of the SNR applied. While the slope of the functions is similar for NH and CI listeners ($t_{1,26}$ = 0.11, $p$ = 0.91) the value for LE7 as the proxy for moderate effort (i.e., 7 ESCU) is significantly different ($t_{1,26}$ = 3.2, $p$ = 0.004). As shown in the figure both group-mean functions are shifted by about 3 dB SNR given the same ESCU-value or about 3 ESCU given the same SNR.

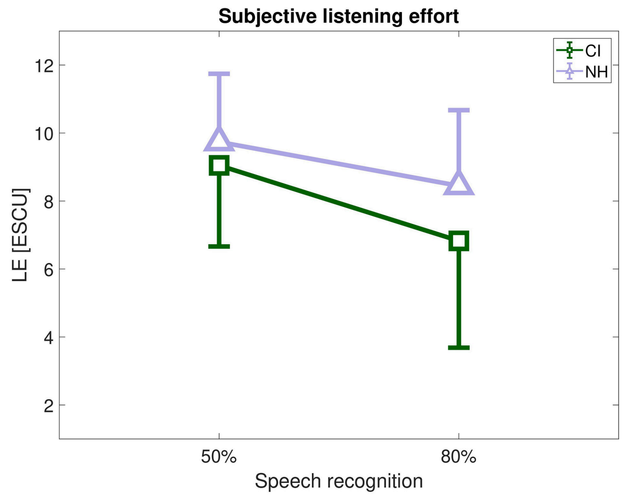


**FIGURE 3 |** Listening effort (LE) in CI recipients (green) and NH listeners (violet) at SNRs associated with 50 and 80% speech recognition. ESCU = Effort Scaling Categorical Units.

By using the estimated speech recognition functions (see **Figure 1**), individual LE-scores for 50% and 80% speech recognition, denoted as LE50 and LE80 were determined (see **Figure 3**). Mean listening effort was about 9–10 ESCU ("considerable" to "very much effort") for 50% speech recognition and around 7–9 ESCU ("moderate" to "considerable effort") for 80 % recognition. A rmANOVA with speech recognition (50%, 80%) as within-subjects variable and listener group (CI, NH) as between-subjects variable revealed a significant main effect of speech recognition ($F_{1,26}$ = 130.35, $p < 0.001$, $\eta_p{}^2 = 0.83$) and a significant speech recognition by group interaction ($F_{1,26}$ = 11.81, $p$ = 0.002, $\eta_p{}^2 = 0.31$). The interaction mirrored the impression

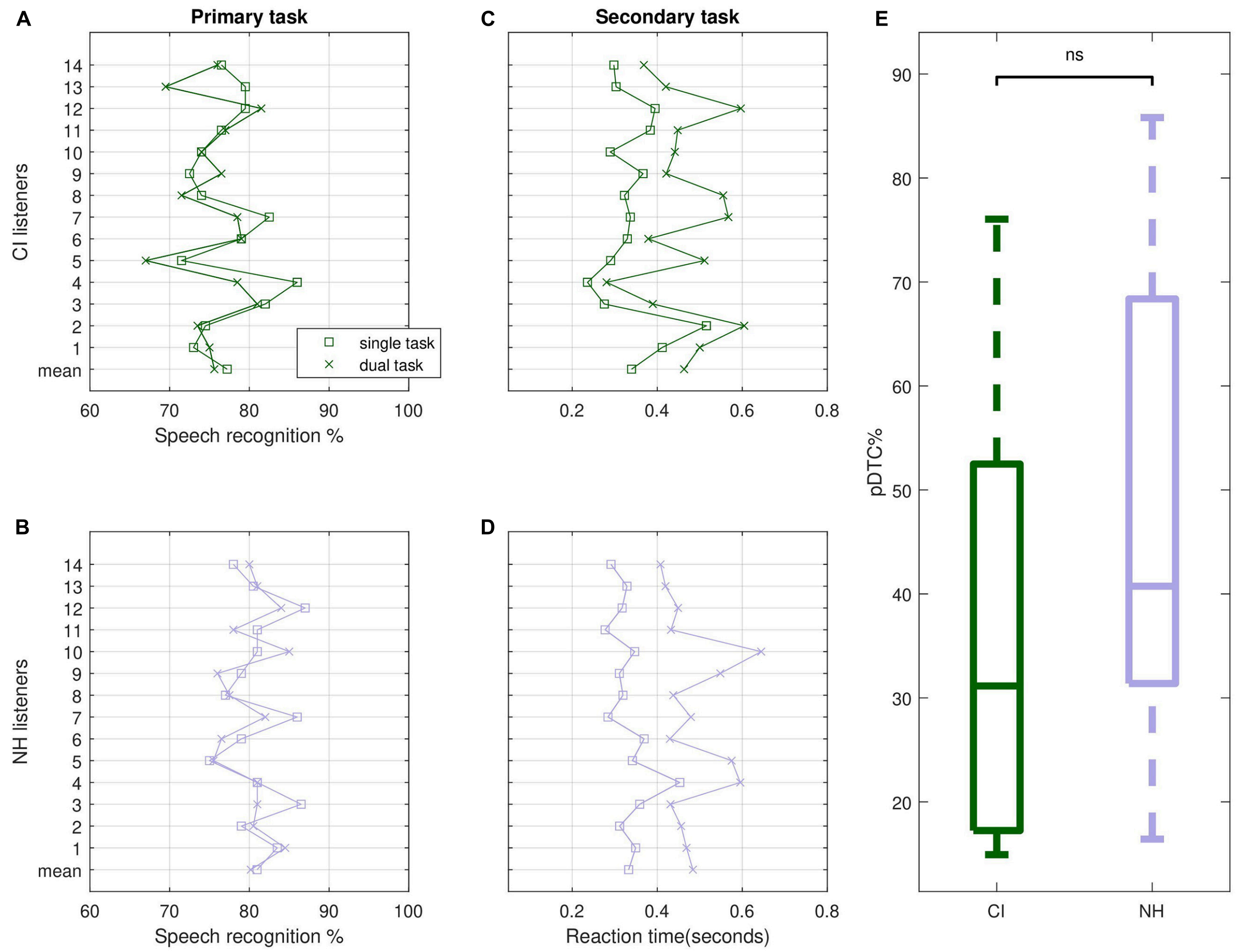


**FIGURE 4 |** Outcome of the dual-task paradigm. **(A,B)** Primary task (speech recognition) in CI and NH listeners, **(C,D)** secondary task (reaction time) in CI and NH listeners, **(E)** proportional dual-tasks costs for the secondary task. CI, cochlear implant recipients; NH, normal-hearing listeners. Squares = single-task, crosses = dual-task.

of **Figure 3** that CI and NH listeners rated LE relatively similar at 50% but NH perceived somewhat higher LE at 80%. However, *post hoc* tests rendered this group difference insignificant ($t_{1,26} = -1.94$, $p = 0.064$).

## Objective Listening Effort

The primary task of the dual-task paradigm showed that the goal to target a speech recognition of about 80% was met in both listener groups (**Figures 4A,B**). Apart from single cases (CI05, CI13) this held for both, performance in the single-task and the dual-task condition. A rmANOVA with task (single, dual) as within-subjects variable and listener group (CI, NH) as between-subjects variable revealed a significant main effect of task ($F_{1,26} = 4.85$, $p = 0.037$, $\eta_p{}^2 = 0.16$) and a significant main effect of group ($F_{1,26} = 9.56$, $p = 0.005$, $\eta_p{}^2 = 0.27$). Speech recognition was higher in the single-task than in the dual-task (79.1 ± 4.2% vs. 77.7 ± 4.1%) and in the NH compared to the CI listeners (80.5 ± 3.1% vs. 76.7 ± 4.2%). Since our aim was to capture LE by dual-task costs in the secondary task, as outlined above, a performance difference in the primary task could be critical. However, despite statistical significance this difference did not influence outcome, as proportional dual-task costs for the primary task amounted to only about 2%, when calculated in analogy to formula (2). Furthermore, based on the psychometric functions of dual-task paradigms given in Wu et al. (2016), it is assumed that the small performance difference between CI and NH listeners of about 4% in the primary task did not affect costs in the secondary task.

Reaction times in the secondary task were highly variable and appear to show a clear delay in all cases, when assessed in the dual-task (see **Figures 4C,D**). Subjecting the data to a rmANOVA with task (single, dual) as within-subjects variable and listener group (CI, NH) as between-subjects variable revealed a significant main effect of task ($F_{1,26} = 110.30$, $p < 0.001$, $\eta_p{}^2 = 0.81$). Corresponding proportional dual-task costs are shown in **Figure 4E**. Comparing the costs between both listener groups revealed no significant difference ($U$-Test, $z = 1.15$, $p = 0.27$).

## Listening Effort and Cognitive Functions

The listeners of both groups were assessed in terms of their processing speed, cognitive flexibility, and working memory capacity using the Trail making Test (Version A and B) and the German version of the Reading span test (Carroll et al., 2015). Outcome is given in **Table 2**.

The CI recipients revealed two outliers for the outcome of TMT-B. Groups were compared using $U$-tests that did not show any significant difference for the tests applied (all $p > 0.45$).

**Table 3** shows the Spearman's correlation coefficients of the neuropsychological test outcome and the listening effort measures across both groups. LE80 was taken as the proxy for subjective listening effort and pDTC% as the proxy for objective listening effort, both reflecting the demands associated with 80% speech recognition. Age of the listeners was also considered as it is assumed to be associated with cognition. Indeed, TMT-A, TMT-B, and RST showed a significant correlation with age. As expected, older listeners were slower in both Trail making tests A and B and showed worse recall in the WMC test. Furthermore, the three cognitive metrics were significantly correlated demonstrating that they do not represent completely unrelated domains. This also held when the two outliers (TMT-B) were removed.

However, both LE80 and pDTC% did not reveal any significant correlation with the outcome of the neuropsychological tests nor with age. Moreover, the two LE outcome measures were not significantly associated with each other suggesting that they tap into different dimensions of the listening effort construct.

## DISCUSSION

The aim of this study was to compare measures of listening effort and speech recognition in CI recipients and age-matched normal-hearing listeners and to gain information for potential clinical applications and implications. To this end, methods that potentially may be used in clinical assessments were considered. We hypothesized that CI recipients show increased effort due to the limitations of CI-mediated sound transmission. Alternatively, it could be suspected that CI and NH listeners exhibit comparable listening effort once speech recognition performance of the participants is balanced. Furthermore, we expected that individual cognitive abilities may mediate listening effort.

**TABLE 2 |** Outcome of the neuropsychological tests regarding processing speed (TMT-A), executive control (TMT-B), and working memory capacity (RST).

| | | Minimum | Maximum | Median | Mean | Std.-Dev. |
|---|---|---|---|---|---|---|
| CI-recipients ($n$ = 14) | TMT-A [s] | 16 | 76 | 32.8 | 39.9 | 19.1 |
| | TMT-B [s] | 36 | 393 | 72.0 | 105.7 | 105.3 |
| | RST [%] | 11.1 | 66.6 | 42.5 | 39.6 | 14.5 |
| NH-listeners ($n$ = 14) | TMT-A [s] | 21 | 77 | 35.9 | 40.7 | 18.6 |
| | TMT-B [s] | 21 | 128 | 68.6 | 76.2 | 31.6 |
| | RST [%] | 24.1 | 61.1 | 41.6 | 40.8 | 8.8 |

**TABLE 3 |** Spearman's rank correlations and significance levels (two-tailed) of the outcome of the listening effort measures (LE80, pDTC%) and neuropsychological tests (TMT-A, TMT-B, RST), as well as age, Asterisk depict significant correlations.

| | | LE80 | pDTC% | TMT-A | TMT-B | RST | Age |
|---|---|---|---|---|---|---|---|
| LE80 | $r_{sp}$ | 1.000 | −0.100 | −0.104 | −0.021 | 0.105 | −0.332 |
| | p | | 0.612 | 0.598 | 0.916 | 0.595 | 0.085 |
| pDTC% | $r_{sp}$ | | 1.000 | 0.169 | −0.221 | −0.151 | 0.225 |
| | p | | | 0.389 | 0.258 | 0.443 | 0.250 |
| TMT-A | $r_{sp}$ | | | 1.000 | 0.805** | −0.556** | 0.695** |
| | p | | | | 0.000 | 0.002 | 0.000 |
| TMT-B | $r_{sp}$ | | | | 1.000 | −0.529** | 0.556** |
| | p | | | | | 0.004 | 0.002 |
| RST | $r_{sp}$ | | | | | 1.000 | −0.583** |
| | p | | | | | | 0.001 |
| Age | $r_{sp}$ | | | | | | 1.000 |
| | p | | | | | | |

## Speech Recognition in Noise

Paramount to our examination of LE was that individual speech recognition performance in noise was known. Therefore, speech recognition functions were estimated. As expected, the functions revealed better performance in the NH than the CI listeners. This manifested in both, speech recognition thresholds and slope of the functions. The latter was shallower for the CI users, that is, they did not benefit from increasing the SNR to the same amount as the NH listeners. This confirms results by MacPherson and Akeroyd (2014) who found a trend of decreasing slope with increasing hearing impairment. Moreover, Sobon et al. (2019) reported a significant negative correlation between slope and SRT in NH listeners, but only for a two-talker speech masker. In general, one single SRT (typically associated with 50% recognition) may thus not fully acknowledge speech recognition problems over a wider range of SNRs. However, the decrease in slope of about 1% per dB SRT in the CI listeners might be helpful for estimating performance at different SNRs. From a practical background this indicates that listeners with poor SRTs may gain less from any change in SNR offered by the signal processing in hearing aids or cochlear implants (cf. MacPherson and Akeroyd, 2014).

Thus, from a clinical perspective it seems advisable to determine not only the SRT but also the slope. According to Brand and Kollmeier (2002) this is basically feasible by using a test list of at least 30 sentences. These "extra costs" appear to be acceptable in the framework of clinical routine where typically at least 20 sentences (in the case of matrix sentences after training) are used. Hence, the proposed method of assessing both, SNR and slope might give valuable extra information, especially when trying to relate other measures (such as listening effort outcome) to individual speech recognition, as will be discussed in the following.

## Subjective Listening Effort

Assessing subjective listening effort, e.g., *via* ACALES, appears to be easily applicable in clinical routine. Methodological demands and time consumption are moderate. Determining listening effort including a brief orientation phase takes about 6–8 min. Clear instructions provided, the procedure appears to be a good representation of what it intends to measure. Thus, it may be assumed that it reveals high face validity. In terms of reliability, Krueger et al. (2017) reported a high intraclass correlation above 0.9 when using the olnoise masker. However, since each listener might have his or her own subjective effort construct, it is not entirely clear whether individual outcome mirrors the same underlying dimensions and whether results can be directly compared with each other. Potentially as a consequence, estimated LE showed high interindividual variability in both, CI and NH listeners.

ACALES assesses subjective LE relative to adaptive variations in SNR. This has the advantage that the entire range from "no" to "extreme effort" is covered. When relating LE to SNR there was indeed a significant difference between the listener groups. NH participants showed about 3 ESCU lower listening effort ratings for the same SNR. However, this comparison might be misleading if the association of SNR with speech recognition is unknown. In the present study this association could be estimated based on the individual psychometric functions of the participants. When similar performance was assumed, both groups did not differ significantly with respect to LE. Nevertheless, a significant speech recognition by group interaction was found reflecting that CI users exhibited lower effort at 80% performance relative to the NH listeners (see **Figure 3**). Despite *post hoc* tests rendered this difference insignificant ($p = 0.064$) it deserves further discussion. In general, it is not exactly clear which factors contribute to the individual estimation of listening effort. However, it is conceivable that the subjectively perceived level of the speech signal relative to the noise is taken into account. Due to the shallower speech recognition function in the CI recipients SNR improved more than in the NH listeners when targeting 80% recognition instead of 50%. This would be in line with the observation of a larger decrease in ESCU in the CI users than in the NH listeners.

## Objective Listening Effort

Assessing listening effort objectively typically assumes high methodological and technical demands, as it is the case with electroencephalography, pupillometry, electrodermal activity or heart rate variability (cf. Bernarding et al., 2013; Holube et al., 2016; Mackersie and Calderon-Moultrie, 2016; Winn et al., 2018). In terms of behavioral measurements an alternative are dual-task paradigms which consist of a primary and a secondary task. The reliability of dual-task paradigms appears to be satisfactory, as Picou and Ricketts (2014) reported a test-retest correlation of 0.79 when using a "simple" secondary task comparable to that of the present study. However, in contrast to the subjective estimation it has to be taken into account, that time consumption is about three times higher (20–25 min), since three test lists have to be administered successively.

The primary task was recognizing speech at a SNR associated with 80% performance. This level was considered in order to make the task demanding but to avoid low performance that might be detrimental to these paradigms due to cognitive overload (see Wu et al., 2016) and also to better reflect everyday listening where intelligibility is mostly high or approaches ceiling. The results presented above confirm that 80% recognition was related to substantial subjective effort. Ideally, the performance in the primary task is constant across all test conditions since the proxy for LE is expected to emerge in the secondary task. Our statistical analysis of the primary task outcome revealed significant condition- and group-effects. However, these differences were in a range of only a few percent and are assumed not to play a critical role regarding the task load. Thus, the goal of keeping the primary task relatively constant across listeners and tasks and capturing the effect of dual-task costs in the secondary task appears to be met.

Significant proportional dual-task costs reflecting listening effort could be shown in the secondary task. Costs showed large interindividual differences but both listener groups did not differ significantly which also supports the idea that LE is similar when comparable speech recognition is assumed. In this study, we applied a simple reaction-time based secondary task providing 20 RTs across one test list. This is a relatively low number potentially affecting the quality of the outcome. However, when assessing split-half reliability (i.e., trials 1–10 vs. trials 11–20) the correlation was high ($r_p \geq 0.8$, $p < 0.001$) for both the primary and the dual-task. Moreover, calculating the average RTs across groups revealed very similar results, regardless of whether the first or second half of trials was used.

The choice of the secondary is generally critical. On the one hand it must not be too demanding in order to avoid performance shifts across tasks ("trade-off") and on the other hand it must not be too simple because of the then missing task load. In our case, the choice of a relatively simple visual paradigm appears to be appropriate, since the primary task outcome remained largely stable and load effects clearly surfaced in the secondary task. However, a secondary task requiring more processing depth might be even more sensitive. Picou and Ricketts (2014) compared different secondary tasks, involving a simple and a complex visual reaction time paradigm as well as a semantic paradigm, requiring to understand the word presented in the primary task. Whereas the visual reaction time paradigms both reflected the effect of background noise on LE the latter showed larger effects sizes and thus might better reflect more subtle mechanisms of effort. Further, Hsu et al. (2020) modified the depth of processing in the secondary task by asking children with CIs to judge whether the word presented was an animal (lower level of semantic processing) or whether the animal was dangerous (higher level). However, both secondary tasks appeared to reflect the increased load associated with adding noise (i.e., SNR of 3 dB) relative to listening in quiet.

## Association With Cognition and Age

Three cognitive domains (processing speed, executive control and working memory capacity) potentially associated with recognizing speech and listening effort in adverse acoustic situations were considered. No significant group effects were found. This does not support the expectation that hearing impaired persons show lower cognitive abilities compared to age-matched normal-hearing listeners (e.g., Lin et al., 2013). However, as expected, the outcome of the cognitive tests was correlated with age. Nevertheless, none of the cognitive metrics nor age was significantly associated with subjectively (i.e., LE80) or objectively (pDTC%) assessed listening effort. This finding was unexpected, given the theoretical rationale that effortful listening depletes limited cognitive resources, as proposed by the ELU- and the FUEL-model.

Reports on the correlation of listening effort outcome and cognitive abilities are relatively scarce. Harvey et al. (2017) found that cognitive functions predict listening effort performance during complex tasks in NH listeners. Furthermore, Hua et al. (2014) showed that participants with better cognitive flexibility reported less perceived listening effort. In contrast, Brännström et al. (2018) reported no significant association of measures of WMC and cognitive flexibility with subjectively perceived effort. However, they found a positive correlation of listening effort and inhibitory control. This result was surprising, given that better inhibitory control was associated with higher perceived effort. In listeners provided with cochlear implants, Perreau et al. (2017) also did not find an association of WMC and LE in a dual-task paradigm, but age and LE were correlated. However, as recently pointed out by Francis and Love (2020), LE suggests a complex and possibly "unresolvable" interaction between the commitment of processing resources on the one hand and the response to their deployment on the other hand.

The proxies of subjective and objective listening effort also did not show a significant relation with each other. While some examinations report correlations for single factors (e.g., Holube et al., 2016; Picou and Ricketts, 2018) this is generally in line with a number of studies showing a lack of correspondence between objective and subjective measures of listening effort (e.g., Fraser et al., 2010; Zekveld et al., 2010; Gosselin and Gagné, 2011) and is consistent with the assumption that measures of LE are multidimensional (McGarrigle et al., 2014; Alhanbali et al., 2019). In this context, Lemke and Besser (2016) distinguish between perceived listening effort and processing load. Following this view applying the ACALES procedure addresses perceived LE whereas the dual-task rather reflects the latter. As pointed out by Lemke and Besser (2016), a listening situation might pose high processing load but must not necessarily be perceived as effortful, and vice versa.

## General Discussion

Including listening effort in the assessment of hearing disorders could add a dimension that has not yet been covered by clinical auditory measurements. It could also provide information regarding rehabilitative measures such as the use of specific signal processing or training programs. As discussed above the two measurements of LE applied in this study appear to tap into different domains of the listening effort framework. Both, estimating subjectively perceived listening effort, e.g., *via* ACALES as well as the dual-task paradigm do not require much technological or organizational resources and can be readily integrated using standard speech audiometric material. Another important clinical criterion is the time required to perform the measurement. In this respect the adaptive ACALES procedure appears to be better suited than a the dual-task paradigm, which contains three successive test lists. As a matter of fact, however, extra information can only be gained when additional time is allowed.

Independent from the method used we hypothesized that CI listeners reveal larger LE compared to NH subjects. This was indeed the case when subjective LE was related to the SNR. However, it did not hold when balancing performance across listener groups. This is in line with Hughes and Galvin (2013) who also demonstrated similar LE in adolescent CI recipients and normal-hearing subjects when similar speech recognition was considered.

In general, a close connection of LE and speech recognition performance could be demonstrated. It is tempting to review some recent studies on listening effort in cochlear implant recipients in the light of the present findings. For instance, Perreau et al. (2017) assessed LE subjectively as well as objectively in different groups of CI users and a control group of normal-hearing listeners. The objective measure of LE based on a dual-task paradigm including a reaction-time metric. The authors considered six different SNR-conditions revealing speech recognition scores from around 60% to near perfect. Across the SNR conditions they found larger reduction in LE for the NH compared to the CI listeners. However, considering the steeper psychometric function of normal-hearing listeners as described above, this finding may be explained by their larger increment in performance for a given SNR increase than for the CI recipients.

The effect of a specific sound processing algorithm (i.e., "soft voice") on speech recognition and listening effort was examined by Stronks et al. (2020). The algorithm aims at improving speech recognition at low sound levels by removing internal noise of the device. LE was assessed objectively by pupillometry and subjectively by scaling. Whereas pupillometry did not reveal any effect of the processing algorithm, it had a positive effect on subjectively perceived effort at a speech level of 33 dB SPL (SNR = −5 dB). This was also the level where the algorithm improved speech recognition to the largest extent, giving evidence for a close connection of performance and LE. Consequently, the authors stated that performance measures themselves might be a valid predictor of listening effort. Thus, as outlined in the present study, effects on LE might be difficult to interpret if the underlying speech recognition performance is unknown.

In terms of clinical applications this also raises the question in which cases LE measurements actually provide extra information over commonly used speech audiometry. Given the typical time limitations in clinical assessments this question is crucial. In the present study it could be shown that at least over a range of 50 to 80% speech recognition a close connection between performance and LE can be found. Moreover, no differences in LE between CI and NH listeners were found once performance was accounted for. Most of the studies that assumed larger LE for listeners with hearing loss referred to everyday listening, that is, situations typically including positive SNRs and high speech intelligibility (Smeds et al., 2015). In this regard the matrix-test reveals limited ecological validity, since the SRTs determined are often in a negative SNR-range. The functions presented in **Figure 1** show that all NH listeners show perfect speech recognition at positive SNRs whereas some of the CI users approach asymptote at higher signal-to-noise ratios. Thus, it is plausible that CI recipients show increased effort at these ecologically more valid SNRs. This is also confirmed when looking at the association of LE and signal-to-noise ratio depicted in **Figure 2**. This suggests that assessing LE might provide more information when it is not assessed at 50 or 80% speech intelligibility but rather when speech recognition is near or at ceiling. Here, LE stills shows considerable inter-individual variability though effort is lower than at intermediate speech recognition. However, sustained effort could still yield substantial fatigue (Hornsby et al., 2016). Thus, even differences in low effort may have practical consequences for everyday life. Moreover, particular signal processing schemes such as noise reduction algorithms may not affect intelligibility but could be efficient regarding the reduction of effort.

## CONCLUSION

There is increasing need for measures that capture effects of speech perception beyond speech audiometry. This is due to advances in rehabilitation technology and the fact that challenges in everyday communication are not fully covered by common audiometric tests. One construct that promises valuable information is the effort associated with recognizing speech. Here, we compared the results of two potentially clinically suited methods in groups of listeners with cochlear implants and normal hearing. Both measurements revealed highly variable results that were not significantly related to different cognitive abilities or age. Moreover, the outcome of the two tests was not correlated with each other suggesting that they tap into different dimensions of the effort construct. Also, we did not find any significant difference in LE between the two listener groups, once performance was equalized by adjusting individual SNRs. A limitation of the study was that the sample size of the two groups was small and thus might not have been sufficient to detect small effects. However, LE was strongly correlated with speech recognition at least when assessed subjectively. Thus, when examining LE it is highly recommended to take possible performance differences into account, e.g., by determining both, SRT and slope of the psychometric function. Due to the strong association of effort and speech recognition it is suggested that LE-assessment is more instructive when performance is near or at ceiling. Here, the large inter-individual variability in listening effort could give information beyond speech audiometry and would also consider the range of more ecological signal-to-noise ratios.

## 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 the Ethics Committee of the Medical Faculty, University of Cologne. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

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

## FUNDING

This study received funding from the Marga-und-Walter-Boll Foundation (Reference 210-10-13) as well as in part from Cochlear Europe (Reference IIR-684). ACALES was kindly provided by Hoerzentrum Oldenburg GmbH. The funders were 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

We are grateful to the reviewers for their valuable comments on an earlier version of the manuscript.

## REFERENCES

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47(Suppl. 2), S53–S71. doi: 10.1080/14992020802301142

Alhanbali, S., Dawes, P., Lloyd, S., and Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear.* 38, e39–e48. doi: 10.1097/AUD.0000000000000361

Alhanbali, S., Dawes, P., Millman, R. E., and Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear Hear.* 40, 1084–1097. doi: 10.1097/AUD.0000000000000697

Başkent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., et al. (2016). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: how and to what degree can it be achieved? *Trends Hear.* 20, 1–16. doi: 10.1177/2331216516670279

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., and Corona-Strauss, F. I. (2013). Neural correlates of listening effort related factors: influence of age and hearing impairment. *Brain Res. Bull.* 91, 21–30. doi: 10.1016/j.brainresbull.2012.11.005

Boisvert, I., Reis, M., Au, A., Cowan, R., and Dowell, R. C. (2020). Cochlear implantation outcomes in adults: a scoping review. *PLoS One* 15:e0232421. doi: 10.1371/journal.pone.0232421

Brand, T., and Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J. Acoust. Soc. Am.* 111, 2801–2810.

Brännström, K. J., Karlsson, E., Waechter, S., and Kastberg, T. (2018). Listening effort: order effects and core executive functions. *J. Am. Acad. Audiol.* 29, 734–747. doi: 10.3766/jaaa.17024

Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., and Meister, H. (2015). Development of a German reading span test with dual task design for application in cognitive hearing research. *Int. J. Audiol.* 54, 136–141. doi: 10.3109/14992027.2014.952458

Clark, G. M. (2015). The multi-channel cochlear implant: multi-disciplinary development of electrical stimulation of the cochlea and the resulting clinical benefit. *Hear Res.* 322, 4–13. doi: 10.1016/j.heares.2014.08.002

Desjardins, J. L., and Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear Hear.* 34, 261–272. doi: 10.1097/AUD.0b013e31826d0ba4

Dryden, A., Allen, H. A., Henshaw, H., and Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic literature review and meta-analysis. *Trends Hear.* 21, 1–21. doi: 10.1177/2331216517744675

Francis, A. L., and Love, J. (2020). Listening effort: are we measuring cognition or affect, or both? Wiley interdisciplinary reviews. *Cogn. Sci.* 11:e1514. doi: 10.1002/wcs.1514

Fraser, S., Gagné, J. P., Alepins, M., and Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: the effects of providing visual speech cues. *J. Speech Lang. Hear. Res.* 53, 18–33. doi: 10.1044/1092-4388(2009/08-0140)

Gagné, J. P., Besser, J., and Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm. *Trends Hear.* 21:2331216516687287. doi: 10.1177/2331216516687287

Gosselin, P. A., and Gagné, J. P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *Int. J. Audiol.* 50, 786–792. doi: 10.3109/14992027.2011.599870

Harvey, J., von Hapsburg, D., and Seeman, S. (2017). Cognitive function predicts listening effort performance during complex tasks in normally aging adults. *Noise Health* 19, 254–262. doi: 10.4103/nah.NAH_83_16

Hick, C. B., and Tharpe, A. M. (2002). Listening effort and fatigue in school-age children with and without hearing loss. *J. Speech Lang. Hear. Res.* 45, 573–584. doi: 10.1044/1092-4388(2002/046)

Holube, I., Haeder, K., Imbery, C., and Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends Hear.* 20:2331216516667734. doi: 10.1177/2331216516667734

Hornsby, B. W., Naylor, G., and Bess, F. H. (2016). A taxonomy of fatigue concepts and their relation to hearing loss. *Ear Hear.* 37(Suppl 1), 136S–144S. doi: 10.1097/AUD.0000000000000289

Hsu, B. C., Vanpoucke, F., Langereis, M., Dierckx, A., and van Wieringen, A. (2020). Age-related changes in listening effort for children and teenagers with normal hearing and cochlear implants. *Ear Hear.* 42, 506–519. doi: 10.1097/AUD.0000000000000953

Hua, H., Emilsson, M., Ellis, R., Widén, S., Möller, C., and Lyxell, B. (2014). Cognitive skills and the effect of noise on perceived effort in employees with aided hearing impairment and normal hearing. *Noise Health* 16, 79–88. doi: 10.4103/1463-1741.132085

Hughes, K. C., and Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int.* 14, 121–129. doi: 10.1179/1754762812Y.0000000009

Hughes, S. E., Hutchings, H. A., Rapport, F. L., McMahon, C. M., and Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: a grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear.* 39, 922–934. doi: 10.1097/AUD.0000000000000553

Hughes, S. E., Rapport, F., Watkins, A., Boisvert, I., McMahon, C. M., and Hutchings, H. A. (2019). Study protocol for the validation of a new patient-reported outcome measure (PROM) of listening effort in cochlear implantation: the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI). *BMJ Open* 9:e028881. doi: 10.1136/bmjopen-2018-028881

Kaandorp, M. W., Smits, C., Merkus, P., Festen, J. M., and Goverts, S. T. (2017). Lexical-access ability and cognitive predictors of speech recognition in noise in adult cochlear implant users. *Trends Hear.* 21:2331216517743887. doi: 10.1177/2331216517743887

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Krueger, M., Schulte, M., Brand, T., and Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. *J. Acoust. Soc. Am.* 141:4680. doi: 10.1121/1.4986938

Lemke, U., and Besser, J. (2016). Cognitive load and listening effort: concepts and age-related considerations. *Ear Hear.* 37(Suppl. 1), 77S–84S. doi: 10.1097/AUD.0000000000000304

Lin, F. R., Yaffe, K., Xia, J., Xue, Q. L., Harris, T. B., Purchase-Helzner, E., et al. (2013). Hearing loss and cognitive decline in older adults. *JAMA Intern. Med.* 173, 293–299. doi: 10.1001/jamainternmed.2013.1868

Mackersie, C. L., and Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: heart rate variability and skin conductance. *Ear Hear.* 37(Suppl. 1), 118S–125S. doi: 10.1097/AUD.0000000000000305

MacPherson, A., and Akeroyd, M. A. (2014). Variations in the slope of the psychometric functions for speech intelligibility: a systematic survey. *Trends Hear.* 18:2331216514537722. doi: 10.1177/2331216514537722

Mattingly, J. K., Castellanos, I., and Moberly, A. C. (2018). Nonverbal reasoning as a contributor to sentence recognition outcomes in adults with cochlear implants. *Otol. Neurotol.* 39, e956–e963. doi: 10.1097/MAO.0000000000001998

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., et al. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int. J. Audiol.* 53, 433–440. doi: 10.3109/14992027.2014.890296

Moberly, A. C., Harris, M. S., Boyce, L., Vasil, K., Wucinich, T., Pisoni, D. B., et al. (2018). Relating quality of life to outcomes and predictors in adult cochlear implant users: are we measuring the right things? *Laryngoscope* 128, 959–966. doi: 10.1002/lary.26791

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., and Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. *Int. J. Audiol.* 48, 684–691.

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., et al. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: a systematic review. *Ear Hear.* 38, 267–281. doi: 10.1097/AUD.0000000000000396

Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., and Corts, D. (2017). Listening effort measured in adults with normal hearing and cochlear implants. *J. Am. Acad. Audiol.* 28, 685–697. doi: 10.3766/jaaa.16014

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the Framework for Understanding Effortful Listening (FUEL). *Ear Hear.* 37(Suppl. 1), 5S–27S. doi: 10.1097/AUD.0000000000000312

Picou, E. M., and Ricketts, T. A. (2014). The effect of changing the secondary task in dual-task paradigms for measuring listening effort. *Ear Hear.* 35, 611–622. doi: 10.1097/AUD.0000000000000055

Picou, E. M., and Ricketts, T. A. (2018). The relationship between speech recognition, behavioural listening effort, and subjective ratings. *Int. J. Audiol.* 57, 457–467. doi: 10.1080/14992027.2018.1431696

Picou, E. M., Ricketts, T. A., and Hornsby, B. W. (2011). Visual cues and listening effort: individual variability. *J. Speech Lang. Hear. Res.* 54, 1416–1430. doi: 10.1044/1092-4388(2011/10-0154)

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Percept. Mot. Skills* 8, 271–276. doi: 10.2466/pms.1958.8.3.271

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *J. Acoust. Soc. Am.* 136, 2642–2653. doi: 10.1121/1.4897398

Rönnberg, J., Holmer, E., and Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int. J. Audiol.* 58, 247–261. doi: 10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rönnberg, N., Rudner, M., Lunner, T., and Stenfelt, S. (2014). Assessing listening effort by measuring short-term memory storage and processing of speech in noise. *Speech Lang. Hear.* 17, 123–132. doi: 10.1179/2050572813Y.0000000033

Rosemann, S., Gießing, C., Özyurt, J., Carroll, R., Puschmann, S., and Thiel, C. M. (2017). The contribution of cognitive factors to individual differences in understanding noise-vocoded speech in young and older adults. *Front. Hum. Neurosci.* 11:294. doi: 10.3389/fnhum.2017.00294

Sanchez-Cubillo, I., Perianez, J. A., Adrover-Roig, D., Rodriguez- Sanchez, J. M., Rios-Lago, M., Tirapu, J., et al. (2009). Construct validity of the Trail Making Test: role of task switching, working memory, inhibition/interference control, and visuomotor abilities. *J. Int. Neuropsychol. Soc.* 15, 438–450.

Smeds, K., Wolters, F., and Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J. Am. Acad. Audiol.* 26, 183–196. doi: 10.3766/jaaa.26.2.7

Sobon, K. A., Taleb, N. M., Buss, E., Grose, J. H., and Calandruccio, L. (2019). Psychometric function slope for speech-in-noise and speech-in-speech: effects of development and aging. *J. Acoust. Soc. Am.* 145, EL284–E290. doi: 10.1121/1.5097377

Stenbäck, V., Marsja, E., Hällgren, M., Lyxell, B., and Larsby, B. (2021). The contribution of age, working memory capacity, and inhibitory control on speech recognition in noise in young and older adult listeners. *J. Speech Lang. Hear. Res.* 64, 4513–4523. doi: 10.1044/2021_JSLHR-20-00251

Stronks, H. C., Apperloo, E., Koning, R., Briaire, J. J., and Frijns, J. (2020). SoftVoice improves speech recognition and reduces listening effort in cochlear implant users. *Ear Hear.* 42, 381–392. doi: 10.1097/AUD.0000000000000928

Völter, C., Oberländer, K., Carroll, R., Dazert, S., Lentz, B., Martin, R., et al. (2021). Nonauditory functions in low-performing adult cochlear implant users. *Otol. Neurotol.* 42, e543–e551. doi: 10.1097/MAO.0000000000003033

Wagener, K., Brand, T., and Kollmeier, B. (1999). Entwicklung und evaluation eines satztests in deutscher Sprache II: optimierung des Oldenburger satztests (Development and evaluation of a sentence test in German language II: optimization of the Oldenburg sentence test). *Z. Audiol.* 38, 44–56.

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear.* 36, e153–e165. doi: 10.1097/AUD.0000000000000145

Winn, M. B., Wendt, D., Koelewijn, T., and Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear.* 22:2331216518800869. doi: 10.1177/2331216518800869

Wu, Y. H., Stangl, E., Zhang, X., Perkins, J., and Eilers, E. (2016). Psychometric functions of dual-task paradigms for measuring listening effort. *Ear Hear.* 37, 660–670. doi: 10.1097/AUD.0000000000000335

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear.* 31, 480–490. doi: 10.1097/AUD.0b013e3181d4f251

Zhan, K. Y., Lewis, J. H., Vasil, K. J., Tamati, T. N., Harris, M. S., Pisoni, D. B., et al. (2020). Cognitive functions in adults receiving cochlear implants: predictors of speech recognition and changes after implantation. *Otol. Neurotol.* 41, e322–e329. doi: 10.1097/MAO.0000000000002544

**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 in Neuroscience

ORIGINAL RESEARCH
published: 08 March 2022
doi: 10.3389/fnins.2022.787153

# Evaluating Spatial Hearing Using a Dual-Task Approach in a Virtual-Acoustics Environment

**Marina Salorio-Corbetto[1,2,3]*, Ben Williges[1], Wiebke Lamping[1], Lorenzo Picinali[2†] and Deborah Vickers[1†]**

[1] SOUND Laboratory, Cambridge Hearing Group, Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom, [2] Audio Experience Design, Dyson School of Design Engineering, Imperial College London, London, United Kingdom, [3] Wolfson College, Cambridge, United Kingdom

OPEN ACCESS

**Edited by:**
Isabelle Boisvert,
The University of Sydney, Australia

**Reviewed by:**
Yonghee Oh,
University of Florida, United States
Karyn Galvin,
The University of Melbourne, Australia

***Correspondence:**
Marina Salorio-Corbetto
ms878@cam.ac.uk

†These authors share senior authorship

**Specialty section:**
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

**Received:** 30 September 2021
**Accepted:** 11 January 2022
**Published:** 08 March 2022

**Citation:**
Salorio-Corbetto M, Williges B, Lamping W, Picinali L and Vickers D (2022) Evaluating Spatial Hearing Using a Dual-Task Approach in a Virtual-Acoustics Environment. Front. Neurosci. 16:787153. doi: 10.3389/fnins.2022.787153

Spatial hearing is critical for communication in everyday sound-rich environments. It is important to gain an understanding of how well users of bilateral hearing devices function in these conditions. The purpose of this work was to evaluate a Virtual Acoustics (VA) version of the Spatial Speech in Noise (SSiN) test, the SSiN-VA. This implementation uses relatively inexpensive equipment and can be performed outside the clinic, allowing for regular monitoring of spatial-hearing performance. The SSiN-VA simultaneously assesses speech discrimination and relative localization with changing source locations in the presence of noise. The use of simultaneous tasks increases the cognitive load to better represent the difficulties faced by listeners in noisy real-world environments. Current clinical assessments may require costly equipment which has a large footprint. Consequently, spatial-hearing assessments may not be conducted at all. Additionally, as patients take greater control of their healthcare outcomes and a greater number of clinical appointments are conducted remotely, outcome measures that allow patients to carry out assessments at home are becoming more relevant. The SSiN-VA was implemented using the 3D Tune-In Toolkit, simulating seven loudspeaker locations spaced at 30° intervals with azimuths between −90° and +90°, and rendered for headphone playback using the binaural spatialization technique. Twelve normal-hearing participants were assessed to evaluate if SSiN-VA produced patterns of responses for relative localization and speech discrimination as a function of azimuth similar to those previously obtained using loudspeaker arrays. Additionally, the effect of the signal-to-noise ratio (SNR), the direction of the shift from target to reference, and the target phonetic contrast on performance were investigated. SSiN-VA led to similar patterns of performance as a function of spatial location compared to loudspeaker setups for both relative localization and speech discrimination. Performance for relative localization was significantly better at the highest SNR than at the lowest SNR tested, and a target shift to the right was associated with an increased likelihood of a correct response. For word discrimination, there was an interaction between SNR and word group. Overall, these outcomes support the use of virtual audio for speech discrimination and relative localization testing in noise.

**Keywords: spatial hearing, bilateral cochlear implants, binaural performance, dual task, remote testing, speech in noise, functional testing**

## INTRODUCTION

Speech testing plays a crucial role in the assessment of hearing function, including the evaluation of outcomes with hearing devices such as hearing aids or cochlear implants. There are a variety of speech testing materials, ranging from closed-set words to open-set sentence formats presented in quiet or in noise. A limitation of many of the current speech tests is that the listening skills assessed by the task are often different from those required for everyday communication environments. For instance, many tests were designed with a fixed speech source location and a co-located masker. This is the case for the Automated McCormick Toy Test (Summerfield et al., 1994), the Speech Reception in Noise Test [SPRINT (Cord et al., 1992; Brungart et al., 2017)], the Words in Noise Test [WIN (Wilson et al., 2007)], the Quick Speech in Noise Test (Killion et al., 2004), the Bamford-Kowal-Bench Speech in Noise Test [BKB-SIN (Etymotic Research Inc., 2005)], and the AzBio sentences lists (Spahr et al., 2012), among others. Conversely, in everyday environments, social interaction typically involves several talkers and sources of noise scattered around the listener. For communication to be successful, the listener needs to identify where the talker of interest is located and switch their focus rapidly as conversation unfolds.

It has been shown that tests using multi-talker babble or a single competing talker are sensitive to hearing status (Phatak et al., 2019), particularly if they target the use of “dip listening” – the ability to detect a signal in a fluctuating masker, which depends on accurate encoding of temporal fine structure information (Moore, 2014) – or quantify spatial release from masking (SRM), – the improvement in the detection of a signal in background noise arising from the spatial separation of the target signal and the background (Bronkhorst, 2000; Litovsky, 2012). Although tests of SRM can provide important diagnostic information about spatial hearing, they require that a speech-identification task is performed repeatedly as the location of the speech or noise is varied (Litovsky, 2012; Bizley et al., 2015), which makes them time-consuming. Moreover, unlike in real communication environments, the speech sources used in SRM testing have a fixed location at either the front or the sides of the listener.

In light of these limitations, Bizley et al. (2015) developed the Spatial Speech in Noise (SSiN) test as a tool for simultaneously evaluating SRM, localization, and speech discrimination performance in a background of multi-talker babble noise. The SSiN uses speech signals appropriate for adults and children, targeting discrimination of specific phonetic contrasts: complex vowel, simple vowel, initial consonant, and final consonant. These contrasts are represented by groups of four words each, so that testing is done in a closed-set discrimination paradigm. For example, for complex vowel, the four words within the group are “pale,” “peel,” “pile,” and “pool” (**Table 1**). The test features a speech-discrimination task in which the listener needs to report back two words within the group, the *reference* word and the *target* word, which are presented in succession. Simultaneously, listeners engage in a relative-localization task requiring that they report whether the target word was presented from the right or from the left of the location of the reference word. This dual-task approach and the use of multi-talker babble as a background noise were chosen to represent the challenges of listening in a complex communication environment. Unlike in typical SRM test setups, the locations of the sources of noise can be varied within the test session or across versions, as will be described later, and the speech-source locations change from trial to trial. Further work was carried out by Ahnood (2017) and Parmar et al. (2018) in order to adapt the SSiN for use with people with hearing aids and cochlear implants. Modifications included increasing the spacing between loudspeakers and restricting the number of noise sources during the task in order to make the task feasible to listeners with hearing loss.

**TABLE 1 |** Word groups by target phonetic contrast and word items within each group.

| Target phonetic contrast | Word items | | | |
|---|---|---|---|---|
| Complex vowel (Vc) | Pale | Pool | Pile | Peel |
| Simple vowel (Vs) | Hoot | Heat | Heart | Hurt |
| Initial consonant (Ci) | Chalk | Talk | Fork | Stork |
| Final consonant (Cf) | Cheat | Cheese | Cheap | Cheek |

While the SSiN is an efficient way to simultaneously assess speech discrimination and relative localization, a key element of its setup is the use of a loudspeaker array simulating a AB-York Crescent of Sound (Kitterick et al., 2011) to deliver the stimuli. Implementations based on loudspeaker arrays are costly both in terms of material and spatial requirements. Additionally, this test require a face-to-face visit. The constraints imposed by the current COVID-19 pandemic have accelerated the development and adoption of tele-audiology practices (Ayas et al., 2020; Saunders and Roughley, 2020; Parmar et al., 2021). Remote-health applications that enable users to complete diagnostic tests and submit them to their clinical departments are very much in demand.

One solution to the spatial and economic costs of multi-loudspeaker arrays, and a response to the demand for remote clinical testing, is to use binaural spatialization to render complex listening environments which can be delivered to the listener using a pair of headphones (Cuevas-Rodríguez et al., 2019). The most common implementation of the binaural spatialization technique is based on the Head-Related Transfer Function (HRTF), which embeds localization cues – such as Interaural Time Differences (ITDs), Interaural Level Differences (ILDs), and spectral cues – within the original sound stimuli (Blauert, 1997). The capabilities of binaural spatialization for generating complex soundscapes are virtually unlimited in terms of the number and location of sound sources and their relative distance, as well as the characteristics of the simulated room (e.g., large halls, small studios, etc.) (Cuevas-Rodríguez et al., 2019). Additionally, the requirements for playback devices are simple. It is possible to use a standard pair of headphones connected to the computer audio output, or wireless streaming for hearing devices, for the delivery of the sounds. For these reasons, binaural spatialization could have a major impact when applied to audiological testing (Pausch and Fels, 2019; Keidser et al., 2020).

In spite of their great potential, there are only a few examples of clinical-audiology applications that use binaural spatialization. For instance, the Listening in Spatialized Noise Sentences test [LiSN (Cameron and Dillon, 2008)] assesses stream segregation by adaptively estimating the speech reception threshold (SRT) for sentences in a competing background. Pitch cues (identity of the target talker vs. identity of the talker/s in the background) and/or spatial cues (co-located or ±90° azimuth separation) are varied during the test. Another example of a virtual-audio clinical-audiology application is the Auditory Speech Sound Evaluation (A§E®, © P.J. Govaerts, Antwerp, Belgium) ILD Sound Localization Test, which uses two loudspeakers to simulate thirteen spatial locations by introducing ILDs on a 4000-Hz narrow band of noise (Otoconsult Helpdesk, 2021).

The SSiN has some advantages over these examples. First, it uses smaller intervals than the LiSN for spatial-discrimination testing. Second, it uses more meaningful stimuli than the A§E ILD sound localization test, with a wider frequency range. However, it still has the limitation of requiring a complex set-up. A virtual-audio version of the SSiN, the SSiN-VA, was implemented. SSiN-VA retains the SSiN capabilities for testing speech discrimination and relative localization while minimizing any space and equipment requirements. The aim of this project was to determine whether the patterns of responses as a function of spatial location obtained with the SSiN-VA are similar to those previously obtained with the SSiN test for normal-hearing listeners. Our hypothesis was that the SSiN-VA leads to patterns of word discrimination and relative localization similar to those previously obtained with the SSiN. We investigated this hypothesis by conducting the SSiN-VA with 12 normal-hearing participants. Our predictions, based on existing SSiN data, were that, for relative localization, performance would deteriorate at the lateral locations relative to performance at the midline and that, for word discrimination, performance would be best at the lateral locations and lowest around the midline. Additionally, we hypothesized that performance for both word discrimination and relative localization increases the higher the signal-to-noise ratio (SNR) at which the test was conducted. This is a novel aspect of this work, as the influence of SNR on performance has not been assessed for the SSiN. Further, assessing the effect of SNR on SSiN-VA outcomes was of interest as there was no knowledge of the difficulty of the task for the participants, given the virtual setup. Lastly, we hypothesized that:

For relative localization:

a) Performance is similar across word groups as their overall audibility is equivalent.
b) The effect of SNR is similar across spatial locations of the speech sources (no interaction between SNR and the spatial location of the speech).
c) There is no direction bias in the responses of the participants. In other words, a correct response is equally likely for trials where the target shifts to the left and trials where the target shifts to the right. If a direction bias were found, it would be investigated whether the bias is present regardless of the location of the speech sources (no interaction between spatial location and direction of the shift) and for all SNRs (no interaction between SNR and direction of the shift).

For word discrimination:

a) Word discrimination varies across word groups as the spectral cues required for correct discrimination of each group may differ in their vulnerability to being masked by the babble noise.
b) The order of the words (i.e., whether a word is target or reference) does not influence performance.
c) Changing the location of the speech source (azimuth) might lead to changes in the SNR at each ear, and this affects performance for different word groups unevenly (interaction between word group and azimuth).
d) Increasing the SNR improves word discrimination regardless of azimuth (no significant interaction between SNR and azimuth).
e) The effect of SNR is stronger for word groups where the speech sounds key to the phonetic contrast is lowest in level, such as the initial-consonant and the final-consonant groups (interaction between SNR and word group).

Finally, the patterns of responses for SSiN-VA were graphically compared to those obtained with a dataset obtained with a loudspeaker spatial setup similar to the one simulated here.

## MATERIALS AND METHODS

### Participants

Twelve participants (eight female, four male) with normal hearing were tested. Their median age was 26 years, ranging from 21 to 52 years (mean 28.58, SD = 8.73). All participants had air-conduction hearing thresholds for octave frequencies in the range 250–8000 Hz equal to or better than 20 dB HL, or a maximum of one frequency with threshold equal to 25 dB HL, as measured with an Interacoustics Affinity audiometer in a quiet room.

The experiment designs for preliminary work were reviewed and approved by the Joint Research Compliance Office at Imperial College (Ref. 19IC5073). The main experiments were reviewed and approved by the Cambridge Psychology Research Ethics Committee (Ref. 2019.093).

### Implementation of the Spatial Speech in Noise-Virtual Acoustics Test

As mentioned, the test used here, the SSiN-VA, was an adaptation of the SSiN Test developed by Bizley et al. (2015). The basic structure of the SSiN-VA was the same as that for the SSiN: In each trial, a reference word was presented from one of the loudspeaker locations. The reference word was followed by a target word, which was presented from an adjacent loudspeaker location. Simultaneously, sixteen-male-talker babble (Huckvale, 1989) was presented to the listener. The listener was required to provide a speech-discrimination response by selecting the reference word and the target word from four buttons, each corresponding to one word. Additionally, the listener provided a relative-localization response by indicating in which direction the location of the target word shifted relative to the reference word. This was done by using one of two buttons labeled "left" and "right," respectively. The test used speech material taken from a closed-set children's speech discrimination test, the Chear Auditory Perception Test [CAPT (Marriage et al., 2011; Vickers et al., 2018)]. Each word belonged to one of four closed-set groups. Each group contained four words which possessed a particular type of phonetic contrast; the words differed in a complex vowel (pale, pool, pile, or peel), a simple vowel (hoot, heat, heart, or hurt), initial consonant (chalk, talk, fork, or stork), or a final consonant (cheat, cheese, cheap, or cheek), as explained above and shown in **Table 1**.

There are some differences between the original SSiN test and the SSiN-VA other than the use of headphones instead of loudspeakers. These changes were introduced in order to make the test more feasible for users with hearing loss. For the SSiN-VA, the number of spatial locations was reduced from from 13 to seven, and the spacing between the sources used in a given trial was doubled compared to the first implementation of the SSiN. Thus, the SSiN-VA used azimuths corresponding to −90°, −60°, −30°, 0°, 60°, 30°, and 90° (**Figure 1**). Intervals of 15°, such as those used in the SSiN, may be too small for people with hearing loss to be able to perform the relative-localization task above chance (Ahnood, 2017; Parmar et al., 2018). In addition, for the SSiN-VA, the babble was constantly delivered from four spatial locations: −60°, −30°, 60°, and 30°, instead of simultaneously from all loudspeaker locations as in Bizley et al. (2015). Delivering the babble from all loudspeaker locations may make the test too challenging for people with hearing loss (Ahnood, 2017; Parmar et al., 2018). This prompted other researchers to reduce the number of babble source for this task. For instance, Ahnood (2017) delivered the noise from either the −60° and −30° locations, or the 30° and 60° locations (i.e., from two sources at a time, alternative from the right or the left hemispace). Note that these are the same locations that are used here, but with only two sources within the same hemispace active at a given trial. Although this maximizes the amount of SRM that can be obtained by increasing the distance between the sources of the speech and the sources of the babble for some of the trials, the number of trials needs to be large enough to be able to accurately represent both noise-location configurations. Because the ultimate objective of the SSiN-VA is for it to be used clinically, it was decided that the noise would be consistently delivered from the reduced set of locations used by Ahnood (2017) but in a simultaneous manner from all four sources. This made it possible to reduce the number of trials collected and simplify the study design.

The SSiN-VA prototype was created using MaxMSP (Cycling'74, 2021) and the 3D Tune-In Toolkit (Cuevas-Rodríguez et al., 2019), specifically its Virtual Studio Technology (VST) implementation (Picinali et al., 2019a). One instance of the VST plugin, loaded with a KEMAR mannequin HRTF from the SADIE II database (Armstrong et al., 2018), was used to spatialize each individual virtual loudspeaker. No room-acoustics simulation was performed, therefore the spatialization was fully anechoic. The ITDs were individualized for each participant by measuring their head circumference and inputting it in the 3D Tune-In Toolkit rendering engine. More details about how the spatialization was performed accounting for this measure can be found in Cuevas-Rodríguez et al. (2019). A head tracker was used to update the locations of the virtual loudspeakers every 12 ms in order to ensure that the rendered virtual sound field was anchored to the surrounding space rather than rotating with the head of the listener, as it happens in real environments when listening to audio reproduced from an array of loudspeakers. More details about the HRTF interpolation processes implemented in order to simulate the movement of the head relative to the virtual loudspeakers can be found in Cuevas-Rodríguez et al. (2019). The importance of accounting for head movements when reproducing binaural signals is well documented in the literature (Begault et al., 2001). Even though participants were instructed to look at the front, it was decided to implement head tracking in order to make the experience as close as possible to the original SSiN test, as it is known that small head movements can have a dramatic impact on SRM, significantly improving performance (Grange and Culling, 2016). The timing of the sequence playback was arranged as follows: babble onset at 0 s, first word at 0.5 s, second word at 2 s, and babble offset at 3.5 s. Each word was approximately 1 s long.

## Equipment for the Spatial Speech in Noise-Virtual Acoustics Test

Stimuli were presented using a MacBook Pro via Sennheiser HD-600 headphones. An Apple iPhone 5 was used as head tracker, and was mounted on the top of the headphones. The app GyrOSC was used to send the head-tracking data through WiFi to the MacBook Pro via Open Sound Control (OSC, Freed, 1997).

## Procedure for the Spatial Speech in Noise-Virtual Acoustics Test

### Calibration and Presentation Level

Scaling of the word stimuli was performed by calculating the root-mean-square (RMS) levels of the steady-state portions of the vowels within the words as identified independently by two researchers using Praat software, version 6.1.14 (Boersma and Weenink, 1992). Where discrepancies occurred, a third researcher was involved in discussion. A MATLAB (The MathWorks Inc., 2019) script was used to adjust the RMS levels of the word stimuli so that the levels of the steady-state portions of the vowels were equal across words. Appropriate scaling of the background noise was performed taking into account the number of sources in order to achieve the same RMS level as for the word stimuli. Presentation levels were calibrated using a Tektronix MDO3024 Mixed Domain Oscilloscope using the headphone sensitivity data to calculate the voltage required to deliver the sound level required for the calibration noise. Because the desired playback level for the word stimuli was 52 dB SPL when the SNR was specified as 0, the RMS level of the calibration noise was set 20 dB above the RMS level of the word stimuli, and the calibration noise was played back at 72 dB SPL. The level of the multi-talker babble was kept at 52 dB SPL throughout the task (consistent with Bizley et al.'s implementation). The level of the

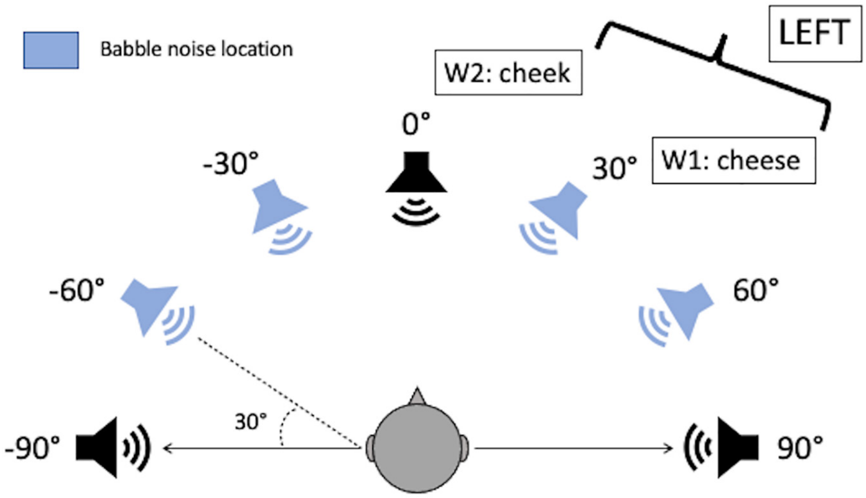


**FIGURE 1 |** Diagram of the simulated loudspeaker locations around the listener showing seven sound sources separated at 30° azimuth intervals. An example of a trial is given where the sources of the reference word (W1) and the target word (W2) are represented on the diagram, and the correct relative-localization answer is given. The diagram shown to the participant was identical except that no indication of the spatial location of the babble or the azimuth were given.

speech was varied across runs in order to collect data for three different SNRs as described below.

### Speech Reception Threshold Determination

First, the speech reception threshold (SRT) for each participant was determined by presenting words from a simulated azimuth of 0° (i.e., from the frontal location) while the multi-talker babble was simultaneously delivered from −60°, −30°, 60°, and 30°. In each trial, one word, randomly selected from the sixteen used in the test, was presented. Participants were shown the discrimination group that contained the correct word. For example, if the word presented was "peel," participants were shown the words "pool," "pile," "pale," and "peel." Participants were required to click on the word that they heard. The SNR was 0 dB in the first trial (calibration details given in section "Calibration and Presentation Level"), and was adaptively varied in 2-dB steps following a one-up one-down technique in subsequent trials. The test stopped after eight reversals were obtained. The SRT was calculated as the average SNR at the last six reversals.

Once the SRT was measured, the SNRs for three conditions were calculated: (1) the individually measured speech recognition threshold (SRT), determined as explained above; (2) SRT + 3 dB; (3) SRT + 6 dB. This was to address the aim of determining whether there was an effect of SNR on performance for both relative localization and word discrimination.

### The Spatial Speech in Noise-Virtual Acoustics Task

Before testing started, participants were introduced to the task by being shown a diagram of the simulated loudspeakers around the listener. They were told that they might have the impression that words came from the locations shown in the diagram, and that in each trial two words would be presented. The second word would come from either the right or the left with respect to the first word. Their task was to report the two words that they heard, in order, and the location of the second word with respect to the first one. Examples were given using cards with words written on them, as shown in **Figure 1**, which were moved around the diagram to simulate possible trials and verify that the participant had understood the task. Next, approximately six trials of the task were presented to the participant at SNR = 0. Participants used the interface shown in **Figure 2** to provide their answers. After this, the participant was asked whether they had understood the task. If they confirmed that they had, testing began.

The three test conditions were administered in three blocks of two runs each. Each run took approximately 12–15 min. Blocks were presented in a pseudorandom order, using all possible orders across the twelve participants who took part. Each run was tested using a 96-trial list. Each trial was characterized by a reference word, a target word, and a simulated location for each of them. There were eight trials for each reference-target location pair. Each of the four word-discrimination groups was played twice from each reference-target location pair. Responses where the participant pressed buttons repeatedly or pressed extra buttons led to inaccurate logging. These trials were discarded prior to analysis. One random additional trial was presented in each run due to an implementation flaw. The responses to these trials were not discarded as the additional run was not associated with a particular condition. The total number of trials presented to each participant was 582 (97 trials * 6 runs). After discarding spurious trials, participants contributed an average of 577 trials each to the analysis phase (range 561–582).

## Analyses

### Statistical Analysis

Statistical analysis was based on trial-by-trial responses. The response variables were binary, with two possible outcomes for "correct response": yes or no. The *glmer* function within the *lme4* package version 1.127 (Bates et al., 2015), in R version 4.1.0

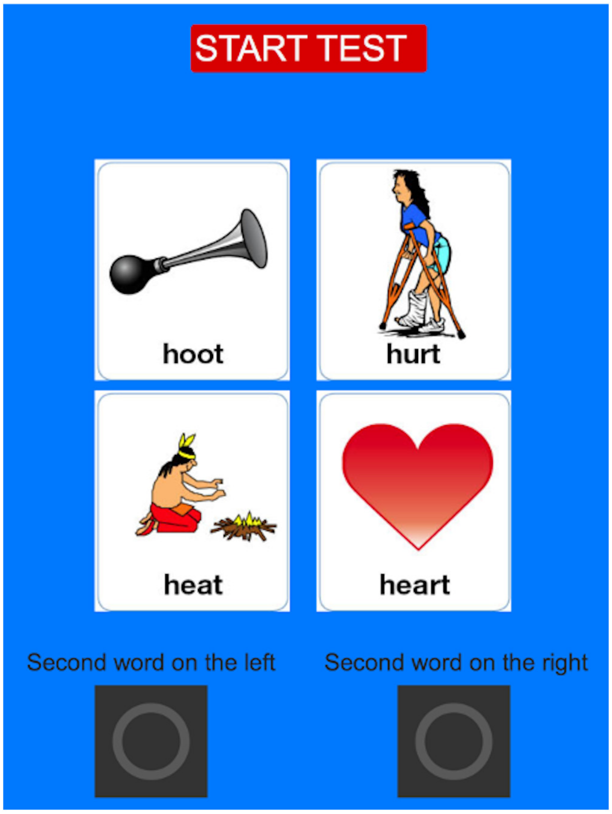


**FIGURE 2 |** Test interface. Participants clicked on each of two words presented in each trial and on the "**right**" or "**left**" buttons to indicate the location of the target word relative to the reference word.

(R Core Team, 2021), was used to fit a mixed-effects maximum likelihood binary logistic multilevel model separately for relative localization and word discrimination. The aim of the analysis was to determine whether the patterns of word discrimination and relative localization were similar to those previously obtained with the SSiN test. If this were the case, the outcomes of the analysis would be that performance for relative localization and speech discrimination is predicted by the spatial location of the speech sources. For relative localization it was expected that the likelihood of a correct response grew with increasing proximity of the speech sources to the midline. For speech discrimination, the opposite pattern was expected, i.e., that the likelihood of a correct response decreased with increasing proximity of the speech sources to the midline. The effect of word group and SNR as predictor for each task and the effect of the direction of the shift from reference to target in the relative-localization task was also assessed. In addition, some interactions between the predictors were investigated, as described below.

For relative localization, the "mean location" was defined as the average location of the pair of spatial locations of the target and the reference word within each trial, following Bizley et al. (2015), Ahnood (2017), and Parmar et al. (2018). The rationale for this is that, for a given pair of simulated loudspeakers, the participants would have had to make a localization judgment based on binaural localization cues of equal magnitude, albeit with opposite directions (Bizley et al., 2015). The model included a random intercept by participants in order to control for the non-independence of the data (Winter, 2019). The model was progressively built up by successively including the following predictors as fixed effects: "mean location" (−75°, −45°, −15°, 15°, 45°, or 75°), SNR (0, 3, or 6 dB above the SRT), direction of the shift from the target to the reference word (right or left), and word group (simple vowel, complex vowel, initial consonant, or final consonant). The following interactions were investigated: SNR × Mean Location, Direction × SNR, Direction × Mean Location.

For word discrimination, the model had a random intercept by participants to account for the fact that the participants were repeatedly tested. Because some participants were well above the 50%-word discrimination mark for the easiest condition (due to overestimation of the SRT), it was assumed that the slopes for SNR would vary across participants. Thus, a random slope for SNR was included. Next, fixed effects for SNR, azimuth, word order (i.e., whether the word was the target of the reference word), and word group were introduced one by one to build up the model. Interaction terms were included for word group × SNR, word group and azimuth, and azimuth and SNR.

The predictors of each model had their variance inflation factors (VIFs) calculated to ensure that multi-collinearity was not present. Models were compared by performing likelihood ratio tests. If any two models compared were not statistically different, the less complex model was chosen. *Post hoc* pairwise comparisons were carried out using the *multcomp* package (Hothorn et al., 2008) and Bonferroni corrections were applied.

### Comparison With Data Previously Collected With Loudspeakers

Outcomes were plotted and compared with data previously collected by Ahnood (2017) while completing an MSc dissertation at University College London, supervised by Jennifer Bizley and author Deborah Vickers. This dataset was chosen because it was obtained using the same number and distribution of azimuths as in the present work in a similar population. Ahnood (2017) tested 12 normal-hearing adults using an implementation of the SSiN test which delivered the background babble alternatively from two loudspeakers placed *either* at −60° and −30° azimuth *or* at 30° and 60° azimuth. In other words, in each trial, the background babble came either from the left or from the right of the listeners. Conversely, in our implementation, the background babble was symmetrically delivered from these same four loudspeaker locations in all trials. In spite of the differences in the location of the noise sources relative to the speech sources, a comparison across these datasets can be insightful as to whether the SSiN-VA leads to similar patterns of spatial hearing compared with the SSiN. It is expected that the shape of the performance-by-location function is more similar across datasets for the relative localization data. The speech discrimination outcomes are likely to be more strongly influenced by the spatial separation between the speech sources and the noise sources with respect to the listener's ears.

# RESULTS

**Figure 3** shows relative-localization and word-discrimination performance for each of the SNRs at which the participants were tested. Relative localization performance is plotted as a function of the mean target-reference location. For each of the SNRs tested, the function has the shape of an inverted *U*. This means that performance tended to be better at the midline than at the lateral locations. For word discrimination, performance tended to be slightly better at the lateral locations than at the midline (i.e., followed a *U*-shaped pattern), although this trend was more evident for the responses obtained with SNR = SRT. The effect of SNR can also be seen in this figure. For relative localization, responses varied somewhat with SNR; but for speech discrimination, the effect of SNR led to large improvements in performance. Statistical analyses were conducted to determine whether these trends were significant in order to test the hypotheses stated in the introduction.

## Statistical Analysis

### Relative Localization

For relative-localization performance, it was hypothesized that performance would resemble the pattern typically obtained with loudspeakers, with relative localization being better near the midline than at the lateral locations. Thus, the mean location of the target and reference words was expected to have an impact on performance. This was confirmed by comparing a random intercept only model with one where mean location was added as a fixed effect [$c^2(5) = 177.86$, $p < 0.001$]. SNR was hypothesized to affect performance. This was confirmed by adding SNR as a fixed effect, which significantly improved the model's fit [$c^2(2) = 9.42$, $p = 0.0090$]. The direction of the spatial shift from the reference word to the target word was thought not to have an impact on performance. However, its addition as a fixed effect significantly improved the model's fit [$c^2(1) = 23.85$, $p < 0.001$], indicating a direction bias. Word group was not expected to influence performance, which was confirmed [$c^2(3) = 6.76$, $p = 0.0798$]. The impact of SNR did not vary across mean locations [$c^2(10) = 10.33$, $p = 0.4123$], and the direction bias effect did not vary across SNRs [$c^2(2) = 1.64$, $p = 0.4414$] or mean locations [$c^2(5) = 10.08$, $p = 0.0730$]. In summary, relative localization performance was predicted by the mean location of the source and the target word, the SNR, and the direction of the spatial shift from the reference to the target word. Performance did not significantly differ across word groups. No significant interactions were found between SNR and mean locations, and between direction and SNR or mean location.

**Table 2** summarizes the odds ratios, 95% confidence intervals (CIs), and *p*-values of the final model. As the variables were treatment-coded, the intercept represents the likelihood of a correct response when all variables are set to 0: mean location −75°, SNR = SRT, direction left, and word group = 1, complex vowels. The intercept was −0.04 (SE = 0.16, $p = 0.7984$, odds ratio = 0.96). This indicates that, for this condition, the likelihood of a correct response was only slightly lower than the likelihood of an incorrect response. All other estimates of the model are referenced to this condition.

Assessing how the odds of a correct response varied when shifting mean location from left to right is helpful to characterize the shape of the performance function. This made it possible to test the hypothesis that performance would be better at the midline than at the lateral locations. When the mean location

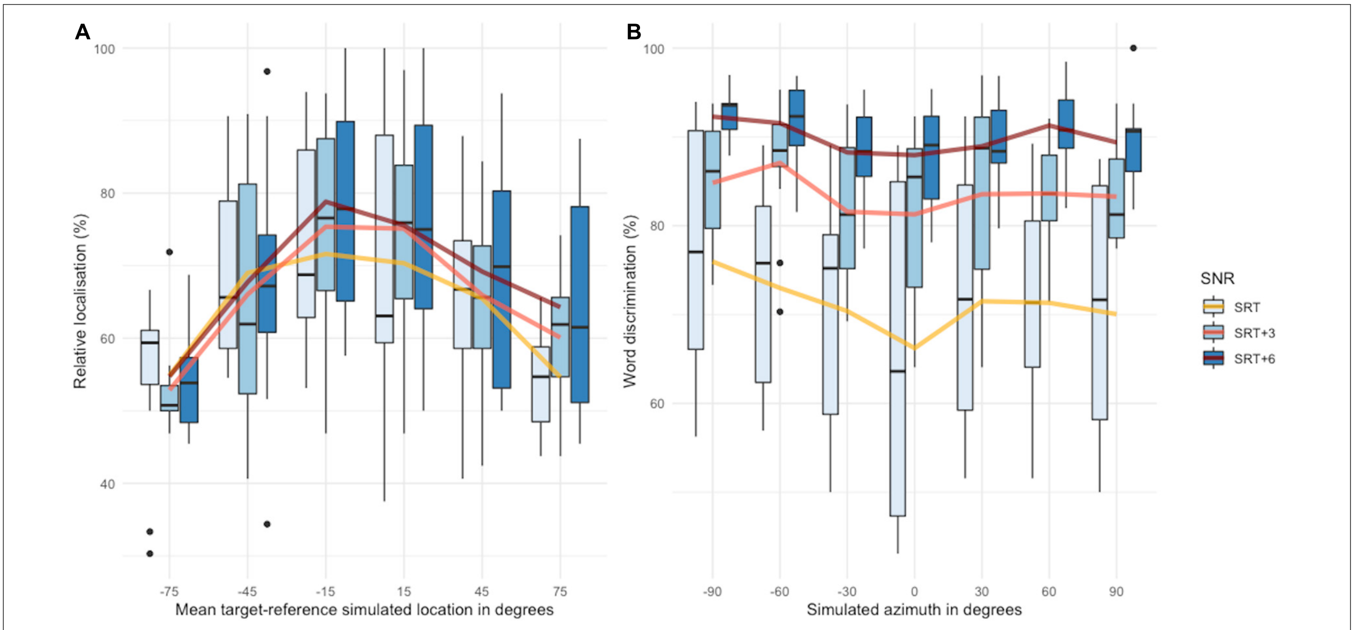


**FIGURE 3 |** SSiN-VA performance for relative localization **(A)** and word discrimination **(B)**. Boxplot colors indicate the SNR at which the test was done. The black lines within the boxplots represent the median. Mean scores are joined with a thick line across spatial locations, separately for each SNR. Figure made using the packages ggplot2 (Wickham, 2016) and cowplot (Wilke, 2020) in R (R Core Team, 2021).

**TABLE 2 |** Outcomes of the statistical analysis for SSiN-VA relative-localization performance.

| | Relative Localization Performance | | |
|---|---|---|---|
| ***Predictors*** | ***Odds Ratios*** | ***CI*** | ***p*** |
| (Intercept) | 0.96 | 0.71–1.31 | 0.798 |
| Mean Location −75° | *Reference* | | |
| Mean Location −45° | 1.82 | 1.53–2.17 | **<0.001** |
| Mean Location −15° | 2.71 | 2.26–3.24 | **<0.001** |
| Mean Location 15° | 2.49 | 2.08–2.98 | **<0.001** |
| Mean Location 45° | 1.76 | 1.48–2.09 | **<0.001** |
| Mean Location 75° | 1.27 | 1.07–1.51 | **0.006** |
| SNR SRT | *Reference* | | |
| SNR SRT + 3 dB | 1.08 | 0.95–1.22 | 0.250 |
| SNR SRT + 6 dB | 1.22 | 1.07–1.38 | **0.002** |
| Direction left | *Reference* | | |
| Direction right | 1.29 | 1.17–1.43 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ \text{Participant}}$ | 0.23 | | |
| ICC | 0.06 | | |
| $N_{\text{Participant}}$ | 12 | | |
| Observations | 6924 | | |
| Marginal $R^2$/Conditional $R^2$ | 0.040/0.102 | | |

*Table generated using the package sjPlot (Lüdecke, 2021). Bold values correspond to statistically significant outcomes.*

of the target and reference was −45° or 45°, the odds of a correct response significantly increased by 1.82 and 1.76 times, respectively, compared to −75°. This means that for mean locations −45° and 45°, a correct response was 1.75 and 1.70 times more likely than an incorrect response, respectively. Greater increases, by 2.71 and 2.49 times with respect to the intercept, respectively, were found when mean location −75° was compared with −15° and 15°. This suggests that for mean locations −15° and 15°, a correct response was 2.60 and 2.39 times more likely than an incorrect response. Finally, and against expectations, when the mean location of the target and reference was 75°, the odds of a correct response increased significantly, by 1.27 times, with respect to −75°. Thus, a correct response for mean location 75° was 1.23 times more likely than an incorrect response. In spite of this asymmetry, these outcomes are overall consistent with the expected inverted-*U* shape of relative localization as a function of mean location. The odds reported here were transformed into percentages and are illustrated in **Figure 4**.

Comparing the odds of a correct response across SNRs is helpful to evaluate the hypothesis that performance increases with increasing SNR. The odds of a correct response for relative localization significantly increased by 1.22 times when the SNR was raised by 6 dB with respect to the SRT. This means that at SNR = SRT + 6 dB, a correct response was 1.17 times more likely than an incorrect response, compared to 0.96 times for SNR = SRT. Although this increase is small, it confirms one of our hypotheses that performance would improve with increasing SNR. An increase of the SNR by 3 dB above the SRT failed to significantly increase the odds of a correct response. The odds reported here were transformed into percentages and are illustrated in **Figure 5**.

Comparing the odds of a correct response for each direction of the spatial shift from reference to target suggested that when the shift was toward the right, participants were more likely to obtain a correct answer. The odds of a correct response increased by 1.29 times with respect to a shift toward the left, making a correct response 1.24 times more likely than an incorrect one when the shift from reference to target was toward the right. This indicates a bias in this direction, contrary to what was hypothesized in the introduction, i.e., that there would not be a direction bias. **Figure 6** shows the reported odds transformed into percentages.

*Post hoc* analysis was performed in order to compare average performance across pairs of mean locations, SNRs, word groups, and shift directions, using the *ghlt* function within the *multcomp* package (Hothorn et al., 2008). Bonferroni corrections were applied to account for repeated testing. Outcomes are shown in **Table 3**. These comparisons revealed that performance at the most eccentric mid locations (−75° and 75°) was significantly lower than performance at the most central mid locations of −15° and 15°, which in turn were not significantly different from each other. Additionally, performance at −45° and 45° was significantly lower than performance at −15°, but not significantly different from performance at 15°. Performance at −75° was significantly lower than performance at 45° but not significantly lower than performance at −45°. Again, the pattern was consistent with an inverted-*U* shape, although, statistically, there were some asymmetries in the data. For SNR, *post hoc* pairwise comparisons indicated that, across mean locations, word groups, and shift directions, performance was significantly lower for SNR = SRT than for SNR = SRT + 6. For the direction of the shift, the comparison between left and right continued to be significant.

## Word Discrimination

For word discrimination, it was hypothesized that performance would resemble the pattern typically obtained with loudspeakers, with word discrimination being worse near the midline than at the lateral locations. Thus, the spatial location of the source (azimuth) was expected to be influential. This was confirmed, as a random-intercept-random-slope only model (random intercept by participant and random slope by SNR) was significantly worse than an identical model which had azimuth as a fixed effect to predict the discrimination outcome [$c^2(6) = 35.42$, $p < 0.001$]. Performing the task at different SNRs was thought to influence performance, which was the case [inclusion of SNR as fixed effect, $c^2(2) = 26.20$, $p < 0.001$]. As expected, different word groups led to varying levels of word discrimination performance [word group, $c^2(3) = 962.03$, $p < 0.001$]. To assess whether the order of presentation of the words within each trial (i.e., whether the word was target or reference) was associated with an increased likelihood of a correct response, word order was included. However, this failed to improve the predictions of the model [$c^2(1) = 1.51$, $p = 0.2192$]. Increasing SNR should lead to better word discrimination independent of azimuth, which was the case [no significant interaction between SNR and azimuth,

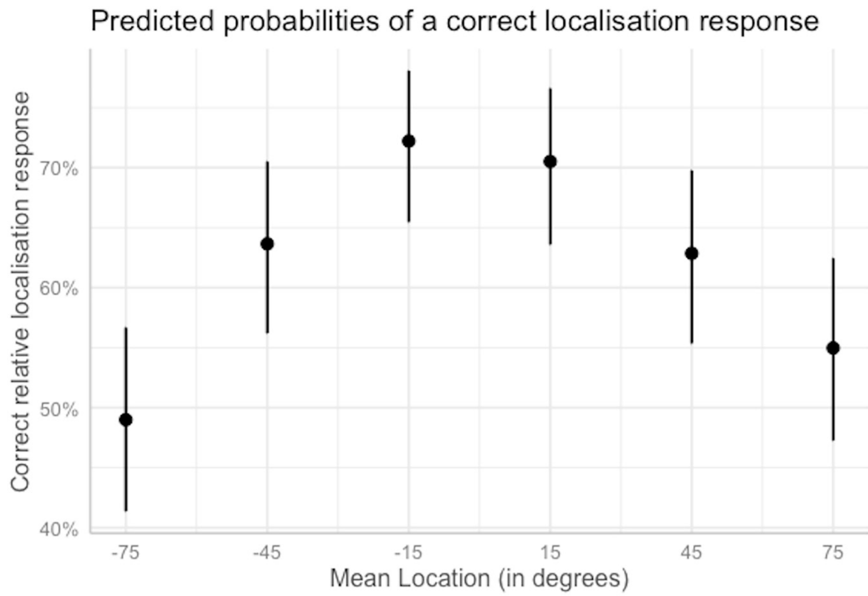


**FIGURE 4 |** Predicted probabilities of a correct localization response as a function of the mean location of the target and reference words when other predictors are held at their reference level. Plot obtained using the sjPlot package (Lüdecke, 2021). Error bars indicate 95% confidence intervals.

$c^2(12) = 6.91$, $p = 0.8632$]. As the different word groups might be more or less susceptible to masking by the level of the babble noise, the interaction term of SNR × word group was included, resulting in improved fit [$c^2(6) = 119.37$, $p < 0.001$]. Presenting words from varying azimuths changes the SNR at each ear, thus possibly making some word groups easier to understand than others. However, inclusion of the interaction azimuth × word group significantly worsened the model's fit [$c^2(6) = 119.37$, $p < 0.001$]. In summary, performance for word discrimination was predicted by the location of the speech source (azimuth), and by SNR and word group. Additionally, there was an interaction between SNR and word group. Thus, the effects of these two factors need to be considered jointly, as will be done below. The effect of SNR or word group did not vary with azimuth. The order of presentation of the words within each trial did not influence performance.

**Table 4** reports the odds ratios, 95% CIs, and $p$-values for the final model. As the variables were treatment-coded, the intercept represents the likelihood of a correct response when all variables were set to the reference level for each one of them (azimuth = −90°, SNR = SRT, word group = 1, complex vowels). The intercept was 1.53 (SE = 0.23, $p < 0.001$, odds ratio = 4.62). Therefore, for this condition, a correct word identification response was 4.62 times more likely than an incorrect response. Again, assessing how the odds of a correct response vary as a function of azimuth is useful to characterize the shape of the performance function. Having the speech coming from −60° or 60° did not significantly affect the odds of a correct response compared to −90°. However, having the speech coming from −30° or 30° significantly decreased the odds of a correct response, compared to −90°, by 0.70 and 0.78 times, respectively, making a correct response 3.23 and 3.60 times more likely than an incorrect response, respectively. Further, when the speech came from 0°, the odds of a correct response decreased even further, by 0.63 times, making a correct response 2.91 times more likely than an incorrect one. Finally, contrary to prior expectations, the odds of a correct response were significantly lower for the rightmost location in space (azimuth = 90°) compared to the left-most location in space (azimuth = −90°) by 0.75 times, making a correct response 3.47 times more likely than an incorrect one. Overall, these outcomes were consistent with the expected $U$-shaped performance function, although displaying an asymmetry in performance between −90° and 90°.

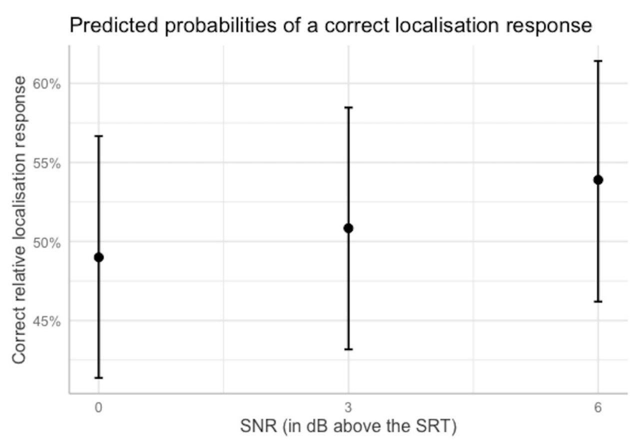


**FIGURE 5 |** Predicted probabilities of a correct localization response as a function of the SNR at which the stimuli were presented when other predictors were held constant at their reference level. Plot obtained using the sjPlot package (Lüdecke, 2021). Error bars indicate 95% confidence intervals.

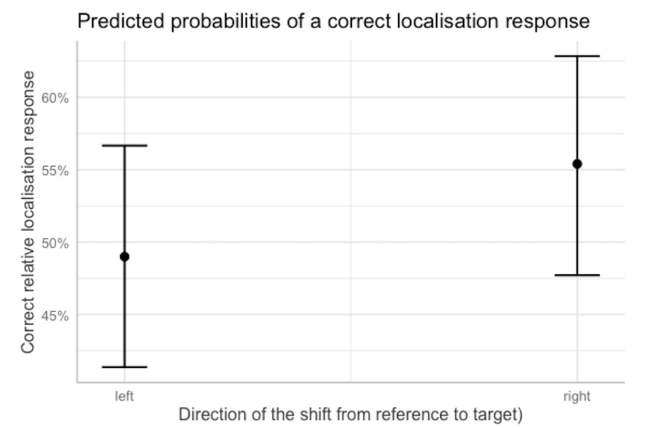


**FIGURE 6 |** Predicted probabilities of a correct localization response as a function of the direction of the shift from reference to target word, when other predictors were held constant at their reference level. Plot obtained using the sjPlot package (Lüdecke, 2021). Error bars indicate 95% confidence intervals.

**TABLE 3 |** *Post hoc* analysis for relative localization.

| Comparison | Estimate | Std. Error | *z*-value | Pr(> \| z\|) |
|---|---|---|---|---|
| **Mean Location** | | | | |
| −75° vs. −45° | −0.60 | 0.09 | −6.77 | **<0.001** |
| −75° vs. −15° | −1 | 0.09 | −10.76 | **<0.001** |
| −75° vs. 15° | −0.91 | 0.09 | −9.96 | **<0.001** |
| −75° vs. 45° | −0.57 | 0.09 | −6.41 | **<0.001** |
| −75° vs. 75° | −0.24 | 0.09 | −2.77 | 0.083 |
| −45° vs. −15° | −0.39 | 0.09 | −4.16 | **<0.001** |
| −45° vs. 15° | −0.31 | 0.09 | −3.31 | **0.014** |
| −45° vs. 45° | 0.03 | 0.09 | 0.38 | 1 |
| −45° vs. 75° | 0.36 | 0.09 | 4.04 | **<0.001** |
| −15° vs. 15° | 0.08 | 0.10 | 0.86 | 1 |
| −15° vs. 45° | 0.43 | 0.09 | 4.54 | **<0.001** |
| −15° vs. 75° | 0.76 | 0.09 | 8.13 | **<0.001** |
| 15° vs. 45° | 0.35 | 0.09 | 3.69 | **0.003** |
| 15° vs. 75° | 0.67 | 0.09 | 7.31 | **<0.001** |
| 45° vs. 75° | 0.3 | 0.09 | 3.68 | **0.003** |
| **SNR** | | | | |
| SRT vs. SRT + 3 dB | −0.7 | 0.06 | −1.15 | 0.749 |
| SRT vs. SRT + 6 dB | −0.20 | 0.06 | −3.04 | **0.007** |
| SRT + 3 dB vs. SRT + 6 dB | −0.12 | 0.06 | −1.89 | 0.177 |
| **Direction** | | | | |
| left vs. right | −0.26 | 0.05 | −4.87 | **<0.001** |

*P-values were Bonferroni corrected. Bold values correspond to statistically significant outcomes.*

The reported odds were transformed into percentages and are illustrated in **Figure 7**.

Increasing the SNR from SRT to SRT + 3 dB and from SRT to SRT + 6 dB led to a greater likelihood of a correct response, consistent with our hypothesis of better performance the higher the SNR. Changes in word group from complex vowels to simple vowels and from complex vowels to final consonant significantly decreased the likelihood of a correct response.

**TABLE 4 |** Outcomes of the statistical analysis for SSiN-VA word-discrimination performance.

| | Word Discrimination Performance | | |
|---|---|---|---|
| ***Predictors*** | ***Odds Ratios*** | ***CI*** | ***p*** |
| (Intercept) | 4.62 | 2.95–7.22 | **<0.001** |
| Azimuth −90° | *Reference* | | |
| Azimuth −60° | 0.96 | 0.78–1.18 | 0.701 |
| Azimuth −30° | 0.70 | 0.57–0.86 | **0.001** |
| Azimuth 0° | 0.63 | 0.51–0.77 | **<0.001** |
| Azimuth 30° | 0.78 | 0.63–0.96 | **0.017** |
| Azimuth 60° | 0.82 | 0.67–1.01 | 0.069 |
| Azimuth 90° | 0.75 | 0.60–0.95 | **0.018** |
| SNR SRT | *Reference* | | |
| SNR SRT + 3 dB | 3.02 | 2.24–4.07 | **<0.001** |
| SNR SRT + 6 dB | 6.68 | 4.53–9.84 | **<0.001** |
| Word Group $V_c$ | *Reference* | | |
| Word Group $V_s$ | 0.76 | 0.62–0.92 | **0.005** |
| Word Group $C_i$ | 1.22 | 1.00–1.50 | 0.056 |
| Word Group $C_f$ | 0.37 | 0.31–0.45 | **<0.001** |
| SNR SRT + 3 dB: Word Group $V_s$ | 0.96 | 0.69–1.35 | 0.832 |
| SNR SRT + 3 dB: Word Group $C_i$ | 0.89 | 0.63–1.27 | 0.522 |
| SNR SRT + 3 dB: Word Group $C_f$ | 0.46 | 0.34–0.63 | **<0.001** |
| SNR SRT + 6 dB Word Group $V_s$ | 1.08 | 0.70–1.67 | 0.712 |
| SNR SRT + 6 dB Word Group $C_i$ | 1.39 | 0.84–2.30 | 0.202 |
| SNR SRT + 6 dB Word Group $C_f$ | 0.28 | 0.20–0.41 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00\ Subject}$ | 0.47 | | |
| $\tau_{11\ Subject.SNR3}$ | 0.09 | | |
| $\tau_{11\ Subject.SNR6}$ | 0.15 | | |
| $\rho_{01}$ | −0.65 | | |
| | −0.82 | | |
| ICC | 0.09 | | |
| $N_{Subject}$ | 12 | | |
| Observations | 13848 | | |
| Marginal $R^2$/Conditional $R^2$ | 0.237/0.304 | | |

*Table generated using the package sjPlot (Lüdecke, 2021). Bold values correspond to statistically significant outcomes.*

However, it should be noted that, as treatment coding was used and a significant interaction between SNR and word group was found, it is not possible to assess the effects of SNR and word group separately. The interaction was explored using *post hoc* analysis as detailed below.

*Post hoc* analysis (**Table 5**) was performed to compare performance across azimuths, SNRs, and word groups using the *ghlt* function within the *multcomp* package (Hothorn et al., 2008), and Bonferroni corrections were applied to account for repeated testing. Comparing average performance for different azimuths across SNRs and word groups, indicated that the likelihood of a correct word-discrimination response was significantly higher at the left most location, −90°, than at −30° and at the midline. Additionally, performance at the midline was significantly lower than at −60° and 60°. Performance at −60° was also significantly higher than at −30°. These outcomes further support the

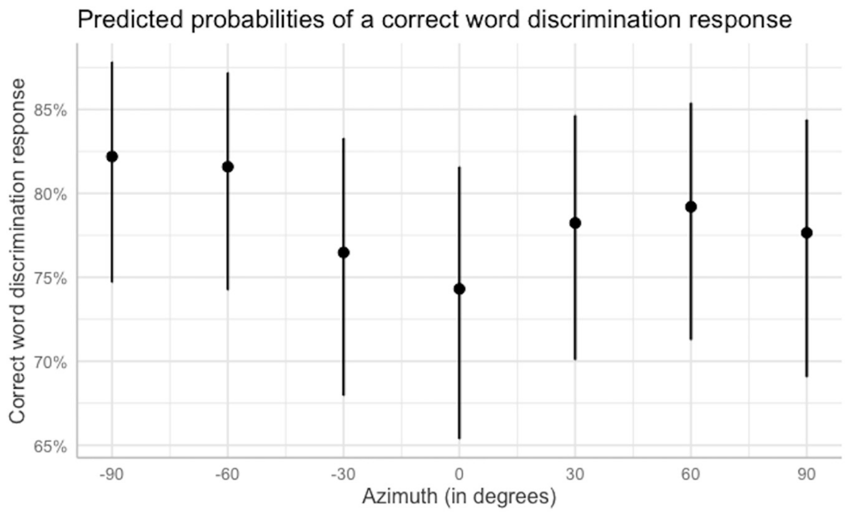


**FIGURE 7 |** Predicted probabilities of a correct word-discrimination response as a function of azimuth when other predictors are held at their reference level. Error bars indicate 95% confidence intervals. Plot obtained using the sjPlot package (Lüdecke, 2021).

hypothesis of a *U*-shape for performance as a function of azimuth. No other comparisons were statistically significant. Exploration of the significant interaction (**Table 6**) between SNR and word group via pairwise comparisons across averaged variable levels indicated that, although there was an overall increase in performance with increasing SNR, performance for the final-consonant group was significantly lower than that for each of the other discrimination groups at each of the SNRs tested. Additionally, performance for the initial-consonant group was significantly higher than that for the simple-vowel group only at the lowest SNR. No other comparisons of word groups within each SNR were statistically significant. **Figure 8** shows the predicted probabilities for each word group as a function of the SNR at which testing was conducted.

**TABLE 5 |** *Post hoc* analysis for the effect of azimuth on word discrimination.

| Azimuth | | | | |
|---|---|---|---|---|
| **Comparison** | **Estimate** | **Std. Error** | ***z*-value** | **Pr(> \| z\|)** |
| −90° vs. −60° | 0.04 | 0.11 | 0.38 | 1.000 |
| −90° vs. −30° | 0.35 | 0.10 | 3.35 | **0.017** |
| −90° vs. 0° | 0.47 | 0.10 | 4.50 | **<0.001** |
| −90° vs. 30° | 0.25 | 0.11 | 2.38 | 0.367 |
| −90° vs. 60° | 0.19 | 0.11 | 1.82 | 1.000 |
| −90° vs. 90° | 0.28 | 0.12 | 2.37 | 0.372 |
| −60° vs. −30° | 0.31 | 0.08 | 3.69 | **0.005** |
| −60° vs. 0° | 0.43 | 0.08 | 5.14 | **<0.001** |
| −60° vs. 30° | 0.21 | 0.08 | 2.47 | 0.284 |
| −60° vs. 60° | 0.15 | 0.09 | 1.78 | 1.000 |
| −60° vs. 90° | 0.24 | 0.10 | 2.38 | 0.365 |
| −30° vs. 0° | 0.12 | 0.08 | 1.46 | 1.000 |
| −30° vs. 30° | −0.10 | 0.08 | −1.23 | 1.000 |
| −30° vs. 60° | −0.16 | 0.08 | −1.92 | 1.000 |
| −30° vs. 90° | −0.07 | 0.10 | −0.67 | 1.000 |
| 0° vs. 30° | −0.22 | 0.08 | −2.69 | 0.150 |
| 0° vs. 60° | −0.28 | 0.08 | −3.38 | **0.015** |
| 0° vs. 90° | −0.18 | 0.10 | −1.85 | 1.000 |
| 30° vs. 60° | −0.06 | 0.08 | −0.70 | 1.000 |
| 30° vs. 90° | 0.03 | 0.10 | 0.34 | 1.000 |
| 60° vs. 90° | 0.09 | 0.10 | 0.91 | 1.000 |

*P-values were Bonferroni corrected. Bold values correspond to statistically significant outcomes.*

## Comparison With Data Previously Collected Using Loudspeakers

**Figure 9** shows a graphical comparison of the data collected in this study for SNR = SRT with that collected by Ahnood (2017). The SSiN-VA data is plotted separately against the responses of

**TABLE 6 |** *Post hoc* analysis of the interaction between SNR and Word Group.

| SNR | Comparison | Estimate | Std. Error | *z*-value | Pr(> \| z\|) |
|---|---|---|---|---|---|
| SRT | $V_c$ vs. $V_s$ | 0.28 | 0.10 | 2.80 | 0.342 |
| | $V_c$ vs. $C_i$ | −0.20 | 0.10 | −1.91 | 1.000 |
| | $V_c$ vs. $C_f$ | 0.99 | 0.10 | 10.29 | **<0.001** |
| | $V_s$ vs. $C_i$ | −0.48 | 0.10 | −4.69 | **<0.001** |
| | $V_s$ vs. $C_f$ | 0.71 | 0.09 | 7.65 | **<0.001** |
| | $C_i$ vs. $C_f$ | 1.19 | 0.10 | 12.05 | **<0.001** |
| SRT + 3 dB | $V_c$ vs. $V_s$ | 0.31 | 0.14 | 2.27 | 1.000 |
| | $V_c$ vs. $C_i$ | −0.08 | 0.15 | −0.57 | 1.000 |
| | $V_c$ vs. $C_f$ | 1.76 | 0.12 | 14.46 | **<0.001** |
| | $V_s$ vs. $C_i$ | −0.40 | 0.14 | −2.84 | 0.296 |
| | $V_s$ vs. $C_f$ | 1.45 | 0.11 | 12.95 | **<0.001** |
| | $C_i$ vs. $C_f$ | 1.84 | 0.12 | 14.91 | **<0.001** |
| SRT + 6 dB | $V_c$ vs. $V_s$ | 0.20 | 0.20 | 0.99 | 1.000 |
| | $V_c$ vs. $C_i$ | −0.53 | 0.24 | −2.24 | 1.000 |
| | $V_c$ vs. $C_f$ | 2.24 | 0.16 | 14.05 | **<0.001** |
| | $V_s$ vs. $C_i$ | −0.72 | 0.23 | −3.16 | 0.106 |
| | $V_s$ vs. $C_f$ | 2.05 | 0.15 | 13.67 | **<0.001** |
| | $C_i$ vs. $C_f$ | 2.77 | 0.20 | 14.06 | **<0.001** |

*P-values were Bonferroni corrected. Comparisons are reported in this table only for pairs of word groups at each SNR. Bold values correspond to statistically significant outcomes.*

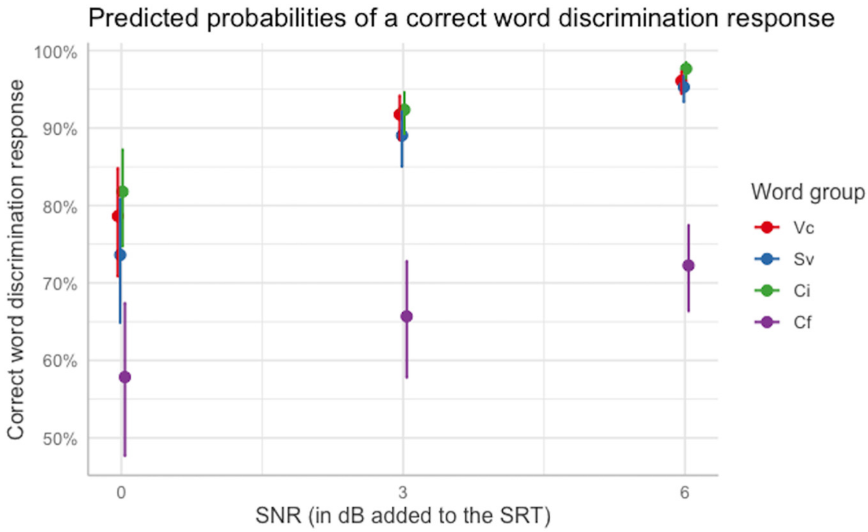


**FIGURE 8 |** Interaction between word group and SNR. Performance for the final-consonant group was significantly lower than for other word groups across all SNRs, but performance for the simple-vowel group was significantly lower than performance for initial consonant for the lowest SNR only, accounting for a significant interaction between word group and SNR. Error bars indicate 95% confidence intervals. Plot obtained using the sjPlot package (Lüdecke, 2021).

Ahnood's participants for trials in which the noise sources were located in the same hemispace than the speech and for those in which the noise sources were located in the opposite hemispace than that of the speech. For relative localization, performance is quite close across datasets for both comparisons (same hemispace or opposite hemispace). For Ahnood's dataset, there is a trend toward relative localization performance to be better at the most lateral locations when the noise was delivered from the opposite hemispace with respect to the speech, but overall, both curves are quite close to the data collected using SSiN-VA. For word discrimination, Ahnood's outcomes vary considerably depending on whether the noise was located on the same hemispace than the speech or on the opposite hemispace. This is expected as locating the noise on the opposite hemispace would have maximized the distance between the speech and the noise sources, which in turn would have had an impact on the effective SNR at each ear. This would have increased the SRM achieved by the participants, improving word-discrimination outcomes. For the trials where Ahnood delivered the nose from the same hemispace as the speech, her participants performed better than those using SSiN-VA at the most lateral locations. For the trials where Ahnood delivered the noise from the same hemispace as the speech, her participants performed much worse than those tested here, except at the midline. These patterns are likely to be largely accounted for by the spatial separation of the speech and the noise, which was different across datasets, rather than by differences related to the virtual nature of the stimuli used here.

## DISCUSSION

The aim of this study was to determine if the patterns of responses obtained with the SSiN-VA are similar to those obtained with loudspeaker implementations. Localization performance was predicted by mean location, SNR, and the direction of the shift from reference to target. For word discrimination, performance was predicted by azimuth, SNR, and word group and a significant interaction between word group and SNR was found. In what follows these results are discussed in more detail.

### Shape of the Performance Functions

As for loudspeaker data, relative localization followed the pattern of an inverted-$U$ shape and word discrimination followed a $U$-shaped pattern (Bizley et al., 2015; Ahnood, 2017), confirming that SSiN-VA leads to patterns of performance similar to those previously found with loudspeaker implementations. This makes sense, because as azimuth (or mean location) is varied, the availability of cues for each task, relative localization and word discrimination, changes. The relative levels of the signals arriving at each ear (ILDs) and their relative timing (ITDs) increases the further away the sources are from the midline. Additionally, the relative SNRs across ears change. In other words, the availability of binaural cues and binaural effects such as binaural summation or binaural squelch varies across spatial locations.

The inverted $U$-shape pattern of our data reproduces the outcomes found by Bizley et al. (2015) and Ahnood (2017) using the same stimuli, albeit with a different spatial location of the noise sources. As pointed out by Bizley et al. (2015), this pattern was also observed in a previous study where broadband noise was used (Wood and Bizley, 2015). In the same study, as well as in previous work (Butler, 1986), using spectrally restricted stimuli led to a more marked decrease in performance at the most lateral spatial locations. The improvement of relative localization performance around the midline is consistent with the idea that, because ILDs are roughly proportional to the sine of the azimuthal angle, horizontal localization errors should increase monotonically with increasing azimuth

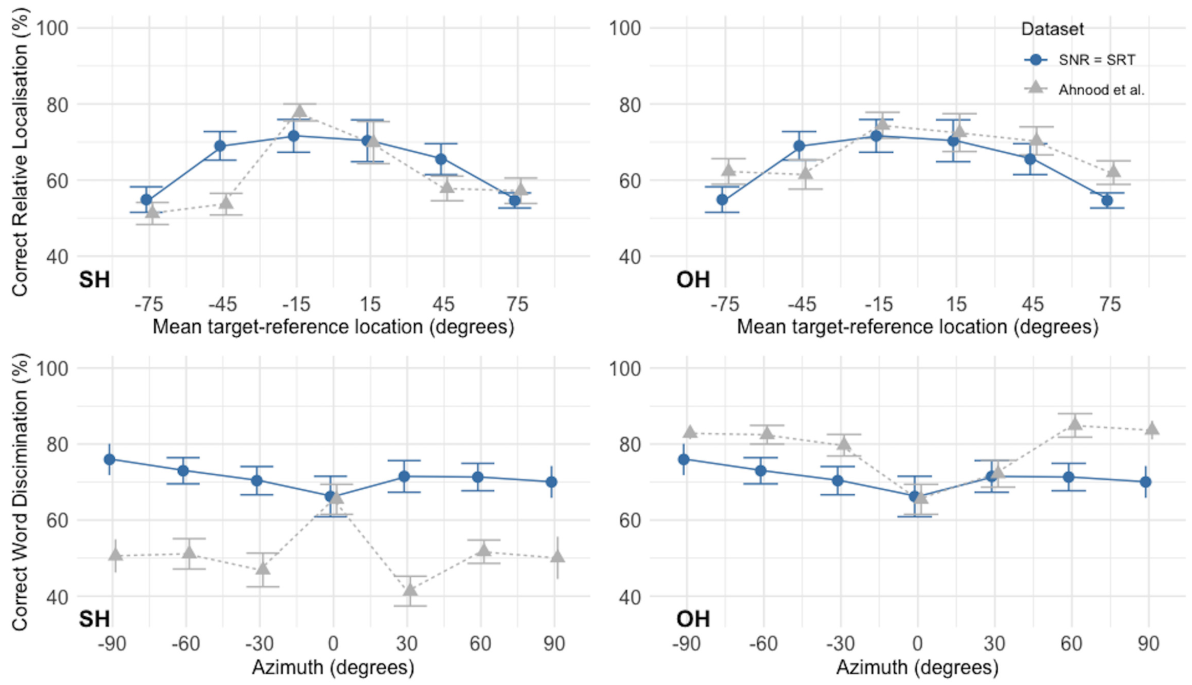


**FIGURE 9 |** Comparison of the outcomes reported here for SNR = SRT with the outcomes obtained by Ahnood (2017) for normal-hearing participants using a loudspeaker setup. The outcomes of the present study are displayed by blue circles joined by continuous line. Ahnood's results are displayed by gray triangles joined with dotted line. Error bars indicate standard errors. Because Ahnood, unlike us, used an asymmetric noise configuration, our data are compared separately with Ahnood's outcomes for trials where the noise sources were located on the same hemispace than the source of the speech (SH, noise in the same hemispace), and for trials where the noise was located on the hemispace opposite to the speech (OH, noise in the opposite hemispace). Outcomes for word discrimination at the midline are plotted on both panels and were not separated by noise location. Figure made using the packages ggplot2 (Wickham, 2016) and cowplot (Wilke, 2020) in R (R Core Team, 2021).

(Middlebrooks and Green, 1991). This is similar to the outcomes reported by Bizley et al. (2015) for one of their experiments in which they used loudspeakers separated at 15° intervals, although these authors did not find a significant effect of azimuth on relative-localization performance for another experiment where they used a 30°-interval separation for most of their loudspeakers. However, the type of analysis carried out here is different to that carried out by Bizley et al. (2015), and there may be power differences across studies underlying the discrepancy.

The *U*-shaped pattern observed for word discrimination is also consistent with the outcomes obtained with the original implementation of the SSiN (Bizley et al., 2015), and with other published work. For example, Laitakari and Laitakari (1997) also reported improved speech discrimination at 90°compared to 0°. In conditions where the background noise is symmetrically distributed around the midline, the advantage of lateral locations with respect to the midline for speech discrimination may arise from factors such as the "better-ear glimpsing" effect (Glyde et al., 2013), where information from the ear with better SNR is used to make sense of the speech. This difference in SNR across ears is partly underpinned by the head-shadow effect. Thus, this effect is likely to help to improve performance for speech sources that are away from the midline. ITDs are likely to contribute too, as they can be used to achieve binaural unmasking of the low-frequency portions of a signal (Hawley et al., 2004).

In spite of our data having the expected shape in terms of performance as a function of spatial location, there were some asymmetries, i.e., performance to the left and the right of the midline was sometimes significantly different. This trend is also apparent in loudspeaker data such as that reported by Bizley et al. (2015) and Ahnood (2017). It is possible that this is due to noise in the data arising from individual performance. This should be taken into account when interpreting clinical outcomes.

Overall, these results are encouraging and suggest that the use of this virtual implementation of the SSiN leads to similar patterns of responses to loudspeaker implementations. The next steps in the development of the prototype are to manipulate different parameters in order to achieve varying levels of difficulty. This is key to the clinical implementation of SSiN-VA, as the test is conceived as a flexible tool able to test a wide range of clinical populations with diverse spatial-listening skills.

## Effect of the Signal-to-Noise Ratio on Each of the Tasks

Unlike previous work with the SSiN test, the effect of SNR on performance was measured. SNR had a strong effect on speech-discrimination performance, with each 3-dB increase in SNR leading to a significant improvement in word recognition. However, this effect should not be interpreted in isolation, as there was an interaction between SNR and word group (this will be discussed below). The impact of SNR on relative localization was lower, with a trend toward improving performance with increasing SNR, but where a 6-dB increase in SNR above the SRT led to significantly improved performance, a 3-dB increase did not. The greater impact of SNR on word recognition than on relative localization may have arisen from the fact that speech discrimination requires the audibility of specific parts of the two speech signals. There would have been instances where audibility was appropriate for detection but not for discrimination. Conversely, relative localization is still possible even if audibility is not enough for discrimination.

## Effect of Word Group on Each of the Tasks

Contrary to Bizley et al. (2015), who reported that relative localization was better for the final-consonant group, no effect of word group on relative-localization performance was found. However, there was a trend in the same direction in our data which did not reach statistical significance. Bizley et al. (2015) proposed that the listener, knowing that the discrimination cue is at the end, has more time to listen for the localization cues before focusing on the discrimination cues.

There was a significant effect of word group for the word-discrimination task, also in contrast with the findings of Bizley et al. (2015). Here, the effect of word group on word discrimination can be largely accounted for by the much lower likelihood of a correct response for the final-consonant group at each SNR. The consonants in this group were plosive and fricative consonants, characterized by their predominantly high-frequency energy. These phonemes might be more vulnerable to being masked by the babble noise, depending on the babble level. For example, the relative amplitude of the formant transition F3 of the voiced fricative /z/ (as in “cheese”) with respect to the adjacent vowel is about −16.3 dB (Jongman et al., 2000). The other consonants in the final-consonant group were the voiceless stops /p/, /t/, /k/. These phonemes are characterized by a brief silence followed by a brief burst, which is an important discrimination cue (Kapoor and Allen, 2012). This burst would have been susceptible to being easily masked by the background babble. Additionally, initial consonants are typically more discriminable than final consonants due to their lower level and shorter duration, and due to the higher amount of information present in the consonant-vowel formant transition compared to the vowel-consonant transition (Redford and Diehl, 1999). There was also a difference between the initial-consonant group and the simple-vowel group which was only significant at the lowest SNR, so that at the SRT condition, correct word discrimination was less likely for the simple-vowel discrimination group than for the initial-consonant group. The interaction of SNR with word group suggests that the audibility of the cues played a role, as the simple-vowel-initial consonant difference was significant only at the lowest SNR.

In spite of the possible impact of these acoustic and perceptual differences across word groups, it cannot be ruled out that at least part of this effect was underpinned by deviations of the headphone frequency response from the free field response. This could explain the conflicting findings for the effect of word group on each of the tasks across this study and Bizley et al. (2015). The effect of the transducers will be investigated in future research, as it is necessary to be aware of any limitations imposed by the transducers before the test is generalized for clinical use.

## Direction Bias

Unexpectedly, when the target word shifted to the right of the reference word, a correct relative-localization response was more likely. Ocklenburg et al. (2010) reported that participants carrying out a sound-localization task showed a bias toward the opposite side to the dominant hand when they pointed at the source of the sound using their hand or their head. Here, participants did not point at the source of the stimuli but used a computer interface where they had buttons to click on, labeled “left” and “right” (as shown in **Figure 2**). Participants were not asked whether they were right- or left-handed, but it is reasonable to assume that most of them would have been right-handed. It is difficult to compare across these studies because the nature of the localization task (absolute vs. relative localization) and the mode of giving a response differed. However, our results show an effect that appears to be in conflict with what was reported in the literature. Inspection of the individual data suggested that the bias was large for one subject and much smaller for others. Three subjects showed the opposite pattern (bias to the left) and three other subjects showed very small differences across shift directions.

## Location of the Noise Sources

One of the possible parameters for adjustment in future versions of the test is the location of the noise sources. The location of the noise sources has an impact on the shape of the performance function, especially for word discrimination. This is evident from **Figure 9**, which compares the data collected here with Ahnood’s dataset. For Ahnood’s dataset, speech discrimination was strongly affected by the location of the noise, as speech discrimination is highest when the spatial separation between speech source and babble noise is maximized, and vice versa (Hirsh, 1950). Thus, Ahnood’s participants’ responses were expected to show greater differences across spatial locations compared to our participants as, in the case of the latter, symmetrical maskers were used. Our participants would have been more reliant on “glimpsing” (Glyde et al., 2013), i.e., on extracting information during short-term improvements in SNR which, with symmetrical maskers, will occur alternatively at one ear or the other (Brungart and Iyer, 2012). The differences in patterns of response across the datasets should be more evident for the comparison between word-discrimination functions than for the comparison between relative-localization functions, as localization performance is relatively independent from source-masker spatial separation. **Figure 9** supports these predictions. Ahnood’s noise configuration was shown to lead to different patterns of responses for normal-hearing participants and cochlear-implant users (Ahnood, 2017). The potentially large separations between speech and babble source are better suited to test SRM but would have required us to increase the number of trials in our test. As the purpose of this work was to determine if the virtual implementation led to similar patterns of responses than the original loudspeaker implementation, a symmetric configuration was used, similarly to Bizley et al. (2015), although the number of sources was reduced. Having a symmetrical configuration allowed the use of a simpler design. Moving forward toward a clinical implementation of the test, a direct comparison of a loudspeaker setup and this virtual implementation using different noise configurations (symmetric and asymmetric) is due to be carried out.

### Limitations of the Auralization Technique

Auralization using HRTFs that are not individualized may lead to inaccurate sound localization and issues with externalization (Stitt et al., 2019). Furthermore, some training might be needed in order to achieve performance at similar levels than with individualized HRTFs (Blum et al., 2004; Parseihian and Katz, 2012; Steadman et al., 2019; Stitt et al., 2019). HRTFs include a spectral component which is used for front-back judgments and localization along the vertical axis, and an interaural component given by the ITDs and ILDs. Here the size of the head of each participant was used to personalize ITDs. Stitt et al. (2019) found that, even when head-circumference-based ITDs are used, the localization performance error increases to 15.5°–19.4° from the 9.3°–12.5° measured from a control group using individual HRTFs. Parseihian and Katz (2012) reported an increase from 13°–16° to 17°–25° between a group with individualized HRTFs and groups with non-individualized HRTFs. Training using a VR videogame (Steadman et al., 2019) or providing proprioceptive feedback (Parseihian and Katz, 2012; Stitt et al., 2019) did not significantly improve lateral angle judgments. The relative-localization task performed here requires spatial discrimination with 30° resolution. As this is generally larger than the average errors encountered by these investigators, it is likely that performance would have been similar with individualized HRTFs. However, there may be individual cases where the introduced error makes it hard to give a relative-localization response. Informal feedback given by a few participants during the task was consistent with some front-back confusions and with reports of the two words originating from the same source. The impact of using non-individualized HRTFs with participants who have hearing loss should be investigated.

### Limitations of the Study

A small sample size of 12 participants with normal hearing was used. This is similar to previous studies with loudspeakers (Bizley et al., 2015; Ahnood, 2017). Testing a larger sample of participants including examining the effects of age, hearing status, and co-existence of other disabilities on the user experience with virtual audio might be of interest in order to optimize SSiN-VA for a wide range of users.

Other limitations of the present study are that all data were collected with SSiN-VA and that no data were obtained with a loudspeaker setup, and that the existing dataset used for comparison was generated using a noise-location configuration different than that used here. As explained in the Results section, this should lead to some differences in the patterns of responses, especially for the speech-discrimination task. Further work with SSiN-VA will address this issue by directly comparing both setups using the same noise configuration.

## CONCLUSION

The findings reported in this study support the use of virtual-audio to develop clinical-audiology applications to assess spatial listening skills. The SSiN-VA led to similar patterns of responses than SSiN for speech discrimination and relative localization as a function of the spatial location of the sound sources. This suggests that binaural spatialization has the potential to make a step change in the clinical testing of spatial hearing abilities, making it possible to inexpensively assess the benefits of different hearing devices, such as bilateral hearing aids, bilateral or bimodal cochlear implants, and devices used to help people with unilateral hearing loss or single-sided deafness. This approach also increases clinical efficiency because testing can be carried out in the home if necessary.

Simplifying the equipment and space requirements to conduct reliable tests that assess complex listening skills, including the development of home-testing versions, ultimately increases the equality of access to hearing care across geographical location and improves the quality of care, enhancing the experience of the patients and their families.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in Apollo, the online repository of the University of Cambridge, following the link: doi: 10.17863/CAM.76227.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Cambridge Psychology Research Ethics Committee (Ref. 2019.093) and Joint Research Compliance Office at Imperial College (Ref. 19IC5073). The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MS-C, LP, and DV designed the experiments. MS-C collected the data and wrote the first draft of the manuscript. LP developed all versions of the SSiN-VA app. BW and MS-C calibrated the speech stimuli. All authors contributed to the analysis of results and to the writing of the manuscript.

## FUNDING

MS-C was funded by Imperial Confidence in Concept, Imperial Biomedical Research Centre (BRC). DV and MS-C were funded by a Programme Grants for Applied Research (NIHR201608).

MS-C, BW, and DV were funded by the Medical Research Council (MRC) United Kingdom, Grant code MR/S002537/1.

## ACKNOWLEDGMENTS

Preliminary results for a previous prototype of the SSiN-VA were presented at the International Congress on Acoustics (Picinali et al., 2019b). We thank Jenny Bizley for kindly providing the dataset collected by Mana Ahnood which was used for comparison with our data. We are grateful to Bhavisha Parmar and Jenny Bizley for providing information about the original version of the Spatial Speech Test and for helpful discussion, and to Lindsey Van Yper and Jaime Undurraga for their insightful discussion about statistics. We also thank Tom Baer and Brian C. J. Moore for their advice on headphone calibration.

## REFERENCES

Ahnood, M. (2017). *Simultaneous Assessment of Speech Identification and Spatial Discrimination: A Potential Testing Approach for Bilateral Cochlear Implant Users*.Thesis submitted as partial fulfilment of the requirements for the MSc in Advanced Audiology. London: University College London. .

Armstrong, C., Thresh, L., Murphy, D., and Kearney, G. (2018). A perceptual evaluation of individual and non-individual HRTFs: a case study of the SADIE II database. *Appl. Sci.* 8:2029. doi: 10.3390/app8112029

Ayas, M., Ali Al Amadi, A. M. H., Khaled, D., and Alwaa, A. M. (2020). Impact of COVID-19 on the access to hearing health care services for children with cochlear implants: a survey of parents. *F1000Res.* 9, 690–690. doi: 10.12688/f1000research.24915.1

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* 67, 1–48.

Begault, D. R., Wenzel, E. M., and Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.* 49, 904–916.

Bizley, J. K., Elliott, N., Wood, K. C., and Vickers, D. A. (2015). Simultaneous assessment of speech identification and spatial discrimination: a potential testing approach for bilateral cochlear implant users? *Trends Hear.* 19:2331216515619573. doi: 10.1177/2331216515619573

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge: MIT Press.

Blum A, Katz BF, Warusfel O. (2004). "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training," in *Proceeding of the CFA/DAGA*, Strasbourg.

Boersma, P., and Weenink, D. (1992). *Praat: Doing Phonetics by Computer (Version 5.1.3.1)*. Available online at: http://www.praat.org/ (accessed May 2, 2020).

Bronkhorst, A. W. (2000). The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acta Acust. United Acust.* 86, 117–128.

Brungart, D. S., and Iyer, N. (2012). Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. *J. Acoust. Soc. Am.* 132, 2545–2556. doi: 10.1121/1.4747005

Brungart, D. S., Walden, B., Cord, M., Phatak, S., Theodoroff, S. M., Griest, S., et al. (2017). Development and validation of the Speech Reception in Noise (SPRINT) Test. *Hear. Res.* 349, 90–97. doi: 10.1016/j.heares.2017.01.008

Butler, R. A. (1986). The bandwidth effect on monaural and binaural localization. *Hear. Res.* 21, 67–73. doi: 10.1016/0378-5955(86)90047-x

Cameron, S., and Dillon, H. (2008). The listening in spatialized noise test: an auditory processing disorder study. *J. Am. Acad. Audiol.* 19, 377–391. doi: 10.3766/jaaa.17.5.2

Cord, M. T., Walden, B. E., and Atack, R. M. (1992). *Speech Recognition in Noise Test (SPRINT) for H-3 Profile*. Bethesda, MD: Walter Reed Army Medical Center.

Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., et al. (2019). 3D tune-in toolkit: an open-source library for real-time binaural spatialisation. *PLoS One* 14:1–37. doi: 10.1371/journal.pone.0211899

Cycling'74 (2021). *Tools for Sounds, Graphics and Interactivity*. Available online at: https://cycling74.com (accessed May 6, 2021).

Etymotic Research Inc. (2005). *BKB-SIN Test Manual*. Elk Grove Village, IL: Etymotic Research Inc.

Freed, A. (1997). "Open sound control: a new protocol for communicating with sound synthesizers," in *Proceedings of the International Computer Music Conference (ICMC)*, Thessaloniki.

Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., and Cameron, S. (2013). The effect of better-ear glimpsing on spatial release from masking. *J. Acoust. Soc. Am.* 134, 2937–2945. doi: 10.1121/1.4817930

Grange, J. A., and Culling, J. F. (2016). The benefit of head orientation to speech intelligibility in noise. *J. Acoust. Soc. Am.* 139, 703–712. doi: 10.1121/1.4941655

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *J. Acoust. Soc. Am.* 115, 833–843. doi: 10.1121/1.1639908

Hirsh, I. J. (1950). The relation between localization and intelligibility. *J. Acoust. Soc. Am.* 22, 196–200. doi: 10.1121/1.1906588

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Huckvale, M. (1989). *SCRIBE Database*. Available online at: https://www.phon.ucl.ac.uk/resource/scribe/ (accessed June 10, 2021).

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108, 1252–1263.

Kapoor, A., and Allen, J. B. (2012). Perceptual effects of plosive feature modification. *J. Acoust. Soc. Am.* 131, 478–491. doi: 10.1121/1.3665991

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear Hear.* 41(Suppl. 1), 5S–19S. doi: 10.1097/AUD.0000000000000944

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116(4 Pt 1), 2395–2405. doi: 10.1121/1.1784440

Kitterick, P. T., Lovett, R. E. S., Goman, A. M., and Summerfield, A. Q. (2011). The AB-York crescent of sound: an apparatus for assessing spatial-listening skills in children and adults. *Cochlear Implants Int.* 12, 164–169. doi: 10.1179/146701011X13049348987832

Laitakari, K., and Laitakari, J. (1997). New computerized Finnish speech in noise test and binaural hearing. *Acta Otolaryngol. Suppl.* 117, 74–76. doi: 10.3109/00016489709124086

Litovsky, R. Y. (2012). Spatial release from masking. *Acoust. Today* 8, 18–25.

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science. R Package Version 2.8.8, [Internet]*. Available online at: https://CRAN.R-project.org/package=sjPlot. (accessed November 26, 2021).

Marriage, J. E., Vickers, D. A., Baer, T., and Moore, B. C. J. (2011). "Using speech perception measures to guide the choice of amplification," in *A Sound Foundation Through Early Amplification 2010*, eds R. C. Seewald and J. M. Bamford (Staefa: Phonak), 273–279.

Middlebrooks, J. C., and Green, D. M. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.* 42, 135–159.

Moore, B. C. J. (2014). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*. Singapore: World Scientific, 1–182.

Ocklenburg, S., Hirnstein, M., Hausmann, M., and Lewald, J. (2010). Auditory space perception in left- and right-handers. *Brain Cogn.* 72, 210–217. doi: 10.1016/j.bandc.2009.08.013

Otoconsult Helpdesk (2021). *How to Perform the A§E Sound Localization Test?*. Available online at: https://support.otoconsult.com/support/solutions/articles/3000089586-how-to-perform-the-a-e-sound-localization-test- (accessed July 8, 2021).

Parmar, B., Beukes, E., and Rajasingam, S. (2021). The impact of COVID-19 on provision of UK audiology services & on attitudes towards delivery of telehealth services. *Int. J. Audiol.* Online ahead of print doi: 10.1080/14992027.2021.1921292

Parmar, B., Vickers, D. A., Ahnood, M., and Bizley, J. K. (2018). "The spatial speech test for assessing binaural hearing," in *Proceedings of the International Hearing-Aid Research Conference (IHCON)*, Lake Tahoe, CA.

Parseihian, G., and Katz, B. F. G. (2012). Rapid head-related transfer function adaptation using a virtual auditory environment. *J. Acoust. Soc. Am.* 131, 2948–2957. doi: 10.1121/1.3687448

Pausch, F., and Fels, J. (2019). MobiLab – A mobile laboratory for on-site listening experiments in virtual acoustic environments. *Acta Acust. United Acust.* 105, 875–887. doi: 10.3813/aaa.919367

Phatak, S. A., Brungart, D. S., Zion, D. J., and Grant, K. W. (2019). Clinical assessment of functional hearing deficits: speech-in-noise performance. *Ear Hear*. 40, 426–436. doi: 10.1097/AUD.0000000000000635

Picinali, L., Hrafnkelsson, R., and Reyes-Lecuona, A. (2019a). "The 3D tune-in toolkit VST binaural audio plugin," in *Proceeding of the Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio* (New York, NY: Audio Engineering Society).

Picinali, L., Salorio-Corbetto, M., and Vickers, D. (2019b). *Spatial Release from Masking Assessment in Virtual Reality for Bilateral Cochlear Implants Users*. Aachen: Universitätsbibliothek der RWTH Aachen.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Redford, M. A., and Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *J. Acoust. Soc. Am.* 106, 1555–1565. doi: 10.1121/1.427152

Saunders, G. H., and Roughley, A. (2020). Audiology in the time of COVID-19: practices and opinions of audiologists in the UK. *Int. J. Audiol.* 60, 255–262. doi: 10.1080/14992027.2020.1814432

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., et al. (2012). Development and validation of the AzBio sentence lists. *Ear Hear.* 33, 112–117. doi: 10.1097/aud.0b013e31822c2549

Steadman, M. A., Kim, C., Lestang, J.-H., Goodman, D. F., and Picinali, L. (2019). Short-term effects of sound localization training in virtual reality. *Sci. Rep.* 9, 1–17. doi: 10.1038/s41598-019-54811-w

Stitt, P., Picinali, L., and Katz, B. F. G. (2019). Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. *Sci. Rep.* 9:1063. doi: 10.1038/s41598-018-37873-0

Summerfield, Q., Palmer, A. R., Foster, J. R., Marshall, D. H., and Twomey, T. (1994). Clinical evaluation and test-retest reliability of the IHR-McCormick automated toy discrimination test. *Br. J. Audiol.* 28, 165–179. doi: 10.3109/03005369409086564

The MathWorks Inc. (2019). *Matlab R2019b Update 1*. Natick, MA: The MathWorks Inc.

Vickers, D. A., Moore, B. C. J., Majeed, A., Stephenson, N., Alferaih, H., Baer, T., et al. (2018). Closed-set speech discrimination tests for assessing young children. *Ear Hear.* 39, 32–41. doi: 10.1097/AUD.0000000000000528

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer-Verlag. Available online at: https://ggplot2.tidyverse.org

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R Package Version 1.1.1.* Available online at: https://CRAN.R-project.org/package=cowplot (accessed December 30, 2020).

Wilson, R. H., Carnell, C. S., and Cleghorn, A. L. (2007). The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *J. Am. Acad. Audiol.* 18, 522–529.

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Milton Park: Routledge.

Wood, K. C., and Bizley, J. K. (2015). Relative sound localisation abilities in human listeners. *J. Acoust. Soc. Am.* 138, 674–686. doi: 10.1121/1.4923452

**Author Disclaimer:** The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**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 in Psychology

REVIEW
published: 10 March 2022
doi: 10.3389/fpsyg.2022.786347

Check for updates

# Listening-Based Communication Ability in Adults With Hearing Loss: A Scoping Review of Existing Measures

**Katie Neal[1], Catherine M. McMahon[1,2], Sarah E. Hughes[3,4,5†] and Isabelle Boisvert[2,6*]**

[1] Department of Lingustics, Macquarie University, Sydney, NSW, Australia, [2] Hearing, Macquarie University, Sydney, NSW, Australia, [3] Centre for Patient Reported Outcome Research, Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom, [4] National Institute of Health Research (NIHR), Applied Research Collaboration (ARC), West Midlands, United Kingdom, [5] Faculty of Medicine, Health and Life Science, Swansea University, Swansea, United Kingdom, [6] Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

OPEN ACCESS

**Edited by:**
Jerker Rönnberg,
Linköping University, Sweden

**Reviewed by:**
Erin Margaret Picou,
Vanderbilt University Medical Center,
United States
Christian Füllgrabe,
University College London,
United Kingdom

***Correspondence:**
Isabelle Boisvert
isabelle.boisvert@sydney.edu.au

**†Present address:**
Sarah E. Hughes,
Birmingham Health Partners Centre
for Regulatory Science and
Innovation, University of Birmingham,
Birmingham, United Kingdom

**Specialty section:**
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

**Received:** 30 September 2021
**Accepted:** 31 January 2022
**Published:** 10 March 2022

**Citation:**
Neal K, McMahon CM,
Hughes SE and Boisvert I (2022)
Listening-Based Communication
Ability in Adults With Hearing Loss:
A Scoping Review of Existing
Measures.
Front. Psychol. 13:786347.
doi: 10.3389/fpsyg.2022.786347

**Introduction:** Hearing loss in adults has a pervasive impact on health and well-being. Its effects on everyday listening and communication can directly influence participation across multiple spheres of life. These impacts, however, remain poorly assessed within clinical settings. Whilst various tests and questionnaires that measure listening and communication abilities are available, there is a lack of consensus about which measures assess the factors that are most relevant to optimising auditory rehabilitation. This study aimed to map current measures used in published studies to evaluate listening skills needed for oral communication in adults with hearing loss.

**Methods:** A scoping review was conducted using systematic searches in Medline, EMBASE, Web of Science and Google Scholar to retrieve peer-reviewed articles that used one or more linguistic-based measure necessary to oral communication in adults with hearing loss. The range of measures identified and their frequency where charted in relation to auditory hierarchies, linguistic domains, health status domains, and associated neuropsychological and cognitive domains.

**Results:** 9121 articles were identified and 2579 articles that reported on 6714 discrete measures were included for further analysis. The predominant linguistic-based measure reported was word or sentence identification in quiet (65.9%). In contrast, discourse-based measures were used in 2.7% of the articles included. Of the included studies, 36.6% used a self-reported instrument purporting to measures of listening for communication. Consistent with previous studies, a large number of self-reported measures were identified ($n$ = 139), but 60.4% of these measures were used in only one study and 80.7% were cited five times or fewer.

**Discussion:** Current measures used in published studies to assess listening abilities relevant to oral communication target a narrow set of domains. Concepts of communicative interaction have limited representation in current measurement. The lack of measurement consensus and heterogeneity amongst the assessments limit comparisons across studies. Furthermore, extracted measures rarely consider the broader linguistic, cognitive and interactive elements of communication. Consequently, existing measures may have limited clinical application if assessing the listening-related skills required for communication in daily life, as experienced by adults with hearing loss.

**Keywords: listening, communication ability, hearing loss, adults, scoping review, outcome measure**

## INTRODUCTION

Communication forms the foundation of social interaction. For adults, communication is recognised as a critical component to adapting and adjusting to aging, essential to maintaining independence and personal relationships, performing social roles and functions, making decisions and having control over life quality (Heinrich et al., 2016). While language use and structure change across the life span, conversational skills are generally preserved in typically aging adults (Shadden, 1988). Aging, however, is associated with an increased prevalence of conditions that affect communication, of which hearing loss is the most prevalent (Wallhagen and Pettengill, 2008). The effect of impaired communication is linked to several aspects of social relationships and psychological well-being. For example, Palmer et al. (2019) demonstrated that communication impairment is an independent predictor for reduced social integration and participation, increased levels of loneliness and depression, and reduced social self-efficacy. Findings from this work are not isolated, Keidser et al. (2015) and Sung et al. (2016) emphasise the importance of communication as the conduit for social connection and its associated health and well-being impacts.

Oral communication is dynamic, spanning multiple interconnected domains of hearing, listening, language and cognition and is overlayed by contextual nuances that make up real-world communication. Listening experiences underpin the development and use of this dynamic complex (Nittrouer, 2002; Kuhl and Rivera-Gaxiola, 2008); hence, disruptions in listening caused by hearing loss can have broad impacts across this communication complex. The significant gap between traditional measures of hearing loss, such as hearing thresholds, and the pervasive expression of its effects across oral communication and social participation for an individual (Ambert-Dahan et al., 2018; Lin, 2020) fails to provide individuals (or their hearing healthcare professionals) with an understanding of one's full communication capacity (Manchaiah, 2017).

For adults with hearing loss, listening and communication ability are rarely measured in the context they are experienced (Beechey et al., 2019). From a diagnostic and device fitting perspective, standards are principally and necessarily focused toward measures of *hearing impairment* that enable a comparable numeric representation of hearing acuity. Assessments such as audiometric threshold measures provide a sensitive and valid representation of changes within the auditory pathway. However, these measures are associated with the integrity of the peripheral auditory pathway, thereby separating hearing from its role as part of a complex brain network, one that both precedes and provides the basis for listening (Stewart and Arnold, 2018). Clinically, the limitations of hearing measurement are commonly addressed with the inclusion of speech audiometry, which requires the listener to repeat single words or brief sentences. While also sensitive to changes in auditory function, speech-based measures involve the engagement of components of the complex brain network of listening, such as attention and linguistic knowledge. It is therefore logical to infer that this type of assessment adequately reflects the requirements of listening for communication.

Effective communication relates to the complex and interwoven systems that enable adults to participate, ask and answer questions, comment and understand indirect and often abstract language. To achieve this, adults need to be competent across the linguistic, social, and cognitive complexes that define and constitute communication. Additionally, real-world processing of acoustic information is strongly influenced by environmental, linguistic, contextual and production (speaker) factors (Gifford and Revit, 2010; Klatte et al., 2010). These factors affect the interpretation of speech signals and require cognitive mechanisms to engage, compensate and resolve frequent ambiguity (Rönnberg et al., 2013; Guediche et al., 2014; Baskent et al., 2016). Understanding this relationship has become an increasingly important consideration in the field of hearing, as listeners vary significantly in their ability to understand speech in complex environments and traditional audiological assessment can only partly explain this variation (Pichora-Fuller, 2003; Anderson and Kraus, 2010; Rönnerg et al., 2016).

Defining listening function in terms of a dynamic communicative complex has broad implications for both the individual and clinical practice. A reductionist conceptualisation of listening focussed on hearing impairment not only limits our understanding of how listening is experienced for an individual but may also fail to demonstrate the impacts of hearing impairment as a social, health and economic priority (Deloitte Access Economics, 2017; World Health Organization, 2017). In general, clinical audiology services are increasingly aware of the need to adapt hearing evaluations toward a more person-centred ideal (Boisvert et al., 2017). Measures that fully explore and provide an understanding of an individual's needs and prognosis in relation to different audiological interventions, however, seem to be lacking, which can affect the adoption and development of technology and rehabilitation programs (Rudner, 2016; Hughes et al., 2017).

The concern about the limitations of existing measures to adequately assess communication function in adults with hearing loss is not new (Cox et al., 2000; Moberly et al., 2018a). It is unclear, however, how knowledge of these limitations has influenced recent studies that assess functional abilities in adults with hearing loss. While self-report instruments have been identified as measures that could bridge assessment gaps

(Rivera et al., 2019; Shao et al., 2020), the constructs of listening and communication do not appear to be well conceptualised within existing self-reported measures for adults with hearing loss. In view of this, this scoping review aimed to identify measures used in recently published studies to evaluate skills that are necessary for oral communication in adults with hearing loss, and to map these measures in relation to constructs of listening and communication to assess potential gaps or biases in measurement.

# METHODOLOGY

This study used a systematic scoping review approach guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis extension for Scoping Reviews [PRISMA-ScR; 22] (Tricco et al., 2018).

## Eligibility Criteria

Published studies were included in this review if participants were adults (18 years and over) who reported or had been identified as having any hearing difficulty. The assessments used within the study had to meet the following criteria: 1) linguistic-based measurement relevant to oral communication, AND 2) behavioural or self-report measures of listening abilities, with listening ability defined as the conscious processing and response to an auditory stimulus. Cognitive assessments that included an auditory function element in the assessment of abilities [for example: Montreal Cognitive Assessment (MoCA)] were also included. Studies measuring vestibular function, tinnitus or hyperacusis (classified as additional symptoms as opposed to hearing or listening ability), device output measures, measures of hearing sensitivity only (e.g., detection thresholds), detection-based localisation, physiological or anatomical measures, and music-based measures that did not include a behavioural or self-reported linguistic measure of listening ability relevant to oral communication were excluded. To focus the review on listening assessments that were more likely to be used with hard-of-hearing adult participants, studies that included both paediatric and adult data were excluded as were studies with a sample size of fewer than ten hard-of-hearing adults.

## Information Sources

A systematic search of databases [Medline (Ovid), EMBASE (Ovid), Web of Science Core Collection (Web of Science) and Google Scholar] was initially performed in September 2018 and repeated in December 2019. This combination of four databases was selected in accordance with Bramer et al. (2017) findings which demonstrated a retrieval performance of 98.3% for systematic searches using this combination. Search terms and strategy were devised and supported with the assistance of a research librarian at Macquarie University. Keyword and related MeSH terms relevant to 'oral-communication', 'listening' and 'hearing' were combined with terms associated with 'hearing loss' and 'measurement'. The search strategy was limited by year of publication (2008-current) to focus on contemporary studies, and avoid duplication with a previous comprehensive systematic review of hearing outcome measures (Granberg et al., 2014). Publication language was limited to English; however, the assessment language was not restricted in the search criteria. The final search strategy applied with Medline (Ovid) is shown in **Supplementary Material 1**. The results of the searches were uploaded into the reference management software, Endnote X9.2 (Clarivate Analytics, Boston, MA, United States). Duplicates were removed and the remaining abstracts imported into Covidence (Covidence[1]) online systematic review management software. Deduplication was repeated in Covidence to ensure all duplicate records were removed prior to screening.

## Selection of Sources of Evidence

The main author (KN) and two research assistants (RF, RK) were involved in the screening of studies against the eligibility criteria. Each study was independently screened by a minimum of two reviewers. An initial screening of titles and abstracts was conducted to remove records of studies that were out of scope for this review. Full-text screening was conducted for the remaining records. Excluded records were labelled with a reason for their exclusion. Reviewers flagged any study that did not clearly meet the inclusion or exclusion criteria. Reasons for ambiguity, such as studies that indicated audiological or functional assessment but did not specify the measurements used, were labelled accordingly and retained or removed following a discussion between the reviewers. Persistent discrepancies at all stages were managed in consultation with a third reviewer (IB), with final decisions regarding study inclusion or exclusion reached through consensus-based discussion. Because this review aimed to identify measures used within published studies, critical appraisal of the methodological quality of the included studies was not considered relevant to the aims of the review and not undertaken.

## Data Charting Process

All eligible studies were charted independently by two members of the review team. Percentage agreement was used to determine inter-rater agreement and consistency. This was set as a minimum of 90% agreement, that is, 10% or less of charted items being categorised as a conflict (McHugh, 2012). Unclear or ambiguous information about measures used within a study was clarified by retrieving and reviewing the source measure (for example, the specific questionnaire used within a study).

### Coding Framework and Data Items

Data charting focused on extracting details of the assessment measures used in each study and study-specific information. Charting of assessment measures began by using the study tags within Covidence, and the charting of items was further refined using Microsoft Excel (2020). A bespoke coding framework to support data-charting was developed and piloted with 300 studies before being refined. All piloted studies were rescreened by two reviewers (KN, RF) to ensure that the refined coding scheme captured the relevant components. The coding

[1] http://www.covidence.org

**TABLE 1 |** Coding framework to categorise each measure used within the included studies.

| Study charting | Subcategory |
|---|---|
| Assessment Measures | |
| Detection (based response) | |
| Phoneme | Independent<br>Extracted from longer form stimuli |
| Word/sentence | |
| Word/sentence context | Quiet<br>Noise |
| Word/sentence auditory hierarchy | Detection<br>Discrimination<br>Recognition<br>Comprehension<br>None |
| Discourse | |
| Linguistic unit | Acceptable noise level judgement<br>Paralinguistic cues<br>Phonology<br>Semantic/Syntactic<br>Suprasegmental<br>Suprasegmental - Tonal language |
| Self-report measure | |
| Self-report assessment name | |
| Self-report category | Auditory<br>Non-Auditory<br>Unclear<br>Condition specific<br>Generic<br>Modifiable |
| Cognitive measure | |
| Cognitive measure assessment name | |
| Cognitive measure administration | Auditory<br>Non-auditory |
| Cognitive measure neurocognitive domain and/or type | DSM-5 Complex attention<br>DSM-5 Executive function<br>DSM-5 Learning & memory<br>DSM-5 Language<br>DSM-5 Social cognition<br>DSM-5 Perceptual-motor function<br>Unspecified<br>Screening<br>Diagnostic |

framework (**Table 1**) was designed to categorise measures as: (1) measures of linguistic constructs of functional listening relevant for communication; (2) self-report measures; and (3) cognitive measures.

For linguistic measures, key categories were derived initially based on a hierarchy of language unit components (i.e., from phonemes to discourse) and the level of auditory processing required (Estabrooks et al., 2020). Levels within the auditory hierarchy were defined as speech detection (the awareness of speech sounds), speech discrimination (the detection of changes in the acoustic stimuli), speech identification (the *recognition* of speech sounds, no semantic processing required; repetition of the stimuli), and speech comprehension (attaching meaning to the acoustic stimuli) (Erber, 1982; Thibodeau, 2007). Additional characteristics such as stimulus complexity (i.e., presented in quiet or in noise) were also extracted.

Charting of self-report measures identified hearing-specific measures as well as generic self-report measures that stated or implied the inclusion of auditory items relating to oral communication and functional language use. Charting included characteristics of the self-report measures such as single item, study-specific versus existing measure, and administration mode. Study-specific refers to measures that have been specifically developed or adapted (from existing formal assessment measures) for the purpose of a specific study. Formal measure describes previously published self-report assessments that are used within clinical studies and audiology clinics. All formal self-report measures where included irrespective of the extent of any psychometric evaluation of their measurement properties. When available, the target construct of study-specific measures [e.g., quality of life (QoL) or disability measurement] was extracted. For studies using published questionnaires, this information was reported based on the original description of the assessment, and classified into health status outcome domains. Health status domains reflect the status of individuals, in terms of conditions, functioning, and well-being. Categorisation into health status outcomes was derived from the principal description by the developers of respective measures, or from the description in the included studies from which the data was extracted (Barker et al., 2015; Madans and Webster, 2015). All accessible self-report measures, excluding study-specific measures, were sourced from the studies' attached appendices, original development papers or through correspondence with authors, for the items (individual questions) of each measure to be extracted for further analysis.

Cognitive measures that included a functional auditory element were identified and coded according to the six neurocognitive domains specified in The Diagnostic and Statistical Manual of Mental Disorders ($5^{th}$ ed.; DSM-5; American Psychiatric Association, 2013). The six principal domains as stated in the DSM-5 are: complex attention, executive function, learning and memory, language, social cognition and perceptual-motor function. The methods sections of the included articles were used to clarify the targeted cognitive domains for any tests that could be administered in more than one way. For example, the digit span test can be used to assess either forward or backward recall, which relate to different neurocognitive domains. Cognitive *screening* tests, which typically assess multiple domains, and studies in which three or more domain-specific diagnostic measures were used were categorised as *multi-domain measures (screening)* or *multi-domain measures (diagnostic)* respectively. The code "Unspecified" was used when studies did not provide sufficient information to determine the cognitive domain associated with the measures used. Publication details (year of publication), assessment language (English or Non-English), the dataset country of origin, study sample size, and hearing devices used by participants were also charted.

## Data Synthesis

Descriptive analyses were conducted to: (1) provide an overview of the types and frequency of measures used for the assessment of listening and communication in clinical studies; (2) determine if the representation of measurement types changed across time; and to (3) compare the content of assessments and their underlying constructs in comparison with broader constructs of functional listening and communication as described in the literature. Measures using speech-based stimuli were categorised according to: (1) a language unit hierarchy from the phonemic unit (minimal) to the discourse unit (maximal), and (2) an auditory hierarchy from speech detection (minimal) to comprehension (maximal). Division into these units was chosen to reflect the broad terms used to identify speech-based assessment material, the associated complexities related to appraising the details of the stimulus used (phoneme, word, sentence, discourse), and what was measured in relation to the task requested from the listener (imitation or comprehension). The distinction between imitation and comprehension, the targeted language unit and the auditory context (quiet/noise) represents different levels of listening complexity and engagement of cognitive mechanisms (Rodd et al., 2012; Moberly and Reed, 2019), factors key to determining the relationship of these measures to functional listening and communication. Data analyses and figures were prepared using a combination of Tableau Public (Tableau Public[2]) and Microsoft Excel (2020).

# RESULTS

## Included Studies

Details of search results and screening processes are shown in the Preferred Reporting Items for Systematic Reviews (PRISMA) diagram (**Figure 1**). From 16,069 records identified through the database and grey literature search, 6,948 duplicates were removed. The remaining 9,121 studies' titles and abstracts were reviewed against the inclusion criteria. Of these, 6,273 studies were excluded. A full text screening of the 2,848 potentially eligible studies resulted in an additional 269 exclusions, leaving 2,579 studies which included adults with hearing difficulties and contained a linguistic measurement relevant to oral communication.

## Study Characteristics

Overall, the number of studies that met the inclusion criteria increased during the period assessed (see **Figure 2**). Data originated from 41 countries with the United States of America ($n$ = 719 articles; 27.9%), the United Kingdom ($n$ = 196 articles; 7.6%), and Netherlands ($n$ = 172 articles; 6.7%) being the most represented. Two hundred and eighty-three studies (11.0%) presented data collected across multiple countries. Grouping by continent revealed that most publications originated from Europe ($n$ = 1023, 39.7%) followed by the Americas ($n$ = 928, 36.0%). Within the 2,579 included studies, 34.6% ($n$ = 892) of the measures were presented in a language other than English. Participant numbers ranged from 10 (minimum specified in inclusion criteria) to 7,210,535. Most studies used a small number of participants with a group sample size of 10-25 participants accounting for 31.5% and 26–100 participants for 35.7% of studies, respectively. Larger population-based studies ($n > 1000$) were represented 10.1% of the included studies.

## Characteristics of Measures Used Within the Included Studies

In total, 6,714 discrete assessment measures were extracted from the 2,579 included studies and charted in relation to the type of measure used (**Figure 3**) and their linguistic properties (**Table 2**). Detection-based responses [indicating the presence or absence of stimuli (tonal or other)] though not targeted for this review, were found in 74.7% of the included studies ($n$ = 1927/2579).

## Speech-Based Measures

The majority of studies ($n$ = 2178/2579, 84.5%) included a word or sentence measure, which accounted for 32.4% ($n$ = 6714) of the total measures identified. The most frequently used language unit was word or sentence identification presented in quiet (WSQ) ($n$ = 1699/2579; 65.9%) followed by word or sentence identification in noise (WSN) ($n$ = 1407/2579; 54.6%). Discourse-based measures, that extend beyond a single sentence and reflect the form and function of language in the social context, had the smallest representation with only 2.7% ($n$ = 69/2579) of studies. One-hundred and fifty-nine studies (6.2%) used a phonemic (smallest language unit) measure. The phoneme-based measures were from studies that specifically stated the use of phonemes as an individual measure or directly reported on phonemic outcomes as a separate language unit derived from word or sentence stimuli. The upper part of **Figure 3** illustrates the different categories of measures that used a speech-based stimuli.

When charting the word and sentence measures in relation to the auditory hierarchy (**Table 2A**), a high representation of speech *recognition* measures was found ($n$ = 1968; 90.4%) in comparison to measures of speech *comprehension* ($n$ = 72; 3.3%). Studies that used multiple levels of measurement, such as speech discrimination and speech comprehension, were categorised according to the highest auditory hierarchy level represented by the measures. Speech discrimination was used in 6.1% ($n$ = 132) of the studies and only five studies (0.2%) used word or sentence stimuli as a speech detection task.

A few studies reported on linguistic measurement aspects complementary to, or as a related functional characterisation of, speech-based stimuli ($n$ = 165/2579; 6.4%). Acceptable noise level judgement (ANLJ) tests that used speech material as the target stimuli were included in this grouping. **Table 2B** displays the other linguistic measures, found in 165 articles, categorised into their related linguistic domain. Suprasegmental features were assessed most often (35.8%; $n$ = 59/165), including both non-tonal (28.5%; $n$ = 47/165) and tonal languages (7.3%; $n$ = 12/165). Paralinguistic cues (aspects of spoken communication that add emphasis and meaning but are not in words, such as gesture and body language, conversational proximity, mood) were assessed the least (9.7%, $n$ = 16/165).

## Cognitive Measures

Measures of cognition were found in 13.3% ($n$ = 343) of all included studies (**Table 2D**). Eighty-seven studies (25.4%) used

[2]https://public.tableau.com/

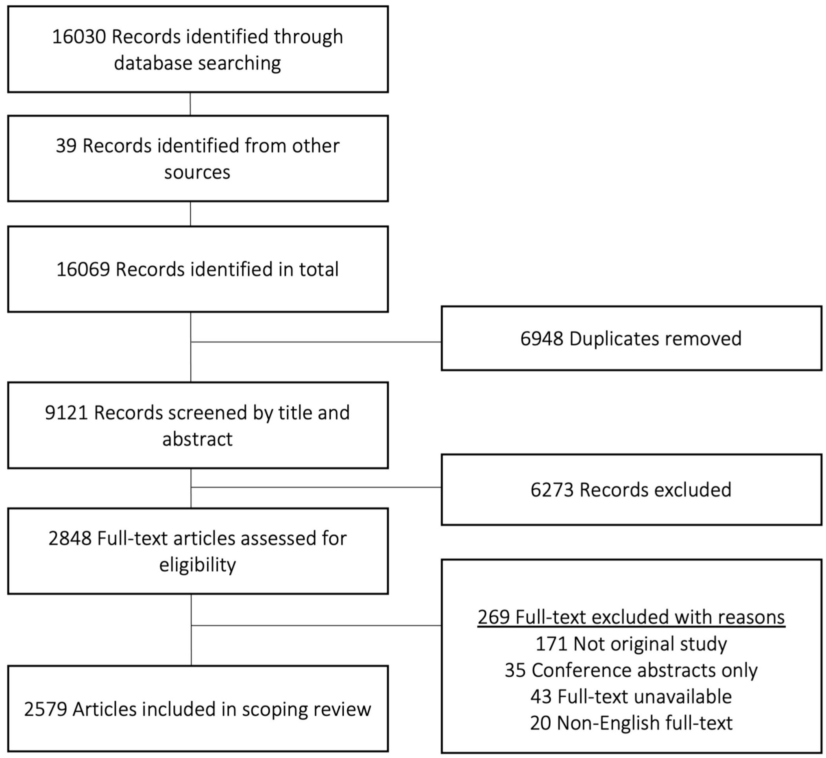


**FIGURE 1 |** Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (Moher et al., 2009).

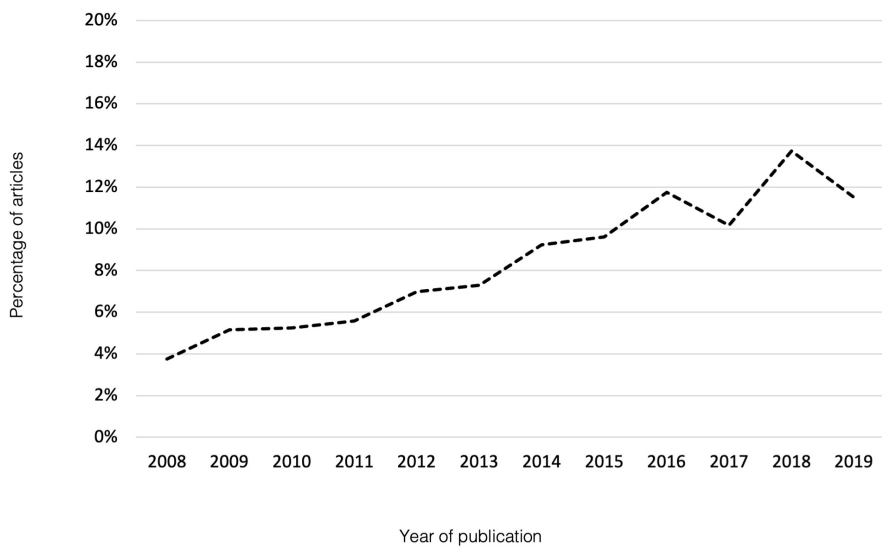


**FIGURE 2 |** Percentage of included articles by year of publication 2008–2019.

a cognitive measure that targeted a single cognitive domain. Multi-domain diagnostic cognitive measures were reported most commonly ($n$ = 135/343; 39.4%), with screening measures (single measures that assess multiple cognitive domains) used in 33.5% studies ($n$ = 115). The most frequently used cognitive screening measure was the Mini-Mental State Examination (MMSE) (Folstein et al., 1975; Lacritz and Hom, 1996). All reported cognitive measures were categorised into the target neurocognitive domains per the DSM-5. According to DSM-5 categorisation, 41.4% of studies ($n$ = 142) included a specific measure of executive function (which encompasses planning, decision making, working memory, responding to feedback/error

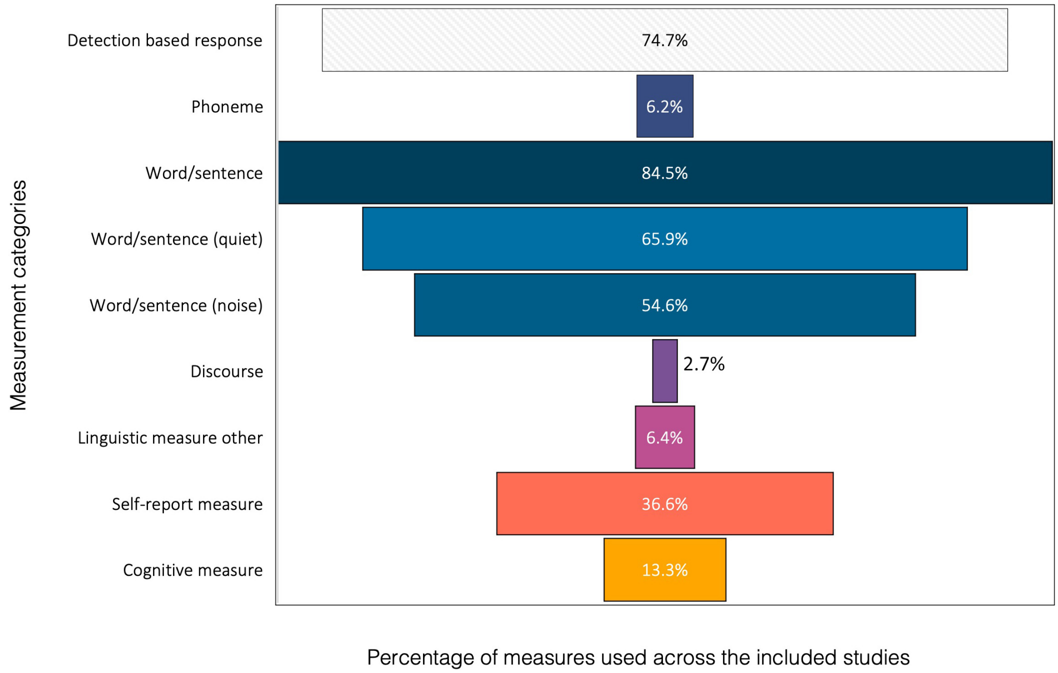


**FIGURE 3 |** Percentage of assessment measures (total *n* = 6714) by category (vertical axis) identified in the included studies (*n* = 2579 studies). Charting categorisation details for specific measurement categories (Word/Sentence; Linguistic units; Self-report measures; and Cognitive measures) are presented in **Table 2**.

correction, overriding habits/inhibition and mental flexibility). Measures of complex attention (including evaluation of sustained attention, divided attention, selective attention and processing speed) were present in 30.0% (*n* = 103/343) of studies utilising cognitive measures. Measures of language were used in 20.4% of studies (*n* = 71/343) and measures of learning and memory in 14.6% (*n* = 50/343) of studies. Measures of social cognition (such as assessment of emotion and theory of mind) were limited, with 2.0% (*n* = 7/343) of studies reporting measures related to this DSM-5 domain. Six studies, labelled as "unspecified," did not state the specific cognitive measure used or provided inadequate methodological information, preventing DSM-5 domain allocation during data charting.

## Self-Report Measures

One or more self-report measures were used in 945 of all included studies (36.6%; *n* = 945/2579). Including all previously published self-report measures (study-specific questionnaires, as well as single-question self-report measures), a total of 1306 self-report measures were found across 945 studies. A total of 139 previously published self-report measures, classified as either condition-specific (76.9%; *n* = 107/139) or generic (23.0%; *n* = 32/139), were extracted and subsequently categorised in terms of health status outcomes, based on Barker et al. (2015), Madans and Webster (2015) (**Table 2C**). These domains included: (1) communication; (2) device benefit; (3) disability; (4) health; (5) physiological; (6) psychological; (7) quality of life, and (8) other. As ambiguity exists in relation to definitions for constructs such as disability and quality of life, a number of self-report measures were found to cover multiple constructs. Detailed discussion relating to this issue is beyond the scope of this review but interested readers can refer to Eyssen et al. (2011), Milton (2013) for more information. For this review, disability was used as an umbrella term to encompass impairments, activity limitations, and participation restrictions as linked constructs (World Health Organization, 2001).

Condition-specific (auditory) disability represented 31.2% (*n* = 408/1306) of self-report measures, followed by measures of device benefit (21.2%; *n* = 277/1306). Measures targetting communication as the primary construct accounted for 1.3% (*n* = 18/1306) of the self-report measures used. Over 70% (*n* = 664/945) of studies used a single self-report measure, 20.0% (*n* = 189/945) used two self-report measures, 7.1% (*n* = 68/945) three self-report measures, 2.0% (*n* = 19/945) four self-report measures, and 0.4% (*n* = 4/945) used four or more self-report measures. Of the formal self-report measures identified across studies, the majority *n* = 84/139 (60.4%) were used in a single study. In total, 80.5% (*n* = 112/139) of formal measures were cited five times or fewer, indicating a lack of consistency in the selection of self-report measures in clinical studies. Measures designed explicitly for a study (i.e., study-specific) were the self-reported measures used in most studies (*n* = 315/945; 33.3%). The most frequently used psychometrically validated measures were the Speech, Spatial and Qualities of hearing (SSQ) scale (Gatehouse and Noble, 2004), the Abbreviated Profile of Hearing Aid Benefit [APHAB; (Cox and Alexander, 1995)] and the Hearing Handicap Inventory for the Elderly [HHIE; (Ventry and Weinstein, 1982)].

**TABLE 2 |** A. Word and sentence measures by auditory hierarchy; B. Linguistic measures by linguistic domain; C. Self-report measures by health status domain; and D. Cognitive measures by neuropsychological cognitive domain. Word/sentence measures are depicted as a total group (Word/sentence) and by presentation in either quiet [Word/sentence (quiet)] or noise [Word/sentence (noise)]. Percentages exceed 100% due to multiple measures used within studies.

| | N | % |
|---|---|---|
| **A Word/sentence by auditory hierarchy (*n* = 2178 ST)** | | |
| Speech detection | 5 | 0.2% |
| Speech discrimination | 132 | 6.1% |
| Speech recognition | 1968 | 90.4% |
| Speech comprehension | 72 | 3.3% |
| **B Linguistic units by domain (*n* = 165 ST)** | | |
| ANLJ | 30 | 18.2% |
| Paralinguistic | 16 | 9.7% |
| Phonology | 17 | 10.3% |
| Semantic/syntactic | 43 | 26.1% |
| Suprasegmental (non-tonal language) | 47 | 28.5% |
| Suprasegmental (tonal language) | 12 | 7.3% |
| **C Self-report by health status domain (*n* = 945 ST)** | | |
| Communication | 18 | 1.3% |
| Device benefit | 277 | 21.2% |
| Disability (condition specific) | 408 | 31.2% |
| Disability (generic) | 50 | 3.8% |
| Health | 36 | 2.7% |
| Other | 2 | 0.1% |
| Physiological | 1 | 0.1% |
| Psychological | 64 | 4.9% |
| Quality of Life | 135 | 10.3% |
| **D Cognitive measures by domain (*n* = 343 ST)** | | |
| Complex attention | 103 | 30.7% |
| Executive function | 142 | 41.4% |
| Learning & memory | 50 | 14.6% |
| Language | 70 | 20.4% |
| Social cognition | 7 | 2.0% |
| Perceptual-motor function | 14 | 4.1% |
| Unspecified domain | 6 | 1.7% |
| Single domain | 87 | 25.4% |
| Multidomain diagnostic assessment | 135 | 39.4% |
| Screening (multidimensional) | 115 | 33.5% |

## Representation of Assessment Measures Within Individual Studies

To assess whether the makeup of communication-relevant measures used in published studies had changed over time, the number of measures, categorised by measurement type, used in studies per year was graphed (**Figure 4**). While the total number of publications increased over time (**Figure 2**), the distribution of measures by measure type remained relatively consistent. Word and sentence measures, specifically measures in quiet, were the most frequently used assessment measure each year. When measures were grouped by measurement type, comparison of measures across years demonstrated the relatively narrow range of variability within groupings. There was less than ten percent variation between the lowest and highest percentage of measurement group by type for all categories. The exception was word and sentence measures in noise (WSN) which varied from 47.5% to 66.5%. The limited variation found in the representation of cognitive measurement across years was unexpected. The recent developments in the field of cognitive hearing science, which highlights the intrinsic role of cognition in listening (Arlinger et al., 2009; Lunner et al., 2020), and the publication of studies that showed a relationship between hearing loss and neurocognitive disorders such as dementia (Lin et al., 2011; Livingston et al., 2017), would intuitively have promoted an increase in the use of cognitive measures. The data extracted in this review suggests, however, that there was a proportional increase in the use of all types of measures relevant to listening and communication. Articles published in 2019 had the highest percentage of self-report measures with 42.8% (*n* = 127/297) of included studies using some form of self-report. Discourse measures were the most infrequently used form of measurement (range = 0.8% – 5.6%) across publication years.

# DISCUSSION

This scoping review identified and examined measures used within recently published studies to evaluate listening skills for oral communication in adults with hearing loss. In particular, using a linguistic perspective, the review provides a useful categorisation system to evaluate the capacity for existing measures to represent everyday communication as experienced by adults with hearing loss. Results from this review suggest that measures used to assess listening abilities target a narrow set of domains, limited predominantly to measures of speech detection and recognition at the word or sentence level, and that preferences for outcome measure selection have remained relatively constant for the last decade. Furthermore, despite these measurement preferences, there remains a lack of consensus within published studies regarding the selection of measures that target the complexities of listening and communication. The persistent focus on detection-based measures and the limited use of measures assessing complex/higher-level listening abilities suggests that current measures may not be evaluating those listening constructs of most relevance to adults with hearing loss when they are listening in the communication situations of everyday life.

## Measurement Bias – The Prevalence of Detection Measures

The prevalence of detection-based measures in the included articles points to a focus within outcome studies to undertake measurement at the level of impairment (i.e., hearing) and not at the level of disability or handicap (World Health Organization, 2001). These findings suggest that within published studies, assessment of hearing is conceptualised as an isolable function that is independent or disconnected

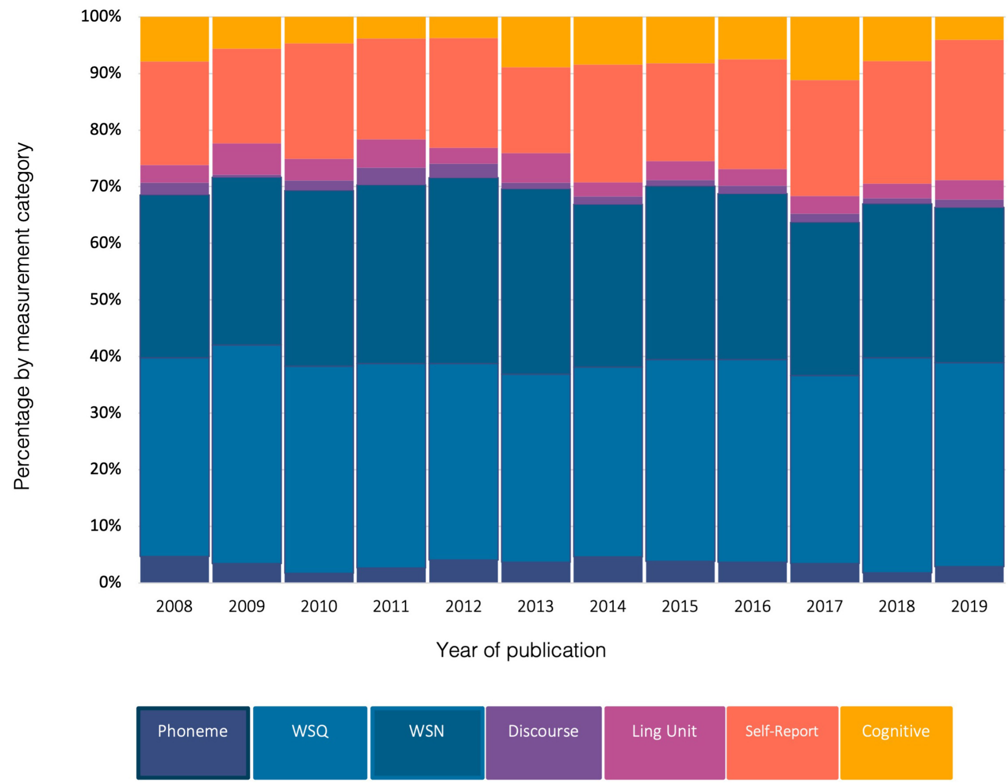


**FIGURE 4 |** Assessment measure, linguistic categories by year of publication. Total percentages exceed 100% due to multiple measures used within studies.

from its role in listening and communication (Manchaiah et al., 2019). While detection-based measures are valuable for classifying hearing levels, they provide limited information about functional listening ability in communicative contexts. Evidence indicates that detection-based measures do not provide information beyond hearing sensitivity (Engdahl et al., 2013; Fredriksson et al., 2016; Musiek, 2017; Phatak et al., 2019). For example, two listeners with the same audiometric thresholds can have different speech-in-noise performance (Gifford et al., 2007), and many individuals report significant hearing difficulties that are not reflected in hearing threshold measurement (Füllgrabe et al., 2015; Bakay et al., 2018; Barbee et al., 2018; Vermiglio et al., 2018). In contrast, a large study exploring access barriers to hearing intervention in older adults, found that 40% of adults with audiometrically measurable hearing loss did not report a hearing difficulty (Sawyer et al., 2020). The review findings indicate that, despite these well-evidenced shortcomings, measurement at the level of detection continues to be the dominant assessment measure reported in published studies of adults with hearing loss.

## Measurement and Language Units

Beyond detection, measures using words or sentences as stimuli were the measures most frequently identified in this review. This finding is consistent with an earlier systematic review of outcome measures in hearing loss in which word-level speech recognition measures with and without noise comprised the largest measurement group (Granberg et al., 2014). The high representation of word and sentence measures found in this review was expected as words or sentences represent the primary language unit to which contextual, linguistic and cognitive modifications are applied. The high prevalence of word and sentence-based measures has also been reported in a scoping review of outcome measures used to assess adults with cochlear implants (Boisvert et al., 2020).

The high proportion of word or sentence measures identified in this review is problematic, however, because, similar to detection-based measures, limitations also exist when using word and sentence stimuli, particularly in quiet conditions. For example, word and sentence stimuli administered in a quiet environment are prone to ceiling effects, correlate poorly with reports of listening abilities, and have low ecological validity (Firszt et al., 2004; Best et al., 2016b; Musiek, 2017). For example, a study appraising speech perception protocols for cochlear implant users demonstrated that, when tested in quiet, 28% ($n$ = 206) of participants achieved the maximum score of 100% (Gifford et al., 2008). While measures of speech perception are expected to correlate with each other, this study also found poor agreement between scores achieved in quiet and those achieved in noise for both monosyllables and sentences. Individual performance in quiet was not predictive of performance when measures used speech-based material in noise. Perhaps more significant from a functional perspective, difficulty listening in noise, not in quiet, is one of the most frequently reported auditory symptoms and a defining feature of adult hearing loss (Arlinger, 2009; Hughes et al., 2018; Pang et al., 2019).

## Language Units and Auditory Context

Attempts to address measurement limitations when using word and sentence in quiet stimuli frequently involves changes to the stimulus complexity (Klatte et al., 2010). Within this scoping review, noise was the most frequent modifier of word or sentence complexity. When viewed in relation to the challenges associated with listening in noise as reported by adults with hearing loss (Arlinger, 2009), the extensive use of words and sentences in noise measures have high face validity. The inclusion of noise in word and sentence-based measures has been found to: (1) contribute to a higher degree of diagnostic accuracy for the challenges of listening in noise (Vermiglio et al., 2018); (2) minimise ceiling effects associated with assessment of word and sentence recognition undertaken in quiet (Gifford and Revit, 2010); and (3) involve the engagement of additional cognitive mechanisms required to interpret degraded auditory input (Hwang et al., 2017). Therefore, changing stimulus complexity through the addition of noise may be a more realistic assessment of hearing and listening ability. There are, however, other considerations that may influence the representativeness of these measures. For instance, despite the preservation of some characteristics, the artificial noise generated as part of clinical testing protocols has little in common with the dynamic and reverberant acoustic environments encountered in everyday life (Weisser and Buchholz, 2019). Behaviours related to communicating in noise, such as speaker volume and physical proximity adaptations, are similarly not accounted for in existing measures. The adaptative behaviours of speakers assist with managing communication in varying noise levels and, therefore, may affect an individual's varying capacity for communication in these environments (Beechey et al., 2019). The preference for the addition of noise to create representative measurement in the included studies suggests a reductive approach to measurement that does not account for the impact and importance of cognitive and higher-level linguistic factors on interpersonal communication.

## Measurement Units and Communication

Current outcome measures use language unit boundaries (phoneme, word and sentence) to create discrete independent measurement units. Attempting to represent communication *via* these unit boundaries implies that these independent units are present and measurable in continuous speech streams. However, natural speech and language is not easily divisible into distinct, and seamlessly recognisable components (Walsh, 2011). The imperfections, deletions and ill-defined boundaries that are present in spontaneous communication, provide rich information used to contextualise and clarify spoken communication between communication partners (Podlubny et al., 2018). Dysfluencies, prosodic shifts and fillers support natural conversation, acting as recognisable markers in speech to signify the need for repetition or request for clarification between speakers (Corley and Stewart, 2008). These features are supportive communicative tactics, but current unit-based (phoneme, word and sentence) measures, either do not represent these features, or classify them as inaccurate responses that are scored accordingly, contrary to their supportive communicative function. From this perspective, reductive unit measures, such as phonemes, words and sentences, lack the dynamic and multimodal elements that define interactive communication as experienced by a listener in the real world.

## Auditory Hierarchy and Comprehension

Charting word and sentence measures in relation to the auditory hierarchy demonstrated the disproportionate representation of measures considered to assess speech recognition. Classification within the auditory hierarchy is valuable when considering the capacity of speech perception measures to characterise listening and communication ability. As a measure of auditory ability, speech recognition measures, which are typically based on clients repeating the individual items they hear, require limited linguistic prosessing and do not represent comprehension of the presented stimuli. Auditory comprehension, extracting meaning from auditory input, is crucial for oral communication competence. Extracting meaning from words and sentences changes the speech paradigm to engage a variety of linguistic (e.g., lexical, syntactic and phonological) and cognitive (e.g., working memory, attention, processing speed) mechanisms (Macdonald, 2017). The changes in load and task associated with comprehension enable more direct measurement of higher-level speech processes that are central to functional communication. Comprehension measures in this review sought to clarify these mechanisms relative to hearing loss and included, for example: processing structurally complex sentences and degraded speech (Carroll et al., 2016); neural activation in speech understanding (Zhou et al., 2018); suitability of dynamic speech materials to capture features specific to conversation (Best et al., 2016a); and the influence of syntactic form on plausibility (Amichetti et al., 2016). Interestingly, studies comparing measures of recognition and comprehension suggest that existing comprehension paradigms in the assessment of listening may be inadequate (Best et al., 2018).

The complex and continuous process of auditory comprehension in the listening situations of daily life is reliant on mechanisms that enable accurate interpretation of dynamic inferential and contextual information (Doedens and Meteyard, 2019), as well as socio-cognitive contributions such as theory of mind and self-regulation (Worthington, 2018). Without representation of these dynamic and dependent elements, measures of auditory comprehension may have a reduced capacity to represent real world communication ability. Similarly, and as suggested by the findings of this scoping review, the continued preference for studies to utilise measures of speech recognition, maintains a focus on reductive instruments that are unable to measure the complex processes of auditory comprehension and its contribution to day-to-day communication.

## Cognitive Assessment Measures

The operationalisation of listening and communication is dependent on cognition (Wolvin and Wiley, 2010). Cognitive mechanisms are required to attend to, make sense of, and remember auditory information – the prerequisite functions of listening and communication (Rost, 2016). Measures of cognition are therefore relevant to understanding the processing and individual expression of listening and communication (Lunner et al., 2009). For example, a recent study using a hearing-impairment simulation demonstrated that hearing loss does indeed impact cognitive-test performance, and this is not only due to reduced audibility (Füllgrabe, 2020).The studies identified in this review used measures of cognition for a variety of purposes: (1) to understand relationships between cognitive domains and listening (Amichetti et al., 2013; Ferguson and Henshaw, 2015; Keidser et al., 2016); (2) to account for variance in listening ability that is not identified within standard audiological measures (Kronenberger et al., 2014; Kaandorp et al., 2017; Moberly et al., 2018b); (3) as an indicator of neurocognitive function (Gates et al., 2008; Wong et al., 2014; Dawes et al., 2015); and (4) to determine if targetting cognition assists with rehabilitation (Pichora-Fuller and Levitt, 2012; Castiglione et al., 2016; Nkyekyer et al., 2018). As with language and communication, measurement of cognition as a separate and discrete function is complex. The included studies have addressed this complexity with multi-domain diagnostic assessments aimed at clarifying how cognitive ability is impacted by hearing loss (Mosnier et al., 2015; Claes et al., 2018). While studies that included domain-specific and multi-domain cognitive assessments are driving our understanding of cognition in relation to listening, language, and hearing loss (Rönnberg et al., 2019), they are underrepresented in this review. A significant proportion of studies exploring cognition used screening measures, which have noted limitations as the primary form of assessment. Raymond et al. (2020) systematic review of cognitive screening with adults with post-lingual hearing loss confirmed the frequent use of screening assessments such as the MMSE and the Montreal Cognitive Assessment (MoCA), both of which are reliant on auditory components. The authors note that based on the available evidence, these auditory components may have a deleterious effect on scores for adults with hearing loss. Therefore, poor performance may be an indication of poor cognition, poor audibility for instruction, or increased effort for listening, which is known to impact working memory and recall (Wayne et al., 2016). Adaptations to screening measures to adjust for auditory components have also proven problematic, as the removal and modification of items can directly influence the pass/fail status (sensitivity) (Parada et al., 2020) and these modifications may not yet have been formally validated (Dawes et al., 2019; Raymond et al., 2020).

Classification of the extracted outcome measures into linguistic categories indicated that standard measures used to assess hearing and listening were not designed to assess basic information-processing operations of listening and communication. In regard to functional communication, product or output measures, such as speech perception, may misrepresent the experience of listening with hearing loss by ignoring the cognitive involvement required in the task (Moberly et al., 2018a). Consequently, currently available outcome measures do not capture the functional variability that is evidenced in adults with hearing loss. The measures most frequently used in the included studies did not appear to capture the cognitive involvement required to attend to and process speech information or the effort intrinsic in communication adaptation and compensation (Hughes et al., 2017; Peelle, 2018). These measures also do not reflect what listening and communication mean for the individual driven by the motivation and need for social connectedness (Hughes et al., 2018).

## Self-Report Measures

The self-report measures included in this review were described in relation to health status outcomes and the number of times each measure was cited in clinical studies. Consistent with previous reviews of self-report in hearing loss, a large number of self-report measures were identified (Granberg et al., 2014; Akeroyd et al., 2015). Outcomes from these reviews showed that the majority of measures were not used repeatedly in clinical studies (Granberg et al., 2014; Akeroyd et al., 2015; Barker et al., 2015). The complications of many different self-report measures used infrequently across studies are compounded by the large number of studies that used a bespoke, or study-specific, self-report measure. This lack of consistency has the potential to constrain cross-study comparison and prevent data aggregation, limiting the use of data beyond an individual study. The pervasive impact of hearing loss may account for the diversity of targeted health status domains in self-reports. This diverse representation (e.g., disability, device benefit, QoL) may provide some explanation as to why so many self-report measures have been developed (Vas et al., 2017). Similarly, it may also reflect the inability of current measures to address the targeted health domains effectively (Barker et al., 2017). The volume and prevalence of self-report measures, however, suggests that criteria for selecting an appropriate measure is not evident, and currently no single standard measure is widely adopted in clinical studies (Akeroyd et al., 2015).

## Study Limitations

There were a number of limitations associated with this review. Despite the use of a comprehensive search strategy, it is possible that some studies were not included due to abstracts not indicating the use of linguistic-based measurement. Studies published in languages other than English that did not have an accompanying translation were not considered. As such, a language bias is present in this review. Excluding studies with sample sizes smaller than ten subjects potentially limited the extraction of all relevant measures. In addition to language, potential country specific bias may also reflect the legislative and policy contexts that mandate the inclusion of particular measures for use in the included studies. The high number of studies and broad country representation helped to address these biases. From a semantic and cognitive perspective, the terminology used to define measure types was indistinctive. Without clarification into levels within the auditory hierarchy, categorisation based on the level of speech processing assessed by the measure was not possible. Using linguistic categorisation to chart the extracted outcome measures presented limitations related to exploring language and communication from a compartmentalised perspective (Walsh, 2011). Given that functional communication encompasses components across multiple domains of the communication complex, utilisation of a theoretical framework may lead to reductionist conceptualisations of communication with hearing loss. Recognising this limitation of current conceptualisations of functional communication may help us understand why current measurement limitations exist. Finally, the allocation of self-report measures according to health domains may not accurately reflect the intended content of a measure's items. Manchaiah et al. (2019) study on content validity and readability in self-report measures of hearing disability demonstrated substantial variability in domain measurement. For example, measures were described as measures of disability; however, item analysis indicated the targetting of a number of additional constructs. The findings of this review, supported by Manchaiah et al. (2019) study lend support to the assertion that, without a rigorous evaluation of a measure's content validity, it may not be possible to understand fully the conceptual coverage provided by a measure's items (Terwee et al., 2018).

### Implications for Clinical Practice and Future Research Directions

This review outlines limitations in measures of listening and communication when these are viewed from a functional perspective. Findings from this review provide a reference to describe how outcome measures relate to the components of functional listening in daily life. This information could be used in clinical practice and research to provide a more nuanced evaluation of the listening abilities of adults with hearing loss. The reductive approach to measurement described in this review may account for the contrast between what is measured and the priorities and perspectives of adults living with hearing loss (Sawyer et al., 2020). The review findings may also assist in addressing the possible disconnect between people's understanding of hearing loss and its relationship to communication. While this work provides insight into the potential domains that may be relevant for measurement of functional communication, additional investigation is required to match these theoretical foundations to the communication experiences of adults with hearing loss. For example, qualitative approaches, when applied to understanding functional communication from the perspective of deaf and hard-of-hearing adults, may identify missing links within the listening and communication complex or provide insight into the weighting of different domains and items within that complex. Consultation with stakeholder groups, including adults with hearing loss and clinicians, to corroborate and extend the review findings could provide valuable insights on their usefulness leading to recommendations for policy and practice. This information could then inform the development and selection of outcome measures that better align to the lived experience of adults with hearing loss. Future work is required to evaluate the psychometric properties (the validity and reliability) of new and existing outcome measures in line with the target construct to be measured and the proposed context of use. Further work must also consider the costs (e.g., time, equipment and training required) in comparison to the benefits of selecting and implementing specific outcome measures within clinical or research contexts.

## CONCLUSION

Real-life communication is quick, responsive, dynamic, continuous and unpredictable. To be an effective communicator we need not only language, but the ability to incorporate and understand language in the context of others and the complexity that they bring with them. Listening is the foundation of oral communication, but there is currently no consensus on how to best represent and measure the complexities of everyday listening for communication in audiological clinical practice. By categorising the included outcome measures in terms of the complexity of the stimuli used, the participant's response required for the task, as well as the domains targeted within self-reports and cognitive measures that are relevant for listening, this scoping review highlighted both the reductive approach to measurement and the large and heterogenous pool of assessments available to measure functional listening in adults with hearing loss. Without consideration of the broader linguistic, cognitive and interactive elements of communication, measures cannot adequately capture the complex way adults with hearing loss experience listening for communication. To effectively represent functional listening, it will be necessary to expand how audiological measurement is conceptualised and undertaken to ensure functional listening for communication is measured in the context in which it is experienced and from the perspective of those who experience it.

## AUTHOR CONTRIBUTIONS

KN conceived and led the study as part of her masters of research thesis supervised by IB and CM and charted the data and prepared the drafts of the manuscript, tables, and figures with critical guidance from IB, CM, and SH. KN and IB conducted the systematic searches. All authors provided substantial input into the final draft submitted.

## FUNDING

SH receives funding from the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) West Midlands, UK Research and Innovation (UKRI), and declares personal fees from Aparito Limited and Cochlear Limited outside the submitted work. 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 thank Raneem Kazal and Rachael Ferguson for their help as research assistants for this review.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.786347/full#supplementary-material

## REFERENCES

Akeroyd, M. A., Wright-Whyte, K., Holman, J. A., and Whitmer, W. M. (2015). A comprehensive survey of hearing questionnaires: how many are there, what do they measure, and how have they been validated? *Trials* 16:26. doi: 10.1186/1745-6215-16-S1-P26

Ambert-Dahan, E., Laouenan, C., Lebredonchel, M., Borel, S., Carillo, C., Bouccara, D., et al. (2018). Evaluation of the impact of hearing loss in adults: validation of a quality of life questionnaire. *Eur. Annals Otorhinolaryngology-Head Neck Dis.* 135, 25–31. doi: 10.1016/j.anorl.2017.09.003

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, D.C.: American Psychiatric Association.

Amichetti, N. M., White, A. G., and Wingfield, A. (2016). Multiple solutions to the same problem: utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Front. Psychol.* 7:789. doi: 10.3389/fpsyg.2016.00789

Amichetti, N., Stanley, R., White, A., and Wingfield, A. (2013). Monitoring the capacity of working memory: executive control and effects of listening effort. *Memory Cogn.* 41, 839–849. doi: 10.3758/s13421-013-0302-300

Anderson, S., and Kraus, N. (2010). Sensory-Cognitive interaction in the neural encoding of speech in noise: a review. *J. Am. Acad. Audiol.* 21, 575–585. doi: 10.3766/jaaa.21.9.3

Arlinger, S. (2009). Negative consequences of uncorrected hearing loss—a review. *Int. J. Audiol.* 42, 17–20. doi: 10.3109/14992020309074639

Arlinger, S., Lunner, T., Lyxell, B. P.-F., and Katherine. (2009). The emergence of cognitive hearing science. *Scand. J. Psychol.* 50, 371–384. doi: 10.1111/j.1467-9450.2009.00753.x

Bakay, W. M. H., Anderson, L. A., Garcia-Lazaro, J. A., McAlpine, D., and Schaette, R. (2018). Hidden hearing loss selectively impairs neural adaptation to loud sound environments. *Nat. Commun.* 9:4298. doi: 10.1038/s41467-018-06777-y

Barbee, C., James, J., Park, J., Smith, E., Johnson, C., Clifton, S., et al. (2018). Effectiveness of auditory measures for detecting hidden hearing loss and/or cochlear synaptopathy: a systematic review. *Semin. Hear.* 39, 172–209. doi: 10.1055/s-0038-1641743

Barker, A. B., Leighton, P., and Ferguson, M. A. (2017). Coping together with hearing loss: a qualitative meta-synthesis of the psychosocial experiences of people with hearing loss and their communication partners. *Int. J. Audiol.* 56, 297–305. doi: 10.1080/14992027.2017.1286695

Barker, F., Mackenzie, E., Elliott, L., and De Lusignan, S. (2015). Outcome measurement in adult auditory rehabilitation: a scoping review of measures used in randomized controlled trials. *Ear Hear.* 36, 567–573. doi: 10.1097/AUD.0000000000000167

Baskent, D., Clarke, J., Pals, C., Benard, M. R., Bhargava, P., Saija, J., et al. (2016). Cognitive compensation of speech perception with hearing impairment, cochlear implants, and aging: how and to what degree can it be achieved? *Trends Hear*. 20:2331216516670279.

Beechey, T., Buchholz, J. M., and Keidser, G. (2019). Eliciting naturalistic conversations: a method for assessing communication ability, subjective experience, and the impacts of noise and hearing impairment. *J. Speech Lang. Hear. Res.* 62, 470–484. doi: 10.1044/2018_jslhr-h-18-0107

Best, V., Keidser, G., Buchholz, J. M., and Freeston, K. (2016a). Development and preliminary evaluation of a new test of ongoing speech comprehension. *Int. J. Audiol.* 55, 45–52. doi: 10.3109/14992027.2015.1055835

Best, V., Keidser, G., Freeston, K., and Buchholz, J. M. (2016b). A dynamic speech comprehension test for assessing real-world listening ability. *J. Am. Acad. Audiol.* 27, 515–526. doi: 10.3766/jaaa.15089

Best, V., Keidser, G., Freeston, K., and Buchholz, J. M. (2018). Evaluation of the NAL dynamic conversations test in older listeners with hearing loss. *Int. J. Audiol.* 57, 221–229. doi: 10.1080/14992027.2017.1365275

Boisvert, I., Clemesha, J., Lundmark, E., Crome, E., Barr, C., and McMahon, C. M. (2017). Decision-Making in audiology: balancing evidence-based practice and patient-centered care. *Trends Hearing* 21:2331216517706397. doi: 10.1177/2331216517706397

Boisvert, I., Reis, M., Au, A., Cowan, R., and Dowell, R. C. (2020). Cochlear implantation outcomes in adults: a scoping review. *PLoS One* 15:e0232421. doi: 10.1371/journal.pone.0232421

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., and Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Rev.* 6:78. doi: 10.1186/s13643-017-0644-y

Carroll, R., Uslar, V., Brand, T., and Ruigendijk, E. (2016). Processing mechanisms in hearing-impaired listeners: evidence from reaction times and sentence interpretation. *Ear Hear.* 37, e391–e401. doi: 10.1097/AUD.0000000000000339

Castiglione, A., Benatti, A., Velardita, C., Favaro, D., Padoan, E., Severi, D., et al. (2016). Aging, cognitive decline and hearing loss: effects of auditory rehabilitation and training with hearing aids and cochlear implants on cognitive function and depression among older adults. *Audiol. Neurotol.* 21, 21–28. doi: 10.1159/000448350

Claes, A. J., Van de Heyning, P., Gilles, A., Den Brandt, A. H.-V., Van Rompaey, V., and Mertens, G. (2018). Impaired cognitive functioning in cochlear implant recipients over the age of 55 years: a cross-sectional study using the repeatable battery for the assessment of neuropsychological status for hearing-impaired individuals (RBANS-H). *Front. Neurosci.* 12:580. doi: 10.3389/fnins.2018.00580

Corley, M., and Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: the meaning of um. *Lang. linguistics Compass* 2, 589–602. doi: 10.1111/j.1749-818X.2008.00068.x

Cox, R. M., and Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear Hear.* 16, 176–186. doi: 10.1097/00003446-199504000-199504005

Cox, R., Hyde, M., Gatehouse, S., Noble, W., Dillon, H., Bentler, R., et al. (2000). Optimal outcome measures, research priorities, and international cooperation. *Ear. Hear.* 21(4 Suppl.), 106S–115S. doi: 10.1097/00003446-200008001-00014

Dawes, P., Cruickshanks, K. J., Fischer, M. E., Klein, B. E., Klein, R., and Nondahl, D. M. (2015). Hearing-aid use and long-term health outcomes: hearing handicap, mental health, social engagement, cognitive function, physical health, and mortality. *Int. J. Audiol.* 54, 838–844. doi: 10.3109/14992027.2015.1059503

Dawes, P., Pye, A., Reeves, D., Yeung, W. K., Sheikh, S., Thodi, C., et al. (2019). Protocol for the development of versions of the Montreal Cognitive Assessment (MoCA) for people with hearing or vision impairment. *BMJ Open* 9:e026246. doi: 10.1136/bmjopen-2018-026246

Deloitte Access Economics (2017). *The Social and Economic Cost of Hearing Loss in Australia*. Sydney NSW: Hearing Care Industry Association.

Doedens, W. J., and Meteyard, L. (2019). The importance of situated language use for aphasia rehabilitation. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/svwpf

Engdahl, B., Tambs, K., and Hoffman, H. J. (2013). Otoacoustic emissions, pure-tone audiometry, and self-reported hearing. *Int. J. Audiol.* 52, 74–82. doi: 10.3109/14992027.2012.733423

Erber, N. P. (1982). *Auditory Training / Norman P. Erber*. Washington, D.C: Alexander Graham Bell Association for the Deaf.

Estabrooks, W., Morrison, H. M., and MacIver-Lux, K. (2020). *Auditory-verbal Therapy: Science, Research, and Practice*. San Diego, CA: Plural Publishing, Inc.

Eyssen, I. C., Steultjens, M. P., Dekker, J., and Terwee, C. B. (2011). A systematic review of instruments assessing participation: challenges in defining participation. *Arch. Phys. Med. Rehabil.* 92, 983–997. doi: 10.1016/j.apmr.2011.01.006

Ferguson, M. A., and Henshaw, H. (2015). Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. *Front. Psychol.* 6:556. doi: 10.3389/fpg.2015.00556

Firszt, B. J., Holden, K. L., Skinner, W. M., Tobey, A. E., Peterson, L. A., Gaggl, A. W., et al. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear Hear.* 25, 375–387. doi: 10.1097/01.aud.0000134552.22205.ee

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-90026

Fredriksson, S., Hammar, O., Magnusson, L., Kahari, K., and Waye, K. P. (2016). Validating self-reporting of hearing-related symptoms against pure-tone audiometry, otoacoustic emission, and speech audiometry. *Int. J. Audiol.* 55, 454–462. doi: 10.1080/14992027.2016.1177210

Füllgrabe, C. (2020). On the possible overestimation of cognitive decline: the impact of age-related hearing loss on cognitive-test performance. *Front. Neurosci.* 14:454. doi: 10.3389/fnins.2020.00454

Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Gatehouse, S., and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int. J. Audiol.* 43, 85–99. doi: 10.1080/14992020400050014

Gates, G. A., Anderson, M. L., Feeney, M. P., McCurry, S. M., and Larson, E. B. (2008). Central auditory dysfunction in older persons with memory impairment or Alzheimer dementia. *Arch. Otolaryngology-Head Neck Surgery* 134, 771–777. doi: 10.1001/archotol.134.7.771

Gifford, R. H., Bacon, S. P., and Williams, E. J. (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *J. Speech Lang. Hear. Res.* 50:857. doi: 10.1044/1092-4388(2007/060)

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (2008). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol. Neurootol.* 13, 193–205. doi: 10.1159/000113510

Gifford, R., and Revit, L. J. (2010). Speech perception for adult cochlear implant recipients in a realistic background noise: effectiveness of preprocessing strategies and external options for improving speech recognition in noise. *J. Am. Acad. Audiol.* 21, 441-451; quiz 487-448. doi: 10.3766/jaaa.21.7.3.

Granberg, S., Dahlström, J., Möller, C., Kähäri, K., Danermark, B., Linköpings, U., et al. (2014). The ICF core sets for hearing loss - researcher perspective. Part I: systematic review of outcome measures identified in audiological research. *Int. J. Audiol.* 53, 65–76. doi: 10.3109/14992027.2013.851799

Guediche, S., Blumstein, S. E., Fiez, J. A., and Holt, L. L. (2014). Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Front. Systems Neurosci.* 7:126. doi: 10.3389/fnsys.2013.00126

Heinrich, A., Gagne, J. P., Viljanen, A., Levy, D., Ben-David, B. M., and Schneider, B. A. (2016). Effective communication as a fundamental aspect of active aging and well-being: paying attention to the challenges older adults face in noisy environments. *Soc. Inquiry Well-Being* 2, 51–69. doi: 10.13165/SIIW-16-2-1-05

Hughes, E. S., Hutchings, A. H., Rapport, L. F., McMahon, M. C., and Boisvert, M. I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: a grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear.* 39, 922–934. doi: 10.1097/AUD.0000000000000553

Hughes, S. E., Rapport, F. L., Boisvert, I., McMahon, C. M., and Hutchings, H. A. (2017). Patient-reported outcome measures (PROMs) for assessing perceived listening effort in hearing loss: protocol for a systematic review. *BMJ Open* 7:e014995. doi: 10.1136/bmjopen-2016-014995

Hwang, J. S., Kim, K. H., and Lee, J. H. (2017). Factors affecting sentence-in-noise recognition for normal hearing listeners and listeners with hearing loss. *J. Audiol. Otol.* 21, 81–87. doi: 10.7874/jao.2017.21.2.81

Kaandorp, M. W., Smits, C., Merkus, P., Festen, J. M., and Goverts, S. T. (2017). Lexical-Access ability and cognitive predictors of speech recognition in noise in adult cochlear implant users. *Trends Hear.* 21:2331216517743887. doi: 10.1177/2331216517743887

Keidser, G., Rudner, M., Seeto, M., Hygge, S., and Ronnberg, J. (2016). The effect of functional hearing and hearing aid usage on verbal reasoning in a large community-dwelling population. *Ear Hear.* 37, e26–e36. doi: 10.1097/AUD.0000000000000196

Keidser, G., Seeto, M., Rudner, M., Hygge, S., and Ronnberg, J. (2015). On the relationship between functional hearing and depression. *Int. J. Audiol.* 54, 653–664. doi: 10.3109/14992027.2015.1046503

Klatte, M., Lachmann, T., and Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise Health* 12, 270–282. doi: 10.4103/1463-1741.70506

Kronenberger, W. G., Colson, B. G., Henning, S. C., and Pisoni, D. B. (2014). Executive functioning and speech-language skills following long-term use of cochlear implants. *J. Deaf Stud. Deaf Educ.* 19, 456–470. doi: 10.1093/deafed/enu011

Kuhl, P., and Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annu. Rev. Neurosci.* 31, 511–534. doi: 10.1146/annurev.neuro.30.051606.094321

Lacritz, L. H., and Hom, J. (1996). Mini-mental state examination (MMSE) severity ratings and neuropsychological functioning. *Arch. Clin. Neuropsychol.* 11:414. doi: 10.1016/0887-6177(96)83934-83935

Lin, F. R. (2020). Making sense of the senses in aging research. *J. Gerontol. A Biol. Sci. Med. Sci.* 75, 529–530. doi: 10.1093/gerona/glaa028

Lin, F. R., Ferrucci, L., Metter, E. J., An, Y., Zonderman, A. B., and Resnick, S. M. (2011). Hearing loss and cognition in the baltimore longitudinal study of aging. *Neuropsychology* 25, 763–770. doi: 10.1037/a0024238

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 13, 413–446. doi: 10.1016/S0140-6736(17)31363-6

Lunner, T., Alickovic, E., Graversen, C., Ng, E. H. N., Wendt, D., and Keidser, G. (2020). Three new outcome measures that tap into cognitive processes required for real-life communication. *Ear Hear.* 41, 39S–47S. doi: 10.1097/AUD.0000000000000941

Lunner, T., Rudner, M., and Rönnberg, J. (2009). Cognition and hearing aids. *Scand. J. Psychol.* 50, 395–403. doi: 10.1111/j.1467-9450.2009.00742.x

Macdonald, S. (2017). Introducing the model of cognitive-communication competence: a model to guide evidence-based communication interventions after brain injury. *Brain Injury* 31, 1760–1780. doi: 10.1080/02699052.2017.1379613

Madans, J. H., and Webster, K. M. (2015). "Health surveys," in *International Encyclopedia of the Social & Behavioral Sciences*, ed. J. D. Wright (Oxford: Elsevier), 725–730.

Manchaiah, V. (2017). Role of self-reported hearing disability and measured hearing sensitivity in understanding participation restrictions and health-related quality of life: a study with hundred and three older adults with hearing loss. *Clin. Otolaryngol.* 42, 924–926. doi: 10.1111/coa.12758

Manchaiah, V., Granberg, S., Grover, V., Saunders, G. H., and Hall, D. A. (2019). Content validity and readability of patient-reported questionnaire instruments of hearing disability. *Int. J. Audiol.* 58, 565–575. doi: 10.1080/14992027.2019.1602738

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. Med.* 22, 276–282. doi: 10.11613/bm.2012.031

Milton, C. L. (2013). The ethics of defining quality of life. *Nursing Sci. Quarterly* 26, 121–123. doi: 10.1177/0894318413477153

Moberly, A., and Reed, J. (2019). Making sense of sentences: top-down processing of speech by adult cochlear implant users. *J. Speech Lang. Hear. Res.* 62, 2895–2905. doi: 10.1044/2019_JSLHR-H-18-0472

Moberly, A. C., Castellanos, I., and Mattingly, J. K. (2018b). Neurocognitive factors contributing to cochlear implant candidacy. *Otol. Neurotol.* 39, E1010–E1018. doi: 10.1097/mao.0000000000002052

Moberly, A., Castellanos, J. I., Vasil, F. K., Adunka, B. O., and Pisoni, B. D. (2018a). Product" Versus "Process" measures in assessing speech recognition outcomes in adults with cochlear implants. *Otol. Neurotol.* 39, e195–e202. doi: 10.1097/MAO.0000000000001694

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G, and The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097

Mosnier, I., Bebear, J. P., Marx, M., Fraysse, B., Truy, E., Lina-Granade, G., et al. (2015). Improvement of cognitive function after cochlear implantation in elderly patients. *JAMA Otolaryngology– Head Neck Surg.* 141, 442–450. doi: 10.1001/jamaoto.2015.129

Musiek, F. E. (2017). Perspectives on the pure-tone audiogram. *J. Am. Acad. Audiol.* 28, 655–671. doi: 10.3766/jaaa.16061

Nittrouer, S. (2002). From ear to cortex: a perspective on what clinicians need to understand about speech perception and language processing. *Lang. Speech Hear. Serv. Sch.* 33, 237–252. doi: 10.1044/0161-1461(2002/020)

Nkyekyer, J., Meyer, D., Blamey, P. J., Pipingas, A., and Bhar, S. (2018). Investigating the impact of hearing aid use and auditory training on cognition, depressive symptoms, and social interaction in adults with hearing loss: protocol for a crossover trial. *JMIR Res. Protocols* 7:e85. doi: 10.2196/resprot.8936

Palmer, A. D., Carder, P. C., White, D. L., Saunders, G., Woo, H., Graville, D. J., et al. (2019). The impact of communication impairments on the social relationships of older adults: pathways to psychological well-being. *J. Speech Lang. Hear. Res.: JSLHR* 62, 1–21. doi: 10.1044/2018_JSLHR-S-17-0495

Pang, J., Beach, E. F., Gilliver, M., and Yeend, I. (2019). Adults who report difficulty hearing speech in noise: an exploration of experiences, impacts and coping strategies. *Int. J. Audiol.* 58, 851–860. doi: 10.1080/14992027.2019.1670363

Parada, J. C., Hillyer, J., and Parbery-Clark, A. (2020). Performance on the standard and hearing-impaired montreal cognitive assessment in cochlear implant users. *Int. J. Geriatric Psychiatry* 35, 338–347. doi: 10.1002/gps.5267

Peelle, E. J. (2018). Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000494

Phatak, A. S., Brungart, S. D., Zion, J. D., and Grant, W. K. (2019). Clinical assessment of functional hearing deficits: speech-in-noise performance. *Ear Hear.* 40, 426–436. doi: 10.1097/AUD.0000000000000635

Pichora-Fuller, M. K. (2003). Cognitive aging and auditory information processing. *Int. J. Audiol.* 42(Suppl. 2), S26–S22.

Pichora-Fuller, M., and Levitt, H. (2012). Speech comprehension training and auditory and cognitive processing in older adults. *Am. J. Audiol.* 21, 351–357. doi: 10.1044/1059-0889(2012/12-0025)

Podlubny, R. G., Nearey, T. M., Kondrak, G., and Tucker, B. V. (2018). Assessing the importance of several acoustic properties to the perception of spontaneous speech. *J. Acoust. Soc. Am.* 143, 2255–2268. doi: 10.1121/1.5031123

Raymond, M., Barrett, D., Lee, D. J., Peterson, S., Raol, N., and Vivas, E. X. (2020). Cognitive screening of adults with postlingual hearing loss: a systematic review. *Otolaryngology–Head Neck Surg.* 164, 49–56. doi: 10.1177/0194599820933255

Rivera, S. C., Kyte, D. G., Aiyegbusi, O. L., Slade, A. L., McMullan, C., and Calvert, M. J. (2019). The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. *Health Qual. Life Outcomes* 17:156. doi: 10.1186/s12955-019-1220-z

Rodd, J. M., Johnsrude, I. S., and Davis, M. H. (2012). Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cereb. Cortex* 22, 1761–1773. doi: 10.1093/cercor/bhr252

Rönnberg, J., Holmer, E., and Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int. J. Audiol.* 58, 247–261. doi: 10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rönnerg, J., Lunner, T., Ng, E. H., Lidestam, B., Zekveld, A. A., Sorqvist, P., et al. (2016). Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *Int. J. Audiol.* 55, 623–642. doi: 10.1080/14992027.2016.1219775

Rost, M. (2016). *Teaching and Researching Listening / Michael Rost*. New York, NY: Routledge.

Rudner, M. (2016). Cognitive spare capacity as an index of listening effort. *Ear Hear.* 37(Suppl. 1), 69S–76S. doi: 10.1097/AUD.0000000000000302

Sawyer, C. S., Armitage, C. J., Munro, K. J., Singh, G., and Dawes, P. D. (2020). Biopsychosocial classification of hearing health seeking in adults aged over 50 years in England. *Ear Hear.* 41, 1215–1225. doi: 10.1097/AUD.0000000000000839

Shadden, B. B. (1988). Communication behavior and aging: a sourcebook for clinicians. *Ear Hear.* 9:228. doi: 10.1097/00003446-198808000-198808044

Shao, D., Moberly, A. C., and Ray, C. (2020). Quality of life outcomes reported by patients and significant others following cochlear implantation. *Am. J. Audiol.* 29, 404–409. doi: 10.1044/2020_AJA-19-00101

Stewart, M. C., and Arnold, C. L. (2018). Defining social listening: recognizing an emerging dimension of listening. *Int. J. Listening* 32, 85–100. doi: 10.1080/10904018.2017.1330656

Sung, Y. K., Li, L., Blake, C., Betz, J., and Lin, F. R. (2016). Association of hearing loss and loneliness in older adults. *J. Aging Health* 28, 979–994. doi: 10.1177/0898264315614570

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., et al. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual. Life Res.* 27, 1159–1170. doi: 10.1007/s11136-018-1829-1820

Thibodeau, L. (2007). "Speech audiometry," in *Audiology Diagnosis*, 2nd Edn, ed. R. Hosford-Dunn (New York, NY: Thieme Publishing Group), 228–313.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR) : checklist and explanation. *Ann. Intern. Med.* 169, 467–473. doi: 10.7326/M18-0850

Vas, V., Akeroyd, M. A., and Hall, D. A. (2017). A data-driven synthesis of research evidence for domains of hearing loss, as reported by adults with hearing loss and their communication partners. *Trends Hear.* 21:233121651773408. doi: 10.1177/2331216517734088

Ventry, I. M., and Weinstein, B. E. (1982). The hearing handicap inventory for the elderly: a new tool. *Ear Hear.* 3, 128–134. doi: 10.1097/00003446-198205000-198205006

Vermiglio, A., Soli, S. D., and Fang, X. (2018). An argument for self-report as a reference standard in audiology. *J. Am. Acad. Audiol.* 29, 206–222. doi: 10.3766/jaaa.16128

Wallhagen, M. I., and Pettengill, E. (2008). Hearing impairment. *J. Gerontol. Nursing* 34, 36–42. doi: 10.3928/00989134-20080201-20080212

Walsh, R. (2011). Looking at the ICF and human communication through the lens of classification theory. *Int. J. Speech-Lang. Pathol.* 13, 348–359. doi: 10.3109/17549507.2011.550690

Wayne, R. V., Hamilton, C., Huyck, J. J., and Johnsrude, I. S. (2016). Working memory training and speech in noise comprehension in older adults. *Front. Aging Neurosci.* 8:49. doi: 10.3389/fnagi.2016.00049

Weisser, A., and Buchholz, J. M. (2019). Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *J. Acoust. Soc. Am.* 145, 349–349. doi: 10.1121/1.5087567

Wolvin, A. D., and Wiley, I. (2010). *Listening and Human Communication in the 21st century*. Malden, MA: Wiley-Blackwell.

Wong, L. L. N., Yu, J. K. Y., Chan, S. S., and Tong, M. C. F. (2014). Screening of cognitive function and hearing impairment in older adults: a preliminary study. *BioMed Res. Int.* 2014:867852. doi: 10.1155/2014/867852

World Health Organization (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO Press.

World Health Organization (2017). *Global Costs of Unaddressed Hearing Loss and Cost-effectiveness of Interventions: a WHO report, 2017*. Geneva: World Health Organization.

Worthington, D. L. (2018). "Modelling and measuring the cognitive components of listening," in *The Sourcebook of Listening Research: Methodology and Measures*, eds D. L. eds D. L. Worthington and G. Bodie (Hoboken, NJ: John Wiley & Sons, Inc).

Zhou, X., Seghouane, A., Shah, A., Innes-Brown, H., Cross, W., Litovsky, R., et al. (2018). Cortical speech processing in postlingually deaf adult cochlear implant users, as revealed by functional near-infrared spectroscopy. *Trends Hear.* 22:2331216518786850. doi: 10.1177/2331216518786850



**Conflict of Interest:** SH receives funding from the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) West Midlands, UK Research and Innovation (UKRI) and declares personal fees from Aparito Limited and Cochlear Limited outside the submitted work. 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.

The remaining 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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/fnins.2022.789565

# Measuring Speech Intelligibility and Hearing-Aid Benefit Using Everyday Conversational Sentences in Real-World Environments

*Kelly Miles*[1]*, *Timothy Beechey*[2], *Virginia Best*[3] *and Jörg Buchholz*[1]

[1] ECHO Laboratory, Department of Linguistics, Macquarie University, Sydney, NSW, Australia, [2] Hearing Sciences – Scottish Section, School of Medicine, University of Nottingham, Glasgow, United Kingdom, [3] Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, United States

OPEN ACCESS

**Edited by:**
Astrid van Wieringen,
KU Leuven, Belgium

**Reviewed by:**
Inyong Choi,
The University of Iowa, United States
Hartmut Meister,
University of Cologne, Germany
Kirsten Carola Wagener,
Hörzentrum Oldenburg GmbH,
Germany

***Correspondence:**
Kelly Miles
kelly.miles@mq.edu.au

**Specialty section:**
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience

**Received:** 05 October 2021
**Accepted:** 17 February 2022
**Published:** 17 March 2022

**Citation:**
Miles K, Beechey T, Best V and Buchholz J (2022) Measuring Speech Intelligibility and Hearing-Aid Benefit Using Everyday Conversational Sentences in Real-World Environments.
Front. Neurosci. 16:789565.
doi: 10.3389/fnins.2022.789565

Laboratory and clinical-based assessments of speech intelligibility must evolve to better predict real-world speech intelligibility. One way of approaching this goal is to develop speech intelligibility tasks that are more representative of everyday speech communication outside the laboratory. Here, we evaluate speech intelligibility using both a standard sentence recall task based on clear, read speech (BKB sentences), and a sentence recall task consisting of spontaneously produced speech excised from conversations which took place in realistic background noises (ECO-SiN sentences). The sentences were embedded at natural speaking levels in six realistic background noises that differed in their overall level, which resulted in a range of fixed signal-to-noise ratios. Ten young, normal hearing participants took part in the study, along with 20 older participants with a range of levels of hearing loss who were tested with and without hearing-aid amplification. We found that scores were driven by hearing loss and the characteristics of the background noise, as expected, but also strongly by the speech materials. Scores obtained with the more realistic sentences were generally lower than those obtained with the standard sentences, which reduced ceiling effects for the majority of environments/listeners (but introduced floor effects in some cases). Because ceiling and floor effects limit the potential for observing changes in performance, benefits of amplification were highly dependent on the speech materials for a given background noise and participant group. Overall, the more realistic speech task offered a better dynamic range for capturing individual performance and hearing-aid benefit across the range of real-world environments we examined.

**Keywords: speech intelligibility, hearing aid benefit, realistic speech, clinical assessment development, speech in noise, ECO-SiN**

## INTRODUCTION

Among the primary functions of speech-in-noise testing are the prediction of speech intelligibility and device benefit outside the clinic or laboratory conditions in which testing is conducted. However, numerous studies have identified discrepancies between the results of speech testing and self-reported speech understanding and device benefit in everyday settings (Working Group on Speech Understanding, Committee on Hearing, Bioacoustics, and Biomechanics, 1988; Cord et al., 2004; Walden and Walden, 2004; Pronk et al., 2018; Wu et al., 2019). For example, using the Hearing in Noise Test (HINT; Nilsson et al., 1994), Cord et al. (2004) found that benefit from directional microphones measured in the laboratory was not predictive of perceived benefit outside the laboratory. Using the same speech test, Wu et al. (2019) found benefits of directional microphones and digital noise reduction but found no such benefits using self-report scales. Similarly, Walden and Walden (2004) found a lack of evidence for any relationship between aided or unaided QuickSIN (Killion et al., 1998) results and subjective ratings of hearing aid benefit once age was taken into account. Speech tests appear to be particularly prone to overestimating real-world outcomes, often showing overly high word recognition scores at rather low (negative) signal-to-noise (SNR) ratios. Such overestimation is problematic because it can mask the need for further rehabilitation or device optimization and can also disguise rehabilitation and device benefits through ceiling effects. That is, overestimation of speech intelligibility can both underplay and overplay the benefit of interventions. A related problem arises when measuring the speech reception threshold (SRT), in which the SNR is adapted to reach a certain performance point (e.g., 50% correct word identification). Even though the SRT is widely used in clinics, as it is quick and avoids floor and ceiling effects, it results in rather arbitrary test SNRs that are driven by the listener's performance rather than by real-world SNRs.

Overestimation of real-world performance has led researchers to identify the need for more challenging speech tests (Wackym et al., 2007; Gifford et al., 2008). However, common strategies that may be used to increase the difficulty of speech tests tend to result in speech test materials that are less, rather than more, representative of everyday speech signals. For example, testing at highly negative SNRs increases test difficulty but does not reflect conditions in which people usually need to understand speech, or conditions to which hearing aid features such as compression or adaptive beamforming are best suited or are most likely to be in operation. Word or syllable recognition tasks are more challenging than sentence tests (see for example Olsen et al., 1997) but do not provide the many levels of context normally available to the listener. And, speech tests that are paired with concurrent tasks, such as memorization, are more challenging than singleton tasks but do not closely reflect the cognitive load of everyday speech perception, such as procedural memory demands (Caplan, 2016). It is therefore unlikely that making speech tests more difficult in ways that serve to make speech materials less similar to natural speech signals will provide greater external validity or more accurate real-world predictions.

To create speech tests which can provide more generalizable results it is necessary to account for the cause of overestimation of real-world performance, rather than finding arbitrary ways to make speech tests more challenging. A potential cause can be seen if we consider the differences in perceptual cues provided to listeners by clear speech of the type employed in speech test materials, and conversational speech that is frequently encountered in daily life. Like any complex signal originating in the environment, speech signals consist of multiple redundant cues (Brunswik, 1955). These cues are in a probabilistic, rather than a deterministic, relationship with perceptual targets such as articulated speech features or segments (Blumstein and Stevens, 1981; Heald et al., 2016). Speech tests may overestimate real-world speech perception abilities because speech test materials provide much more robust or reliable segmental cues than are available in conversational speech (Payton et al., 1994; Ferguson and Kewley-Port, 2002; Ferguson, 2012; Ferguson and Quene, 2014). In contrast to clear speech, spontaneous, conversational speech is characterized by high rates of phonetic reduction (Johnson, 2004; Ernestus et al., 2015; Tucker and Ernestus, 2016) and relatively high and variable articulation rates (Miller et al., 1984). For example, excised portions of conversational speech are often unintelligible in isolation (Pollack and Pickett, 1963; Winitz and LaRiviere, 1979), indicating that to understand conversational speech, listeners cannot rely on segmental cues to the extent possible when listening to clear speech. As a result, clear speech of the type employed in speech test materials is more intelligible than conversational speech (Krause and Braida, 2004) but less representative.

By this logic, one approach to improving the predictive capabilities of speech testing is to incorporate features of conversational speech, such as phonetic reductions and realistic speech rates, into the test materials. Including features found in conversational speech has the dual benefit of increasing both the difficulty and realism of speech tests. We recently took this approach in developing the Everyday COnversational Sentences in Noise (ECO-SiN) test (Miles et al., 2020). The ECO-SiN materials were derived from interlocutors conversing in different kinds of realistic background noise, presented via open headphones. This naturally led to variations in vocal effort (e.g., Lombard speech; Lombard, 1911) as well as other accommodations in speaking rate and style (Cooke et al., 2014; Beechey et al., 2018). As a result, when ECO-SiN speech is presented in the noise in which it was produced, it sounds natural and avoids mismatches in level and spectra that listeners are sensitive to Hendrikse et al. (2019).

Our expectation is that the naturalistic aspects inherent to the ECO-SiN sentences will make them less intelligible than clearly articulated sentences typical of existing speech tests. However, at the same time, their vocal effort is appropriate for situations involving background noise, which should enhance the SNR at mid to high frequencies (Badajoz-Davila and Buchholz, 2021). The potential speech intelligibility benefit provided by this SNR boost may interact with the hearing status of the listener if hearing loss restricts access to the additional speech information due to limited audibility, temporal fine structure processing, or spatial processing (e.g., Rana and Buchholz, 2018). It is unclear how the combined effect of these different aspects of realistic effortful speech will affect intelligibility, particularly in realistic noise, and how this may interact with hearing loss and non-linear amplification provided by hearing aids.

To better understand the effect of using more realistic speech materials on hearing outcomes, we directly compared the intelligibility of the highly realistic ECO-SiN sentences to that of more traditional sentences when each were presented in six different realistic background noises. The speech and noise signals were presented at their realistic (fixed) levels (and thus SNRs) and performance was quantified by the percentage of words correctly recognized. Our evaluation included young listeners with normal hearing as well as older listeners with hearing loss, who are ultimately the target population for new and more effective approaches to speech testing. Listeners with hearing loss were assessed unaided and aided to also determine the effect of hearing-aid amplification on speech scores. The outcomes of this exploratory study are intended to highlight the advantages (and possible disadvantages) of increasing the realism of the speech materials in the assessment of speech perception in realistic background noise.

## MATERIALS AND METHODS

### Participants

Ten young adults with normal hearing (NH) and 20 older adults with hearing loss were recruited as part of a larger study. All participants reported that they were native Australian-English speakers and had no known cognitive or neurological problems. The NH group had audiometric thresholds below 20 dB HL at all audiometric frequencies between 250 and 8,000 Hz. The requisites for admission into the group with hearing loss were symmetrical sensorineural hearing loss with no more than one audiometric pure-tone threshold differing by more than 10 dB between the ears. Four frequency (0.5, 1, 2, and 4 kHz) average hearing loss (4FAHL) was calculated for each individual, and participant groups were established based on the following criterion according to Clark (1981): mild (20 dB HL $\leq$ 4FAHL $<$ 40 dB HL); moderate (40 dB HL $\leq$ 4FAHL $<$ 55 dB HL), and moderate-severe (55 dB HL $\leq$ 4FAHL $<$ 70 dB HL) hearing loss. For those with mild losses, we used the less fine-grained distinction between slight and mild classifications, as per Jerger and Jerger (1980). This grouping was employed as it is how the on-site audiology clinic categorized patients, and as such, how our recruitment efforts were structured. Descriptive statistics of the participants are summarized in **Table 1**. Using multiple two-sample $t$-tests found no significant differences in age between the three groups with hearing loss ($p > 0.1$) but showed that 4FAHLs were significantly different ($p < 0.05$ using Bonferroni corrections). **Figure 1** (left panel) illustrates the individual audiograms (thin lines, averaged across the ears) and the group averages (thick lines) for each of the groups with hearing loss (mild, moderate, and moderate-severe) along with the individual 4FAHLs (right panel). Participants received monetary gratuity for participating in the study. The study was approved by the Macquarie University Human Research Ethics Committee.

### Sentence Materials

The realistic sentence materials were drawn from the ECO-SiN corpus (cf. Miles et al., 2020). The ECO-SiN corpus comprises 192 naturally spoken sentences, in which four lists of 16 sentences were spoken with three different vocal efforts. The average sentence length is 6.3 words, and an example sentence is “That discovery was like really interesting for me.” In brief, the sentences were extracted from two people engaging in unscripted conversation while they listened to three different realistic background noises from the ARTE database (Weisser et al., 2019b); a church, an indoor café, a busy food court (see **Table 2**) via highly open headphones. The background noises were selected based on the conversational speech levels determined by Weisser and Buchholz (2019). The resultant speech levels corresponded to normal, raised, and loud vocal efforts as described in ANSI-S3.5. (1997). All ECO-SiN sentences presented here were spoken by one Australian-English speaking female talker. The female talker was chosen (as opposed to the other male talker of the ECO-SiN corpus) to provide the best point of comparison with the reference sentences (see below) which are spoken by a female talker.

The more traditional (reference) materials were drawn from a corpus of “BKB-like” sentences created by the Cooperative Research Centre for Cochlear implant and Hearing Aid Innovation (CRC HEAR). These sentences are similar to the original BKB sentences (Bench et al., 1979), however, the BKB-like corpus contains more sentences and was recorded with an Australian-English speaking female. The corpus has 80 lists in total, with each list consisting of 16 sentences. The average sentence length is 4.9 words and a n example sentence is “The clown had a funny face.” The scripted and clearly spoken sentences were produced in a sound-attenuated booth with the intention of being easily understood by 5-year-old children. The average spectrum of the BKB-like sentences is normalized to match the long-term average speech spectrum (LTASS) described by Byrne et al. (1994). The BKB-like sentences (hereafter referred to as BKB sentences) are widely used in research laboratories (e.g., see Dawson et al., 2013; Rana and Buchholz, 2016; Bentsen et al., 2019) and hearing clinics throughout Australia and were therefore considered here as an appropriate reference material.

The average spectrum of the speech materials is shown in **Figure 2** (left panel) for the BKB sentences (black stars) and the ECO-SiN sentences, separately for the normal (blue squares), raised (magenta diamonds), and loud (red circles) vocal effort. The spectra were derived in 3rd-octave bands for an unweighted RMS level of 65 dB SPL and averaged across all available sentences (i.e., the 1,280 sentences of the BKB material and the 64 sentences for each effort level of the ECO-SiN material). Compared to the BKB sentences, the ECO-SiN sentences provide a significant energy boost at mid-frequencies between 800 and 4,000 Hz, which further increases with increasing vocal effort level.

The corresponding temporal modulation spectra of the different speech materials are shown in the right panel of **Figure 2**. The modulation spectra were derived by concatenating all sentences for a given speech material into a single signal, which was then bandpass filtered using an A-weighting filter to focus roughly on the frequency range most relevant for speech perception. The amplitude of the resulting signal was squared, analyzed by a modulation filterbank with one-octave wide filters, and the power in each modulation channel calculated in dB. The resulting modulation spectrum was then normalized to its maximum value for easier comparison across speech materials. The modulation spectra exhibit a modulation bandpass

**TABLE 1 |** Descriptive statistics of the 10 NH participants and 20 participants with hearing loss.

| | NH participants | Participants with hearing loss | | | |
|---|---|---|---|---|---|
| | | All | Mild | Moderate | Moderate-severe |
| Number | 10 | 20 | 6 | 9 | 5 |
| Age (Years) | 23.1 ± 4.7 | 74.2 ± 5.2 | 74.2 ± 4.2 | 71.6 ± 5.2 | 76.8 ± 5.2 |
| 4FAHL (dB HL) | < 20 | 47.0 ± 11.4 | 32.3 ± 3.6 | 48.7 ± 3.9 | 60.0 ± 5.6 |

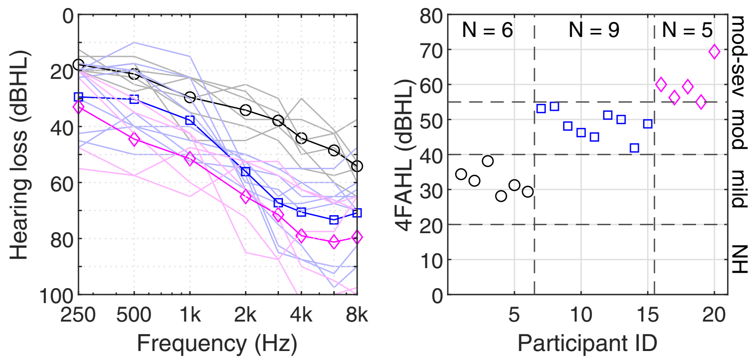


**FIGURE 1 |** Pure-tone audiograms of the participants with hearing loss averaged across ears (left panel) and their corresponding 4FAHL (right panel). The thin lines in the left panel refer to the individual audiograms and the thick lines with symbols to the audiograms averaged within groups.

**TABLE 2 |** Details of the realistic environments and speech materials.

| ID | Environment | Noise level | RT (Sec) | Speech level (dB SPL) | SNR (dB) | Vocal effort | |
|---|---|---|---|---|---|---|---|
| | | (dB SPL) | | | | ECO-SiN | BKB |
| 1 | Office | 58 | 0.2 | 63.4 | 5.4 | Normal | N/A |
| 2 | Church | 62.5 | 1.2 | 65.4 | 2.9 | | |
| 3 | Living room | 66.9 | 0.2 | 67.4 | 0.4 | Raised | |
| 4 | Cafe | 71.4 | 1.1 | 69.3 | −2.1 | | |
| 5 | Dinner party | 75.9 | 0.4 | 71.3 | −4.6 | Loud | |
| 6 | Food court | 80.3 | 1 | 73.3 | −7.1 | | |

*Numbers are rounded.*

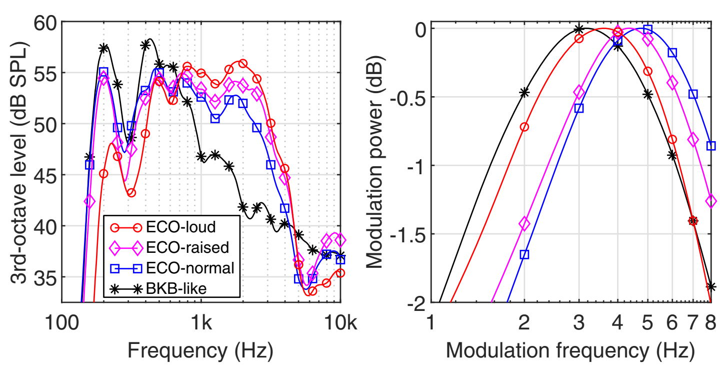


**FIGURE 2 |** Third-octave spectrum (left panel) and modulation spectrum (right panel) for the different speech materials and effort levels at an average broadband level of 65 dB SPL.

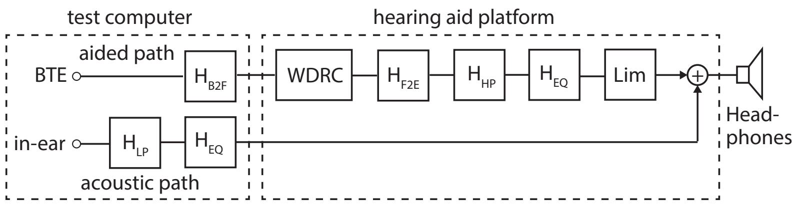


**FIGURE 3 |** Block diagram of the acoustic and aided signal path from the HATS' in-ear and BTE (front) microphone to the headphones for binaural playback. Only the pathway for the left or right ear is shown here.

characteristic with a center frequency that changes across speech materials and effort levels. Considering the center frequency as a rough estimator of the average talking rate, the talking rate in the BKB sentences is the slowest (3.2 Hz) and for the ECO-SiN sentences decreases with increasing vocal effort: normal (4.8 Hz), raised (4.4 Hz), and loud (3.6 Hz).

## Acoustic Environments

The background noises were drawn from the ARTE database (Weisser et al., 2019b), which were recorded with a 62-channel hard-sphere microphone array and encoded into the higher-order Ambisonics (HOA) format. They were then decoded here for simulated playback with the spherical 41-channel loudspeaker array inside the anechoic chamber of the Australian Hearing Hub, Macquarie University. **Table 2** shows the selected environments, their associated noise levels (i.e., the unweighted sound pressure level calculated over the entire recording of 150 s) and reverberation times (RT) in free-field, and the mapping of the ECO-SiN and BKB speech materials to the environments. The environments consisted of (1) an open plan office that was separated into cubicles using acoustically absorptive wall dividers, and contained people typing, chatting, and talking on the phone; (2) a small church with people entering and chatting before service; (3) a small living room with access to a kitchen in the back, with a television presenting commercials and kitchen sounds from the back; (4) an indoor café at medium occupancy with people chatting and diverse kitchen and coffee making noises; (5) a dining room with eight people chatting and laughing over a table and background music; and (6) a very large and noisy food court in a shopping mall at lunch time, which produced a very diffuse and stationary babble-like noise.

The speech levels for the six different environments (see **Table 2**) were derived from Equation 9 of Weisser and Buchholz (2019), who measured realistic SNRs in different realistic environments, including the ones used in the current study. In this equation, the gender-averaged SNR of two talkers sitting at a head-to-head distance of 1 m was considered, and the noise levels were slightly adjusted from their original levels to result in fixed SNR steps of 2.5 dB. To maximize the realism of the ECO-SiN sentences, and thereby to optimize their perceptual integration with the background noise, realistic room reverberation was added by convolving the individual sentences with multi-channel Room Impulse Responses (RIRs). The RIRs were taken from the ARTE database (Weisser et al., 2019b) and measured in the real-world environments with a loudspeaker at a distance of 1.3 m in front of the 62-channel microphone array. As for the noise recordings, the measured RIRs were encoded into the HOA format and decoded for simulated playback with the 41-channel loudspeaker array. Thereby, to compensate for the difference in the measured (1.3 m) and simulated (1 m) source-receiver distance, the direct sound was separated from the individual RIRs using a frequency-dependent time window, amplified such that the broadband direct-sound-to-reverberation energy ratio in free-field was increased by $20 \times \log(1.3\ \text{m}/1\ \text{m}) = 2.3$ dB, and then added back to the RIRs. To reduce the apparent source width of the direct sound, its impulse response was integrated across all 41 loudspeaker channels before it was added back to only the frontal channel of the RIRs. The anechoic BKB sentences were presented only from the frontal position.

Note that the speech levels given in **Table 2** refer to the average broadband free-field levels of the anechoic BKB sentences and the direct-sound only (i.e., anechoic) ECO-SiN sentences. The free-field levels of the reverberant ECO-SiN sentences were slightly higher than the values shown in **Table 2**, the reverberation providing an increase in the effective test SNR by: + 0.8 dB, church: + 0.1 dB, living room: + 1.2 dB, café: + 0.8 dB, dinner party: + 1.6 dB, and food court: + 0.5 dB. For a detailed description of the microphone array recording, HOA encoding and decoding, and the RIR manipulation process see Weisser et al. (2019b).

## Binaural Playback and Hearing-Aid Amplification

The loudspeaker signals for the different noise and speech conditions were transformed into binaural signals by simulating their playback via the 41-channel loudspeaker array to the in-ear microphones of a Bruel and Kjaer (Skodsborg Vej 307, 2850 Naerum, Denmark) type 4128C Head and Torso Simulator (HATS). Additionally, to enable the integration of a pair of hearing aids in the binaural playback, behind-the-ear (BTE) hearing aid satellites were placed above the left and right ear of the HATS. These purpose-built satellites were provided by Sonova AG (Laubisrütistrasse 28, 8712 Stäfa, Switzerland) and included front and rear microphones that were connected to a purpose-built pre-amplifier. The playback simulation path included individual loudspeaker equalization filters as well as measured impulse responses from each of the 41 loudspeakers to the six microphones at the left and right ears of the HATS: two in-ear microphones and four hearing aid microphones. However, only the front hearing aid microphones were used in this study to realize an omni-directional hearing aid input. Further details of the playback simulation process can be found in Weisser and Buchholz (2019).

**Figure 3** illustrates the implemented acoustic and aided signal path from the in-ear and front BTE microphones to the headphones used for binaural playback in the listening tests. Since the signal paths are identical at the left and right ear only one ear is shown here. The acoustic path describes the sound that arrives directly at the listener's ear drum (i.e., the in-ear microphone) and circumvents any hearing aid fitting (or ear mold). This path includes a low-pass filter, $H_{LP}$, to mimic the passive attenuation of the hearing aid fitting as well as a headphone equalization filter, $H_{EQ}$. The equalization filter ensured a flat frequency response of the headphones when measured on the HATS. The aided path describes the signal path from the hearing aid microphone via the hearing aid processing to the headphones. This path includes (1) a BTE microphone to free-field transformation filter, $H_{B2F}$, that removes the acoustic head shadow for a frontal sound source and provides a free-field equivalent output; (2) a multi-channel wide dynamic range compressor (WDRC) as the main hearing aid processing; (3) a free-field to ear-drum transformation filter, $H_{F2E}$, that basically reintroduces the acoustic head shadow for a frontal sound source but as recorded by the in-ear microphone; (4) a high-pass filter to simulate the limited sensitivity of the hearing aid receiver at low frequencies; (5) the same headphone equalization filter used in the acoustic path; and (6) an instantaneously acting broadband limiter, *Lim*, to protect the listener from excessively loud sounds.

A standard desktop computer was used to run the listening tests and to play the different 4-channel speech and noise stimuli via a RME Fireface UC (Audio AG, Am Pfanderling 60, 85778 Haimhausen, Germany) USB sound card to a second desktop computer with an RME Audio Fireface UFX USB sound card. The second computer ran a real-time hearing-aid research platform developed at the National Acoustic Laboratories, Hearing Australia, and presented the (aided) binaural stimuli to the participants via Beyerdynamic (Theresienstrasse 8, 74072 Heilbronn, Germany) DT990 headphones. All stimulus playback was realized at a sampling frequency of 44.1 kHz except for the hearing aid platform, which operated at a sampling frequency of 24 kHz and was band-limited to about 10 kHz.

The low-pass filter, $H_{LP}$, and high-pass filter, $H_{HP}$, shown in **Figure 3** were both realized by second order Butterworth IIR filters with different cut-off frequencies to approximate the acoustic attenuation by an ear mold with a vent size of 1, 2, and 3.5 mm. The cut-off frequencies were 620, 883, and 1,371 Hz for the low-pass filter and 311, 470, and 926 Hz for the high-pass filter. The filters approximated the gain data provided by Dillon (2001, page 127, Figure 5.11) and Dillon (2001, p. 127, Table 5.1), respectively, and presented a wide range of fittings from an almost open fitting (3.5 mm) to an almost closed fitting (1 mm). For each participant with hearing loss, the vent size was selected based on their low-frequency hearing loss (LFHL) as given by their ear-averaged pure-tone threshold at 500 Hz. Based on a discussion with local audiologists, the vent sizes were 3.5 mm for LFHL $\leq$ 20 dBHL, 2 mm for 20 dBHL $<$ LFHL $\leq$ 30 dBHL, and 1 mm for LFHL $>$ 30 dBHL. The WDRC realized basic syllabic compression within 16 independent frequency channels and acted independently across ears. It was fitted to the individual participant (and ear) using the NAL-NL2 gain prescription formula (Keidser et al., 2012). The instantaneous broadband limiter, *Lim*, was part of the sound card of the hearing aid platform and was set to an attack time of 0 ms, a release time of 100 ms, a compression ratio of 6, and a knee-point of 95 dB SPL. The limiter was significantly engaged only for the participants with moderate and moderate-severe losses, and then only in the loudest environments. For NH participants as well as participants with hearing loss in the unaided conditions, materials were presented through the acoustic path only, with the lowpass filter removed (i.e., set to a flat gain of 0 dB; see **Figure 3**). This rather complicated approach of using headphone reproduction with a hearing aid research platform was chosen here over a multi-loudspeaker system with off-the-shelf hearing aids to maximize control of the entire signal path from the acoustic free field through the hearing aid processing to the signals at the listener's ears. Arguably, such a system may also be easier to use within a hearing clinic.

## Procedure

Individual word recall ability was measured in the six realistic acoustic environments using both the realistic ECO-SiN and the more traditional BKB sentence materials at realistic (fixed) noise and speech levels, and thus SNRs (see **Table 2**). The sentences were always presented from the front. The NH participants were tested unaided, and the participants with hearing loss were tested both unaided and aided. Participants were seated together with the test administrator in a sound attenuating test booth with double walls. In each test condition, a 2.3-min-long noise sample was played in a loop and the 16 sentences in a list were presented in random order. Each time a sentence was presented, the participants recalled aloud all the words they heard. The administrator then scored the number of correctly recalled words on a graphical user interface that was invisible to the participant, and a new sentence was played. Preceding each sentence presentation was a 1 kHz beep to signal to the participant that a sentence was about to be played.

The order of the six background noises and the two speech materials (i.e., 12 test conditions) was randomized. These test conditions were blocked for the participants with hearing loss within the unaided and aided conditions due to the required manual reconfiguration of the hearing-aid platform. The two blocks were tested in random order.

# RESULTS

## Speech Intelligibility Scores

**Figure 4** shows mean intelligibility scores in each environment for unaided (top row) and aided (middle row) listening. Within each panel, data are shown for each listener group and for the two speech materials. For NH listeners the intelligibility scores in the quieter environments were all at ceiling and only decreased in the loudest environments. This decrease was more pronounced for the ECO-SiN than the BKB material, leading to generally higher

BKB scores in the louder environments. When listening unaided, all of the participant groups with hearing loss showed higher BKB scores than ECO-SiN scores in all of the environments, but the magnitude of the difference varied with the environment. For listeners with mild loss, the difference increased in the louder environments as the influence of ceiling effects was reduced. For listeners with moderate-severe hearing loss, the opposite pattern was observed, with the difference between BKB and ECO-SiN scores decreasing in the louder environments as floor effects came into play. When amplification was provided for listeners with hearing loss, intelligibility scores generally improved. As for the unaided condition, BKB scores were generally higher than ECO-SiN scores across all environments. Because of the overall shifts in the intelligibility functions, however, the magnitude of the speech material differences varied differently across environments.

To quantify the effect of speech material (BKB vs. ECO-SiN) on unaided and aided speech intelligibility scores, a Bayesian Beta regression model was fitted (Ferrari and Cribari-Neto, 2004) using the R-INLA package (Rue et al., 2017). Intelligibility scores were modeled as proportions as a function of categorical predictor variables for speech material, hearing loss group, and acoustic environment. A random intercept for individual subjects was included to account for repeated measures. The results of this analysis are provided in **Table 3**. Focusing on the contrast between BKB and ECO-SiN scores, for NH listeners, predicted mean scores were significantly higher for BKB than for ECO-SiN sentences in the café, dinner party, and food court environments (but not in the office, living room, or church environments). For listeners with hearing loss, the difference between speech materials was significant in all environments for both unaided and aided conditions.

## Hearing-Aid Benefit

Hearing-aid benefit was calculated by subtracting the unaided speech intelligibility percentage score from the aided speech intelligibility percentage score for each individual, separately for the BKB and ECO-SiN materials, with positive values indicating that amplification provided an improvement in speech intelligibility. Mean benefits are shown in the bottom row of **Figure 4**.

Given the complex behavior of the unaided and aided scores described in section "Speech Intelligibility Scores," the differences between them were also complex and were strongly affected by floor and ceiling effects. The largest aided benefits were observed for the listeners with moderate-severe hearing loss in the quietest environments. In those same environments, ceiling performance tended to reduce or eliminate the measurable benefit for better-performing listeners with milder losses. For the louder environments (e.g., the food court), floor effects meant that benefits of amplification were generally not observed for the listeners with moderate-severe hearing losses. In these louder environments though, better performing listeners who were not at floor demonstrated negative benefits (or "disbenefits"). In some cases, the magnitude of the benefit clearly depended on the type of speech material used.

To quantify the effect of speech material (BKB vs. ECO-SiN) on hearing-aid benefit, a robust regression model with a Student-T noise distribution was fitted to model hearing-aid benefit data which is not constrained to the [0, 1] interval. The results of this analysis are provided in **Table 4**. Focusing again on the differences between BKB and ECO-SiN materials, this analysis found significantly larger ECO-SiN benefits in the office and church environments for the listeners with moderate hearing loss only. In the living room and café environments, benefits were significantly larger for the BKB materials in the listeners with moderate-severe hearing loss. In the dinner party environment, the effect of speech material was significant only for the listeners with mild hearing loss, who showed larger *disbenefits* for the ECO-SiN materials.

## Relationship Between BKB and Everyday Conversational Sentences in Noise Scores and Benefits

**Figure 5** shows individual listener scores for ECO-SiN sentences as a function of their scores for BKB sentences when listening unaided (top row, excludes NH listeners) and with non-linear amplification (bottom row). Consistent with the observations made in section "Speech Intelligibility Scores," the majority of the points lie below the diagonal, indicating that ECO-SiN scores were lower than BKB scores achieved by most individuals.

A quantile regression model was fitted to compare the behavior of the individual ECO-SiN scores as a function of BKB scores in each environment and aiding condition with the predicted slopes describing the relative spread of the distributions of ECO-SiN and BKB scores. Quantile regression was used because it is robust to outliers and makes no assumptions about the underlying distribution of the data. The regression lines in **Figure 5** show predicted median ECO-SiN score as a function of performance on the BKB task. A slope of 1 would indicate that ECO-SiN scores change at the same rate as BKB scores, whereas a slope greater than 1 indicates that ECO-SiN scores change more than BKB scores and a slope less than 1 indicates that ECO-SiN scores change less than BKB scores. A higher rate of change indicates greater spread of scores and a wider distribution, while a lower rate of change indicates that scores are more concentrated within a small range, corresponding to a narrow distribution such as data accumulating at floor (0, i.e., 0%) or ceiling (1, i.e., 100%).

In unaided conditions, ECO-SiN and BKB scores show very similar spreads in the office (slope = 1.1; CI = 0.97, 1.25; $p < 0.001$), church (slope = 1.28; CI = 0.56, 1.85; $p < 0.001$), living room (slope = 0.93; CI = 0.63, 1.51; $p < 0.001$), and café (slope = 1.06; CI = 0.66, 1.69; $p < 0.001$) environments. In the two loudest environments there is a trend toward lower rates of change in ECO-SiN scores relative to BKB scores, with a slope of 0.75 (CI = 0.48, 1.19; $p < 0.001$) in the dinner party environment and a slope of 0.56 (CI = 0.20, 0.95; $p = 0.003$) in

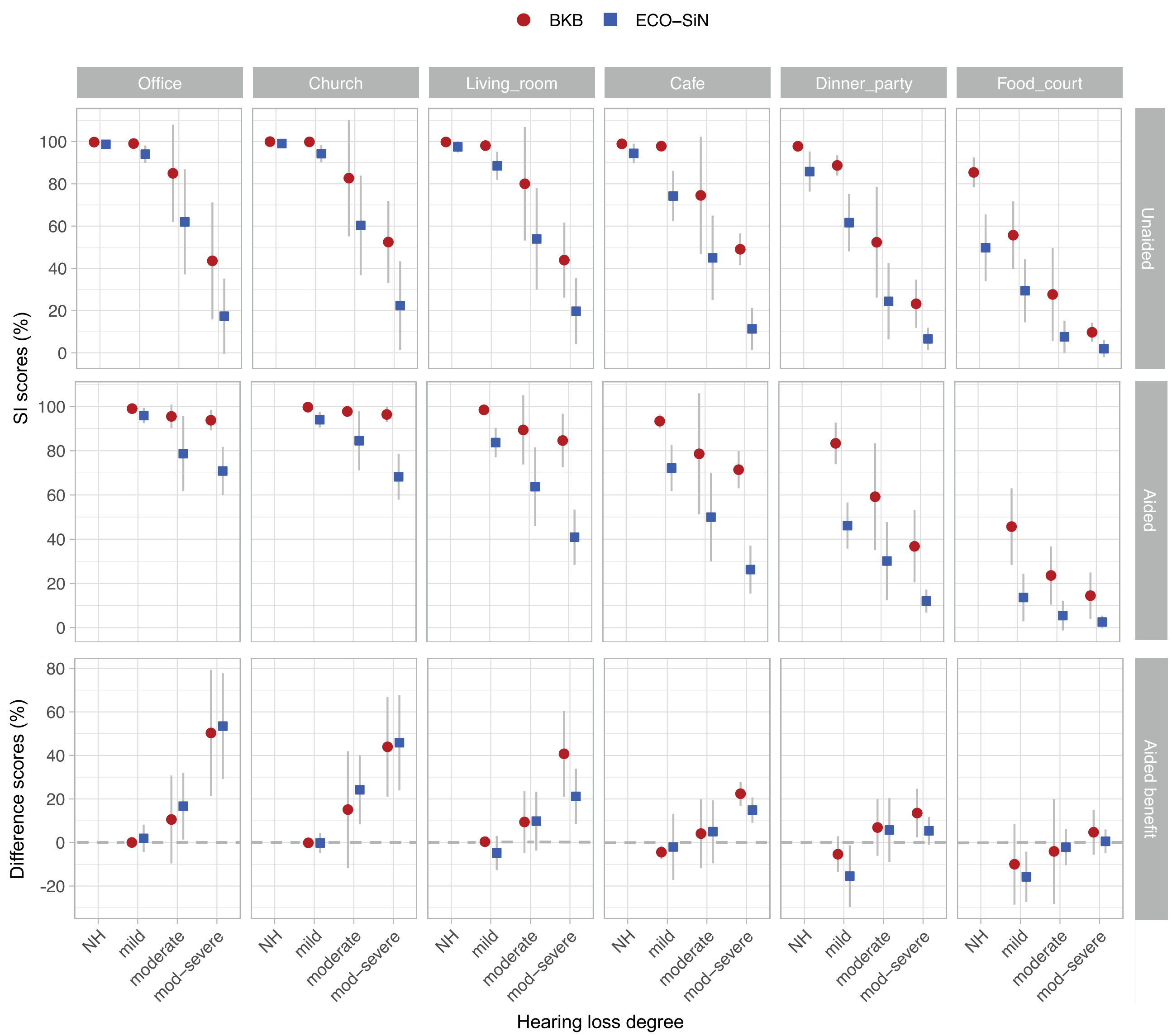


**FIGURE 4 |** Speech intelligibility (SI) scores obtained with BKB and ECO-SiN sentences. Shown are group means in percent correct (error bars show standard deviations). Top row: unaided listening; middle row: aided listening; bottom row: hearing-aid benefit (difference between aided and unaided scores). Note the different *y*-axis scales.

the food court environment. In the aided conditions, a similar trend is seen with slopes becoming progressively shallower in the louder environments. We see high relative rates of change in ECO-SiN scores in the four softest environments including the office (slope = 2.51; CI = 0.20, 3.67; $p$ = 0.005), church (slope = 4.32; CI = 0.42 7.31; $p$ = 0.014), living room (slope = 1.55; CI = −1.73, 5.28; $p$ = 0.39) and café (slope = 1.77; CI = −1.5, 5.56; $p$ = 0.33). Very low relative rates of change in ECO-SiN scores occurred in the two loudest environments including the dinner party (slope = 0.46; CI = −2.58, 3.91; $p$ = 0.79) and the food court (slope = 0.39; CI = −3.06, 3.83; $p$ = 0.84).

A significant relationship can be observed between the individual ECO-SiN and BKB scores for all environments when individuals were unaided, and in the quietest environments when aided. Hence, within many of the individual test conditions, a linear model can reasonably well predict the individual ECO-SiN scores from the corresponding BKB scores. However, this is not the case across the different environments and aiding conditions, where a far more complicated relationship exists between the two speech materials. Hence, knowing a BKB score in a single test condition does not allow prediction of the individual score in another environment nor the

**TABLE 3 |** Results of the statistical analysis comparing intelligibility scores for the two types of speech materials.

| | Degree of HL | Aiding | Difference | Low 95% CI | Upper 95% CI |
|---|---|---|---|---|---|
| Office | NH | Unaided | 0.55 | −1.10 | 2.44 |
| Office | Mild | Aided | 6.17* | 0.84 | 13.77 |
| Office | Mild | Unaided | 5.58* | 0.61 | 12.96 |
| Office | Moderate | Aided | 17.63* | 8.13 | 29.73 |
| Office | Moderate | Unaided | 33.17* | 17.39 | 49.09 |
| Office | Moderate-severe | Aided | 24.16* | 7.32 | 43.66 |
| Office | Moderate-severe | Unaided | 33.69* | 14.30 | 54.09 |
| Church | NH | Unaided | 0.59 | −0.82 | 2.26 |
| Church | Mild | Aided | 5.31* | 0.79 | 11.95 |
| Church | Mild | Unaided | 5.72* | 1.31 | 12.52 |
| Church | Moderate | Aided | 12.88* | 5.57 | 22.92 |
| Church | Moderate | Unaided | 28.19* | 12.12 | 44.74 |
| Church | Moderate-severe | Aided | 33.14* | 16.38 | 52.30 |
| Church | Moderate-severe | Unaided | 35.98* | 13.26 | 57.58 |
| Living room | NH | Unaided | 1.54 | −0.55 | 4.22 |
| Living room | Mild | Aided | 14.29* | 5.17 | 26.89 |
| Living room | Mild | Unaided | 6.71* | 0.32 | 15.51 |
| Living room | Moderate | Aided | 25.33* | 11.66 | 40.69 |
| Living room | Moderate | Unaided | 28.17* | 11.90 | 44.72 |
| Living room | Moderate-severe | Aided | 43.84* | 21.74 | 63.85 |
| Living room | Moderate-severe | Unaided | 24.57* | 2.01 | 46.80 |
| Cafe | NH | Unaided | 3.77* | 0.22 | 8.49 |
| Cafe | Mild | Aided | 18.36* | 3.99 | 34.99 |
| Cafe | Mild | Unaided | 18.13* | 7.29 | 32.37 |
| Cafe | Moderate | Aided | 26.37* | 8.75 | 43.65 |
| Cafe | Moderate | Unaided | 31.01* | 12.22 | 48.55 |
| Cafe | Moderate-severe | Aided | 43.28* | 18.28 | 63.98 |
| Cafe | Moderate-severe | Unaided | 35.87* | 13.83 | 57.20 |
| Dinner Party | NH | Unaided | 6.19* | 0.49 | 13.16 |
| Dinner Party | Mild | Aided | 35.22* | 14.32 | 55.21 |
| Dinner Party | Mild | Unaided | 19.05* | 1.96 | 37.72 |
| Dinner Party | Moderate | Aided | 24.54* | 4.31 | 43.34 |
| Dinner Party | Moderate | Unaided | 25.05* | 7.26 | 42.33 |
| Dinner Party | Moderate-severe | Aided | 22.36* | 2.37 | 43.46 |
| Dinner Party | Moderate-severe | Unaided | 17.60* | 3.94 | 34.55 |
| Food court | NH | Unaided | 24.16* | 8.98 | 39.43 |
| Food court | Mild | Aided | 26.89* | 6.11 | 47.14 |
| Food court | Mild | Unaided | 21.07 | −3.04 | 43.89 |
| Food court | Moderate | Aided | 14.67* | 6.38 | 25.18 |
| Food court | Moderate | Unaided | 16.89* | 7.81 | 28.26 |
| Food court | Moderate-severe | Aided | 5.56 | −1.69 | 15.25 |
| Food court | Moderate-severe | Unaided | 8.18* | 1.66 | 18.29 |

*Significant differences at the $p < 0.05$ level are indicated with an asterisk.*

benefit provided by non-linear amplification. This is highlighted by the slopes (and distributions) that change drastically across the different test conditions (i.e., across panels in **Figure 5**) and are insignificant for the louder aided conditions.

Also shown in **Figure 5** (bottom row) is the hearing-aid benefit measured using ECO-SiN sentences plotted as a function of the equivalent benefit measured using BKB sentences. This display illustrates the fact that when performance scores are at or near ceiling there is reduced scope to detect performance improvements. Visual inspection of the scatter plots reveals clustering of data around zero on the BKB benefit scale (*x*-axis) in the three softest environments: the office, church, and living room. Clustering around zero on the *x*-axis was less clear in the café and dinner party environments. In the loudest environment, the food court, there was instead evidence of clustering of data around zero on the ECO-SiN benefit scale (*y*-axis).

**TABLE 4 |** Results of the statistical analysis of the hearing-aid benefits.

| Speech | Noise | Degree of HL | Mean | 0.025 quant | 0.975 quant |
|---|---|---|---|---|---|
| BKB | Office | Mild | 0.458 | −5.934 | 7.004 |
| ECO-SiN | Office | Mild | −0.357 | −7.651 | 7.386 |
| BKB | Office | Moderate | 5.297 | −1.436 | 12.551 |
| ECO-SiN | Office | Moderate | 18.321 | 7.825 | 28.488 |
| BKB | Office | Moderate-severe | 45.190 | 28.966 | 60.202 |
| ECO-SiN | Office | Moderate-severe | 47.299 | 34.756 | 61.061 |
| BKB | Church | Mild | −0.158 | −7.207 | 6.870 |
| ECO-SiN | Church | Mild | 0.226 | −7.488 | 7.926 |
| BKB | Church | Moderate | 5.536 | −1.172 | 12.925 |
| ECO-SiN | Church | Moderate | 22.083 | 14.138 | 30.298 |
| BKB | Church | Moderate-severe | 43.211 | 28.015 | 57.273 |
| ECO-SiN | Church | Moderate-severe | 49.562 | 38.695 | 60.268 |
| BKB | Living room | Mild | 0.452 | −6.597 | 7.538 |
| ECO-SiN | Living room | Mild | −7.927 | −15.391 | −0.464 |
| BKB | Living room | Moderate | 6.428 | −1.564 | 15.105 |
| ECO-SiN | Living room | Moderate | 8.907 | −1.146 | 18.075 |
| BKB | Living room | Moderate-severe | 39.235 | 26.273 | 51.370 |
| ECO-SiN | Living room | Moderate-severe | 22.024 | 11.707 | 31.259 |
| BKB | Cafe | Mild | −4.199 | −11.397 | 2.991 |
| ECO-SiN | Cafe | Mild | −7.460 | −16.114 | 1.707 |
| BKB | Cafe | Moderate | 5.827 | −1.727 | 13.547 |
| ECO-SiN | Cafe | Moderate | 6.181 | −1.939 | 14.478 |
| BKB | Cafe | Moderate-severe | 22.463 | 13.925 | 31.043 |
| ECO-SiN | Cafe | Moderate-severe | 13.522 | 4.797 | 22.182 |
| BKB | Dinner party | Mild | −5.410 | −14.003 | 2.685 |
| ECO-SiN | Dinner party | Mild | −22.513 | −32.531 | −11.524 |
| BKB | Dinner party | Moderate | 5.629 | −2.129 | 13.866 |
| ECO-SiN | Dinner party | Moderate | 6.339 | −2.037 | 15.016 |
| BKB | Dinner party | Moderate-severe | 16.276 | 6.652 | 25.044 |
| ECO-SiN | Dinner party | Moderate-severe | 6.113 | −2.016 | 14.251 |
| BKB | Food court | Mild | −14.831 | −23.760 | −5.193 |
| ECO-SiN | Food court | Mild | −17.038 | −27.093 | −7.317 |
| BKB | Food court | Moderate | −0.918 | −9.081 | 8.115 |
| ECO-SiN | Food court | Moderate | −0.622 | −7.273 | 6.120 |
| BKB | Food court | Moderate-severe | 6.482 | −3.485 | 15.887 |
| ECO-SiN | Food court | Moderate-severe | 0.314 | −8.616 | 9.243 |

# DISCUSSION

## Summary and Implications of Results

In this study we demonstrated that by using sentences embedded in a range of real-world environments, with their natural SNRs, the overall difficulty of a speech-in-noise test can be varied in a meaningful way. This means that by selecting the right environment a useful operating point (where scores are away from both ceiling and floor) can be found for listeners across a wide range of hearing abilities. Depending on the specific purpose, the test environment may be selected based on the individual's hearing loss, their reported speech-in-noise problem, or the relevance of a test environment (e.g., see Mansour et al., 2021). Furthermore, we demonstrated that, within our framework, the choice of speech materials not only affected the realism of the stimuli but also changed the difficulty of the listening task. Specifically, we found that highly realistic sentences from the ECO-SiN corpus resulted in lower speech intelligibility scores overall, as compared to the clearly spoken BKB sentences. We note that this result is broadly consistent with the results of a number of studies that have demonstrated that clear speech is more intelligible than conversational speech in noise for both NH and listeners with hearing loss (Picheny et al., 1985, 1989; Payton et al., 1994; Uchanski et al., 1996; Krause and Braida, 2004, 2009; Krause and Panagiotopoulos, 2019). We also found that while BKB scores were able to reasonably well predict ECO-SiN scores within a given test condition (e.g., regression lines in **Figure 5**), this linear relationship was weaker in the aided conditions in the louder background noises. In addition to this, the relationship between the different speech materials and the

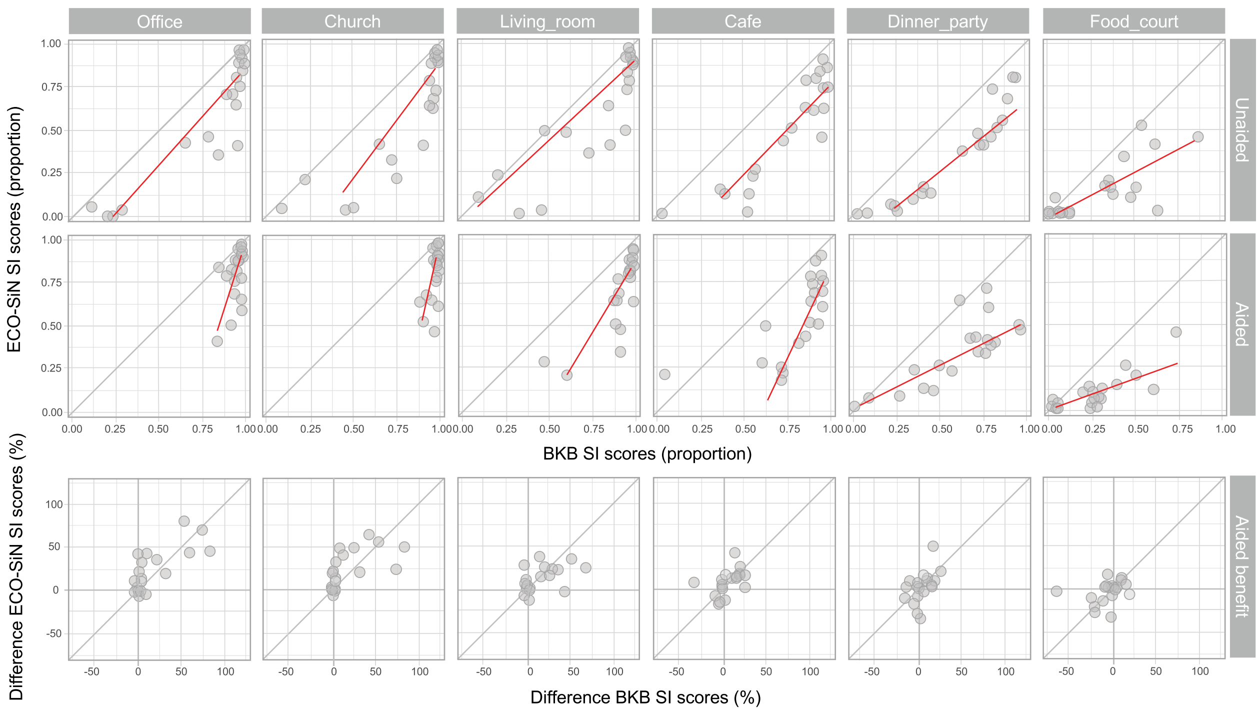


**FIGURE 5 |** Individual speech intelligibility (SI) scores for ECO-SiN as a function of BKB. Top row: unaided listening; middle row: aided listening; bottom row: hearing-aid benefit (difference between aided and unaided scores).

aiding conditions demonstrated the complexity of predicting one score from another when making comparisons across the different environments.

This ability to vary the operating point within real-world speech testing (by selecting the right environment) has important consequences if the aim is to examine the effect of a particular intervention. In our study, this point was made for the case of non-linear hearing-aid amplification. Because intelligibility scores varied substantially across environments, degree of hearing loss, and speech material, so too did the ability to measure a benefit of amplification. For instance, as shown in **Figure 4**, there was no aided benefit in the office and church environment (for either kind of speech material) for listeners with mild hearing loss. This was because the unaided and aided scores were all at ceiling. Similarly, there was no aided benefit for the listeners with moderate and moderate-severe hearing loss in the food court environment (for either kind of speech material) because both sets of scores were at or near floor. These two examples highlight there are limits on how much benefit/disbenefit (operationalized as the increase or decrease in words correctly understood) that can be measured for a given listener group in a given environment (or SNR). On top of this, we saw an impact of the chosen speech materials on speech scores and hence on hearing-aid benefits. For example, **Figure 5** (bottom left) shows that hearing-aid benefits clustered around zero for the BKB sentences in the quieter listening environments, while benefits were observable with ECO-SiN sentences. To summarize, hearing-aid benefit depends heavily on both the environment and on the speech materials used. If the goal is to understand how much a particular listener will benefit from amplification in a particular environment (or range of environments), then we argue that the ECO-SiN test at realistic SNRs provides the most meaningful estimate.

Within the constraints of our measurement approach, two main observations could be made regarding hearing-aid benefit. First, the aided benefit was largest for the listeners with the most severe hearing loss in the quietest conditions. The listeners showed the lowest unaided intelligibility scores in these conditions and thus, had also the largest opportunity to receive a benefit from hearing-aid amplification. This observation is in agreement with previous studies showing greater aided benefit with greater hearing loss (McArdle et al., 2012; Woods et al., 2015) and greater aided benefit when sentences were presented in quiet compared to noise (Mendel, 2007). In addition, it is very likely that their intelligibility scores were limited by reduced audibility, which is the main aspect of hearing loss that can be compensated by hearing-aid amplification. A second observation is that *negative* benefits were observed for the listeners with mild hearing loss in the louder environments. In these conditions, where the overall SNR is negative, speech audibility is not expected to play a significant role because the main limitation is the presence of the noise. Accordingly, it is unsurprising that amplification did not provide any strong improvement in intelligibility. Moreover, the distorting effects of compression, limiting and/or microphone placement may have had a negative impact on intelligibility by reducing the effective SNR at the listener's ears (e.g., see Cubick et al., 2018; Mansour et al., 2022).

## Challenges Associated With Conversational Sentences

So why are ECO-SiN sentences more challenging to understand than BKB sentences under similar conditions? Based on the long-term average spectra shown in the left panel of **Figure 2**, we may have expected the opposite result. Specifically, the increasing vocal intensity of the ECO-SiN sentences coincides with increased spectral tilt (Lu and Cooke, 2009) and a boost in mid-frequency energy relative to the BKB sentences. This frequency region is particularly relevant for understanding speech (see ANSI-S3.5., 1997) and thus could have produced a speech-intelligibility benefit for the ECO-SiN sentences that increases with increasing vocal effort. On the other hand, the right panel of **Figure 2** shows that ECO-SiN sentences also contain higher modulation frequencies on average relative to BKB sentences, especially for normal and raised vocal efforts. This difference, which corresponds loosely to a faster speaking rate, may explain the increased difficulty of the ECO-SiN materials. A similar conclusion was reached by Badajoz-Davila and Buchholz (2021) who demonstrated that speech intelligibility was systematically lower when comparing the ECO-SiN sentences to BKB sentences in realistic background noise for individuals with cochlear implants. While it is known that accelerated speech interacts with speech intelligibility (Wingfield et al., 1984; Adams and Moore, 2009) if the performance difference was purely driven by speaking rate, it would be expected that intelligibility would be similar between the loud ECO-SiN vocal effort and the BKB sentences (e.g., **Figure 2**), however, this was not the case. There may have been additional differences between the ECO-SiN and BKB materials that are relevant here but were not explicitly analyzed, such as differences in formants or vowel space (Bradlow et al., 1996), vowel duration (Lu and Cooke, 2009), or fundamental frequency (f0) and f0 variations (Summers et al., 1988).

Another explanation for the differences in performance measured for the different speech materials in certain environments is that the complexities of the noise may have differentially interacted with the speech materials (cf. Weisser et al., 2019a, for an in-depth discussion on acoustic complexity). For example, some background noises may contain informational masking due to competing speech (e.g., advertisements are playing on a TV in the living room background noise, people are talking over a table in the dinner party background noise), which may have interfered more strongly with the conversational ECO-SiN sentences. In addition, it is well known that amplitude modulations in background noises afford individuals the ability to listen in the dips (Hopkins and Moore, 2009), and it might be that this process is more efficient for clearly spoken sentences than for natural sentences with highly unpredictable structures. It is also possible that the BKB sentences "pop-out" of the background noise more than the ECO-SiN sentences as they are incongruent with the noise in which they were presented (Hendrikse et al., 2019). Conversely, ECO-SiN sentences may blend into the realistic background noise and be harder to selectively attend. In addition, recall that the ECO-SiN sentences were also combined with reverberation that matched the realistic virtual sound environments in which they were presented. While this was done to maximize the realism of the ECO-SiN materials, adding reverberation can result in decreased speech intelligibility (Helfer and Wilber, 1990; Gordon-Salant and Fitzgibbons, 1993; Shi and Doherty, 2008).

## Limitations and Outlook

The primary reason for assessing speech intelligibility in the clinic and laboratory is to provide insight about an individual's hearing ability in their everyday lives. However, developing more realistic speech intelligibility assessments and maintaining a level of experimental control often requires a trade-off. For example, here we used more realistic speech material from the ECO-SiN corpus and compared the sentences to BKB sentences which are typical of the materials used for speech intelligibility testing in laboratories and clinics. While the addition of realism in speech materials is a positive step for increasing realism in speech testing in order to better predict real-world performance, the sentence recall task itself is still highly unrealistic compared to how individuals communicate in the real-world. In this regard, it is important to note that many of the characteristics of natural conversational speech which are expected to benefit speech intelligibility may do so only in the full context of the task of natural conversation. For example, natural speech contains intonation that affects intelligibility (Binns and Culling, 2007; Miller et al., 2010) but also carries information such as talker emotion and cognitive state which may serve to disambiguate meaning in active conversations. It is unclear to what extent such indexical information is useful in a simple sentence repetition task with an unfamiliar talker. In real conversations, listeners can also benefit from discourse context, visual cues, shared knowledge and experience with a conversation partner, repetitions, or clarifications (Beechey et al., 2020). Accordingly, the fact that the ECO-SiN sentences were challenging to understand out of context does not mean they would necessarily be so problematic within the context of a conversation.

There is a growing body of research that aims to increase the realism of speech testing in a variety of ways (Keidser et al., 2020). For example, Best et al. (2016) evaluated a question-and-answer model based on the Helen test (Ludvigsen, 1974) which has an inherent comprehension component tapping cognitive processes used for communication in the real-world, and includes variable target talkers which mimics spatial processing required when communicating in groups in the real-world. Others have used a referential task where interactive conversations can be monitored (Beechey et al., 2019; Weisser and Buchholz, 2019). Another relevant set of studies is exploring how head orientation and movement in realistic environments intersects with speech intelligibility (Hadley et al., 2019; Hendrikse et al., 2019; Weisser et al., 2021). The inclusion of visual information in speech intelligibility testing is an area of active investigation (Devesse et al., 2020; Llorach et al., 2021) and is the next step planned for the ECO-SiN materials.

Another limitation was introduced by the applied hearing-aid platform, which mainly provided non-linear amplification and only considered an omni-directional microphone input. State-of-the-art hearing aids provide more refined implementations of compression and limiting and more advanced signal processing features, such as directional microphones and (bilateral) adaptive beamforming (e.g., Kates, 2008). Including such advanced features may have helped to overcome the negative hearing-aid benefit observed for the listeners with mild hearing loss in the louder noise environments, and potentially even provided a positive benefit. Hence, future evaluations should include state-of-the-art hearing aids to understand their benefit in the different realistic conditions and compare the results to the benefits experienced in the real world.

## 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 the Macquarie University Human Research Ethics Committee and the Australian Hearing Human Research Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

KM: design and conceptualization, data curation, analysis, and writing the manuscript. TB: design and conceptualization, analysis, and writing the manuscript. VB: analysis and writing the manuscript. JB: design and conceptualization, analysis, writing the manuscript, and supervision. All authors contributed to the article and approved the submitted version.

## FUNDING

The authors acknowledge the financial support of the HEARing CRC, established and supported under the Cooperative Research Centres Program (an initiative of the Australian Government), the William Demant Foundation, and Sonova AG. VB was supported by a grant from NIH/NIDCD (R01DC015760).

## ACKNOWLEDGMENTS

We wish to thank all of the participants who made this research possible.

## REFERENCES

Adams, E. M., and Moore, R. E. (2009). Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners. *J. Am. Acad. Audiol.* 20, 28–39. doi: 10.3766/jaaa.20.1.3

ANSI-S3.5. (1997). *American National Standards Methods for the Calculation of the Articulation Index (ANSI S3. 5-1997)*. America: Acoustical Society of America.

Badajoz-Davila, J., and Buchholz, J. M. (2021). Effect of Test Realism on Speech-in-noise Outcomes in Bilateral Cochlear Implant Users. *Ear Hear.* 42, 1687–1698. doi: 10.1097/AUD.0000000000001061

Beechey, T., Buchholz, J. M., and Keidser, G. (2018). Measuring communication difficulty through effortful speech production during conversation. *Speech Commun.* 100, 18–29.

Beechey, T., Buchholz, J. M., and Keidser, G. (2019). Eliciting Naturalistic Conversations: A Method for Assessing Communication Ability, Subjective Experience, and the Impacts of Noise and Hearing Impairment. *J. Speech Lang. Hear. Res.* 62, 470–484. doi: 10.1044/2018_JSLHR-H-18-0107

Beechey, T., Buchholz, J. M., and Keidser, G. (2020). Hearing aid amplification reduces communication effort of people with hearing impairment and their conversation partners. *J. Speech Lang. Hear. Res.* 63, 1299–1311. doi: 10.1044/2020_JSLHR-19-00350

Bench, J., Kowal, A., and Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br. J. Audiol.* 13, 108–112. doi: 10.3109/03005367909078884

Bentsen, T., Mauger, S. J., Kressner, A. A., May, T., and Dau, T. (2019). The impact of noise power estimation on speech intelligibility in cochlear-implant speech coding strategies. *J. Acoust. Soc. Am.* 145, 818–821. doi: 10.1121/1.5089887

Best, V., Streeter, T., Roverud, E., Mason, C. R., and Kidd, G. Jr. (2016). A Flexible Question-and-Answer Task for Measuring Speech Understanding. *Trends Hear.* 20:2331216516678706. doi: 10.1177/2331216516678706

Binns, C., and Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *J. Acoust. Soc. Am.* 122, 1765–1776. doi: 10.1121/1.2751394

Blumstein, S. E., and Stevens, K. N. (1981). Phonetic features and acoustic invariance in speech. *Cognition* 10, 25–32. doi: 10.1016/0010-0277(81)90021-4

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Commun.* 20, 255–272. doi: 10.1016/S0167-6393(96)00063-5

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* 62, 193–217. doi: 10.1037/h0047470

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., et al. (1994). An international comparison of long−term average speech spectra. *J. Acoust. Soc. Am.* 96, 2108–2120.

Caplan, D. (2016). "Working Memory and Sentence Comprehension" in *Neurobiology of Language*. eds G. Hickok and S. L. Small (San Diego: Academic Press). 633–645. doi: 10.1016/b978-0-12-407794-2.00051-1

Clark, J. G. (1981). Uses and abuses of hearing loss classification. *ASHA* 23, 493–500.

Cooke, M., King, S., Garnier, M., and Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Comput. Speech Lang.* 28, 543–571.

Cord, M. T., Surr, R. K., Walden, B. E., and Dyrlund, O. (2004). Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. *J. Am. Acad. Audiol.* 15, 353–364. doi: 10.3766/jaaa.15.5.3

Cubick, J., Buchholz, J. M., Best, V., Lavandier, M., and Dau, T. (2018). Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners. *J. Acoust. Soc. Am.* 144, 2896–2905. doi: 10.1121/1.5078582

Dawson, P. W., Hersbach, A. A., and Swanson, B. A. (2013). An adaptive Australian sentence test in noise (AuSTIN). *Ear Hear.* 34, 592–600. doi: 10.1097/AUD.0b013e31828576fb

Devesse, A., van Wieringen, A., and Wouters, J. (2020). AVATAR Assesses Speech Understanding and Multitask Costs in Ecologically Relevant Listening Situations. *Ear Hear.* 41, 521–531. doi: 10.1097/AUD.0000000000000778

Dillon, H. (2001). *Hearing Aids*. New York: Thieme.

Ernestus, M., Hanique, I., and Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *J. Phonet.* 48, 60–75. doi: 10.1016/j.wocn.2014.08.001

Ferguson, S. H. (2012). Talker differences in clear and conversational speech: vowel intelligibility for older adults with hearing loss. *J. Speech Lang. Hear. Res.* 55, 779–790. doi: 10.1044/1092-4388(2011/10-0342)

Ferguson, S. H., and Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 112, 259–271. doi: 10.1121/1.1482078

Ferguson, S. H., and Quene, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *J. Acoust. Soc. Am.* 135, 3570–3584. doi: 10.1121/1.4874596

Ferrari, S., and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *J. Appl. Statist.* 31, 799–815.

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (2008). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol. Neurootol.* 13, 193–205. doi: 10.1159/000113510

Gordon-Salant, S., and Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *J. Speech Lang. Hear. Res.* 36, 1276–1285. doi: 10.1044/jshr.3606.1276

Hadley, L. V., Brimijoin, W. O., and Whitmer, W. M. (2019). Speech, movement, and gaze behaviours during dyadic conversation in noise. *Sci. Rep.* 9:10451. doi: 10.1038/s41598-019-46416-0

Heald, S., Klos, S., and Nusbaum, H. (2016). *Understanding Speech in the Context of Variability*. United States: Academic Press. 195–208.

Helfer, K. S., and Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *J. Speech Lang. Hear. Res.* 33, 149–155. doi: 10.1044/jshr.3301.149

Hendrikse, M. M. E., Llorach, G., Hohmann, V., and Grimm, G. (2019). Movement and Gaze Behavior in Virtual Audiovisual Listening Environments Resembling Everyday Life. *Trends Hear.* 23:2331216519872362. doi: 10.1177/2331216519872362

Hopkins, K., and Moore, B. C. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *J. Acoust. Soc. Am.* 125, 442–446. doi: 10.1121/1.3037233

Jerger, S., and Jerger, J. (1980). Low frequency hearing loss in central auditory disorders. *Otol. Neurotol.* 2, 1–4.

Johnson, K. (2004). "Massive reduction in conversational American English" in *Proceedings of the Workshop on Spontaneous Speech: Data and Analysis*. (United States: Ohio State University).

Kates, J. M. (2008). *Digital Hearing Aids*. United States: Plural publishing.

Keidser, G., Dillon, H., Carter, L., and O'Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends Amplif.* 16, 211–223. doi: 10.1177/1084713812468511

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: What it is, why it matters, and how to advance it. *Ear Hear.* 41:5S. doi: 10.1097/AUD.0000000000000944

Killion, M., Schulein, R., Christensen, L., Fabry, D., Revit, L., Niquette, P., et al. (1998). Real-world performance of an ITE directional microphone. *Hear. J.* 51, 24–39.

Krause, J. C., and Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *J. Acoust. Soc. Am.* 115, 362–378. doi: 10.1121/1.1635842

Krause, J. C., and Braida, L. D. (2009). Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. *J. Acoust. Soc. Am.* 125, 3346–3357. doi: 10.1121/1.3097491

Krause, J. C., and Panagiotopoulos, A. P. (2019). Speaking clearly for older adults with normal hearing: The role of speaking rate. *J. Speech Lang. Hear. Res.* 62, 3851–3859. doi: 10.1044/2019_JSLHR-H-19-0094

Llorach, G., Kirschner, F., Grimm, G., Zokoll, M. A., Wagener, K. C., and Hohmann, V. (2021). Development and evaluation of video recordings for the OLSA matrix sentence test. *Int. J. Audiol.* 1–11. Epub online ahead of print. doi: 10.1080/14992027.2021.1930205

Lombard, E. (1911). Le signe de l'elevation de la voix. *Ann. Mal. L'Oreille Larynx* 37, 101–119.

Lu, Y., and Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Commun.* 51, 1253–1262. doi: 10.1121/1.4979927

Ludvigsen, C. (1974). Construction and evaluation of an audio-visual test, the Helentest. *Scand. Audiol. Suppl.* 3, 67–75.

Mansour, N., Marschall, M., May, T., Westermann, A., and Dau, T. (2021). Speech intelligibility in a realistic virtual sound environment. *J. Acoust. Soc. Am.* 149, 2791–2801. doi: 10.1121/10.0004779

Mansour, N., Marschall, M., Westermann, A., May, T., and Dau, T. (2022). The effect of hearing aid dynamic range compression on speech intelligibility in a realistic virtual sound environment. *J. Acoust. Soc. Am.* 151, 232–241. doi: 10.1121/10.0008980

McArdle, R. A., Killion, M., Mennite, M. A., and Chisolm, T. H. (2012). Are two ears not better than one? *J. Am. Acad. Audiol.* 23, 171–181.

Mendel, L. L. (2007). Objective and subjective hearing aid assessment outcomes. *Am. J. Audiol.* 16, 118–129. doi: 10.1044/1059-0889(2007/016)

Miles, K. M., Keidser, G., Freeston, K., Beechey, T., Best, V., and Buchholz, J. M. (2020). Development of the Everyday Conversational Sentences in Noise test. *J. Acoust. Soc. Am.* 147:1562. doi: 10.1121/10.0000780

Miller, J. L., Grosjean, F., and Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica* 41, 215–225. doi: 10.1159/000261728

Miller, S. E., Schlauch, R. S., and Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *J. Acoust. Soc. Am.* 128, 435–443. doi: 10.1121/1.3397384

Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am.* 95, 1085–1099. doi: 10.1121/1.408469

Olsen, W. O., Van Tasell, D. J., and Speaks, C. E. (1997). The Carhart Memorial Lecture, American Auditory Society, Salt Lake City, Utah 1996. Phoneme and word recognition for words in isolation and in sentences. *Ear Hear.* 18, 175–188. doi: 10.1097/00003446-199706000-00001

Payton, K. L., Uchanski, R. M., and Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *J. Acoust. Soc. Am.* 95, 1581–1592. doi: 10.1121/1.408545

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *J. Speech Lang. Hear. Res.* 28, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1989). Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. *J. Speech Lang. Hear. Res.* 32, 600–603. doi: 10.1044/jshr.3203.600

Pollack, I., and Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Lang. Speech* 6, 165–171.

Pronk, M., Deeg, D. J., and Kramer, S. E. (2018). Explaining discrepancies between the Digit Triplet Speech-in-Noise Test Score and self-reported hearing problems in older adults. *J. Speech Lang. Hear. Res.* 61, 986–999. doi: 10.1044/2018_JSLHR-H-17-0124

Rana, B., and Buchholz, J. M. (2016). Effect of audibility, frequency region, and hearing loss on better-ear glimpsing. *J. Acoust. Soc. Am.* 140, 3437–3437. doi: 10.1121/1.4971078

Rana, B., and Buchholz, J. M. (2018). Effect of audibility on better-ear glimpsing as a function of frequency in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 143, 2195–2206. doi: 10.1121/1.5031007

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annu. Rev. Statist. Appl.* 4, 395–421.

Shi, L. F., and Doherty, K. A. (2008). Subjective and objective effects of fast and slow compression on the perception of reverberant speech in listeners with hearing loss. *J. Speech Lang. Hear. Res.* 51, 1328–1340. doi: 10.1044/1092-4388(2008/07-0196)

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *J. Acoust. Soc. Am.* 84, 917–928.

Tucker, B. V., and Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *Ment. Lexicon* 11, 375–400. doi: 10.1075/ml.11.3.03tuc

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *J. Speech Lang. Hear. Res.* 39, 494–509. doi: 10.1044/jshr.3903.494

Wackym, P. A., Runge-Samuelson, C. L., Firszt, J. B., Alkaf, F. M., and Burg, L. S. (2007). More challenging speech-perception tasks demonstrate binaural benefit in bilateral cochlear implant users. *Ear Hear.* 28, 80S–85S. doi: 10.1097/AUD.0b013e3180315117

Walden, T. C., and Walden, B. E. (2004). Predicting success with hearing aids in everyday living. *J. Am. Acad. Audiol.* 15, 342–352. doi: 10.3766/jaaa.15.5.2

Weisser, A., and Buchholz, J. M. (2019). Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *J. Acoust. Soc. Am.* 145, 349. doi: 10.1121/1.5087567

Weisser, A., Buchholz, J. M., and Keidser, G. (2019a). Complex Acoustic Environments: Review, Framework, and Subjective Model. *Trends Hear.* 23:2331216519881346. doi: 10.1177/2331216519881346

Weisser, A., Buchholz, J. M., Oreinos, C., Badajoz-Davila, J., Galloway, J., Beechey, T., et al. (2019b). The ambisonic recordings of typical environments (ARTE) database. *Acta Acust. U. Acust.* 105, 695–713.

Weisser, A., Miles, K., Richardson, M. J., and Buchholz, J. M. (2021). Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments. *J. Acoust. Soc. Am.* 149:2896. doi: 10.1121/10.0004774

Wingfield, A., Lombardi, L., and Sokol, S. (1984). Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation. *J. Speech Lang. Hear. Res.* 27, 128–134. doi: 10.1044/jshr.2701.128

Winitz, H., and LaRiviere, C. (1979). Factors contributing to the recovery of monosyllabic words excerpted from natural speech. *J. Phonet.* 7, 225–233.

Woods, D. L., Arbogast, T., Doss, Z., Younus, M., Herron, T. J., and Yund, E. W. (2015). Aided and unaided speech perception by older hearing impaired listeners. *PLoS One* 10:e0114922. doi: 10.1371/journal.pone.0114922

Working Group on Speech Understanding, Committee on Hearing, Bioacoustics, and Biomechanics (1988). Speech understanding and aging. *J. Acoust. Soc. Am.* 83, 859–895.

Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., DeVries, S., and Oleson, J. (2019). Efficacy and Effectiveness of Advanced Hearing Aid Directional and Noise Reduction Technologies for Older Adults With Mild to Moderate Hearing Loss. *Ear Hear.* 40, 805–822. doi: 10.1097/AUD.0000000000000672

**Conflict of Interest:** The authors declare that this study received funding from Sonova AG. 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.

# APPENDIX

**Figure A1** shows the long-term spectrum in third-octave levels (left column), temporal envelope (center column), and modulation spectrum (right column) for the six different acoustic environments that were derived in free-field. The spectrum and modulation spectrum were derived as described in section "Sentence Materials" considering the entire 150 s long noise signals. The temporal envelopes were derived by normalizing the noise waveforms to an RMS value of one, applying an A-weighting bandpass filter, squaring, and temporal convolution with a 0.5 s long Hann window. The figure panels show 30 s long examples of the resulting envelopes in dB.

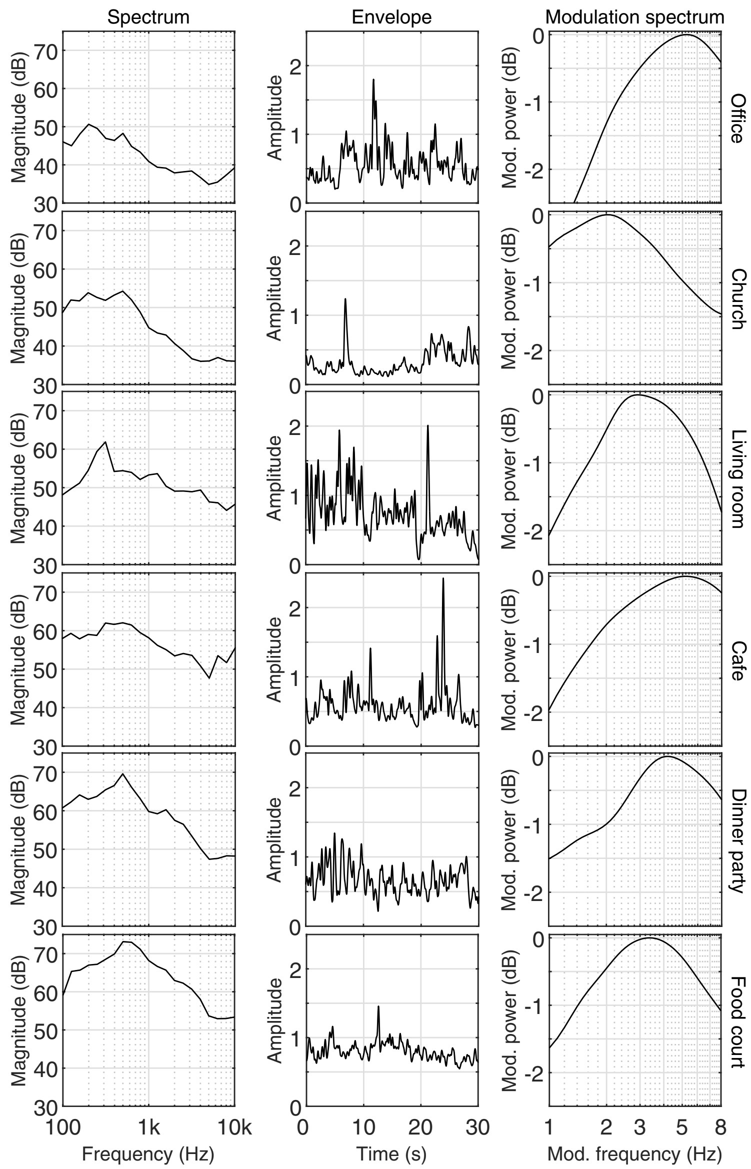


**FIGURE A1 |** Long-term spectrum in third-octave levels (left column), temporal envelope (30 s example, center column), and modulation spectrum (right column) for the six different acoustic environments derived in free-field.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 29 April 2022
doi: 10.3389/fnins.2022.830768

Check for updates

# Cochlear Implant Awareness: Development and Validation of a Patient Reported Outcome Measure

*Laura M. Markodimitraki*[1,2]*, *Inge Stegeman*[1,2] *and Hans G. X. M. Thomeer*[1,2]

[1] *Department of Otorhinolaryngology and Head and Neck Surgery, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands,* [2] *University Medical Center Utrecht Brain Center, University Medical Center Utrecht, Utrecht, Netherlands*

**Background:** Surgical success of cochlear implantation is usually measured through speech perception and quality of life questionnaires. Although these questionnaires cover a broad spectrum of domains, they do not evaluate the consciousness of wearing a cochlear implant (CI) and how this impacts the daily life of patients. To evaluate this concept we aimed to develop and validate a standardized patient reported outcome measure (PROM) for use in cochlear implant users.

**Methods:** Development and evaluation of the COchlear iMPlant AwareneSS (COMPASS) questionnaire was realized following the COSMIN guidelines in three phases: (1) item generation, (2) qualitative pilot study to ensure relevance, comprehensiveness, comprehensibility, and face validity, and (3) quantitative survey study for the assessment of reliability (test-retest) with 54 participants.

**Results:** Nine domains of CI awareness were identified through literature research and interviews with experts and patients. These resulted in the formulation of 18 items which were tested with a pilot study, after which 3 items were deleted. The final 15-item COMPASS questionnaire proved to have good validity and satisfactory reliability. The intraclass correlation coefficient calculated for items with continuous variables ranged from 0.66 to 0.89 with seven out of eight items scoring above the acceptable level of 0.7. The Cohen's kappa calculated for items with nominal variables ranged from −0.4 to 0.78 with 11 (sub)items out of 15 scoring above fair to good agreement. Measurement error analysis for items with continuous variables showed a mean difference of −2.18 to 0.22. The calculated 95% limits of agreement for these items revealed no statistically significant difference between the two administered questionnaires. For items with nominal variables, the percentages of agreement calculated, ranged between 0 and 95%, and 83.3 and 96.6% for positive and negative agreement, respectively.

**Conclusion:** The COMPASS questionnaire is a valid and reliable PROM for evaluating the cochlear implant awareness, and it can be easily used in routine clinical practice.

**Keywords: cochlear implants, cochlea, cochlear implantation, patient reported outcome measure (PROM), sensorineural hearing loss (SNHL), neurotology**



***Edited by:***
*Todd Andrew Ricketts, Vanderbilt University, United States*

***Reviewed by:***
*Paul Merkus, Amsterdam University Medical Center, Netherlands*
*Ronald Pennings, Radboud University Nijmegen Medical Centre, Netherlands*

***Correspondence:**
*Laura M. Markodimitraki*
*l.m.markodimitraki-3@umcutrecht.nl*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *07 December 2021*
***Accepted:*** *11 April 2022*
***Published:*** *29 April 2022*

***Citation:***
*Markodimitraki LM, Stegeman I and Thomeer HGXM (2022) Cochlear Implant Awareness: Development and Validation of a Patient Reported Outcome Measure. Front. Neurosci. 16:830768. doi: 10.3389/fnins.2022.830768*

## INTRODUCTION

Cochlear implants (CI's) are currently the only effective treatment for auditory rehabilitation for patients with severe-to-profound bilateral sensorineural hearing loss (SNHL) with poor speech perception. Since the introduction of this medical device in the 1970s, great advancements have been made regarding the functionality and hardware design. The internal part of the implant, the receiver/stimulator (R/S) device that resides under the skin behind the pinna of the ear, has undergone technological improvements resulting in thinner implants with smaller footprints (Carlson et al., 2012). Comfort of the external parts of the CI use has increased over the years with more discrete designs and lighter speech processors that allow patients to wear their implant throughout the day. Most importantly, the speech perception results have increased greatly, improving quality of life of patients with hearing loss (Gaylor et al., 2013; Loeffler et al., 2014; McRackan et al., 2018).

Despite the wealth of knowledge and research regarding speech perception results and health-related quality of life of CI recipients, little is known about the CI-experience and -awareness by patients. We define awareness of having a cochlear implant as "the state of mind or situation in which the patient is physically conscious he or she is wearing a cochlear implant and how this consciousness impacts their daily life." There are patient-reported outcome measures (PROMs) assessing CI use such as the Cochlear Implant Management Skills (CIMS-self) survey and the Nijmegen Cochlear Implantation Questionnaire (NICQ) (Hinderink et al., 2000; Bennett et al., 2017). The CIMS-self focuses on device management exclusively, and the NICQ assesses health-related quality-of-life by how sound and speech perception limits a CI-recipient in their daily life. However, these PROMs do not evaluate the (physical) impact of a CI, thus they may fail to capture cochlear implant awareness topics in daily life that are of importance from patient perspective. To our knowledge, no CI-specific PROM has been developed yet that included patients in item development, following the standards of the Patient Reported Outcomes Measurement Information System (PROMIS) or the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) (McRackan et al., 2018).

Cochlear implant awareness could be important for speech recognition results and quality of life of CI recipients. Studies have shown that wear time of the CI affects speech recognition outcomes in pediatric and adult patients (Gagnon et al., 2020; Holder et al., 2020). In addition, previous research on hearing aids has shown that fit and comfort are the second most important factors contributing to non-use of hearing aids (McCormack and Fortnum, 2013). Specifically, the satisfaction of patients with comfort of use, burden during daily activities, sleep disturbances related to location of the implant in relation to the preferred sleeping position, pain, or other discomfort caused by the implant are all contributing factors to reduced wear time. Moreover, there might be an underestimation of the prevalence of above mentioned problems in CI recipients, especially when a significant increase in hearing and communication is achieved using the CI. The benefits of the CI could suppress the concomitant inconvenience that accompanies wearing the processor.

In order to assess the physical awareness of the cochlear implant, we aimed to develop and validate a patient reported outcome measure (PROM) questionnaire.

## MATERIALS AND METHODS

This development and validation study was conducted between December 2019 and April 2021 at the University Medical Center (UMC) Utrecht, in compliance with the principles of the Declaration of Helsinki. This study was exempt from approval of an ethics committee under Dutch law. Exemption was granted by the local ethical committee (Institutional Review Board of the UMC Utrecht) (METC protocol 19-722/C). A three-stage procedure for development and validation of the patient reported outcome measure (PROM) was conducted, in accordance with the COSMIN guidelines (see **Figure 1**; Mokkink et al., 2010). Participants were recruited at the time of routine control at the CI center UMC Utrecht, and through an open e-mail invitation to patients registered in the CI database of the UMC Utrecht sent by their attending physician. Written informed consent was obtained from all participants.

### Construction of the Concept

We aimed to develop the COchlear iMPlant AwareneSS (COMPASS) questionnaire to assess the awareness of having a cochlear implant as previously defined. The PROM development group consisted of a otorhinolaryngologist, an epidemiologist and a junior researcher. This questionnaire was designed for adult, Dutch speaking, CI recipients. The instrument was developed to be used as a self-administered evaluation tool, in daily clinical practice, for clinical studies, and for comparison within patients over time (possible evolution of awareness). The questionnaire was designed to detect issues in different categories, specifically concerning the external parts of the CI (speech processor and transmitter) and the internal part (the receiver/stimulator device). In order to assess CI awareness, different domains were identified. It is important to distinguish the situation of awareness and how burdensome the awareness is. Therefore, the questionnaire should consist of multiple choice items as well as scale items to measure the burden. With the results of the questionnaire, health care professionals should be able to identify problems that can be solved by adapting the hardware or by counseling.

### Phase 1: Item Generation

Qualitative data were obtained by a literature review, a series of interviews with seven specialists in cochlear implantation care, including an otorhinolarygologist, speech therapists and audiologists, and individual interviews with a sample ($n = 7$) of CI recipients were conducted, to identify and select relevant aspects of CI awareness. Included patients were adult CI recipients that were using their implant for at least 1 year prior to inclusion in order to have adequate experience with everyday use of their implant to contribute to data collection. The semi-structured

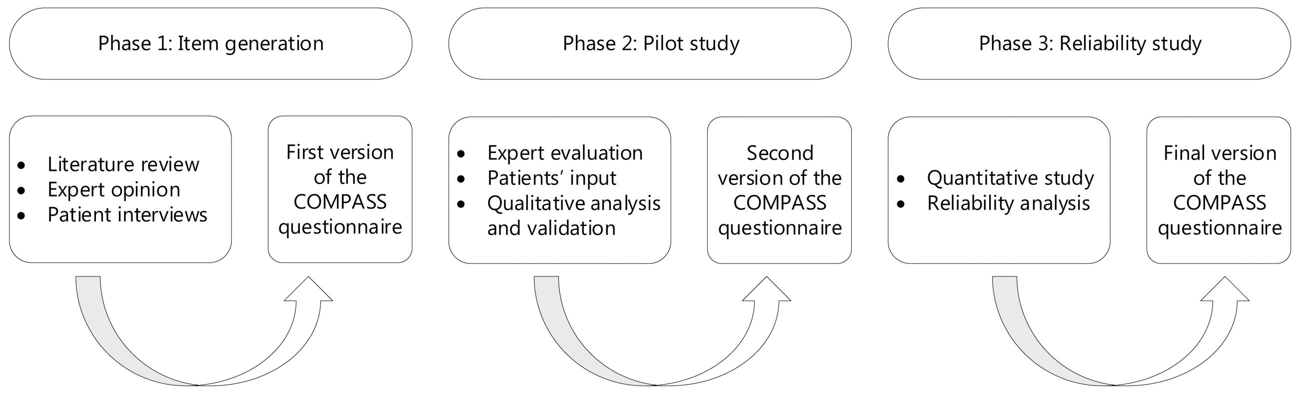


**FIGURE 1 |** The three phase procedure for development and validation of the COMPASS questionnaire.

interviews of approximately 1 h each were recorded and were conducted by a trained investigator (LM) using an interview guide (see **Supplementary Material**). The recorded interviews were then transcripted verbatim. Content analysis was performed independently by two researchers (LM and IS), by coding the transcripts and then grouping the codes into thematic categories. Data collection was continued until saturation was reached. The emerging domains as well as the pertinence of the findings were discussed within the research group until consensus was reached. The questionnaire is based on a formative model, the indicators (items) define the value of CI awareness (the construct measured).

## Phase 2: Pilot Study (Cognitive Debriefing Test)

A pilot study was conducted to assess the content validity of the questionnaire, the comprehensibility and comprehensiveness. The above mentioned experts in the field of cochlear implantation evaluated the content, wording, format, answer options, and intelligibility. Changes were made appropriately. The evaluated questionnaire was administered to ten adult CI patients that were using their implant for at least 3 months prior to inclusion in order to have adequate experience with everyday use of their implant to contribute to data validation. Participants filled out the questionnaire while "thinking aloud," followed by a semi-structured interview with open-ended questions (see **Supplementary Material**) that were audio-recorded. This interview was conducted to capture information on the participant's understanding of the instructions, the intended meaning and clinical relevance of each item, the response options, patients opinion regarding the questionnaire and missing concepts. The time required to fill out the questionnaire was also recorded. Adjustments were made to the questionnaire based on these interviews.

## Phase 3: Reliability Study

A quantitative study was conducted to assess the reliability of the final version of the COMPASS questionnaire. The questionnaire was administered twice to 54 adult CI patients, thereby meeting the COSMIN criteria of participants necessary for quantitative validation (Mokkink et al., 2010). These CI patients were using the CI for at least 3 months prior to inclusion in order to have adequate experience with everyday use of their implant to contribute to data validation. Two weeks after the participants filled out and returned the questionnaire, they were send and filled out the same questionnaire again. The questionnaire was distributed on paper or digitally through Castor EDC (version 1.6, Ciwit B.V., Amsterdam, Netherlands), an electronic data capture platform, depending on the patients' preferences.

## Data Analysis

Data was analyzed using IBM SPSS Statistics for Windows (version 26.0.0.1; IBM Corp., Armonk, NY, United States). Reliability (test-retest) was calculated using the interclass correlation coefficient (ICC) for continuous scores and Cohen's Kappa with standard error and 95% confidence interval for nominal scores. We used the two-way random effect model with interaction for the absolute agreement between single scores to calculate the ICC with 95% confidence interval. This model was chosen because time is a relevant factor for the test-retest assessment, and because the results will be generalized beyond the study points. Also the participants are assumed to be stable for the construct of interest across the two time points (Qin et al., 2019). Values $> 0.70$ are generally considered as good (Nunnally and Bernstein, 1994). However, the ICC should be interpreted with the sample variability in mind. Therefore, we calculate the range of scores per item to illustrate the homogeneity of the subjects. Small inter-subject variability results in a depress of the ICC (Weir, 2005). To interpret the values of kappa we used the criteria by Fleiss et al.: values $< 0.40$ represent poor agreement, 0.41–0.75 fair to good and $\geq 0.75$ represent excellent agreement (Fleiss et al., 2003).

Measurement error, the systematic and random error of an individual patient's score that is not attributed to true changes in the construct to be measured, was assessed by Bland-Altman plots with the 95% limits of agreement for continuous scores, and the positive and negative percentage agreement for nominal scores.

## Scoring the COMPASS

The final version of the COMPASS questionnaire consisted of 15 items. These were divided into two subdomains: external and internal device domains. The external device (speech processor and transmitter) domain and the internal device (receiver/stimulator) domain consisted of seven and eight items, respectively. Items were either multiple choice or visual analogue scale questions. Each item had a maximum score of 5, with a total maximum score of 75. A higher COMPASS score represented a higher awareness level.

# RESULTS

## Phase 1: Item Generation

Domains of awareness that were identified through literature search were bulging of the implant under the skin, discomfort or pain caused by the implant and sleep disturbances related to the implant. Domains identified through expert interviews were pain caused by the speech processor and transmitter, problems with wearing glasses, satisfaction with the position of the transmitter on the head, and interference of the external implant with daily activities and with wearing head covers (such as helmets). These domains were all mentioned by patients during the interviews in addition to problems with the transmitter coil (magnet falling off or being too strong). These domains of awareness were included in the first draft of the questionnaire. The domains most frequently mentioned were pain caused by the speech processor and/or magnet (mentioned by five out of seven participants), fear or discomfort caused by the external implant falling off the ear, and feeling a bulge where the internal implant resides under the skin (both mentioned by four participants). In order to measure these domains, 18 items were formulated. These items assessed the presence of the domains contributing to awareness and the burden that it created for the patient. Eight dichotomous (yes/no) items assessed the presence of domains; one multiple choice item assessed the ideal position of the transmitter according to the patient; seven visual analogue scale (VAS) items assessed the burden of these domains and two VAS item assessed pain caused by the external parts of the CI and in the area of operation.

## Phase 2: Pilot Study (Cognitive Debriefing Test)

A pilot study was conducted with 10 CI patients (see **Table 1** for characteristics of the participants). The mean time to complete the questionnaire was 5 min and 21 s (range 3:10–9:40). Based on the results of the item analysis and the cognitive debriefing test small revisions to the questionnaire items and response options were made to ensure comprehensibility and comprehensiveness. Four items measuring interference of the CI with daily activities that overlapped and two items measuring interference of the CI with wearing glasses were fused into two items, one multiple choice item including all activities that the CI could pose troubles with wearing glasses, and one visual analogue scale item measuring burden experienced by these problems. Two items assessing satisfaction with the position of the CI were removed that were deemed not specific for identifying the underlying issue that causes CI awareness. Thus the scoring results of these items would not be helpful for the clinician using this PROM. Two items assessing sleep disturbance caused by the implant were split into four items to increase specificity of the domain by assessing change of sleep position and awareness of the implant while lying on the operated side of the head. Lastly, one item was added to include more complaints other than pain, as suggested by the CI patients. Thus, the number of items was reduced to 15 (see **Table 2** and **Supplementary Figure 1**). Additionally, the lay-out of the paper questionnaire was adapted based on the suggestions of the CI patients.

## Phase 3: Reliability Study

We included 54 participants in the reliability study. A total of 52 participants (96.3%) filled out and returned both questionnaires. The unilaterally implanted study group had a wide age range (18–82 years) with an average age of 65 years (see **Table 1** for

**TABLE 1 |** Characteristics of study participants per study phase.

| Characteristics | Phase 1<br>*n* = 7 | Phase 2<br>*n* = 10 | Phase 3<br>*n* = 52 |
|---|---|---|---|
| Age, mean (SD) [range] | 68.6 (7.3) [62–80] | 60.7 (14.3) [31–76] | 65 (12.9) [18–82] |
| *Sex, No. (%)* | | | |
| Male | 3 (42.9) | 6 (60.0) | 35 (67.3) |
| Female | 4 (57.1) | 4 (40.0) | 17 (32.7) |
| *CI model, No. (%)* | | | |
| Cochlear | 4 (57.1) | 4 (40.0) | 25 (48.1) |
| Advanced bionics | 2 (28.6) | 3 (30.0) | 6 (11.5) |
| MED-EL | 1 (14.3) | 2 (20.0) | 18 (34.6) |
| Oticon Medical | 0 | 1 (10.0) | 3 (5.8) |
| *Operation side* | | | |
| Right | 5 (71.4) | 3 (30.0) | 26 (50.0) |
| Left | 1 (14.3) | 4 (40.0) | 26 (50.0) |
| Bilateral | 1 (14.3) | 3 (30.0) | 0 |
| CI use (months), mean (SD) [range] | 100 (88.0) [13–253] | 56.9 (74.7) [3–220] | 30 (44.1) [3–234] |

**TABLE 2 |** COMPASS questionnaire items, answer options, and scoring calculations (not original lay-out).

| No | Items | Answer options | Scoring calculation |
|---|---|---|---|
| 1. | When I wear headgear (hat/cap/helmet/head scarf), I have to remove the transmitter (magnet). | ∘ Yes<br>∘ No (go to question 3)<br>∘ Not applicable for me. I never wear head gear (go to question 3) | Yes: 5 points<br>No/Not applicable: 0 points |
| 2. | If yes, how bothersome do you find having to remove the transmitter? | Not bothersome Extremely bothersome | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 3. | The transmitter (magnet) sometimes falls off my head. | ∘ Yes<br>∘ No (go to question 5) | Yes: 5 points<br>No: 0 points |
| 4. | If yes, how bothersome do you find that the transmitter (magnet) sometimes falls from your head? | Not bothersome Extremely bothersome | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 5. | The speech processor and transmitter (magnet) have inhibited me in the following activities:<br>*(more than one option can be chosen)* | • Work<br>• Sport<br>• Transport (e.g., bicycling or driving)<br>• Social activities<br>• Wearing glasses (regular glasses/reading glasses/sunglasses)<br>• None of the above (go to question 7) | Each multiple choice item: 1 point<br>None of the above: 0 points<br>*Calculation:*<br>Maximum 5 points |
| 6. | If yes, how bothersome do you find that the speech processor and transmitter (magnet) inhibits you? | Not bothersome Extremely bothersome | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 7. | When lying with my head on the operated side, I feel the cochlear implant under the skin. | Yes<br>No (go to question 9) | Yes: 5 points<br>No: 0 points |
| 8. | How bothersome do you find that you feel the cochlear implant under the skin when lying on it? | Not bothersome Extremely bothersome | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 9. | I adjusted my sleeping position after the implantation because I want to avoid lying with my head on the operated side. | ∘ Yes<br>∘ No (go to question 11) | Yes: 5 points<br>No: 0 points |
| 10. | If yes, how bothersome do you find adjusting your sleeping position. | Not bothersome Extremely bothersome | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 11. | I feel a protrusion where the cochlear implant resides under the skin. | ∘ Yes<br>∘ No (go to question 13) | Yes: 5 points<br>No: 0 points |
| 12. | If yes, how bothersome do you find feeling a protrusion where the cochlear implant resides under the skin? | Not bothersome Extremely bothersome | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 13. | How much pain have you had due to wearing the speech processor and the transmitter (magnet)? | No pain Unbearable pain | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points |
| 14. | I have had the following symptoms in the area of the operation.<br>(*more than one option can be chosen)* | ∘ Pain<br>∘ Numbness<br>∘ Itchiness<br>∘ None<br>∘ Other: __________________ | Each multiple choice item: 1,25 points<br>None: 0 points<br>*Calculation:*<br>Maximum 5 points |
| 15. | *Fill out this question if you answered "Pain" in question 14. If not you can skip this question.*<br>How much pain have you had in the area op operation. | No pain Unbearable pain | Visual analogue scale: 0–10<br>*Calculation:*<br>score/2 = maximum 5 points<br>Total score: maximum 75 |

*Disclaimer: This is a translation of the original Dutch questionnaire for the purpose of this manuscript only. Please refrain from using in the English language without validation. Instructions: With this questionnaire we aim to assess how much your life is affected in the last month by having a cochlear implant. Mark the answer that best resembles your situation, or click and hold the bar to move on the scale. Filling out the questionnaire will take approximately 10 min.*

demographics of the reliability study participants). Most of the population was male (67.3%). On average, the participants had been using the CI for 30 months (range 3–234 months).

Regarding the reliability analysis, the ICC, which represent reproducibility for the visual analogue scale items, ranged from 0.66 to 0.89 with only one item not meeting the acceptable level of 0.7, namely the item assessing the impact of the transmitter falling off the ear (see **Table 3** for all ICC values with 95% confidence intervals). The Cohen's kappa that was calculated for nominal items ranged from −0.4 to 0.78, with six (sub)items out of 15 scoring above fair to good agreement and five (sub)items scoring excellent agreement. The two multiple choice items (number five

**TABLE 3 |** Reliability and measurement error analysis for visual analogue scale items (continuous data).

| No | Items | Sample size | Reliability analysis | | Mean difference (SD) | 95% Limits of agreement | |
|---|---|---|---|---|---|---|---|
| | | | ICC | 95% CI | | Lower limit | Upper limit |
| 2 | Impact of taking off transmitter | 17 | 0.73 | 0.24–0.9 | 0.15 (3.40) | −6.51 | 6.82 |
| 4 | Impact of falling of transmitter | 30 | 0.66 | 0.28–0.84 | 0.13 (3.50) | −6.73 | 6.99 |
| 6 | Impact of speech processor and transmitter inhibiting activities | 30 | 0.86 | 0.7–0.93 | 0.22 (2.08) | −3.85 | 4.29 |
| 8 | Impact of feeling the cochlear implant under the skin | 16 | 0.79 | 0.41–0.93 | −0.72 (2.61) | −5.83 | 4.39 |
| 10 | Impact of adjustment sleep position | 5 | 0.84 | −0.87–0.98 | −2.18 (2.58) | −7.24 | 2.88 |
| 12 | Impact of feeling the protrusion | 41 | 0.88 | 0.78–0.94 | −0.17 (1.34) | −2.79 | 2.46 |
| 13 | Pain due to wearing the speech processor and the transmitter | 52 | 0.89 | 0.81–0.94 | −0.20 (0.97) | −2.10 | 1.70 |
| 15 | Amount of pain in the operation area | 8 | 0.83 | 0.1–0.97 | 0.26 (1.59) | −2.86 | 3.38 |

*Items are numbered in accordance with the COMPASS questionnaire.*

**TABLE 4 |** Reliability and measurement error analysis for checkbox and multiple choice items (nominal data).

| No | Items | | Cohen's kappa | Standard error | 95% CI | Agreement (%) | |
|---|---|---|---|---|---|---|---|
| | | | | | | Positive | Negative |
| 1 | Taking off transmitter to wear headgear | | 0.7 | 0.09 | 0.53–0.88 | 82.8 | 90.9 |
| 3 | Transmitter falls of head | | 0.73 | 0.09 | 0.55–0.91 | 87.3 | 85.7 |
| 5 | Speech processor and transmitter inhibiting activities: | i. Work | −0.40 | 0.02 | −0.44–0.36 | 0 | 96.0 |
| | | ii. Sports | 0.51 | 0.16 | 0.19–0.82 | 58.8 | 92.0 |
| | | iii. Transport | 0.24 | 0.23 | −0.22–0.69 | 28.6 | 94.8 |
| | | iv. Social activities | 0.19 | 0.21 | −0.22–0.60 | 25.0 | 93.8 |
| | | v. Glasses | 0.63 | 0.12 | 0.39–0.86 | 73.3 | 89.2 |
| | | vi. None of the above | 0.62 | 0.11 | 0.40–0.83 | 81.5 | 80.0 |
| 7 | Feeling the cochlear implant under the skin while lying on it | | 0.76 | 0.10 | 0.57–0.96 | 82.8 | 93.3 |
| 9 | Adjustment of sleep position | | 0.78 | 0.15 | 0.49–1.07 | 80.0 | 97.9 |
| 11 | Feeling the protrusion of the cochlear implant | | 0.78 | 0.10 | 0.58–0.99 | 95.0 | 83.3 |
| 14 | Symptoms in the area of operation | i. Pain | 0.77 | 0.13 | 0.51–1.02 | 80.0 | 96.6 |
| | | ii. Numbness | 0.77 | 0.13 | 0.51–1.02 | 62.5 | 96.6 |
| | | iii. Itchiness | 0.56 | 0.16 | 0.24–0.87 | 87.9 | 93.2 |
| | | iv. None | 0.67 | 0.11 | 0.46–0.88 | 61.5 | 94.5 |

*Items are numbered in accordance with the COMPASS questionnaire.*

and fifteen), contained the four subitems that had poor agreement kappa values, with one subitem on inhibition of work due to the speech processor and transmitter scoring a negative value of −0.40 implying that there was no effective agreement between the two questionnaires on this item (see **Table 4** for all Cohen's kappa values with standard error and 95% confidence intervals).

The mean difference for items of continuous variables was −2.18 to 0.22. The 95% limits of agreement (LoA) revealed no statistically significant difference between the two administered questionnaires in all continuous variables (zero is included in each interval) (see **Table 3**). We observed higher mean differences with wider 95% LoA for items with smaller sample sizes (see **Supplementary Figure 2** for Bland-Altman plots). Percentages of agreement ranges between 0 and 95%, and 83.3 and 96.6% for positive and negative agreement, respectively. The positive agreement percentage showed the widest range, with the multiple choice items number three and eight scoring the lowest values (see **Table 4**).

## DISCUSSION

The purpose of the study was to develop and validate a PROM to assess CI awareness, thus the state of mind or situation in which the patient is physically conscious he or she is wearing a cochlear implant and how this consciousness impacts their daily life. The COMPASS questionnaire was developed following the COSMIN guidelines (Mokkink et al., 2010) for development of PROMs and was based on expert opinion and patient interviews, pilot tested with a cognitive interview study, and validated by administrating it to a population of CI recipients. We tested the content validity (comprehensibility, comprehensiveness, and relevance), and reliability of the questionnaire. The COMPASS questionnaire consists of 15 items and showed fair to excellent test-retest reliability for almost all items and measurement error analysis revealed no systematic or random errors of the score per patient. The lowest reliability and positive agreement scores were calculated for the activities impeded by the speech processor and transmitter; in particular work, transport, and social activities. This could suggest that any restrictions caused by the external part of the CI during these particular activities, varies over time, even in the short test-retest time period of 2 weeks.

We believe that prospective assessment of CI awareness using a PROM, can provide more accurate information on any existing problems. We know that hearing aid issues such as discomfort and handling problems, are common amongst users of these medical devices, one study reporting a prevalence of 98% (McCormack and Fortnum, 2013; Bennett et al., 2020). However, some patients might experience problems with their hearing aids, though do not report them to their clinician (Bennett et al., 2019). One study on cochlear implant recipient issues, reported that the majority of patients included in the study (89.8%), had at least one CI device handling problem (Bennett et al., 2017). Previous studies using patient reported outcome measures also found a high prevalence of other adverse events, such as change of taste. Lloyd et al. (2007) and Mikkelsen et al. (2017) reported changes of taste after surgery in 16.9 and 45% of CI patients, respectively. The COMPASS questionnaire could be used by clinicians to assess issues caused by the external and internal components of the CI that contribute to awareness of the cochlear implant. These issues could be solved by counseling or arranging accessories such as an adjustment of magnet power. Moreover, the location of the implant in relation to the ear pinna might be adjusted likewise (cap wearing interferes with superior implant positioning).

The questionnaire fills in the gap and responds to the needs of the implantees that experience negative effects of the presence of the subperiosteal implant. Cochlear implants have undergone tremendous developments in the last decades regarding shape, hardware volume and intrinsic technical refinements. The different manufacturers produce R/S device aspects that are quite divers. One of the interesting developments is the significant reduction in implant volume, that might decrease implant protrusion visible at the level of the skin. Moreover, this might prevent the surgeon to drill a bony well in the temporal cortex as beforehand with the older implant the gold standard has been to drill a well, to tackle this issue. To our knowledge, there is little evidence thus far available regarding the influence of implant volumes reduction or the effects of drilling or not drilling a bony well, on CI awareness of a patient and implant related complaints. Our developed questionnaire meets these goals. Items assessing burden by issues caused by the internal device such as protrusion of the skin, sleep disturbances due to the implant or problems with headgear, could be rectifiable post-implantation by revision surgery (and re-positioning the implant), however, it might be advisable to perform the implantation correctly during primary implantation. Therefore the COMPASS questionnaire could be used in clinic to assess the impact of different surgical methods for positioning and fixation of the R/S device on CI awareness.

A limitation of this study is the study population sample used for development and validation of the questionnaire, which was recruited from a single center. This could introduce selection bias, however, participants were operated by several CI surgeons with different surgical techniques. Also, assessment of the criterion validity of the COMPASS questionnaire could not be executed. After extensive literature research, we were unable to find validated outcome measures assessing CI use as defined in this study. Furthermore, despite our hypothesis that there are indeed differences of CI awareness between groups, it was impossible to execute this validation step. We expect that patients operated with different fixation techniques of the R/S device will differ in CI awareness. However, in our center we only use one fixation technique (the bony bed technique), and thus we could not compare these groups. Lastly, all four CI device brands were represented in the study population, and patients included in the study had sufficient experience with using the CI to contribute to the study.

In conclusion, the COMPASS questionnaire has good reliability and validity. Combining this PROM with clinical findings may assist in the routine follow up of patients with CI. Furthermore, it can be used as an endpoint in a clinical study, to evaluate different surgical techniques and its effect on awareness.

## 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 the local ethical committee (Institutional Review Board of the UMC Utrecht) (METC protocol 19-722/C). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

All authors contributed equally to draft and revised the manuscript and approved the final submission.

## FUNDING

Oticon Medical directly funded the Ph.D. research project of LM, *via* the University Medical Center Utrecht.

## ACKNOWLEDGMENTS

We thank the audiologists, speech therapists, and patients who participated in this study for their input to the questionnaire.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.830768/full#supplementary-material

## REFERENCES

Bennett, R. J., Jayakody, D. M. P., Eikelboom, R. H., and Atlas, M. D. (2017). Self-reported cochlear implant management skills: development and validation of the self-administered Cochlear Implant Management Skills (CIMS-self) survey. *Clin. Otolaryngol.* 42, 164–171. doi: 10.1111/coa.12713

Bennett, R. J., Kosovich, E. M., Stegeman, I., Ebrahimi-Madiseh, A., Tegg-Quinn, S., and Eikelboom, R. H. (2020). Investigating the prevalence and impact of device-related problems associated with hearing aid use. *Int. J. Audiol.* 59, 615–623. doi: 10.1080/14992027.2020.1731615

Bennett, R. J., Laplante-Lévesque, A., and Eikelboom, R. H. (2019). How do hearing aid owners respond to hearing aid problems? *Ear Hear.* 40, 77–87. doi: 10.1097/AUD.0000000000000595

Carlson, M. L., Driscoll, C. L. W., Gifford, R. H., and McMenomey, S. O. (2012). Cochlear implantation: current and future device options. *Otolaryngol. Clin. North Am.* 45, 221–248. doi: 10.1016/j.otc.2011.09.002

Fleiss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. Hoboken, NJ: John Wiley & Sons.

Gagnon, E. B., Eskridge, H., and Brown, K. D. (2020). Pediatric cochlear implant wear time and early language development. *Cochlear Implants Int.* 21, 92–97. doi: 10.1080/14670100.2019.1670487

Gaylor, J. M., Raman, G., Chung, M., Lee, J., Rao, M., Lau, J., et al. (2013). Cochlear implantation in adults: a systematic review and meta-analysis. *JAMA Otolaryngol. Head Neck Surg.* 139, 265–272.

Hinderink, J. B., Krabbe, P. F., and Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen cochlear implant questionnaire. *Otolaryngol. Head Neck Surg.* 123, 756–765.

Holder, J. T., Dwyer, N. C., and Gifford, R. H. (2020). Duration of processor use per day is significantly correlated with speech recognition abilities in adults with cochlear implants. *Otol. Neurotol.* 41, e227–e231. doi: 10.1097/MAO.0000000000002477

Lloyd, S., Meerton, L., Di Cuffa, R., Lavy, J., and Graham, J. (2007). Taste change following cochlear implantation. *Cochlear Implants Int.* 8, 203–210. doi: 10.1179/cim.2007.8.4.203

Loeffler, C., Aschendorff, A., Burger, T., Kroeger, S., Laszig, R., and Arndt, S. (2014). Quality of life measurements after cochlear implantation. *Open Otorhinolaryngol. J.* 4, 47–54.

McCormack, A., and Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *Int. J. Audiol.* 52, 360–368. doi: 10.3109/14992027.2013.769066

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., et al. (2018). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *Laryngoscope* 128, 982–990. doi: 10.1002/lary.26738

Mikkelsen, K. S., Ovesen, T., and Swan, C. Z. (2017). Pre- and post-operative dizziness, tinnitus, and taste disturbances among cochlear implant recipients. *J. Laryngol. Otol.* 131, 309–315. doi: 10.1017/S0022215116010008

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual. Life Res.* 19, 539–549. doi: 10.1007/s11136-010-9606-8

Nunnally, J., and Bernstein, I. (1994). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill.

Qin, S., Nelson, L., McLeod, L., Eremenco, S., and Coons, S. J. (2019). Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Qual. Life Res.* 28, 1029–1033. doi: 10.1007/s11136-018-2076-0

Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J. Strength. Cond. Res.* 19, 231–240. doi: 10.1519/15184.1

**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.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 May 2022
doi: 10.3389/fnins.2022.787607

Check for updates

# Defining a Patient-Centred Core Outcome Domain Set for the Assessment of Hearing Rehabilitation With Clients and Professionals

*David Allen*[1,2]*, *Louise Hickson*[3] *and Melanie Ferguson*[1,4,5]

[1] National Acoustic Laboratories, Sydney, NSW, Australia, [2] Department of Linguistics, Macquarie University, Sydney, NSW, Australia, [3] School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, QLD, Australia, [4] Ear Science Institute Australia, Perth, WA, Australia, [5] Curtin enAble Institute, Faculty of Health Sciences, Curtin University, Perth, WA, Australia

OPEN ACCESS

***Edited by:***
*Jennifer Lentz,*
*Indiana University Bloomington,*
*United States*

***Reviewed by:***
*Michael Akeroyd,*
*University of Nottingham,*
*United Kingdom*
*Amy Piper,*
*Indiana University, United States*

****Correspondence:***
*David Allen*
*david.allen@mq.edu.au*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *01 October 2021*
***Accepted:*** *06 April 2022*
***Published:*** *03 May 2022*

***Citation:***
*Allen D, Hickson L and Ferguson M (2022) Defining a Patient-Centred Core Outcome Domain Set for the Assessment of Hearing Rehabilitation With Clients and Professionals. Front. Neurosci. 16:787607. doi: 10.3389/fnins.2022.787607*

**Background:** A variety of outcome domains are currently measured for the assessment of hearing rehabilitation. To date, there is no consensus about which outcome domains should be measured, when they should be measured, and how they should be measured. In addition, most studies seeking to develop core outcome sets and measures for hearing rehabilitation services have primarily focussed on the opinions and expertise of researchers, and, to a lesser extent, clinicians, rather than also involving clients of those services. The principles of experience-based co-design suggest that health services, researchers, and policymakers should come together with clients and their families to design health services and define what metrics should be used for their success.

**Objectives:** This study aimed to seek views and consensus from a range of key stakeholders to define which client-centred self-report outcome domains should be measured, when they should be measured, and how they should be measured, in a national publicly funded hearing rehabilitation scheme. In addition, the study aimed to identify current and future potential mechanisms and systems to standardise the collection of data and reporting of outcomes, to enable comparison across clients and hearing service providers.

**Methods:** Two stakeholder groups participated in a three-round online Delphi process: (1) 79 professional stakeholders involved in the delivery of hearing services in Australia, and (2) 64 hearing rehabilitation services' clients identified by not-for-profit consumer organisations. An initial set of in-person workshops scoped the key issues upon which to develop the initial open-ended questions and subsequent Likert-scale statements addressing these issues. These statements were then distributed to both groups in an online survey. The respondent ratings were summarised, and the summary was returned to respondents along with a second round of the survey. This process was then repeated once more. The five most important outcome domains from both groups were then combined, and a consensus workshop of seven professionals and three client advocates agreed on the top four ranked domains.

**Results:** A range of potential outcome domains were identified as relevant indicators of successful hearing rehabilitation. Communication ability, personal relationships, wellbeing, and participation restrictions were identified as a core outcome domain set that should be measured as a minimum for patients receiving hearing rehabilitation. There was little agreement on the preferred timepoints for collection of outcome measures, with respondents expressing the view that this should be established by research once a set of outcome measures has been selected. However, there was broad agreement that measurements of these domains should be collected at baseline (before the provision of hearing rehabilitation) and no earlier than 3 months following the completion of rehabilitation. Potential benefits and issues with the development of a national outcomes database/collection system were also identified and prioritised, with participants highlighting the importance of valid, high-quality, trustworthy, and comprehensive data collection.

**Conclusion:** These results provide a Core Outcome Domain Set for the self-reported evaluation of hearing rehabilitation and provide important background information for the design of methods to implement them across hearing healthcare systems. However, the wide range of outcome domains identified as potentially providing important additional information and the lack of specific measures to address these domains strongly suggest that there is still more research to be done. Ongoing stakeholder engagement will continue to be vital for future implementation. In addition, further research is required to determine the optimal time following hearing rehabilitation to utilise any particular outcome measure.

**Keywords: outcome assessment (health care), correction of hearing impairment, hearing loss, audiology, patient reported outcome measures (PROMs)**

## INTRODUCTION

Hearing loss is a chronic condition that affects around four million adults in Australia, which represents one in six of the population (Access Economics Pty Ltd, 2015). In addition, hearing loss can have substantial negative consequences, including activity limitations, participation restrictions, stigmatisation, reduced quality of life, and third-party disability (Chia et al., 2007; Wallhagen, 2010; Scarinci et al., 2012; Granberg et al., 2014b; Heffernan et al., 2016; Barker et al., 2017). Furthermore, hearing loss has been associated with depression, cognitive decline, and dementia (Lin, 2011; Dawes et al., 2015).

Auditory rehabilitation aims to address the negative impact of hearing loss and includes a range of interventions. The primary intervention is hearing aids, which have been shown to be clinically effective in terms of listening ability, hearing-related quality of life (i.e., participation) and health-related quality of life (Ferguson et al., 2017). There are other auditory rehabilitation interventions for adults with hearing loss, which include alternative listening devices such as hearables, communication and patient education, and auditory training (Wong and Hickson, 2012; Ferguson et al., 2016; Ferguson et al., 2019). However, systematic reviews on these interventions have identified a lack of high-quality evidence (Henshaw and Ferguson, 2013; Barker et al., 2016; Maidment et al., 2016; Ferguson et al., 2017; Lawrence et al., 2018) in part due to a lack of a "gold standard" outcome measure (Granberg et al., 2014a; Hall et al., 2019).

In order to assess the effectiveness of interventions for adults with hearing loss, irrespective of the intervention type, it is essential to have appropriate and sensitive outcome measures that are relevant to the outcome domains targeted for improvement by auditory rehabilitation (Ferguson and Henshaw, 2015; British Society of Audiology, 2016). These are essential to both measuring an individual's progress toward desired goals, often as a result of an intervention, as well as evaluating the overall effectiveness of audiology services and providers of hearing healthcare. Careful consideration needs to be given to which outcome measures are most fit for purpose. For example, a measure that asks only about specific pre-determined situations may not be relevant to the individual, and may not be compatible with a goal-setting approach to rehabilitation that is person-centred and focussed on the individual (British Society of Audiology, 2016).

One of the major problems with measuring outcomes within auditory rehabilitation is the large number of tools and instruments, including behavioural and self-report measures (Granberg et al., 2014a). In particular, there are a huge number of self-report measures available, with one study identifying 139 hearing-specific questionnaires (Akeroyd et al., 2015). Another major problem is that there is no agreement amongst researchers and clinicians in the field regarding what outcomes should be measured and how they should be measured (PricewaterhouseCoopers Australia, 2017). A systematic review of outcome measures used in research demonstrated the extent of this problem (Granberg et al., 2014a), identifying 51 self-report outcome measurement instruments used across 122 adult hearing loss studies. Of these 51, only 16 instruments had been used in more than one study. It is perhaps not surprising then that a scoping review uncovered considerable heterogeneity in outcome measurement in randomised controlled trials of adult auditory rehabilitation interventions (Barker et al., 2015a).

Many of these measures measure similar underlying constructs, such as hearing device use, benefit, satisfaction, and social participation. In the context of hearing outcomes, these underlying constructs are known as outcome domains. However, even among outcome domains that are in widespread use and seen to be important indicators of successful rehabilitation, such as hearing aid use, there is little consensus around which outcome measures should be used (Perez and Edmonds, 2012). Furthermore, there is an increasing awareness globally that outcome domains that are not solely associated with hearing aid amplification and that address participation restrictions and psychosocial aspects should also be considered, such as wellbeing, identity, and emotion (Bennett et al., 2018; Heffernan et al., 2018a; Bennett et al., 2020; Vercammen et al., 2020). However, many of the most widely used standardised outcome measures, such as the International Outcomes Inventory for Hearing Aids (IOI-HA; Cox et al., 2000), do not address these broader and more recently identified outcome domains.

The evidence is clear that both auditory rehabilitation clinical practice and research lack a single (or even a few) outcome measure that is used widely and consistently and accepted as a "gold standard" instrument. Furthermore, even though there is a large number and variety of measures within the field, clinical trials of adult auditory rehabilitation interventions have overlooked outcomes such as adverse effects and quality of care that may be important to key stakeholders, especially patients, hearing healthcare professionals and commissioners of hearing healthcare (Ferguson et al., 2017). The involvement of these stakeholder groups in the development of such tools is rare, with some exceptions (Heffernan et al., 2018a; Heffernan et al., 2019), as typically it has been researchers alone who have developed outcomes.

A major consequence of an non-standardised approach to outcome measurement is that comparison across different patient cohorts and services is almost impossible. Similarly, within research, it is very difficult to compare and combine the results of different trials that use different measures (for example in systematic reviews with meta-analyses), which results in reduced relevance of the results and increased risk of outcome reporting bias (Ferguson et al., 2017).

Within the Australian hearing healthcare context, hearing services are provided free of charge to over one million people each year through the Hearing Services Program (HSP), primarily through the Voucher Scheme, at a cost of $590 million per annum (Commonwealth of Australia, 2019). The Voucher Scheme provides subsidised hearing services to eligible pensioners, Veterans, service people, and those receiving support for a disability that places their employment at risk (Department of Health, n.d.). Currently, as is seen in many other countries, standardised use of patient-centred outcome measures is not prevalent in Australian hearing healthcare, and typically outputs such as hearing aid uptake are used to measure the success of hearing aids for both clients and service providers (PricewaterhouseCoopers Australia, 2017). Although the importance of measuring client outcomes is highlighted in the regulatory framework of the HSP, typically the Australian-developed Client Orientated Scale of Improvement (COSI; Dillon et al., 1997) or the IOI-HA are used. While the COSI does involve recipients in the development of personalised items, potentially overcoming this limitation of the IOI-HA, its insensitivity makes it unsuitable for the measurement of service outcomes (Dillon et al., 1999).

A Government-commissioned review of the HSP published in 2017 found that the majority of key healthcare stakeholders (i.e., Contracted Service Providers, Device Manufacturers, consumer groups, research organisations) who were consulted agreed that client outcomes were important, however there was no consensus on how they should be measured (PricewaterhouseCoopers Australia, 2017). Four types of measurement methods were identified as in common use—the COSI, the IOI-HA, hearing aid datalogging, and speech testing—but none of these were used consistently. The recommendations from this review were to (i) move quickly toward an outcomes-based model rather than an outputs-based model (i.e., focussing on the number of rehabilitation programmes delivered and devices fitted), (ii) consult with key stakeholders to achieve a consensus on which outcomes should be used and to standardise the approach to measuring these, and (iii) identify how outcomes could be measured across service providers and client groups.

This current study aimed to identify and standardise a Core Outcome Domain Set (CODS): a set of outcome domains that should be used as a minimum standard for the assessment of a health condition (Hall et al., 2018), as well as when and how these domains should be assessed. A CODS can then form the basis for development of a core outcome set (COS): "an agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical trials" (OMET Initiative, n.d.). The development of COSs has grown in stature in over the years and COSs are now a recommended component of clinical trial protocols, Cochrane reviews, and government funding applications (Williamson and Clarke, 2012; Kirkham et al., 2017). Within hearing rehabilitation, a roadmap to develop a COS for tinnitus treatment has been proposed, which stresses that a consensus is needed on *what* outcome domains should be measured, and then *how* this should be measured using an outcomes tool (Hall et al., 2015; Fackrell et al., 2017). The first of these steps is the development of the CODS; it is then the addition of standardised measurement tools that results in an implementable COS. The overall aim of the present study was to achieve the first of these steps to identify a Core Outcome *Domain* Set for self-report within hearing rehabilitation, in the Australian context.

The specific aims of this study were to:

1. Seek views and consensus from a range of key stakeholders to define which client-centred outcome domains should be used, when they should be measured, and how they should be measured, for the assessment of hearing rehabilitation delivered within a national publicly funded hearing rehabilitation scheme.
2. Identify current and future potential mechanisms and systems to standardise the collection of data and reporting of outcomes, to enable comparison across clients and hearing service providers.

# MATERIALS AND METHODS

This study was approved by the Hearing Australia Human Research Ethics Committee. Informed consent was obtained from all participants.

The overall structure of this research study is shown in **Figure 1**. Two groups of participants took part in this study: (i) Professionals, and (ii) Consumers. A scoping workshop and Delphi review was conducted with each group, and a final consensus workshop was conducted with representatives of both groups. The Delphi reviews covered six sections: *Outcome Domains, Time of Collection*, *Methods of Collection, Parties Responsible for Collection, Reason for Collection*, and *National* O*utcomes Database*. Where information from a previous stage was used to inform or develop a subsequent stage, this is denoted by an arrow.

## Method

### Scoping Workshops

Four in-person scoping workshops were used to establish the initial statements for the *Outcome Domains* section of the Delphi reviews. Three workshops were conducted with the Professionals group, one each in Sydney, Melbourne, and Brisbane. One workshop was conducted in Sydney with Consumers from around Australia, with travel costs covered by the research team.

At these workshops, structured brainstorming exercises were used to assist participants to identify a comprehensive long list of domains in which outcomes of hearing rehabilitation might be observed.

For the workshops conducted with the Professional group, the New South Wales Government Human Services Outcomes Framework was used to guide brainstorming (Routledge, 2017). This framework identifies seven broad areas in which outcomes of health and human services interventions might be observed: Education and Skills, Economic, Health, Home, Safety, Empowerment, and Social and Community.

For the workshop conducted with the Consumer group, participants were asked to define "personas" for people involved in hearing services, including people with hearing loss, their family members, clinicians, and policymakers. For each of these personas, participants then brainstormed markers of successful hearing rehabilitation (e.g., "My husband and I doing more things together") and markers of unsuccessful hearing rehabilitation (e.g., "Me getting frustrated by having to repeat things").

The lists of identified outcome domains from the workshops were then combined by the research team, and duplicates were removed.

### Delphi Reviews

A Delphi review is an iterative process in which respondents are asked to complete a series of surveys (rounds), with subsequent rounds including summary information about the responses to the previous round, allowing participants to re-evaluate their previous rating of a statement based on any emerging group consensus (Helmer, 1967). The Delphi technique is useful for building consensus among experts with regard to their areas of expertise (Hsu and Sandford, 2007), and has been used successfully for the development of clinical guidelines and rehabilitation approaches in hearing healthcare (Barker et al., 2015b; Sereda et al., 2015; Ferguson et al., 2018). In the present Delphi Review, each round was conducted *via* an online survey.

Two Delphi reviews, each of three rounds, were conducted with the Professional and Consumer groups separately. These reviews utilised different surveys and slightly different questions, targeted to the two different populations. Where the same question was asked of both groups, summary information from both groups was presented, allowing respondents to use information from both groups of stakeholders in their re-evaluation (see **Figure 1**).

The Delphi reviews covered six sections. Three of these (*Outcome Domains*, *Methods of Collection*, and *National Outcomes Database*) were asked of both groups, and three (*Time of Collection*, *Parties Responsible for Collection*, and *Reason for Collection*) were asked only of the Professionals group, as it was felt by the research team that these would likely be out of the scope of understanding of non-professionals.

For both groups, the standard *agreement* rating item was a five-point Likert item with anchors Strongly agree, Agree, Neither agree nor disagree, Disagree, and Strongly disagree. The standard *importance* rating item was a five-point Likert item with anchors Very important, Important, Neither important nor unimportant, Unimportant, and Very unimportant. The standard *comfortableness* rating item was a five-point Likert item with anchors Very comfortable, Comfortable, Moderately comfortable, Slightly comfortable, and Not at all comfortable. The relevant rating item was chosen for congruence with the statements presented to participants; a full list of statements as presented is available in **Supplementary Material**.

Consensus throughout the Delphi reviews was defined *a priori* as a consensus percentage of 80% or greater for each item, that is the proportion of respondents rating a statement as follows:

- for agreement:

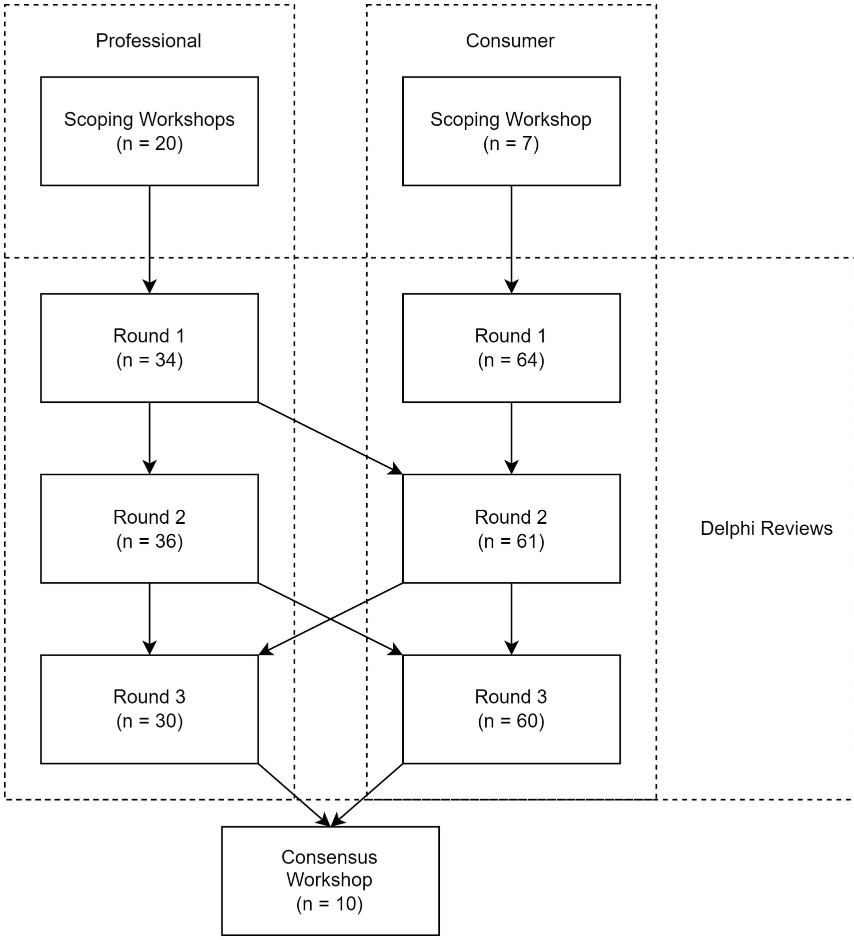


**FIGURE 1 |** Overall structure of the research study, including the number of participants in each workshop and Delphi round. Information that was used to develop subsequent stages is indicated by arrows.

- Agree or Strongly Agree, or
- Disagree or Strongly disagree;

- for importance:
  - Very important or Important, or
  - Unimportant or Very unimportant.

Comfortableness was not used to determine consensus, but only to determine consensus rankings.

To allow for discrimination between items beyond whether they reached consensus or not, consensus rankings were calculated for some items, determined using the Kemeny-Young method (Kemeny, 1959). This method, which has to our knowledge not been used previously in hearing research, generates the consensus ranking(s) of items that have the largest number of total pairwise agreements with the rankings provided by participants. The number of pairwise agreements is calculated by taking every possible pair of items, determining which item is ranked higher in the potential consensus ranking, and then counting the number of participant-provided rankings that also rank that item higher. In some cases, such as where all respondents ranked items in the same order, this ranking is unique, although in some cases the method produces "ties," where there is insufficient information in the data to be able to definitively place one item above another. This method was applied by treating each person's responses as a single ranking, in which those items rated as Strongly Agree (or Very important, or Very comfortable) are ranked above those items rated as Agree (or Important, or Comfortable), and so on.

For example, consider a set of three items which have been ranked by participants. In this case, there are 13 possible orderings to consider as the consensus ranking. With the set of responses described in **Table 1**, the potential consensus ranking Item 1 > Item 2 > Item 3 has three pairwise agreements with Person A (Item 1 > Item 2, Item 2 > Item 3, and Item 1 > Item 3), two agreements with Person B (Item 1 > Item 2 and Item 1 > Item 3), and two agreements with Person C (Item 1 > Item 3 and Item 2 > Item 3) for a total of seven pairwise agreements. If this process is repeated for all 13 possible orderings (13 = 8 total orderings plus 5 combinations involving ties), it can be seen that this ranking (Item 1 > Item 2 > Item 3) has the highest total and is therefore the Kemeny-Young consensus ranking.

**TABLE 1 |** Synthesised responses to illustrate Kemeny-Young method.

| | Person A | Person B | Person C |
|---|---|---|---|
| Item 1 | Strongly agree | Agree | Agree |
| Item 2 | Agree | Disagree | Agree |
| Item 3 | Disagree | Neither agree nor disagree | Disagree |
| | Item 1 > Item 2 > Item 3 | Item 1 > Item 3 > Item 2 | (Item 1 = Item 2) > Item 3 |
| K-Y ranking Item 1 > Item 2 > Item 3 Total of seven pairwise agreements | Three agreements (Item 1 > Item 2; Item 2 > Item 3; Item 1 > Item 3) | Two agreements (Item 1 > Item 2; Item 1 > Item 3) | Two agreements (Item 1 > Item 3; Item 2 > Item 3) |

Determination of a consensus ranking is one of the most difficult problems to solve computationally, as the number of possible consensus rankings (possibly with ties) that need to be checked grows superfactorially (due to ties) with the number of items to be ranked (Bartholdi et al., 1989; Biedl et al., 2006): while there are 13 possible orderings for three items, there are 75 for four items, 541 for five items, approximately 102 million for 10 items, and over 230 trillion for only 15 items. As a result, as the number of items that have been ranked increases, it becomes computationally not feasible to determine the optimal consensus ranking definitively, and heuristic methods that provide computational approximations to the optimal ranking are required. In this study, where small numbers of items were to be ranked (<15), the branch and bound algorithm (which definitively determines the optimal consensus ranking) could be computed in a reasonable time and was used (<4 h; Emond and Mason, 2002). Where a larger number of items were to be ranked, the fast computational approximation developed by Amodio et al. (2016) (stylised "FAST") was used instead.

### Consensus Workshop

A summary of results from the *Outcome Domains*, *Time of Collection*, and *Methods of Collection* sections of the Delphi Reviews was distributed to workshop participants prior to the workshop.

At the online workshop, the top five outcome domains as determined by the Consumer group and the top five outcome domains as determined by the Professional group were discussed. Due to similarities between outcome domains across the groups, seven were identified as separate constructs. These were then presented to the workshop and discussed in detail, to ensure that participants had a coherent shared understanding of each domain.

Participants then separated into two groups to discuss the domains and their importance. Project team staff attended these groups but did not participate in the discussion beyond answering questions about the methodology and the results.

Participants then individually and anonymously ranked the domains from most important to least important. The summary of the individual rankings was presented to the group and discussed until unanimous agreement was reached on a short list of domains that should be recommended for collection from all people receiving hearing services in Australia.

Some discussion was also had regarding methods and time of collection of outcome measures, which was synthesised qualitatively and is summarised below.

### Final Recommendations

The research team met to synthesise the results from all phases of the work, primarily the Consensus Workshop and literature review of potentially applicable outcome measures, into specific interim recommendations for the assessment of hearing rehabilitation in clinical and research practice. Discussion continued until unanimous agreement was reached.

## Participants

### Professionals

This group comprised Professionals involved in the hearing industry in Australia. Potential participants were identified by the research team by brainstorming within each of the categories of hearing researchers, representatives of professional organisations, hearing service organisations, industry organisations such as hearing device manufacturers, and hearing consumer advocacy organisations. This resulted in a list of 59 people who were invited to take part in the study, with 43 (73%) consenting to take part.

Participants were invited to join in one of three initial scoping workshops, which were conducted in Brisbane ($n$ = 6), Melbourne ($n$ = 7), and Sydney ($n$ = 7). Following these scoping workshops, several other potential participants were identified by workshop participants as being people who would be interested in contributing to the Delphi review, and were added to the list, giving a total of 79 potential participants for the Delphi review, of whom 50 (63%) completed at least one round of the Delphi review. Of these, 19 (38%) completed only one round, 16 (32%) completed only two rounds, and 15 (30%) completed all three. Participation in a future round was not contingent on completion of all previous rounds.

All categories used to identify potential participants were represented in the scoping workshops.

### Consumers

Four hearing advocacy organisations active in Australia were identified (Better Hearing Australia Brisbane, Deafness Forum, Hearing Matters Australia, and Soundfair) and invited to nominate one or more representatives who identified as people with hearing loss. Seven representatives across all these groups took part in the initial scoping workshop.

Following the workshop, the organisations were invited to share a link to the first round of the Delphi review with their members, and all did so. Participants who responded to round 1 of the Delphi Review ($n$ = 64) were then invited to participate in rounds 2 and 3. Of these, 55 (86%) completed all three rounds, 6 (9%) completed only two rounds, and 3 (5%) completed only round one. There was higher engagement with the review process in this group that in the Professional group, potentially due to the self-selected nature of the participants.

### Final Consensus Group

A group of potential participants ($n$ = 18) was selected by the research team from those in the Professional group who had responded to at least 2 rounds of the Delphi Review ($n$ = 32). In the selection of these potential participants, the categories used at the potential participant identification stage were considered to ensure broad coverage of the Australian hearing industry. One potential participant from each consumer organisation was then added. The resulting 22 potential participants were then invited to take part in an online workshop (due to COVID-19), with 10 attending.

## Material

The Delphi Reviews were conducted using three rounds of electronic surveys. A summary of the types of responses invited (e.g., ranking of importance) for each section of each survey can be seen in **Table 2**.

In round 1, participants were asked open-ended questions about the topics, which were then synthesised into statements for rating in subsequent rounds.

In rounds 2 and 3, statements were presented for rating using one of the standard items or ranking, along with summary information about responses to any previous rounds of ratings or rankings. In the *Methods of Collection* and *National Outcomes Database* sections, questions asked were similar between the two groups, and so summary information for both groups was presented.

In the *Outcome Domains* section, statements were rated using the standard importance item. As statements had already been determined during the scoping workshops, in Round 1 they were presented for rating, along with an open-ended question to allow participants to add any outcome domains that they felt were missing. In Round 3 participants were also asked to select, in ranked order, the top five domains, which were used to generate a consensus ranking. Due to the large number of outcome domains that reached consensus in Round 1 among Consumer participants, in Round 2 Consumers were presented with only the top 10 outcome domains from the previous round (as determined by the Kemeny-Young method), as well as five additional domains synthesised from open-ended responses. All of these reached consensus as being important, and so in Round 3 Consumer participants were not asked to rate, but only to rank the domains.

In the *Time of Collection* section, Professional participants were asked about the different time points at which outcomes could be collected, and why. Responses were synthesised into four major time points and a set of statements regarding the potential reasons why these time points might be useful. In Rounds 2 and 3, Professional participants were asked to rank the four time points in order of importance, and to rate their agreement with the statements using the standard *agreement* rating item. The ranked time points were used to generate a consensus ranking.

In the *Methods of Collection* section, Professional participants were asked about potential methods that could be used to collect outcomes, and the benefits and drawbacks that might be associated with each method. Professional participants were asked to rank the top five most important methods of collection in both Rounds 2 and 3 to facilitate the recommendation of a single method of collection at the close of the study. Consumer respondents were asked to rate the *comfortableness* of each method in Round 2, and to rank the top five most important methods in Round 3.

In the *Parties responsible for collection* section, Professional participants were asked about different parties (e.g., clinicians, a Government agency, or GPs) who could be responsible for the collection of outcome measures in the Hearing Services Program, and the potential benefits and drawbacks of each these parties undertaking outcomes collection. Statements were rated by Professional participants using the standard *agreement* rating item in subsequent rounds.

In the *Reasons for Collection* section, Professional participants were asked about reasons why different stakeholders might find it important for outcomes to be measured. Statements were rated by Professional participants using the standard *agreement* rating item in subsequent rounds.

In the *National Outcomes Database* section, Professional participants were asked about the potential beneficiaries, benefits, and drawbacks of the development of a national database of outcomes for hearing rehabilitation. The beneficiaries and benefits were synthesised into potential purposes for such a system. The potential purposes and drawbacks were then presented to both groups, and participants were asked to rate them using the standard *importance* rating item. In Round 2, Consumer participants were also asked how *comfortable* they would feel with a range of different stakeholders running a national outcomes database.

Questions asked in each section are available as **Supplementary Material**.

# RESULTS

## Delphi Reviews

Results for this section are shown in **Tables 3–8**. In each table, items are ordered by the consensus ranking, when both consensus ranking and percentage are available, and then by the consensus percentage when ranking is tied (ordered from unanimous agreement to unanimous disagreement). Consensus percentages meeting the predefined criterion (80%) are shown in bold. In some cases, there is disagreement between the ordering implied by the consensus percentages and that obtained using the consensus ranking procedure, as the consensus percentage method treats "agreement" from one participant and "strong agreement" from another as equivalent. As the ranking takes individual preferences between domains into account, ranking should be considered a more accurate measure of consensus preference. However, as the ranking of an item is dependent on preferences for other items in the set, no strict criterion for ranking can be applied.

**TABLE 2 |** Types of response invited for each section of the Delphi review surveys.

| Section | Round 1 | Round 2 | Round 3 |
|---|---|---|---|
| Outcome domains | Professional: Rating of importance and open-ended question | Professional: Rating of importance | Professional: Rating of importance and ranking of importance |
| | Consumer: Rating of importance and open-ended question | Consumer: Rating of importance | Consumer: Ranking of importance |
| Methods of collection | Professional: Open-ended questions | Professional: Ranking of importance | Professional: Ranking of importance |
| | | Consumer: Rating of comfortableness | Consumer: Ranking of importance |
| National Outcomes Database: Purposes and drawbacks | Professional: Open-ended questions | Professional: Rating of importance | Professional: Rating of importance |
| | | Consumer: Rating of importance. | Consumer: Rating of importance |
| National Outcomes Database: Potential stakeholders to run a database | | Consumer: Rating of comfortableness. | |
| Time of collection: Time points | Professional: Open-ended question | Professional: Ranking of importance | |
| Time of collection: Benefits and drawbacks of time points | Professional: Open-ended question | Professional: Rating of agreement | Professional: Rating of agreement |
| Reasons for collection | Professional: Open-ended question | Professional: Rating of agreement | Professional: Rating of agreement |
| Parties responsible for collection | Professional: Open-ended question | Professional: Rating of agreement | Professional: Rating of agreement |

## Outcome Domains

The primary question for this section was "What outcome domains should be measured as markers of success of hearing rehabilitation?"

Results from the Professional group are shown in **Table 3**. For each item, the consensus percentage and the consensus ranking from the final ranking task are shown. The consensus criterion was met for 13 domains and not for two domains.

Most domains that were identified through this process were psychosocial. Notably, the consensus criterion was not reached for the domain "Increased use of hearing aids."

Results from the Consumer group are shown in **Table 4**. There was consensus beyond the predefined criterion on every domain presented. As there were several domains that were excluded after Round 1 and not included for Rounds 2 and 3, it is possible that a subset of these, should they have been presented, may also have reached the predefined criterion for consensus.

## Time of Collection

The primary question for this section was "At what time point(s) should outcome measures be collected, and why?" It was conducted in two parts: by presenting respondents with four specific time points for ranking, and then by presenting a set of statements.

The four specific post-fitting time points identified were, in ranked order from most to least preferred, at 3 months

**TABLE 3 |** Results from the outcome domains section among the professional group.

| Domain | Consensus percentage | Consensus ranking |
|---|---|---|
| Improved communication ability | **100%** | 1 |
| Improved communication in groups | **97%** | 2 |
| Improved personal relationships | **100%** | 3 |
| Improved self-management ability | **87%** | 4 |
| Improved well-being | **87%** | 5 |
| Improved participation in activities | **97%** | 6 |
| Improved social engagement | **90%** | 7 |
| Increased use of hearing aids | 77% | 8 |
| Improved sense of empowerment | **80%** | 9 |
| Increased independence | **87%** | 10 |
| Reduced social isolation | **97%** | 11 |
| Reduced loneliness | **83%** | 12 |
| Reduced listening effort | **97%** | 13 |
| Improved community engagement | **83%** | =14 |
| Improved access to education | 53% | =14 |

*Consensus percentages meeting the consensus criterion are shown in bold.*

**TABLE 4 |** Results from the outcome domains section among the consumer group.

| Domain | Consensus percentage | Consensus ranking |
|---|---|---|
| I can live my life independently | **90%** | 1 |
| I can communicate well with my family | **100%** | 2 |
| I can communicate effectively with people | **98%** | 3 |
| I am able to do the things that I want to do | **95%** | 4 |
| I hear clearly with my hearing aids | **95%** | 5 |
| I can use my hearing aids effectively | **100%** | 6 |
| My hearing impacts less on my family | **98%** | 7 |
| I have the skills I need to communicate | **93%** | 8 |
| I have more control over my hearing | **88%** | 9 |
| I trust my hearing care professional | **93%** | 10 |
| My hearing aids are comfortable | **95%** | 11 |
| I can use the telephone effectively | **88%** | 12 |
| I am better able to hear the TV as a result of my hearing care | **84%** | 13 |
| I am able to participate in the social events that I want | **90%** | 14 |
| I am satisfied with the hearing care I receive | **95%** | 15 |

*Consensus percentages meeting the consensus criterion are shown in bold.*

**TABLE 5 |** Statements from the time of collection section among the professional group.

| Statement | Consensus percentage |
|---|---|
| A baseline measure should be obtained at or prior to fitting of a device to help determine the course of treatment intervention | **93%** |
| A baseline measure should be obtained at or prior to fitting of a device to assess future progress | **93%** |
| The final outcome measure should not be collected any sooner than 3 months as clients may not have acclimatised to their devices | **83%** |
| Outcome measures should be obtained multiple times during a year to assess the course of the rehabilitation intervention | 77% |
| An outcome measure should be obtained at around the 3-months period, as clients struggle with device compliance around this period | 50% |
| Outcome measures are likely to capture a more holistic view if conducted 12-months post fitting | 50% |
| Outcome measures are likely to capture a more holistic view if conducted 6-months post fitting | 47% |

*Consensus percentages meeting the consensus criterion are shown in bold.*

**TABLE 6 |** Rankings of methods of outcomes collection.

| Statement | Professionals consensus ranking | Consumers consensus ranking |
|---|---|---|
| The hearing care professional fills out a questionnaire with the client face to face | 1 | 1 |
| The client fills out a paper questionnaire that is posted to them by their hearing care professional | 2 | =3 |
| The client fills out a questionnaire (paper or electronic) with their GP | 3 | =3 |
| The client fills out a paper questionnaire that is posted to them by their GP | 4 | =3 |
| The client fills out a paper questionnaire (or electronically on a tablet) and returns it to their hearing care professional or the receptionist | 5 | =3 |
| The hearing care professional fills out a questionnaire with the client over the telephone | 6 | =3 |
| The client fills out an online questionnaire that is emailed to them by their hearing care professional | 7 | 2 |

**TABLE 7 |** Statements from the parties responsible for collection section among the professional group.

| Statement | Consensus percentage | Direction |
|---|---|---|
| Outcomes are best collected by the client's own hearing care professional because outstanding problems experienced by the client can be responded to more readily | 60% | Agree |
| Outcome should be collected by a third party independent of the hearing care organisation to avoid the potential for bias | 50% | Agree |
| Clients will be more honest if outcomes are collected by someone independent of their hearing care organisation | 43% | Agree |
| Outcomes are best collected by the client's own hearing care professional because the client is familiar with the hearing care professional and they are familiar with the client | 33% | Agree |
| Outcomes should be collected by hearing advocacy groups because they are less likely to show any bias | 33% | Disagree |
| Outcomes should be collected by hearing care professionals because they are less likely to show any bias | 37% | Disagree |
| A Government body, e.g., the Hearing Services Program is the best placed group to collect outcomes | 50% | Disagree |

*For each statement, the consensus percentage is given, along with the direction in which that consensus percentage was calculated. For example, for the first item, 60% of people rated the item as Agree or Strongly Agree, and a smaller percentage rated it as either Disagree or Strongly Disagree.*

following the fitting, at 6 months following the fitting, at 12 months following the fitting, and at the follow-up appointment (commonly conducted between one and 3 weeks post-fitting in Australia).

The statements and consensus percentages for this section are shown in **Table 5**. The highest consensus percentage at 93% related to the two statements on the use of baseline measures prior to device fitting, which was strongly supported by respondents. There was consensus reaching the criterion for three statements and not for four statements.

## Methods of Collection

The primary question for this section was "What different methods could be used to collect outcome measures?"

Consensus rankings for this section are shown in **Table 6**. Note that the consensus ranking method was unable to distinguish between statements ranked third among respondents in the Consumers group for five of the statements. The orderings obtained from Professional group and from the Consumer group were notably different, with the second preference among Consumers (an online questionnaire emailed by the hearing care professional) being ranked last by the Professional respondent group.

## Parties Responsible for Collection

The primary question for this section was "Thinking of patients being seen for hearing rehabilitation, who could potentially collect outcome measures?"

Statements and consensus percentages for this section are shown in **Table 7**. The predefined consensus criterion was not reached for any of the statements. As the results tended toward disagreement for some of the statements, whether the consensus

**TABLE 8 |** Statements from the reasons for collection section among the professional group.

| Statement | Consensus percentage |
|---|---|
| To provide an evidence base to help inform clinical decision-making | **97%** |
| To inform hearing care professionals as to the need for further interventions for their clients | **97%** |
| To ensure that services offered are providing benefit to clients | **97%** |
| To ensure that hearing care professionals are providing appropriate hearing care services to their clients | **94%** |
| To provide a benchmark against which clinical services can be measured | **94%** |
| To demonstrate whether the Voucher Scheme is positively impacting clients | **94%** |
| To demonstrate the success of the rehabilitation programme for the client | **94%** |
| To enable hearing care organisations to monitor consistency of practice | **90%** |
| To help inform the client's rehabilitation journey and management plan | **87%** |
| To provide evidence for the effective use of government resources | **84%** |
| To help promote a more holistic approach to hearing rehabilitation rather than focus solely on hearing aids | **84%** |
| To enable the hearing care professional to compare management approaches, e.g., when trying a different rehabilitation option | **84%** |
| To help the Government and other funders target poorly performing hearing care organisations for auditing | **81%** |
| To facilitate the identification of hearing care professionals within an organisation who require more training or assistance | 77% |
| To provide population data to health researchers | 74% |

*Consensus percentages meeting the consensus criterion are shown in bold.*

percentage relates to agreement or disagreement is also shown in the table. It should be noted that as "Neither agree nor disagree" was a valid option for respondents, consensus percentage for agreement and consensus percentage for disagreement do not sum to 100%. For example, 33% consensus toward agreement displayed in the table indicates that fewer than 33% of respondents responded "Disagree" or "Strongly disagree," with the remainder responding "Neither agree nor disagree."

### Reasons for Collection

The primary question for this section was "Why might it be important to clinicians providing hearing services/hearing service providers/Government that outcomes are measured?"

The statements and consensus percentages are shown in **Table 8**. Consensus reached the pre-defined criterion on 13 statements and did not reach the criterion on two statements. A broad array of potential reasons for collecting outcome measures with differing beneficiaries was identified, including hearing care professionals, Government, hearing care organisations, and the public.

### National Outcomes Database

The primary questions for this section were: "Are there any people who you think might benefit from a national outcomes database? What are the potential benefits of a national outcomes database? What are the potential drawbacks associated with having a national outcomes database?" Detailed results of this section, including consensus percentages and rankings, are available in **Supplementary Material**, and a summary is provided below.

Both groups agreed that a database should be designed to promote person-centred care, to help determine best practice, and to provide a national standard for hearing care. Both groups also agreed that such a database should be designed to measure the impacts of hearing loss beyond the person themselves, on their partners, family members, and friends. The Consumer group, but not the Professional group, agreed that a database should support clients to choose hearing services and providers and help identify poorly performing clinicians and services.

Both groups were agreed that the accuracy and relevance of measures and the integrity of the data was highly important, with misuse of data by professionals or organisations a significant concern. Professional participants were also concerned with the potential for data breaches and use of the data to justify funding cuts.

Consumer participants felt more comfortable with organisations that might be considered independent from both the hearing industry and Government running a national outcomes database, including independent research organisations, professional associations, and universities.

## Consensus Workshop

The synthesised domains as presented to the consensus workshop are available in **Supplementary Material**.

During initial discussion to ensure that participants understood the domains as presented, participants decided that "Improved participation in activities" should instead refer to reduction of "participation restrictions," as it was felt that it was unreasonable to expect that the provision of hearing rehabilitation alone would result in increased participation by patients in social activities. Rather, participants felt that while hearing rehabilitation could reduce the barriers to participation caused by the hearing loss, the social and psychological effects of long-standing hearing difficulty may result in some continued persistence of patterns of reduced participation for a period of time following any reduction in participation restrictions.

Following this discussion, participants separated into groups to further discuss the domains, and anonymously ranked the domains individually. A summary of the results of individual prioritisation for both groups are shown in **Table 9**.

After prioritisation, group discussion was undertaken to identify which outcome domains were considered core for assessment of hearing services. Discussion focussed on the

**TABLE 9 |** Summary of rankings of individual domains provided by participants in the consensus workshop.

| Domain | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th |
|---|---|---|---|---|---|---|---|
| Improved communication ability | 10 | 0 | 1 | 0 | 0 | 0 | 0 |
| Improved personal relationships | 1 | 4 | 5 | 1 | 0 | 0 | 0 |
| Improved well-being | 0 | 5 | 0 | 3 | 0 | 0 | 1 |
| Reduced participation restrictions | 0 | 1 | 1 | 5 | 3 | 1 | 0 |
| Increased independence | 0 | 1 | 1 | 1 | 7 | 1 | 0 |
| Improved perception of clarity | 0 | 0 | 1 | 0 | 1 | 3 | 6 |
| Improved self-management ability | 0 | 0 | 0 | 1 | 0 | 6 | 4 |

*For each individual domain presented to participants, the table shows the number of participants ranking it first, second, etc. This is the same format in which these data were made available to participants.*

importance of capturing the full breadth of outcomes experienced by patients, the importance of domains that were clearly articulable and comprehensible by patients and clinicians, and on selecting domains that would be directly modifiable by rehabilitation efforts.

Following the discussion, the group decided unanimously that four outcome domains should be recommended as part of a CODS for self-report in hearing rehabilitation in Australia. These were, in order of importance: (1) communication ability, (2) personal relationships, (3) well-being, and (4) participation restrictions. The group stressed that all seven domains presented to the group were important and should be considered for settings where the collection of additional outcomes is possible.

Following the discussion of which outcome domains should be measured following hearing rehabilitation, additional discussion was had regarding the time points at which outcomes should be measured, and which outcome measures should be used. There was no decision made regarding a positive answer to either question, with participants agreeing that these questions should be answered with reference to the research literature. Participants felt that the time of collection may differ between particular outcome measures and should therefore be determined with reference to research relevant to each particular outcome measure.

## Final Recommendations

Three overarching recommendations were made as a primary output of this project. It should be noted that these recommendations were made specifically for the Australian publicly funded hearing rehabilitation context, and with the assumption that outcomes would be collected to facilitate their tracking and improvement over time.

1) Target the outcome domains "communication ability," "personal relationships," "well-being," and "participation restrictions" when assessing hearing rehabilitation.
2) Measure within these domains at baseline and then following the conclusion of the rehabilitation, with a delay of at least 3 months being recommended.
3) Establish an independent body to develop a standardised outcomes instrument and mechanism for outcomes collection.

The detailed recommendations from this project are available in **Supplementary Material**.

# DISCUSSION

## Outcome Domains

This study identified four primary outcome domains that are recommended as part of a core outcome domain set for the self-report evaluation of individual hearing rehabilitation programmes: communication ability, personal relationships, well-being, and participation restrictions.

Many currently available outcome measures used as measures of the success of hearing rehabilitation focus on improvements in communication ability, including the Glasgow Hearing Aid Benefit Profile (GHABP; Gatehouse, 1999), the Abbreviated Profile of Hearing Aid Benefit (APHAB; Cox and Alexander, 1995), and the International Outcomes Inventory for Hearing Aids (IOI-HA; Cox et al., 2000). The delineation of specific subdomains of communication ability in this study (communication with family and communication in groups) suggests that in addition to generalised measures of patients' overall communication ability in their everyday lives, specific measures or subscales highlighting difficulties or successes in these identified areas are also required for a comprehensive assessment of communication ability. Many commonly used measures for the assessment of communication ability validated in hearing rehabilitation do not address these subdomains separably, although some measures include these aspects, such as the SOS-HEAR (Scarinci et al., 2012), the Self-Assessment of Communication (SAC; Schow and Nerbonne, 1982), and the GHABP. A notable exception is the Communication Performance subscale of the Communication Profile for the Hearing Impaired (CPHI; Demorest and Erdman, 1987), which includes several items related to communication both in group social situations and with family around the home. However, these items do not form separable subscales, making it difficult to assess communication on the subdomains in a separable way (Demorest and Erdman, 1986). The instrument also includes items that may be inappropriate for some people, such as hearing in lectures and religious services. Further work may be required to develop an instrument that can assess communication ability both across life as a while and in particularly meaningful situations for a wide variety of people.

The most specific measure of hearing-related participation restriction currently available, the Social Participation Restrictions Questionnaire (SPaRQ), focuses primarily on social participation (Heffernan et al., 2018a,b). Participation restrictions due to hearing loss manifests across a range of kinds of non-social participation, including in employment, education, domestic settings, and political life (Danermark et al., 2013; Granberg et al., 2014c), and so additional measures or expansion of the SPaRQ is likely to be necessary to enable comprehensive assessment.

There currently are no measures validated in hearing rehabilitation for the assessment of personal relationships or general well-being, although general measures of well-being such as the Warwick-Edinburg Mental Well-Being Scale (WEMWBS) are available (Tennant et al., 2007). Application of modern psychometric methods to the WEMWBS has been useful in the derivation of short-form measures with desirable measurement characteristics, suggesting that item inventories of this kind may prove useful as a starting point for the development of shorter, more specific measures of well-being benefit following hearing rehabilitation (Houghton et al., 2017). There is also ongoing work exploring the nature of hearing-specific well-being (Vercammen et al., 2020; Humes, 2021), which may result in measures of well-being that are more sensitive to hearing rehabilitation, although care should be taken when selecting these instruments to ensure that the breadth of well-being as an outcome domain—that it includes all aspects of life, not just hearing—is not lost.

## Collection Considerations

With respect to the time points at which outcomes should be collected, there was disagreement between professional stakeholders in the Australian hearing industry, reflecting the lack of research into the optimal time to collect outcomes measurements. While there is good evidence that auditory ability stabilises quickly after hearing aid fitting (Dawes et al., 2014b), the period following fitting, particularly for new users, is marked by ongoing adjustment during which results obtained from outcome measures may be expected to change (Turner et al., 1996). There are also a variety of personal factors which may affect the rate at which a person adjusts to hearing aids (Dawes et al., 2014a), and it is unclear how those factors may affect different domains of adjustment. As a result, further research will be required to establish the optimal time post-fitting for any particular outcome measure to be applied.

While both Consumer and Professional groups agreed that outcomes could be appropriately collected by hearing care professionals face-to-face with a patient, the difference in preference for online delivery of questionnaires—Consumer respondents ranking it second only to face-to-face collection by a hearing care professional while Professional respondents ranking it as least preferred—suggest that a range of methods are likely to be useful in practice. The principles of experience-based co-design suggest that health services, policymakers, and consumers should be involved in the design of services and the selection of appropriate metrics for their assessment (Donetto et al., 2014), and the results of this study suggest that the methods of collection of those metrics may also be a valuable subject of co-design methods. Further work canvasing views of outcomes collection methods may also identify groups of consumers and service staff who benefit from varying methods of outcome measurement collection, requiring a multi-method approach to implementation into hearing health services.

There was substantial overlap in the identified reasons for collecting outcomes and the purposes of establishing a national database of patient outcomes. Both Consumer and Professional groups highlighted the importance of promoting patient-centred care beyond solely the provision of hearing aids, the comparison of professional practice and outcomes to national benchmarks, the value of outcomes to Government and other funders in developing policy, and potential enhancements in the development of evidence-based hearing care. This overlap suggests that to the participants in this study these two ideas—that outcome measurements should be collected and that outcome measurements should be combined and analysed across health systems—may be conceptually inseparable. Indeed, several of the reasons identified for collecting outcomes, such as the targeting by Government of auditing activities, are likely to only be possible through a centralised outcomes storage and analysis system.

Respondents were clear that centralised outcomes collection systems should be designed to maximise their benefits for various parts of the hearing health system, including hearing healthcare organisations and professionals, other healthcare providers, the public, and policymakers. When designing and implementing these systems, a broad base of stakeholders should be involved in the design and implementation of data products, ensuring their applicability across the health system. Consumers, perhaps unsurprisingly, felt that such a system could and should provide important information to consumers to support their hearing rehabilitation decision-making. The provision of information to consumers alone, however, is not sufficient to ensure that they can use that information to support healthcare decision-making; careful design of the consumer-accessible outputs of these systems will be necessary to ensure that outcomes information can be useful to consumers (Hibbard and Peters, 2003).

In addition to the benefits of aggregating outcome measurements across patients for services and systems evaluation, the results of the present study also highlight the immediate utility of outcome measurements to clinicians as a basis for decisions about the future progress of individual rehabilitation programmes. Making patient outcome measurements available to clinicians and health services may therefore provide an immediate and direct benefit to the care of the patient whose outcomes are being measured. In addition to the use of baseline measurements to support rehabilitation programme planning (such as the determination that a patient may be more likely to benefit from intensive communication training to address difficulties in particular situations, or from a referral to psychological or social support to ameliorate the effects of long-standing participation restriction), outcome measurements may identify patients whose progress has been less than might be expected, prompting additional intervention from the clinician. Providing those results that are available to hearing care professionals and organisations to the patient, both in aggregate and individualised format, may provide significant benefits to clinical practice, improve engagement and uptake by hearing care professionals and service delivery organisations, and smooth the implementation of outcome measurement within hearing health systems.

Concerns relating to the development of a national outcomes database largely related to the validity, quality, trustworthiness, and comprehensiveness of the data stored within it. Both groups expressed concerns that there could be significant potential for interested parties (particularly hearing care professionals or organisations) to modify or misrepresent data in an attempt to appear more favourably in any aggregate results. This concern has implications for the ways in which data may be collected, as methods of data collection that directly involve hearing care professionals or organisations may be viewed as more susceptible to misuse than those that bypass hearing services entirely.

Interestingly, the potential impacts on professionals and organisations of receiving poor outcomes results were not considered important by either Consumer or Professional participants. Publication of health quality data can prompt quality improvement within health systems (Fung et al., 2008), with systematic and structured outcomes an important support to quality improvement activities (Kampstra et al., 2018). Within hearing rehabilitation, improvements in service quality have been seen following the publication of outcomes data in the ongoing quality register of hearing rehabilitation clinics in Sweden (Nationellt Kvalitetsregister Hörselrehabilitering, 2019). This suggests that the ongoing collection and publication of client-centred outcomes may support a move toward improving the quality of hearing rehabilitation.

## Limitations

While the results of the present study do provide important guidance for researchers, clinicians, and policymakers in the selection of outcome domains for the assessment of hearing rehabilitation, the decision in the early stages of the Delphi Review conducted with the Consumer group to restrict to a manageable number of outcome domains does mean that this list should not be considered a comprehensive description of the areas in which consumers of hearing services might experience or seek benefit from hearing rehabilitation. Indeed, several items that reached consensus in the Professional group, including reductions in social isolation and the ability to communicate in groups, were filtered out of consideration by Consumer participants at this stage. In addition, a number of items that might be considered valuable by researchers or clinicians—including feelings of empowerment, improved access to paid and volunteer work, and confidence in the effectiveness of hearing services—were also excluded from further consideration by Consumer participants. As a result, this work does not preclude the usefulness of other domains that have not been listed above, or of measures that assess other aspects of benefit. For example, where a comprehensive assessment of benefit of hearing rehabilitation is desired, the use of a general measure of improvement such as the Clinical Global Impression, which has been adapted for use in hearing rehabilitation (Öberg et al., 2009), may capture a more holistic measure of benefit, supporting the use of these more specific measures. Finally, the Delphi method used in this study involved a self-selected group of participants, and its reliance on internet-delivered text may have posed a barrier to participation for culturally and linguistically diverse Australians, those with cognitive or other difficulties, or low access to technology. Further work is required to ensure that the domains identified are indeed appropriate for the assessment of all consumers of Government-funded hearing services in Australia.

## Strengths

This study includes consumers of hearing services as primary participants in the development of recommendations for the assessment of hearing rehabilitation, which has not previously been done in this field. In addition to providing a possible example to future researchers seeking to include consumers as domain experts in their research, we believe that direct involvement of consumers in research is vital to the principles of patient- and family-centred care. It is also the first in the hearing literature to make use of the Kemeny-Young method for consensus ranking, a data-driven method with useful properties including satisfaction of the Condorcet criterion (that is, it will correctly identify the choice that is preferred over every other choice by most raters should such a choice exist) and the ready availability of "off-the-shelf" algorithms for both its exact computation and heuristic approximation.

## Conclusions

The recommendations from this study define a minimum patient-centred core outcome domain set that should be considered for the assessment of hearing rehabilitation in research and clinical practice. However, there is still significant research required to establish a set of outcome measures suitable for the measurement of each of these outcome domains. The selection of measurement instruments to be associated with a COS is a multi-stage process that will require considerable additional work, particularly given the identified lack of appropriate, validated measures for the identified domains (Prinsen et al., 2016). As part of this process, the measurement properties of developed or identified measures will need to be assessed. Preferably, this should be undertaken using modern psychometric methods such as those utilising Item Response Theory (IRT), which are particularly useful when assessing psychosocial, needs-based aspects of health (Tennant et al., 2004).

In general, these results have identified, through a consensus approach, a core outcome domain set that might be considered for the self-report evaluation of hearing rehabilitation and provide important background information for the design of methods to implement them across hearing healthcare systems. A broader set of self-report outcome domains that researchers and clinicians may also choose to collect in their particular context has also been identified. Furthermore, other outcome areas in addition to self-report, such as behavioural (e.g., speech perception, cognition) and physiological (e.g., electrophysiology) tests, need to be considered before there is a full COS for auditory rehabilitation. For self-report, which was the focus of this study, the range of suggested outcome domains, potential purposes for outcomes collection, and potential concerns with the establishment of centralised national outcomes collection and analysis systems strongly suggest that ongoing stakeholder engagement will be vital for the operationalisation of these results into any hearing healthcare system. In addition, significant further research is required on any selected or developed outcomes measurement instruments to determine the optimal time of outcomes collection following hearing rehabilitation.

## 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.

## ETHICS STATEMENT

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

## AUTHOR CONTRIBUTIONS

MF designed the study. MF and DA designed the Delphi review with subject matter advice from LH. DA managed the data collection, analysed the data, and wrote the manuscript. All authors interpreted the data, reviewed the manuscript and provided critical revision, and approved the submitted version.

## FUNDING

This work was funded by the Australian Government Department of Health.

## ACKNOWLEDGMENTS

Thanks to the project team at the National Acoustic Laboratories who supported this study, including Dr. Paul Jevelle, Margot McLelland, and Jessica Cooper.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.787607/full#supplementary-material

## REFERENCES

Access Economics Pty Ltd (2015). *Listen Hear! The Economic Impact and Cost of Hearing Loss in Australia*. Australia: Access Economics.

Akeroyd, M. A., Wright-Whyte, K., Holman, J. A., and Whitmer, W. M. (2015). A comprehensive survey of hearing questionnaires: how many are there, what do they measure, and how have they been validated? *Trials* 16:26.

Amodio, S., D'Ambrosio, A., and Siciliano, R. (2016). Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the Kemeny axiomatic approach. *Eur. J. Oper. Res.* 249, 667–676. doi: 10.1016/j.ejor.2015.08.048

Barker, A. B., Leighton, P., and Ferguson, M. A. (2017). Coping together with hearing loss: a qualitative meta-synthesis of the psychosocial experiences of people with hearing loss and their communication partners. *Int. J. Audiol.* 56, 297–305. doi: 10.1080/14992027.2017.1286695

Barker, F., MacKenzie, E., Elliott, L., and de Lusignan, S. (2015a). Outcome Measurement in Adult Auditory Rehabilitation: a Scoping Review of Measures Used in Randomized Controlled Trials. *Ear Hear.* 36:7. doi: 10.1097/AUD.0000000000000167

Barker, F., Munro, K. J., and de Lusignan, S. (2015b). Supporting living well with hearing loss: a Delphi review of self-management support. *Int. J. Audiol.* 54, 691–699. doi: 10.3109/14992027.2015.1037019

Barker, F., Mackenzie, E., Elliott, L., Jones, S., and de Lusignan, S. (2016). Interventions to improve hearing aid use in adult auditory rehabilitation. *Cochrane Database Syst. Rev.* 8:CD010342.

Bartholdi, J. J., Tovey, C. A., and Trick, M. A. (1989). The computational difficulty of manipulating an election. *Soc. Choice Welfare* 6, 227–241. doi: 10.1007/BF00295861

Bennett, R. J., Barr, C., Montano, J., Eikelboom, R. H., Saunders, G. H., Pronk, M., et al. (2020). Identifying the approaches used by audiologists to address the psychosocial needs of their adult clients. *Int. J. Audiol.* 60, 104–114. doi: 10.1080/14992027.2020.1817995

Bennett, R. J., Meyer, C. J., Eikelboom, R. H., and Atlas, M. D. (2018). Evaluating Hearing Aid Management: development of the Hearing Aid Skills and Knowledge Inventory (HASKI). *Am. J. Audiol.* 27, 333–348. doi: 10.1044/2018_AJA-18-0050

Biedl, T., Brandenburg, F. J., and Deng, X. (2006). "Crossings and Permutations," in *Graph Drawing*, eds P. Healy and N. S. Nikolov (Heidelberg: Springer), 1–12.

British Society of Audiology (2016). *Practice Guidance - Common Principles of Rehabilitation for Adults in Audiology Services*. United Kingdom: British Society of Audiology.

Chia, E. M., Wang, J. J., Rochtchina, E., Cumming, R. R., Newall, P., and Mitchell, P. (2007). Hearing impairment and health-related quality of life: the Blue Mountains Hearing Study. *Ear Hear.* 28, 187–195. doi: 10.1097/AUD.0b013e31803126b6

Commonwealth of Australia (2019). *Portfolio Budget Statements 2019-20, Budget Related Paper No. 1.9, Health Portfolio", Department of Health*. Sydney, Australia: Commonwealth of Australia.

Cox, R. M., and Alexander, G. C. (1995). The Abbreviated Profile of Hearing Aid Benefit. *Ear Hear.* 16, 176–186. doi: 10.1097/00003446-199504000-00005

Cox, R. M., Hyde, M., Gatehouse, S., Noble, W., Dillon, H., Bentler, R. A., et al. (2000). Optimal Outcome Measures, Research Priorities, and International Cooperation. *Ear Hear.* 21, 106S–115S. doi: 10.1097/00003446-200008001-00014

Danermark, B., Granberg, S., Kramer, S. E., Selb, M., and Möller, C. (2013). The Creation of a Comprehensive and a Brief Core Set for Hearing Loss Using the International Classification of Functioning, Disability and Health. *Am. J. Audiol.* 22:323. doi: 10.1044/1059-0889(2013/12-0052)

Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M., et al. (2015). Hearing Loss and Cognition: the Role of Hearing Aids, Social Isolation and Depression. *PLoS One* 10:e0119616. doi: 10.1371/journal.pone.0119616

Dawes, P., Maslin, M., and Munro, K. J. (2014a). 'Getting used to' hearing aids from the perspective of adult hearing-aid users. *Int. J. Audiol.* 53, 861–870. doi: 10.3109/14992027.2014.938782

Dawes, P., Munro, K. J., Kalluri, S., and Edwards, B. (2014b). Acclimatization to Hearing Aids. *Ear Hear.* 35, 203–212. doi: 10.1097/aud.0b013e3182a8eda4

Demorest, M. E., and Erdman, S. A. (1986). Scale composition and item analysis of the Communication Profile for the Hearing Impaired. *J. Speech Lang. Hear. Res.* 29, 515–535. doi: 10.1044/jshr.2904.535

Demorest, M. E., and Erdman, S. A. (1987). Development of the communication profile for the hearing impaired. *J. Speech Hear. Disord.* 52, 129–143. doi: 10.1044/jshd.5202.129

Department of Health (n.d.). *Am I eligible to receive fully subsidised hearing services?*. Available Online at: https://www.hearingservices.gov.au/wps/portal/hso/site/eligibility/programhelp/eligibility/ (accessed on 24 Jan, 2022).

Dillon, H., Birtles, G., and Lovegrove, R. (1999). Measuring the outcomes of a national rehabilitation program: normative data for the Client Oriented Scale of Improvement (COSI) and the Hearing Aid Users Questionnaire (HAUQ). *J. Am. Acad. Audiol.* 10, 67–79.

Dillon, H., James, A., and Ginis, J. (1997). Client Oriented Scale of Improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids. *J. Am. Acad. Audiol.* 8, 27–43.

Donetto, S., Pierri, P., Tsianakas, V., and Robert, G. (2014). "Experience-based Co-design and healthcare improvement: realising participatory design in the public sector," in *Fourth Service Design and Innovation conference*, (Lancaster, UK: ServDes), 10.

Emond, E. J., and Mason, D. W. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *J. Multi-Criteria Decis. Anal.* 11, 17–28. doi: 10.1002/mcda.313

Fackrell, K., Smith, H., Colley, V., Thacker, B., Horobin, A., Haider, H. F., et al. (2017). Core Outcome Domains for early phase clinical trials of sound-, psychology-, and pharmacology-based interventions to manage chronic subjective tinnitus in adults: the COMIT'ID study protocol for using a Delphi process and face-to-face meetings to establish consensus. *Trials* 18:388. doi: 10.1186/s13063-017-2123-0

Ferguson, M. A., Brandreth, M., Brassington, W., Leighton, P., and Wharrad, H. (2016). A Randomized Controlled Trial to Evaluate the Benefits of a Multimedia Educational Program for First-Time Hearing Aid Users. *Ear Hear.* 37, 123–136. doi: 10.1097/AUD.0000000000000237

Ferguson, M. A., and Henshaw, H. (2015). How Does Auditory Training Work? Joined-Up Thinking and Listening. *Semin. Hear.* 36, 237–249. doi: 10.1055/s-0035-1568985

Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-Jones, M., Barker, F., and Hoare, D. J. (2017). Hearing aids for mild to moderate hearing loss in adults. *Cochrane Database Syst. Rev.* 9:CD012023.

Ferguson, M. A., Leighton, P., Brandreth, M., and Wharrad, H. (2018). Development of a multimedia educational programme for first-time hearing aid users: a participatory design. *Int. J. Audiol.* 57, 600–609. doi: 10.1080/14992027.2018.1457803

Ferguson, M. A., Maidment, D., Henshaw, H., and Heffernan, E. (2019). Evidence-Based Interventions for Adult Aural Rehabilitation: that Was Then, This Is Now. *Semin. Hear.* 40, 068–084. doi: 10.1055/s-0038-1676784

Fung, C. H., Lim, Y.-W., Mattke, S., Damberg, C., and Shekelle, P. G. (2008). Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann. Intern. Med.* 148, 111–123. doi: 10.7326/0003-4819-148-2-200801150-00006

Gatehouse, S. (1999). Glasgow Hearing Aid Benefit Profile: derivation and validation of a client-centered outcome measure for hearing aid services. *J. Am. Acad. Audiol.* 10, 80–103.

Granberg, S., Dahlström, J., Möller, C., Kähäri, K., and Danermark, B. (2014a). The ICF Core Sets for hearing loss – researcher perspective. Part I: systematic review of outcome measures identified in audiological research. *Int. J. Audiol.* 53, 65–76. doi: 10.3109/14992027.2013.851799

Granberg, S., Möller, K., Skagerstrand, Å, Möller, C., and Danermark, B. (2014b). The ICF Core Sets for hearing loss: researcher perspective, Part II: linking outcome measures to the International Classification of Functioning, Disability and Health (ICF). *Int. J. Audiol.* 53, 77–87. doi: 10.3109/14992027.2013.858279

Granberg, S., Pronk, M., Swanepoel, D. W., Kramer, S. E., Hagsten, H., Hjaldahl, J., et al. (2014c). The ICF core sets for hearing loss project: functioning and disability from the patient perspective. *Int. J. Audiol.* 53, 777–786. doi: 10.3109/14992027.2014.938370

Hall, D. A., Haider, H., Kikidis, D., Mielczarek, M., Mazurek, B., Szczepek, A. J., et al. (2015). Toward a Global Consensus on Outcome Measures for Clinical Trials in Tinnitus: report From the First International Meeting of the COMiT Initiative, November 14, 2014, Amsterdam. Netherlands. *Trends Hear.* 19:233121651558027. doi: 10.1177/2331216515580272

Hall, D. A., Kitterick, P., Heffernan, E., Fackrell, K., Lucas, L., and Ferguson, M. A. (2019). How Do We Know That Our Patients Have Benefitted From Our ENT/Audiological Interventions? Presented at the Annual Meeting of ADANO 2016 in Berlin. *Otol. Neurotol.* 40, e474–e481. doi: 10.1097/MAO.0000000000001937

Hall, D. A., Smith, H., Hibbert, A., Colley, V., Haider, H. F., Horobin, A., et al. (2018). The COMiT'ID Study: Developing Core Outcome Domains Sets for Clinical Trials of Sound-, Psychology-, and Pharmacology-Based Interventions for Chronic Subjective Tinnitus in Adults. *Trends Hear.* 22:2331216518814384. doi: 10.1177/2331216518814384

Heffernan, E., Coulson, N. S., and Ferguson, M. A. (2018a). Development of the Social Participation Restrictions Questionnaire (SPaRQ) through consultation with adults with hearing loss, researchers, and clinicians: a content evaluation study. *Int. J. Audiol.* 57, 791–799. doi: 10.1080/14992027.2018.1483585

Heffernan, E., Maidment, D. W., Barry, J. G., and Ferguson, M. A. (2018b). Refinement and Validation of the Social Participation Restrictions Questionnaire: an Application of Rasch Analysis and Traditional Psychometric Analysis Techniques. *Ear Hear.* 40:12. doi: 10.1097/AUD.0000000000000618

Heffernan, E., Coulson, N. S., Henshaw, H., Barry, J. G., and Ferguson, M. A. (2016). Understanding the psychosocial experiences of adults with mild-moderate hearing loss: an application of Leventhal's self-regulatory model. *Int. J. Audiol.* 55, S3–S12. doi: 10.3109/14992027.2015.1117663

Heffernan, E., Habib, A., and Ferguson, M. A. (2019). Evaluation of the psychometric properties of the social isolation measure (SIM) in adults with hearing loss. *Int. J. Audiol.* 58, 45–52. doi: 10.1080/14992027.2018.1533257

Helmer, O. (1967). *Analysis of the Future: The Delphi Method".* California: RAND Corporation.

Henshaw, H., and Ferguson, M. A. (2013). Efficacy of Individual Computer-Based Auditory Training for People with Hearing Loss: a Systematic Review of the Evidence. *PLoS One* 8:e62836. doi: 10.1371/journal.pone.0062836

Hibbard, J. H., and Peters, E. (2003). Supporting Informed Consumer Health Care Decisions: data Presentation Approaches that Facilitate the Use of Information in Choice. *Annu. Rev. Public Health* 24, 413–433. doi: 10.1146/annurev.publhealth.24.100901.141005

Houghton, S., Wood, L., Marais, I., Rosenberg, M., Ferguson, R., and Pettigrew, S. (2017). Positive Mental Well-Being: a Validation of a Rasch-Derived Version of the Warwick-Edinburgh Mental Well-Being Scale. *Assessment* 24, 371–386. doi: 10.1177/1073191115609995

Hsu, C.-C., and Sandford, B. A. (2007). The Delphi Technique: making Sense Of Consensus. *Pract. Assess. Res. Eval.* 12:8.

Humes, L. E. (2021). Development of the SWB-HL: a Scale of the Subjective Well-Being of Older Adults With Hearing Loss. *Front. Psychol.* 12:640165. doi: 10.3389/fpsyg.2021.640165

Kampstra, N. A., Zipfel, N., van der Nat, P. B., Westert, G. P., van der Wees, P. J., and Groenewoud, A. S. (2018). Health outcomes measurement and organizational readiness support quality improvement: a systematic review. *BMC Health Serv. Res.* 18:1005. doi: 10.1186/s12913-018-3828-9

Kemeny, J. G. (1959). Mathematics without Numbers. *Daedalus* 88, 577–591.

Kirkham, J. J., Davis, K., Altman, D. G., Blazeby, J. M., Clarke, M., Tunis, S., et al. (2017). Core Outcome Set-STAndards for Development: the COS-STAD recommendations. *PLoS Med.* 14:e1002447. doi: 10.1371/journal.pmed.1002447

Lawrence, B. J., Jayakody, D. M. P., Henshaw, H., Ferguson, M. A., Eikelboom, R. H., Loftus, A. M., et al. (2018). Auditory and Cognitive Training for Cognition in Adults With Hearing Loss: a Systematic Review and Meta-Analysis. *Trends Hear.* 22:233121651879209. doi: 10.1177/2331216518792096

Lin, F. R. (2011). Hearing Loss and Cognition Among Older Adults in the United States. *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 1131–1136. doi: 10.1093/gerona/glr115

Maidment, D. W., Barker, A. B., Xia, J., and Ferguson, M. A. (2016). Effectiveness of alternative listening devices to conventional hearing aids for adults with hearing loss: a systematic review protocol. *BMJ Open* 6:e011683. doi: 10.1136/bmjopen-2016-011683

Nationellt Kvalitetsregister Hörselrehabilitering (2019). *Årsrapport 2019*. Göteborg: Nationellt kvalitetsregister för öron-, näs- och halssjukvård.

Öberg, M., Wanstrom, G., Hjertman, H., Lunner, T., and Andersson, G. (2009). Development and initial validation of the 'Clinical Global Impression' to measure outcomes for audiological rehabilitation. *Disabil. Rehabil.* 31, 1409–1417. doi: 10.1080/09638280802621408

OMET Initiative (n.d.). *COMET Initiative*. Available online at: https://www.comet-initiative.org/ (accessed April 19, 2022).

Perez, E., and Edmonds, B. A. (2012). A Systematic Review of Studies Measuring and Reporting Hearing Aid Usage in Older Adults since 1999: a Descriptive

Summary of Measurement Tools. *PLoS One* 7:e31831. doi: 10.1371/journal.pone.0031831

PricewaterhouseCoopers Australia (2017). *Review of services and technology supply in the Hearing Services Program: Final Report*. Barangaroo: PricewaterhouseCoopers Australia.

Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., et al. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials* 17:449. doi: 10.1186/s13063-016-1555-2

Routledge, D. (2017). *Human Services Outcomes Framework Guide", Finance, Services & Innovation*. Australia: NSW Government.

Scarinci, N., Worrall, L., and Hickson, L. (2012). Factors associated with third-party disability in spouses of older people with hearing impairment. *Ear Hear.* 33, 698–708. doi: 10.1097/AUD.0b013e31825aab39

Schow, R. L., and Nerbonne, M. A. (1982). Communication Screening Profile: use with Elderly Clients. *Ear Hear.* 3, 135–147. doi: 10.1097/00003446-198205000-00007

Sereda, M., Hoare, D. J., Nicholson, R., Smith, S., and Hall, D. A. (2015). Consensus on Hearing Aid Candidature and Fitting for Mild Hearing Loss, With and Without Tinnitus: delphi Review. *Ear Hear.* 36:13. doi: 10.1097/AUD.0000000000000140

Tennant, A., McKenna, S. P., and Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 7, S22–S26. doi: 10.1111/j.1524-4733.2004.7s106.x

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., et al. (2007). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual. Life Outcomes* 5:63. doi: 10.1186/1477-7525-5-63

Turner, C. W., Humes, L. E., Bentler, R. A., and Cox, R. M. (1996). ). A review of past research on changes in hearing aid benefit over time. *Ear Hear.* 17, 14S–25S. doi: 10.1097/00003446-199617031-00003

Vercammen, C., Ferguson, M., Kramer, S. E., Meis, M., Singh, G., Timmer, B., et al. (2020). Well-Hearing Is Well-Being: a Phonak position statement about why hearing healthcare is vital for healthy living. *Hear. Rev.* 27, 18–22.

Wallhagen, M. I. (2010). The Stigma of Hearing Loss. *Gerontologist* 50, 66–75. doi: 10.1093/geront/gnp107

Williamson, P., and Clarke, M. (2012). The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: its Role in Improving Cochrane Reviews. *Cochrane Database Syst. Rev.* 5:ED000041. doi: 10.1002/14651858.ED000041

Wong, L. L. N., and Hickson, L. (2012). *Evidence-based practice in audiology: evaluating interventions for children and adults with hearing impairment*. San Diego, CA: Plural Publishing.

**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.

Frontiers in Neuroscience
ORIGINAL RESEARCH
published: 06 June 2022
doi: 10.3389/fnins.2022.786939

Check for updates

# The Effectiveness of Unilateral Cochlear Implantation on Performance-Based and Patient-Reported Outcome Measures in Finnish Recipients

*Aarno Dietz[1]*†, Antje Heinrich[2]†, Timo Törmäkangas[3], Matti Iso-Mustajärvi[1], Petrus Miettinen[1,4], Tytti Willberg[4,5] and Pia H. Linder[1]*

[1] *Department of Otolaryngology, Kuopio University Hospital, Kuopio, Finland,* [2] *Division of Human Communication, Development and Hearing, School of Health Sciences, University of Manchester, Manchester, United Kingdom,* [3] *Department of Health Sciences, University of Jyväskylä, Jyväskylä, Finland,* [4] *Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland,* [5] *Department of Otorhinolaryngology – Head and Neck Surgery, Turku University Hospital, Turku, Finland*

OPEN ACCESS

**Edited by:**
*Melanie Ferguson,*
*Curtin University, Australia*

**Reviewed by:**
*Patrick K. A. Neff,*
*Swiss Federal Institute of Technology Lausanne, Switzerland*
*Robert Eikelboom,*
*Ear Science Institute Australia, Australia*
*Richard Charles Dowell,*
*The University of Melbourne, Australia*

***Correspondence:**
*Aarno Dietz*
*aarno.dietz@kuh.fi*

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

**Specialty section:**
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *30 September 2021*
***Accepted:*** *19 April 2022*
***Published:*** *06 June 2022*

**Citation:**
*Dietz A, Heinrich A, Törmäkangas T, Iso-Mustajärvi M, Miettinen P, Willberg T and Linder PH (2022) The Effectiveness of Unilateral Cochlear Implantation on Performance-Based and Patient-Reported Outcome Measures in Finnish Recipients. Front. Neurosci. 16:786939. doi: 10.3389/fnins.2022.786939*

Understanding speech is essential for adequate social interaction, and its functioning affects health, wellbeing, and quality of life (QoL). Untreated hearing loss (HL) is associated with reduced social activity, depression and cognitive decline. Severe and profound HL is routinely rehabilitated with cochlear implantation. The success of treatment is mostly assessed by performance-based outcome measures such as speech perception. The ultimate goal of cochlear implantation, however, is to improve the patient's QoL. Therefore, patient-reported outcomes measures (PROMs) would be clinically valuable as they assess subjective benefits and overall effectiveness of treatment. The aim of this study was to assess the patient-reported benefits of unilateral cochlear implantation in an unselected Finnish patient cohort of patients with bilateral HL. The study design was a prospective evaluation of 118 patients. The patient cohort was longitudinally followed up with repeated within-subject measurements preoperatively and at 6 and 12 months postoperatively. The main outcome measures were one performance-based speech-in-noise (SiN) test (Finnish Matrix Sentence Test), and two PROMs [Finnish versions of the Speech, Spatial, Qualities of Hearing questionnaire (SSQ) and the Nijmegen Cochlear Implant Questionnaire (NCIQ)]. The results showed significant average improvements in SiN scores, from +0.8 dB signal-to-noise ratio (SNR) preoperatively to −3.7 and −3.8 dB SNR at 6 and12 month follow-up, respectively. Significant improvements were also found for SSQ and NCIQ scores in all subdomains from the preoperative state to 6 and 12 months after first fitting. No clinically significant improvements were observed in any of the outcome measures between 6 and 12 months. Preoperatively, poor SiN scores were associated with low scoring in several subdomains of the SSQ and NCIQ. Poor preoperative SiN scores and low PROMs scoring were significantly associated with larger postoperative improvements. No significant association was found between SiN scores and PROMs postoperatively. This study demonstrates significant benefits of cochlear implantation in the performance-based and patient-reported outcomes in an unselected patient sample. The lack of association between performance and PROMs scores postoperatively suggests that both capture unique aspects of benefit, highlighting the need to clinically implement PROMs in addition to performance-based measures for a more holistic assessment of treatment benefit.

**Keywords: cochlear implant, outcome measures, Quality of Life, SSQ, NCIQ, speech perception**

## INTRODUCTION AND PURPOSE OF THE STUDY

The ability to understand speech is the most important application of human hearing. Verbal communication enables us to conduct sophisticated social interaction and social relationships, which are essential to our health and wellbeing. Difficulties recognizing speech in the presence of background noise or in multitalker situations is the most common manifestation of hearing loss (HL; Kramer et al., 1998) and may represent a starting point for a gradually progressing social disconnection. It is therefore not surprising that untreated HL is associated with loss of social activities and autonomy, as well as depression and even cognitive decline (Lin et al., 2011; Loughrey et al., 2018). Given its increasing prevalence and serious socioemotional consequences, HL ranks among the greatest public health challenges globally in the coming decades (World Health Organization [WHO], 2021).

Most mild and moderate HL is rehabilitated with conventional hearing aids; severe and profound HL is commonly treated with cochlear implants (CI). In both cases, the primary goal of rehabilitation is to provide a level of verbal communication that enables satisfactory social interaction and performance in most everyday sound environments and situations (i.e., at home, on the phone, in a car, in a restaurant, at work, etc.), thereby improving the patient's quality of life (QoL).

Predicting the success of cochlear implantation for an individual patient is challenging and requires a holistic approach (Boisvert et al., 2020) beyond measuring aided thresholds, as these do not provide meaningful information about the functional hearing relevant for most everyday hearing situations (Vermiglio et al., 2012). Word and/or sentence perception in quiet have been the most commonly used supra-threshold clinical outcome measures. Numerous studies have shown that cochlear implantation reliably restores sound audibility, thereby enabling speech perception in quiet and non-reverberant surroundings (i.e., sound booth; Gifford et al., 2008; Dietz et al., 2015; Boisvert et al., 2020). However, speech perception tests in quiet are not able to measure functional hearing relevant for most everyday hearing situations. Moreover, speech perception tests in quiet are prone to ceiling effects (Gifford et al., 2008; Dietz et al., 2015). Thus, speech perception tests conducted in background noise are regarded as a more adequate way to measure functionally relevant performance outcomes of cochlear implantation, since background noise better approximates complex listening situations (Holden et al., 2013; Dietz et al., 2015). However, speech-in-noise (SiN) tests have also been criticized for not fully capturing the benefits of cochlear implantation. A more comprehensive assessment may be provided by patient-reported outcome measures (PROMs; McRackan et al., 2019). Although PROMs are often used for hearing aid validation, they are less commonly use for the outcome evaluation of cochlear implantation. This is the case even though PROMs seem well-placed to reflect the impact of the change in hearing performance on a patient's QoL. PROMs are more holistic than performance-based outcome measures in that they assess not only functional aspects of hearing but also hearing-related socioemotional consequences such as social interaction, self-esteem (SE) and emotional wellbeing (Mertens et al., 2020).

One way to categorize outcome measures is to use the International Classification of Functioning, Disability and Health (ICF; World Health Organsation [WHO], 2001). The ICF classifies health according to three domains, namely an individual's body function and structure, their activity limitations, and their participation restrictions (World Health Organsation [WHO], 2001). PROMs typically assess activity limitations and/or participation restrictions. In audiological practice these limitations and restrictions are typically tied to communication. Two PROMs of particular importance for the current study are the Nijmegen Cochlear Implant Questionnaire (NCIQ; Hinderink et al., 2000) and the Speech Spatial and Qualities of Hearing scale (SSQ; Gatehouse and Noble, 2004). The NCIQ was specifically developed for CI users and is currently the most commonly used QoL-questionnaire for this patient group. It has been translated into many languages (Sanchez-Cuadrado et al., 2015; Ottaviani et al., 2016; Santos et al., 2017). The NCIQ assesses mainly participation restrictions related to social and emotional aspects of wellbeing. The SSQ is another PROM that has been previously used with CI users, although it has not been fully validated for this population. The SSQ assesses activity limitations related to speech perception, spatial hearing (SH), and sound quality (SQ) for different everyday situations.

In summary, changes in the patient's activity limitations, participation restrictions, wellbeing, and QoL are rarely assessed in the clinical practice of cochlear implantation, even though these dimensions add unique insights into rehabilitation success beyond performance-based scores.

To provide an objective picture of a patient group, it is paramount to minimize any reporting biases. One way of doing this is to sample patients prospectively without any regard for the success of the intervention or difficulties along the way, and to obtain both pre- and postoperative measures from the same patient. Such a design presents a contrast to many studies that use retrospective and cross-sectional designs where patients are only tested postoperatively. The present study aims to avoid potential reporting bias by using a prospective non-selective longitudinal design that compares the change of QoL, activity limitations, participation restrictions, and SiN perception in adult patients undergoing cochlear implantation.

The primary aim of this study was to investigate performance-based and patient-reported outcome measures (PROMs) after cochlear implantation in an unselected, consecutive Finnish patient cohort undergoing unilateral cochlear implantation. We wanted to understand the benefits of cochlear implantation more fully by investigating the following three questions: (1) what are the changes to communicative ability and QoL in response to cochlear implantation? (2) What is the timeline of change? (3) To what extent do behavioral SiN scores and patient-reported disability scores covary with QoL measures? The ultimate goal was to predict a patient's rehabilitation success more accurately.

# MATERIALS AND METHODS

## Study Design

This was a prospective patient cohort study. Patients referred to the Kuopio University Hospital for unilateral cochlear implantation were given the option to participate. According to the institution's clinical routine, patients were evaluated preoperatively at 2–4 weeks before surgery, at which point they filled out the Finnish NCIQ and the Finnish SSQ. The SiN test was administered within 3 months of the preoperative questionnaire administration and was carried out in the best-aided condition, i.e., according to the device configuration that the patient was using in their everyday life at that point. The postoperative follow-up appointments were scheduled 6 and 12 months after the first fitting of the CI sound processor when patients filled out the questionnaires again and underwent the speech perception test in noise, again in the patient's best-aided condition. Most patients used bilateral hearing aids prior to cochlear implantation; however, many patients stopped using their contralateral hearing aid after implantation.

## Participants

We recruited 134 adult patients referred for cochlear implantation at the Kuopio University Hospital from January 1, 2018 to December 31, 2020. We excluded patients referred for cochlear implantation because of single-sided deafness or referred for sequential bilateral implantation. Other exclusion criteria were a diagnosis of dementia or neurological or other health conditions that severely impair vision or mobility (as judged by the study physician). Patients who, despite their agreement to participate, did not respond to either the preoperative or to one of the postoperative questionnaires were also excluded. A total of 118 patients were included in the analyses. Patient demographics, preoperative pure-tone averages and surgical data are summarized in **Table 1**.

**TABLE 1** | Patient demographics and preoperative unaided pure-tone average (0.5–4 kHz).

| | Mean | Median | Min | Max | SD |
|---|---|---|---|---|---|
| Age (years) | 62.2 | 66.4 | 18 | 88 | 29.5 |
| | **Preoperative PTA $_{(0.5-4\ kHz)}$ (dB HL)** | | | | |
| BEHL | 80.5 | 81.9 | 33.8 | 110 | 18.0 |
| WEHL | 93.5 | 87.5 | 43.8 | 110 | 13.8 |
| **Etiology of hearing loss** | | **(*n*)** | | | **(*n*)** |
| Unknown | | 66 | **Sex** | | |
| Meniere's disease | | 20 | Female | | 53 |
| Otosclerosis | | 6 | Male | | 65 |
| Congenital SNHL | | 17 | | | |
| NSSNHL | | 4 | **Ear** | | |
| SSNHL | | 2 | Right | | 66 |
| Other | | 3 | Left | | 52 |

*BEHL, better ear hearing level; WEHL, worse ear hearing level; SNHL, sensorineural hearing loss; NSSNHL, non-syndromic sensorineural hearing loss; and SSNHL, syndromic sensorineural hearing loss.*

## Tests

### The Finnish Nijmegen Cochlear Implant Questionnaire

The NCIQ comprises 60 questions divided into six subdomains of 10 questions each: basic sound perception (BSP), advanced sound perception (ASP), speech production (SPr), SE, activity (ACT), and SI. The answers to the questions were provided on a 5-point Likert scale (never, rarely, sometimes, often, and always) and scored with values of 0, 25, 50, 75, and 100. Participants also had the possibility to answer: "I don't know" or "not applicable" to a question. These responses led to the exclusion of the question. The final score was the average of all responses and could range from 0 to 100. A higher score represented higher functioning. Separate scores were calculated for each NCIQ subdomain. Because the study was conducted with a Finnish population and an official Finnish translation of the NCIQ does not exist, the NCIQ was custom translated into Finnish.

### The Finnish Speech, Spatial, and Qualities of Hearing Questionnaire

The SSQ comprises 49 questions divided into three subdomains: Speech Perception (SP), Spatial Hearing (SH) and other qualities of Hearing (SQ). The answers were provided on an 11-point Likert scale, ranging from 0 to 10. Answer scores were averaged for a final score and could range from 0 to 10. Separate scores were calculated for each subdomain and for an overall score. Because the study was conducted with a Finnish population and an official Finnish translation of the SSQ does not exist, we adapted the SSQ to Finnish culture and language.

### The Finnish Matrix Sentence Test

The Finnish Matrix Sentence Test (FMST) was used as the SiN test. The FMST uses semantically unpredictable five-word sentences arranged in 20-item test lists (Dietz et al., 2014). The FMST has been validated in CI patients and has been found to be sensitive to subtle changes in hearing performance in normal-hearing participants as well as in CI recipients (Dietz et al., 2014, 2015). The clinical test protocol has been described in detail by Dietz et al. (2015). The background noise consisted of a stationary speech-shaped noise generated from the speech material and was presented at a fixed level of 65 dB SPL. The level of the speech signal changed adaptively to converge to each patient's individual speech reception threshold in noise (SRTN), which is the signal-to-noise ratio (SNR) at which the patient recognizes 50% of the test items correctly. A total of three test lists were presented. The first list was always presented at a fixed SNR of +10 dB (i.e., signal 75 dB SPL, noise 65 dB SPL). The second and third test lists were administered with the adaptive measurement procedure. Only the third list was used as SRTN outcome measure. In patients with very poor hearing (defined here as those who scored <70% at +10 dB SNR), adaptive SRT measurements are not reliable; these patients thus did not undergo adaptive measurements (Dietz et al., 2015; Dingemanse and Goedegebure, 2019). In these cases, we defined a threshold of +5 dB SNR. This procedure resulted in two SiN measurements: (1) perceptual accuracy (in percent) at SNR +10 dB as measured by List 1 and (2) the SNR at 50% perceptual accuracy, i.e., the SRTN as measured by List 3.

### Ethical Considerations

All patients were informed about the study aims and gave their written informed consent. The study complied with the Declaration of Helsinki on biomedical research involving human participants and received ethical approval from the Research Ethics Committee of the Northern Savo Hospital District (1327/2018).

### Data Analysis and Statistics

Descriptive statistics are reported in the form of expected marginal means and 95% confidence intervals from univariate changes across the three time points (preoperative, postoperative 6 months, and postoperative 12 months) for all SiN results (+10 dB SNR and SRTN) and PROMs (NCIQ and SSQ). We also report $p$-values for the tests of differences between time points.

For adaptive SRT measurements in noise (SRTN) we used a Tobit model to account for censoring (Greene, 2003). Censoring refers to a situation in which we do not know the true value of a datapoint, or the observed value is too imprecise for values at or below a threshold and we only know that the true observation was lower than the threshold. Using this model was necessary because we used set values of +5 dB SNR for some of the participants, which then led to inaccuracies in statistical estimates. As a result, we adjusted the SRTN measurements as a left-censored normally distributed variable.

Longitudinal changes were assessed using a univariate and a bivariate latent change score model from the latent change score model framework (McArdle, 2009). In this model, latent variables represent individual changes occurring between time point pairs while adjusting for the baseline measurement. **Supplementary Figures 1, 2** illustrate how these changes were modeled statistically. This model can be seen as an extension of the paired $t$-test over multiple time points (Coman et al., 2013) with the option to relax the assumptions of traditional models of change. We centered all time points on the baseline, i.e., preoperative mean, and thus, the unadjusted means of the change-variables correspond to baseline-adjusted paired $t$-tests for the changes. For univariate assessments we report baseline-adjusted standardized variances of change to enable assessment of change in variability over time. Standardized covariances between change variables are the residual correlations adjusted for baseline measurement. For bivariate assessment of change we report baseline-adjusted standardized covariances across SiN results and across the NCIQ BSP subdomain, NCIQ ASP subdomain, SSQ total, and SSQ subdomain scores. We also computed the unadjusted correlations of SSQ total and subdomain scores with the NCIQ subdomain scores. Using the R programming environment (R Core Team, 2020), mixed model parameters were estimated with package nlme (version 3.1-148) and marginal means and pairwise tests were computed with package emmeans (version 1.5.1, Pinheiro et al., 2020). Change score modeling was conducted in Mplus (version 7.4, Muthén and Muthén, (1998–2015)). Bivariate correlations were estimated and tested with the stats package in R (Length, 2020).

## RESULTS

### Effect of Cochlear Implantation on Speech Perception in Noise and Patient-Reported Outcomes Measures

The mean SiN scores at +10 dB SNR improved significantly from 76% preoperatively to 87 and 90% at 6 and 12 month follow-ups, respectively. Both improvements (pre- to 6 months postoperative and 6–12 months postoperative) were statistically significant (**Table 2**). The mean SRTNs improved significantly from −0.8 dB SNR preoperatively to −3.7 and −3.8 dB SNR at the 6 and 12 month follow-up, respectively, (**Figure 1**).

The NCIQ and SSQ subdomain scores were analyzed using the univariate latent change score model. The results of estimated marginal means from unadjusted mixed models (rather than means of raw data) are shown in **Figure 2**. **Table 2** shows the statistical results. For the NCIQ we observed significant improvements from preoperative testing to the 6 and 12 month follow-ups for all subdomains. We observed no additional significant improvements between the 6 and 12 month follow-ups. However, a slight but statistically significant decline ($p = 0.001$) was observed in the subdomain "social interaction." For the SSQ scores we observed a statistically significant increase from preoperative testing to the 6 month follow-up for all scores. There was no further statistically significant improvement in any of the SSQ subdomains between 6 and 12 month follow-up points.

**Table 2** also shows the regression coefficients for the preoperative adjustment of change scores. Specifically, this analysis shows that poorer preoperative SiN scores at +10 dB SNR and SRTN as well as PROM (NCIQ and SSQ) scores were significantly associated with larger improvements at 6 months follow-up. However, there was no association between the preoperative values and the change occurring between 6 and 12 months for any of the outcome measures.

### Covariance Analysis Between Speech Perception in Noise and Patient-Reported Outcomes Measures

The standardized covariance parameters between SiN scores at constant +10 dB SNR and SRTN as well as PROMs as calculated by the bivariate latent change score model are

**TABLE 2 |** Unstandardized means of change and regression coefficients for preoperative adjustment of change scores, and standardized covariance parameters in univariate latent change score models.

| | Means | | | | | | Regression coefficients | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu_{\Delta PO-6\ m}$ | | | $\mu_{\Delta 6-12\ m}$ | | | Pre-op → $\Delta_{PO-6\ m}$ | | | Pre-op → $\Delta_{6-12\ m}$ | | |
| **Variable** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** |
| SNR+10 dB | 17.02 | 1.69 | **<0.001** | 2.95 | 1.40 | **0.035** | −0.79 | 0.05 | **<0.001** | −0.02 | 0.04 | 0.616 |
| SRT | −2.61 | 0.24 | **<0.001** | −0.42 | 0.19 | **0.027** | −0.66 | 0.08 | **<0.001** | 0.11 | 0.07 | 0.097 |
| NCIQ BSP | 22.45 | 1.49 | **<0.001** | 0.15 | 1.44 | 0.918 | −0.70 | 0.07 | **<0.001** | 0.04 | 0.07 | 0.552 |
| NCIQ ASP | 17.24 | 1.54 | **<0.001** | −0.04 | 1.39 | 0.978 | −0.80 | 0.08 | **<0.001** | 0.14 | 0.07 | 0.059 |
| NCIQ SPr | 7.86 | 1.49 | **<0.001** | 1.40 | 1.41 | 0.321 | −0.53 | 0.08 | **<0.001** | −0.03 | 0.08 | 0.658 |
| NCIQ SE | 13.76 | 1.66 | **<0.001** | 1.65 | 1.35 | 0.221 | −0.61 | 0.10 | **<0.001** | 0.03 | 0.09 | 0.764 |
| NCIQ ACT | 18.24 | 2.10 | **<0.001** | 2.59 | 1.93 | 0.179 | −0.69 | 0.10 | **<0.001** | 0.10 | 0.10 | 0.292 |
| NCIQ SI | 19.18 | 1.94 | **<0.001** | −5.39 | 1.55 | **0.001** | −0.56 | 0.09 | **<0.001** | 0.02 | 0.07 | 0.762 |
| SSQ total | 1.89 | 0.18 | **<0.001** | 0.00 | 0.17 | 1.000 | −0.51 | 0.12 | **<0.001** | 0.07 | 0.11 | 0.541 |
| SSQ SP | 2.11 | 0.22 | **<0.001** | −0.09 | 0.17 | 0.619 | −0.60 | 0.13 | **<0.001** | 0.04 | 0.11 | 0.707 |
| SSQ SH | 1.84 | 0.21 | **<0.001** | −0.26 | 0.21 | 0.220 | −0.54 | 0.12 | **<0.001** | 0.24 | 0.12 | 0.055 |
| SSQ SQ | 1.77 | 0.19 | **<0.001** | 0.20 | 0.19 | 0.313 | −0.56 | 0.11 | **<0.001** | 0.03 | 0.11 | 0.816 |

*SNR+10 dB, speech reception score; SRT, speech reception threshold; NCIQ, Nijmegen Cochlear Implant Questionnaire; SSQ, Speech, Spatial and Qualities of Hearing scale; NCIQ BSP, basic sound perception subdomain of NCIQ; NCIQ ASP, advanced sound perception subdomain of NCIQ; NCIQ SPr, speech production subdomain of NCIQ; NCIQ SE, self-esteem subdomain of NCIQ; NCIQ ACT, activity subdomain of NCIQ; NCIQ SI, social interactions subdomain of NCIQ; SSQ SP, speech perception subdomain of SSQ; SSQ SH, spatial hearing subdomain of SSQ; SSQ SQ, sound quality subdomain of SSQ; pre-op and PO, pre-operative; $\mu_i$, mean change in i; $\sigma_{x,y}$, covariance between x and y; $\Delta_{PO-6\ m}$, change score between pre-operative and 6 months values; $\Delta_{6-12\ m}$, change score between 6 and 12 months values; Est, estimate; SE, standard error; p, p-value. Bold type face indicates $p < 0.05$; Standardization with respect to observed variables. Mean for pre-op measurement was centered to zero.*

presented in **Table 3**. Significant positive covariances were found preoperatively between SiN perception scores at +10 dB SNR and the NCIQ BSP subdomain ($p$ = 0.001) total SSQ and SSQ SP subdomain (both $p \leq 0.001$). Significant negative covariances were found preoperatively between the SiN perception scores at +10 dB SNR and the NCIQ ASP and SSQ SH subdomains (both $p < 0.001$). In terms of change scores (preoperative to 6 months follow-up) significant positive covariances were found between the changes in SiN score at +10 dB SNR and the SSQ total ($p$ = 0.012) and between SiN score and the SSQ SQ subdomain ($p$ = 0.004). In addition, the analyses showed significant covariances between the change scores of SSQ SQ and SiN scores at constant +10 dB SNR 6 month postoperative in both directions of prediction.

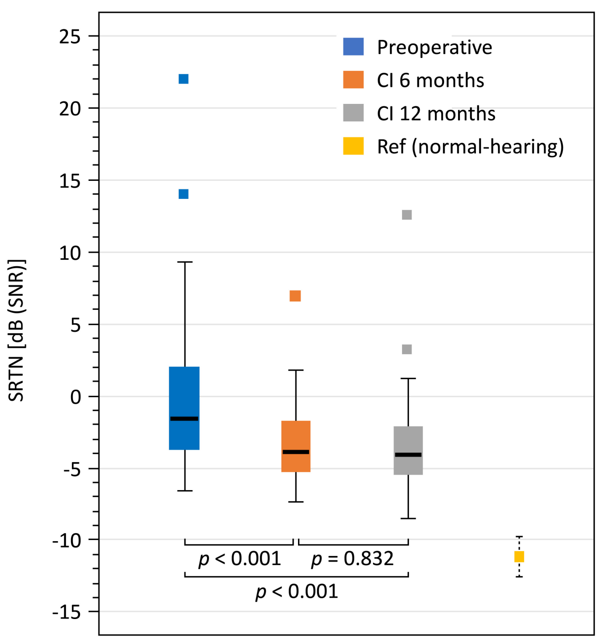


**FIGURE 1 |** Boxplots for preoperative, 6 and 12 months SRTN including normal-hearing reference mean and 95% confidence interval.

The standardized covariance parameters between SRTN results and PROMs presented a very similar picture. There were statistically significant negative covariances preoperatively between SRTN and the NCIQ BSP subdomain ($p < 0.001$), SSQ total score ($p < 0.001$), SSQ SP subdomain ($p$ = 0.049), and SSQ SQ subdomain ($p < 0.001$). There were statistically significant positive covariances preoperatively between SRTN and the NCIQ ASP ($p$ = 0.002) and SSQ SH subdomains ($p < 0.001$; N.B. a negative SRTN indicates better performance). There were no statistically significant covariances between the SRTN and PROM changes from preoperative scores to 6 month postoperative scores.

## Correlations Between Nijmegen Cochlear Implant Questionnaire Subdomains With Speech, Spatial, Qualities of Hearing Questionnaire

Between the SSQ total score and each of the NCIQ subdomains there were statistically significant, moderate-to-strong correlations within all the time points ($r$ = 0.42–0.69, $p < 0.001$)

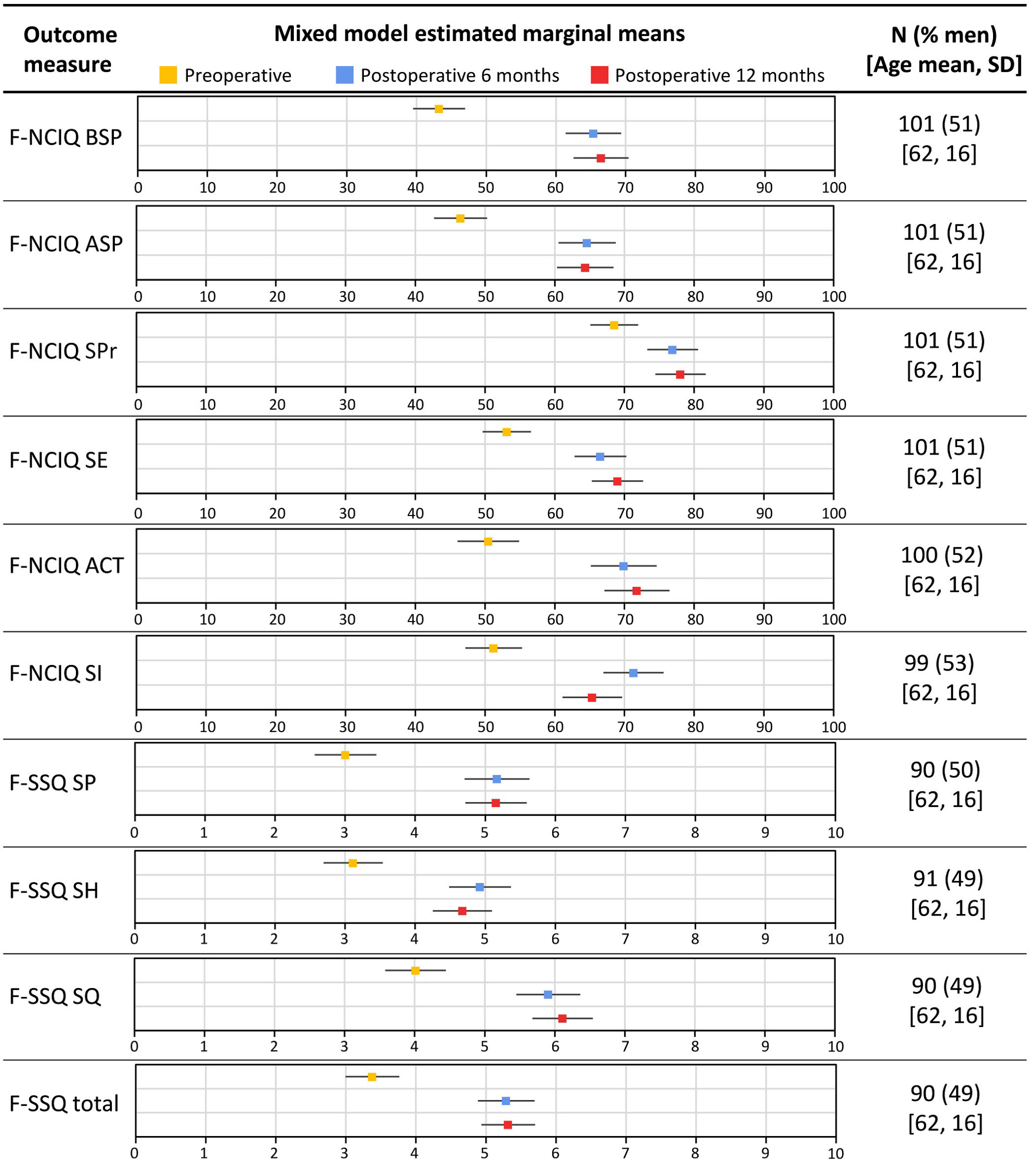


**FIGURE 2 |** Estimated marginal means from unadjusted mixed models for the Nijmegen Cochlear Implant Questionnaire (NCIQ) subdomain scores and the Speech, Spatial and Qualities of Hearing scale (SSQ) subdomain and total score among CI recipients. BSP, basic sound perception; ASP, advanced sound perception; SPr, speech production; SE, self-esteem; ACT, activity; SI, social interactions; SP, speech perception; SH, spatial hearing; SQ, sound quality; and SD, standard deviation. Error bars indicate 95% confidence intervals.

as shown in **Table 4**. The correlations were strongest between SSQ total and the NCIQ ASP subdomain and weakest between SSQ total and the NCIQ SE subdomain.

## Correlation Between Speech Perception in Noise and Patient-Reported Outcomes Measures

Statistically significant correlations between SiN measures (+10 dB SNR and SRTN) and PROMs (NCIQ and SSQ) scores within a time point of measurement were most evident preoperatively (**Table 5**). At the 6 month follow-up, the only statistically significant correlation was between the SiN scores at +10 dB SNR and the NCIQ ASP subdomain. At the 12 month follow-up, the only statistically significant correlations were between SRTN and SSQ total score, and between SiN at +10 dB SNR and SSQ total.

# DISCUSSION

When assessing the clinical outcomes of cochlear implantation, the focus normally lies almost exclusively on performance-based behavioral outcome measures such as speech perception.

**TABLE 3 |** Standardized covariances between Speech-in-noise (SiN) scores and patient reported outcomes (PROMs) (bivariate latent change score models).

| Variables | | $\sigma_{Y_{pre-op},X_{pre-op}}$ | | | $\sigma_{\Delta Y_{PO-6\ m},\Delta X_{PO-6\ m}}$ | | | $\sigma_{\Delta X_{PO-6\ m},\Delta Y_{6-12\ m}}$ | | | $\sigma_{\Delta Y_{PO-6\ m},\Delta X_{6-12\ m}}$ | | | $\sigma_{\Delta Y_{6-12\ m},\Delta X_{6-12\ m}}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***X*** | ***Y*** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** | **Est** | **SE** | ***p*** |
| SNR+10 dB | NCIQ BSP | 0.36 | 0.09 | **0.001** | 0.20 | 0.13 | 0.134 | 0.01 | 0.13 | 0.911 | 0.04 | 0.15 | 0.788 | −0.13 | 0.14 | 0.376 |
| | NCIQ ASP | −0.53 | 0.08 | **<0.001** | −0.04 | 0.16 | 0.788 | −0.09 | 0.22 | 0.670 | 0.18 | 0.19 | 0.348 | −0.16 | 0.30 | 0.592 |
| | SSQ total | 0.49 | 0.08 | **<0.001** | 0.34 | 0.12 | **0.012** | −0.08 | 0.12 | 0.511 | 0.18 | 0.15 | 0.249 | −0.01 | 0.14 | 0.923 |
| | SSQ SP | 0.36 | 0.10 | **0.001** | 0.26 | 0.14 | 0.077 | −0.20 | 0.13 | 0.144 | −0.01 | 0.17 | 0.935 | 0.19 | 0.15 | 0.225 |
| | SSQ SH | −0.52 | 0.08 | **<0.001** | −0.02 | 0.15 | 0.904 | −0.19 | 0.14 | 0.173 | −0.03 | 0.16 | 0.868 | 0.27 | 0.14 | 0.066 |
| | SSQ SQ | 0.14 | 0.10 | 0.189 | 0.43 | 0.13 | **0.004** | −0.35 | 0.13 | **0.017** | −0.17 | 0.17 | 0.305 | 0.31 | 0.14 | **0.039** |
| SRT | NCIQ BSP | −0.50 | 0.08 | **<0.001** | −0.03 | 0.15 | 0.853 | −0.13 | 0.18 | 0.469 | 0.04 | 0.19 | 0.813 | −0.06 | 0.23 | 0.800 |
| | NCIQ ASP | 0.35 | 0.10 | **0.002** | 0.22 | 0.14 | 0.139 | −0.25 | 0.13 | 0.068 | −0.07 | 0.16 | 0.665 | 0.26 | 0.14 | 0.072 |
| | SSQ total | −0.49 | 0.08 | **<0.001** | −0.04 | 0.16 | 0.824 | 0.11 | 0.21 | 0.601 | 0.01 | 0.18 | 0.975 | −0.29 | 0.21 | 0.192 |
| | SSQ SP | −0.21 | 0.10 | **0.049** | −0.17 | 0.16 | 0.304 | 0.08 | 0.22 | 0.714 | −0.04 | 0.19 | 0.839 | −0.23 | 0.23 | 0.335 |
| | SSQ SH | 0.39 | 0.10 | **<0.001** | 0.07 | 0.15 | 0.645 | 0.07 | 0.21 | 0.747 | −0.06 | 0.17 | 0.735 | −0.27 | 0.24 | 0.264 |
| | SSQ SQ | −0.60 | 0.07 | **<0.001** | −0.10 | 0.17 | 0.571 | 0.26 | 0.24 | 0.281 | 0.09 | 0.18 | 0.625 | −0.36 | 0.21 | 0.111 |

*SNR+10 dB, speech reception score; SRT, speech reception threshold; NCIQ, Nijmegen Cochlear Implant Questionnaire; SSQ, Speech, Spatial and Qualities of Hearing scale; NCIQ BSP, basic sound perception subdomain of NCIQ; NCIQ ASP, advanced sound perception subdomain of NCIQ; SSQ SP, speech perception subdomain of SSQ; SSQ SH, spatial hearing subdomain of SSQ; SSQ SQ, sound quality subdomain of SSQ; pre-op and PO, pre-operative;* $\Delta Y_{PO-6\ m}$, *change score between pre-operative and 6 months value for variable Y;* $\Delta Y_{6-12\ m}$, *change score between 6 and 12 months value for variable Y,* $\sigma_{x,y}$, *covariance between x and y. Bold type face indicates* $p < 0.05$.

However, these measures often correlate poorly with perceived benefits (Capretta and Moberly, 2016; McRackan et al., 2019; Vasil et al., 2020). As such, they do not fully address the ultimate goal of cochlear implantation, which is to improve the patients' QoL by restoring their speech perception and providing them with adequate communication skills to fully resume spcial interaction and participation. Using additional PROMs to assess QoL and other relevant subdomains of functioning would help to rectify this; however, PROMs are rarely included in the clinical routine to assess the outcomes of cochlear implantation. One reason for the omission of PROMs in the assessment is the lack of adequate instruments that fulfill modern standards, i.e., are fully psychometrically validated for the patient group. Currently, most PROMs used for CI patients were originally developed for hearing aid users, who have a very different HL profile (mild to moderate HL) and a very different rehabilitation strategy (hearing aids). Instruments developed for a different population and a different rehabilitation strategy are unlikely to adequately assess the experiences of CI users. However, one PROM that was specifically developed for CI users is the NCIQ (Hinderrick et al., 2000).

In this study, we aimed to assess the patient-reported benefits of unilateral cochlear implantation in an unselected Finnish patient cohort of patients with bilateral HL, and compare it to behaviorally assessed speech perception scores. To assess patient-reported benefit we used the NCIQ, a PROM specifically developed for CI-users, and the SSQ. An additional challenge for our study was the lack of PROMs for the use with Finnish patients. Therefore, we adapted the NCIQ and the SSQ to Finnish culture and language.

We found significant improvements in SiN measures (both at a constant SNR of +10 dB and at SRTN), perceived activity limitations and QoL after cochlear implantation. When SiN is measured with a presentation level of +10 dB SNR, the level of the speech signal is so much higher than the noise level that this test condition can be likened to speech perception in quiet. With this in mind, the improvement in SiN scores measured at a constant SNR (+10 dB SNR) fits well with previous results that have shown robust improvements in speech perception in quiet after implantation (Mudery et al., 2017; Zwolan et al., 2020). The improvement we found for SRTN demonstrates that cochlear implantation is also effective in improving patients' speech perception in complex listening environments. The SiN perception results in the present study are comparable to those reported in a previous study in a different cohort of patients at our institution (Dietz et al., 2015). The characteristics and reference values of speech audiometry differ across languages, which makes a direct comparison of the postoperative results in the international context difficult.

The patient-reported improvements show that implantation is not only effective in improving speech perception but also in alleviating activity limitations and improving QoL across a wider range of listening situations. Notably, the benefits measured with the NCIQ and SSQ in this study for a Finnish CI population are similar in magnitude to those recently reported in a systematic review by Andries et al. (2021). This similarity emerged despite possible variations across countries and languages: for example, differences in the indications for cochlear implantation between different countries and healthcare systems, or differences in perceived benefits across different languages. For tonal languages, for example, the impact of cochlear implantation may be more limited, as they rely significantly on the adequate reproduction of spectral and temporal cues, which current CI technology is unable to provide (Wei et al., 2004).

The fact that correlations between SiN results and PROMs decrease for postoperative measurements suggests that the two types of measures assess different aspects of functioning. It

**TABLE 4 |** Pearson correlation for associations between the SSQ total score and component tests of the NCIQ among all participants.

| | Unadjusted model | | | |
|---|---|---|---|---|
| | | | 95% CI | |
| | *n* | *r* | Lower | Upper |
| **Pre-operative** | | | | |
| SSQ total vs. NCIQ BSP | 100 | 0.65 | 0.52 | 0.75 |
| SSQ total vs. NCIQ ASP | 101 | 0.68 | 0.56 | 0.77 |
| SSQ total vs. NCIQ SPr | 101 | 0.48 | 0.32 | 0.62 |
| SSQ total vs. NCIQ SE | 101 | 0.47 | 0.31 | 0.61 |
| SSQ total vs. NCIQ ACT | 98 | 0.53 | 0.37 | 0.66 |
| SSQ total vs. NCIQ SI | 100 | 0.55 | 0.40 | 0.68 |
| **6 months** | | | | |
| SSQ total vs. NCIQ BSP | 78 | 0.53 | 0.35 | 0.67 |
| SSQ total vs. NCIQ ASP | 78 | 0.63 | 0.47 | 0.75 |
| SSQ total vs. NCIQ SPr | 78 | 0.45 | 0.25 | 0.61 |
| SSQ total vs. NCIQ SE | 78 | 0.42 | 0.22 | 0.59 |
| SSQ total vs. NCIQ ACT | 78 | 0.55 | 0.37 | 0.69 |
| SSQ total vs. NCIQ SI | 78 | 0.54 | 0.36 | 0.68 |
| **12 months** | | | | |
| SSQ total vs. NCIQ BSP | 83 | 0.62 | 0.46 | 0.73 |
| SSQ total vs. NCIQ ASP | 83 | 0.69 | 0.56 | 0.79 |
| SSQ total vs. NCIQ SPr | 83 | 0.57 | 0.41 | 0.70 |
| SSQ total vs. NCIQ SE | 83 | 0.55 | 0.38 | 0.68 |
| SSQ total vs. NCIQ ACT | 83 | 0.66 | 0.52 | 0.77 |
| SSQ total vs. NCIQ SI | 83 | 0.63 | 0.48 | 0.74 |

*SSQ total, Speech, Spatial and Qualities of Hearing scale; NCIQ, Nijmegen Cochlear Implant Questionnaire; NCIQ BSP, basic sound perception subdomain of NCIQ; NCIQ ASP, advanced sound perception subdomain of NCIQ; NCIQ SPr, speech production subdomain of NCIQ; NCIQ SE, self-esteem subdomain of NCIQ; NCIQ ACT, activity subdomain of NCIQ; and NCIQ SI, social interactions subdomain of NCIQ; n = sample size available for analysis, r = Pearson correlation, and CI = confidence interval. For all correlations $p < 0.001$.*

**TABLE 5 |** Unadjusted Pearson correlation for associations between SiN perception (+10 dB SNR and SRTN) results and the NCIQ and SSQ among all participants.

| | Unadjusted | | | | |
|---|---|---|---|---|---|
| | | | 95% CI | | |
| | *n* | *r* | Lower | Upper | *p*-value |
| **Pre-operative** | | | | | |
| SNR+10 dB vs. NCIQ BSP | 103 | 0.35 | 0.17 | 0.51 | **<0.001** |
| SNR+10 dB vs. NCIQ ASP | 104 | 0.46 | 0.29 | 0.60 | **<0.001** |
| SNR+10 dB vs. SSQ total | 78 | 0.41 | 0.21 | 0.58 | **<0.001** |
| SNR+10 dB vs. SSQ SP | 59 | 0.08 | −0.18 | 0.33 | 0.561 |
| SNR+10 dB vs. SSQ SH | 58 | −0.01 | −0.26 | 0.25 | 0.959 |
| SNR+10 dB vs. SSQ SQ | 58 | −0.04 | −0.29 | 0.23 | 0.794 |
| SRT vs. NCIQ BSP | 85 | −0.52 | −0.66 | −0.34 | **<0.001** |
| SRT vs. NCIQ ASP | 85 | −0.46 | −0.61 | −0.28 | **<0.001** |
| SRT vs. SSQ total | 65 | −0.46 | −0.63 | −0.24 | **<0.001** |
| SRT vs. SSQ SP | 59 | −0.03 | −0.29 | 0.22 | 0.798 |
| SRT vs. SSQ SH | 58 | −0.12 | −0.36 | 0.15 | 0.385 |
| SRT vs. SSQ SQ | 58 | −0.16 | −0.40 | 0.10 | 0.223 |
| **6 months** | | | | | |
| SNR+10 dB vs. NCIQ BSP | 53 | 0.17 | −0.10 | 0.42 | 0.216 |
| SNR+10 dB vs. NCIQ ASP | 53 | 0.28 | 0.01 | 0.51 | **0.040** |
| SNR+10 dB vs. SSQ total | 50 | 0.11 | −0.17 | 0.38 | 0.439 |
| SNR+10 dB vs. SSQ SP | 33 | −0.07 | −0.40 | 0.28 | 0.705 |
| SNR+10 dB vs. SSQ SH | 32 | −0.05 | −0.39 | 0.30 | 0.782 |
| SNR+10 dB vs. SSQ SQ | 33 | −0.06 | −0.40 | 0.29 | 0.739 |
| SRT vs. NCIQ BSP | 48 | −0.05 | −0.33 | 0.24 | 0.731 |
| SRT vs. NCIQ ASP | 48 | −0.11 | −0.39 | 0.18 | 0.440 |
| SRT vs. SSQ total | 46 | −0.09 | −0.37 | 0.21 | 0.555 |
| SRT vs. SSQ SP | 33 | −0.07 | −0.40 | 0.28 | 0.705 |
| SRT vs. SSQ SH | 32 | −0.05 | −0.39 | 0.30 | 0.782 |
| SRT vs. SSQ SQ | 33 | −0.06 | −0.40 | 0.29 | 0.739 |
| **12 months** | | | | | |
| SNR+10 dB vs. NCIQ BSP | 62 | 0.08 | −0.17 | 0.32 | 0.531 |
| SNR+10 dB vs. NCIQ ASP | 62 | 0.30 | 0.06 | 0.51 | **0.017** |
| SNR+10 dB vs. SSQ total | 61 | −0.01 | −0.26 | 0.24 | 0.943 |
| SNR+10 dB vs. SSQ SP | 32 | 0.14 | −0.22 | 0.46 | 0.456 |
| SNR+10 dB vs. SSQ SH | 32 | −0.24 | −0.54 | 0.12 | 0.190 |
| SNR+10 dB vs. SSQ SQ | 32 | 0.01 | −0.34 | 0.36 | 0.963 |
| SRT vs. NCIQ BSP | 53 | −0.10 | −0.36 | 0.17 | 0.473 |
| SRT vs. NCIQ ASP | 53 | −0.18 | −0.43 | 0.10 | 0.206 |
| SRT vs. SSQ total | 52 | −0.31 | −0.53 | −0.04 | **0.028** |
| SRT vs. SSQ SP | 32 | 0.14 | −0.22 | 0.46 | 0.456 |
| SRT vs. SSQ SH | 32 | −0.24 | −0.54 | 0.12 | 0.190 |
| SRT vs. SSQ SQ | 32 | 0.01 | −0.34 | 0.36 | 0.963 |

*SRT, speech reception threshold; SNR+10 dB, speech reception score; NCIQ, Nijmegen Cochlear Implant Questionnaire; NCIQ BSP, basic sound perception subdomain of NCIQ; NCIQ ASP, advanced sound perception subdomain of NCIQ; SSQ total, Speech Spatial and Qualities of Hearing scale total score; SSQ SP, speech perception subdomain of SSQ; SSQ SH, spatial hearing subdomain of SSQ; SSQ SQ, sound quality subdomain of SSQ; n = sample size available for analysis; r = Pearson correlation; and CI = confidence interval. Bold type face indicates $p < 0.05$.*

is possible that even though SiN tests (such as the FMST) simulate everyday listening situations more accurately than tests performed in quiet, they still fail to capture the manifold hearing environments of everyday life. Future studies will need to examine to what extent speech perception tests in simulated realistic acoustic environments are better able to capture everyday listening and whether this leads to higher correlations with PROMs postoperatively.

The present study demonstrates that the main improvements in speech perception and PROMs took place within the first 6 months postoperatively. This is in line with previous data, which have shown that the main improvements in outcome measures can be seen within the first 6 months after fitting of the sound processor (Lenarz et al., 2012; Zhang et al., 2015; Häußler et al., 2019). Although we observed an additional, statistically significant, improvement in the speech perception tests between 6 and 12 months, the magnitude of improvement was small and within the limits of the test and statistical sensitivity. These results highlight the importance of adequate care and sound processor fitting during the early months of rehabilitation.

There was a slight, but statistically significant, decline of scores in the SI subdomain of the NCIQ between 6 and 12 months postoperative measurements. Based on our clinical experience, we speculate that the decline in the SI subdomain between 6 and 12 months is due to the fact that increased SI after cochlear implantation (observed after 6 months of use) expose patients more often to complex listening situation, which then discloses the limitations of hearing with the CI resulting in a decrease of SI. In addition, patients often stop using their contralateral hearing aids after adaptation to the CI and also often inquire about the possibility of getting a second CI in their contralateral ear.

The regression coefficients showed that lower preoperative PROMs and speech perception results indicated greater changes 6 month postoperatively. This is expected, as CIs can reliably restore the patient's functional hearing to an adequate performance level so that they can have a relaxed conversation in quiet surroundings. Therefore, patients with the most profound HL [i.e., patients who are not able to have any (relaxed) conversation] are more likely to perceive improvements in their hearing as more significant than patients with less severe loss. Patients with less severe HL usually experience problems in complex sound environments, in which the benefits of CIs can be more variable.

Looking at the correlation measures, we found a strong correlation between the NCIQ subdomains of BSP and ASP and the SSQ total score, and a moderate correlation between the remaining subdomains of the NCIQ and the SSQ total score. The SSQ assesses mainly activity limitations associated with hearing whereas the NCIQ is thought to focus more on participation restrictions associated with socioemotional factors. This suggests an existing interconnection between HL and socioemotional issues.

Only a few studies have investigated the relationships between PROMs and SiN measures (fixed SRN and SRTN), and found mainly weak correlations (Hirschfelder et al., 2008; Vasil et al., 2020; West et al., 2020). In the present study, we found statistically significant correlations for some subdomains exclusively at the preoperative assessment (see **Table 5**). Importantly, we found no clinically significant associations or correlations between SiN measures and PROMs for any of the follow-up evaluations, indicating that the cochlear implantation benefits were not fully captured by the SiN test. The baseline-adjusted covariances support this, with performance-based measures showing an association with the SSQ (and its subdomains) and NCIQ (BSP, ASP) only before the intervention. Although, we also found statistically significant covariances between SiN scores (at +10 dB SNR) and some PROMs subdomains postoperatively, these have to be interpreted with caution, as no corresponding significant covariances were observed between these subdomains and the SRTN, which is the more precise and more reliable measure of performance. Taken together, these results confirm that there are other hearing-related CI benefits which are not captured by auditory performance measured with current SiN tests.

## LIMITATIONS AND STRENGTHS OF THE STUDY

Several limitations associated with this study need attention. The original questions of the NCIQ and the SSQ were created by expert opinion, and the psychometric qualities of each question as well as of the questionnaires as a whole are still not fully understood. Further studies are required to fully evaluate these psychometric properties, both in terms of classical test theory (e.g., test-retest reliability, minimal relevant change) and also wider assessment along the COSMIN guidelines (Mokkink et al., 2010), and content and construct validity. Therefore, caution should be applied when interpreting these PROMs. However, as the magnitude of improvement after cochlear implantation measured in this study was substantial, it is beyond any reasonable doubt that these benefits exist.

The strengths of this study are the relatively large cohort of patients and its prospective, longitudinal design, which gives a less biased estimate of population measures than the more commonly used retrospective and cross-sectional designs. In addition, we not only report the change scores in patient-reported and behavioral outcome measures but have also investigated their association, as well as the relationship between baseline-adjusted changes at various time points.

## CONCLUSION

Cochlear implantation significantly improves speech perception, QoL (NCIQ), and self-assessed hearing capabilities (SSQ) in a cohort of unselected Finnish CI recipients. The main improvements were observed within the first 6 months after sound processor activation, indicating the importance of adequate early sound processor fitting. The results highlight the fact that cochlear implantation conveys benefits which go beyond those captured by performance-based outcome measures.

## 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 Research Ethics Committee of the Northern Savo Hospital District (1327/2018). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AD, TW, MI-M, and PL contributed to conception and design of the study. PL and MI-M organized the database. TT performed the statistical analysis and designed the figures and tables. AD and AH wrote the first draft of the manuscript. PL, TW, and PM wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This study was funded by the Academy of Finland (Grant No. 333525 to AD) and North Savo Regional Fund (Grant No. 65202054 to AD). AH is supported by the NIHR Manchester Biomedical Research Centre (NIHR BRC-20007).

## ACKNOWLEDGMENTS

We want to thank all our participating patients and our research assistant Jaana Hyytiäinen for managing the research data. We also want to thank Professor Emmanuel Mylanus from the Radboud University Medical Centre for his permission to translate and apply the Nijmegen Cochlear Implant Questionnaire to the benefit of Finnish CI patients. We wish to express our gratitude to Dr. Sarah Knight for comments on an earlier version of the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.786939/full#supplementary-material

## REFERENCES

Andries, E., Gilles, A., Topsakal, V., Vanderveken, O. M., Van de Heyning, P., Van Rompaey, V., et al. (2021). Review of quality of life assessments after cochlear implantation in older adults. *Audiol. Neurotol.* 26, 61–75. doi: 10.1159/000508433

Boisvert, I., Reis, M., Au, A., Cowan, R., and Dowell, R. C. (2020). Cochlear implantation outcomes in adults: a scoping review. *PLoS One* 15:e0232421. doi: 10.1371/journal.pone.0232421

Capretta, N. R., and Moberly, A. C. (2016). Does quality of life depend on speech recognition performance for adult cochlear implant users? *Laryngoscope* 126, 699–706. doi: 10.1002/lary.25525

Coman, E. N., Picho, K., McArdle, J. J., Villagra, V., Dierker, L., and Iordache, E. (2013). The paired t-test as a simple latent change score model. *Front. Psychol.* 4:738. doi: 10.3389/fpsyg.2013.00738

Dietz, A., Buschermöhle, M., Aarnisalo, A. A., Vanhanen, A., Hyyrynen, T., Aaltonen, O., et al. (2014). The development and evaluation of the finnish matrix sentence test for speech intelligibility assessment. *Acta Otolaryngol.* 134, 728–737. doi: 10.3109/00016489.2014.898185

Dietz, A., Buschermöhle, M., Sivonen, V., Willberg, T., Aarnisalo, A. A., Lenarz, T., et al. (2015). Characteristics and international comparability of the Finnish matrix sentence test in cochlear implant recipients. *Int. J. Audiol.* 54(Suppl. 2), 80–87. doi: 10.3109/14992027.2015.1070309

Dingemanse, G., and Goedegebure, A. (2019). Efficient adaptive speech reception threshold measurements using stochastic approximation algorithms. *Trends Hear.* 23:2331216520919199. doi: 10.1177/2331216520919199

Gatehouse, S., and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int. J. Audiol.* 43, 85–99. doi: 10.1080/14992020400050014

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (2008). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. *Audiol. Neurootol.* 13, 193–205. doi: 10.1159/000113510

Greene, W. H. (2003). *Econometric Analysis*, 5th Edn. Upper Saddle River, NJ: Prentice Hall.

Häußler, S. M., Knopke, S., Wiltner, P., Ketterer, M., Gräbel, S., and Olze, H. (2019). Long-term benefit of Unilateral cochlear implantation on quality of life and speech perception in bilaterally deafened patients. *Otol. Neurotol.* 40, e430–e440. doi: 10.1097/MAO.0000000000002008

Hinderink, J. B., Krabbe, P. F., and Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen cochlear implant questionnaire. *Otolaryngol Head Neck Surg* 123, 756–765.

Hirschfelder, A., Gräbel, S., and Olze, H. (2008). The impact of cochlear implantation on quality of life: the role of audiologic performance and variables. *Otolaryngol. Head Neck Surg.* 138, 357–362. doi: 10.1016/j.otohns.2007.10.019

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear.* 34, 342–360. doi: 10.1097/AUD.0b013e3182741aa7

Kramer, S. E., Kapteyn, T. S., Festen, J. M., and Kramer, S. E. (1998). The self-reported handicapping effect of hearing disabilities. *Audiology* 37, 302–312. doi: 10.3109/00206099809072984

Lenarz, M., Sönmez, H., Joseph, G., Büchner, A., and Lenarz, T. (2012). Long-term performance of cochlear implants in postlingually deafened adults. *Otolaryngol. Head Neck Surg.* 147, 112–118. doi: 10.1177/0194599812438041

Length, R. (2020). *emmeans: Estimated Marginal Means, Aka Least-Squares Means. R Package Version 1.5.1.* Available online at: https://CRAN.R-project.org/package=emmeans (accessed June 23, 2021).

Lin, F. R., Metter, E. J., O'Brien, R. J., Resnick, S. M., Zonderman, A. B., and Ferrucci, L. (2011). Hearing loss and incident dementia. *Arch. Neurol.* 68, 214–220. doi: 10.1001/archneurol.2010.362

Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., and Lawlor, B. A. (2018). Association of age-related hearing loss with cognitive function, cognitive impairment, and dementia: a systematic review and meta-analysis. *JAMA Otolaryngol. Head Neck Surg.* 144, 115–126. doi: 10.1001/jamaoto.2017.2513

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annu. Rev. Psychol.* 60, 577–605. doi: 10.1146/annurev.psych.60.110707.163612

McRackan, T. R., Hand, B. N., Velozo, C. A., and Dubno, J. R. (2019). Association of demographic and hearing-related factors with cochlear implant-related quality of life. *JAMA Otolaryngol. Head Neck Surg.* 145, 422–430. doi: 10.1001/jamaoto.2019.0055

Mertens, G., Andries, E., Claes, A. J., Topsakal, V., Van de Heyning, P., Van Rompaey, V., et al. (2020). Cognitive improvement after cochlear implantation in older adults with severe or profound hearing impairment: a prospective, longitudinal, controlled, multicenter Study. *Ear Hear.* 42, 606–614. doi: 10.1097/AUD.0000000000000962

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J. Clin. Epidemiol.* 63, 737–745. doi: 10.1016/j.jclinepi.2010.02.006

Mudery, J., Francis, R., McCary, H., and Jacob, A. (2017). Older individuals meeting medicare cochlear iplant candidacy criteria in noise but not in quiet: are these patients improved by surgery? *Otol. Neurotol.* 38, 187–191.

Muthén, L. K., and Muthén, B. O. (1998–2015). *Mplus: Version 7.4*. Los Angeles, CA: Muthén & Muthén.

Ottaviani, F., Iacona, E., Sykopetrites, V., Schindler, A., and Mozzanica, F. (2016). Cross-cultural adaptation and validation of the Nijmegen cochlear implant questionnaire into Italian. *Eur. Arch. Otorhinolaryngol.* 273, 2001–2007. doi: 10.1007/s00405-015-3765-8

Pinheiro, J., Bates, D., DebRoy, S., and Deepayan, S. (2020). *nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-148*. Available online at: https://CRAN.R-project.org/package=nlme (accessed June 23, 2021).

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Sanchez-Cuadrado, I., Gavilan, J., Perez-Mora, R., Muñoz, E., and Lassaletta, L. (2015). Reliability and validity of the Nijmegen cochlear implant questionnaire in Spanish. *Eur. Arch. Otorhinolaryngol.* 272, 1621–1625. doi: 10.1007/s00405-014-2983-9

Santos, N. P. D., Couto, M. I. V., and Martinho-Carvalho, A. C. (2017). Nijmegen Cochlear Implant Questionnaire (NCIQ): translation, cultural adaptation, and application in adults with cochlear implants. *Codas* 29:e20170007. doi: 10.1590/2317-1782/20172017007

Vasil, K., Lewis, J., Tamati, T., Ray, C., and Moberly, A. (2020). How does quality of life relate to auditory abilities? A subitem analysis of the Nijmegen cochlear implant questionnaire. *J. Am. Acad. Audiol.* 31, 292–301.

Vermiglio, A. J., Soli, S. D., Freed, D. J., and Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *J. Am. Acad. Audiol.* 23, 779–788. doi: 10.3766/jaaa.23.10.4

Wei, C., Cao, K., and Zeng, F. (2004). Mandarin tone recognition in cochlear-implant subjects. *Hear. Res.* 197, 87–95. doi: 10.1016/j.heares.2004.06.002

West, N. C., Kressner, A. A., Baungaard, L. H., Sandvej, M. G., Bille, M., and Cayé-Thomasen, P. (2020). Nordic results of cochlear implantation in adults: speech perception and patient reported outcomes. *Acta Otolaryngol.* 140, 939–947. doi: 10.1080/00016489.2020.1816656

World Health Organization [WHO] (2021). *World Report on Hearing*. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.

World Health Organsation [WHO] (2001). *International Classification of Functioning, Disability and Health: ICF*. Geneva: World Health Organization.

Zhang, J., Tyler, R., Ji, H., Dunn, C., Wang, N., Hansen, M., et al. (2015). Speech, spatial and qualities of hearing scale (SSQ) and spatial hearing questionnaire (SHQ) changes over time in adults with simultaneous cochlear implants. *Am. J. Audiol.* 24, 384–397. doi: 10.1044/2015_AJA-14-0074

Zwolan, T., Kallogjeri, D., Firszt, J., and Buchman, C. (2020). Assessment of cochlear implants for adult medicare beneficiaries aged 65 years or older who meet expanded indications of open-set sentence Recognition: a multicenter nonrandomized clinical trial. *JAMA Otoloaryngol. Head Neck Surg.* 146, 933–941.

**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.

# Advantages of publishing in Frontiers

**OPEN ACCESS**
Articles are free to read for greatest visibility and readership

**FAST PUBLICATION**
Around 90 days from submission to decision

**HIGH QUALITY PEER-REVIEW**
Rigorous, collaborative, and constructive peer-review

**TRANSPARENT PEER-REVIEW**
Editors and reviewers acknowledged by name on published articles

**Frontiers**
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

**Visit us:** www.frontiersin.org
**Contact us:** frontiersin.org/about/contact

**REPRODUCIBILITY OF RESEARCH**
Support open data and methods to enhance research reproducibility

**DIGITAL PUBLISHING**
Articles designed for optimal readership across devices

**FOLLOW US**
@frontiersin

**IMPACT METRICS**
Advanced article metrics track visibility across digital media

**EXTENSIVE PROMOTION**
Marketing and promotion of impactful research

**LOOP RESEARCH NETWORK**
Our network increases your article's readership