Adaptive hearing aid benefit in children with mild/moderate hearing loss: A registered, double-blind, randomized clinical trial

Abbreviated title: Benefit of adaptive hearing aids in children

Hannah J. Stewart[1, 2], Erin K. Cash[1], Joseph Pinkl[1, 3], Cecilia Nakeva von Mentzer[4, 5], Li Lin[1,6], Lisa L. Hunter[1,6,7,8] and David R. Moore[1, 6, 8, 9], the CCHMC Division of Audiology[7]

1. Communication Sciences Research Center, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
2. Division of Psychology and Language Sciences, University College London, London, UK
3. Department of Research and Development, Gateway Biotechnology Inc., Rootstown, Ohio, USA
4. School of Health Sciences, Örebro University, Örebro, Sweden
5. Department of Neuroscience, Unit for SLP, Uppsala University, 751 23 Uppsala, Sweden
6. Department of Research in Patient Services, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
7. Division of Audiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA. CCHMC Division of Audiology collaborators are listed in the author contribution statement.
8. Department of Otolaryngology, College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA
9. Manchester Centre for Audiology and Deafness, University of Manchester, Manchester, M13 9PL, UK

Correspondence: Hannah J Stewart
Speech, Hearing and Phonetic Sciences
University College London
London WC1N 1PF
h.stewart@ucl.ac.uk

Conflicts of interest and Source of Funding: None declared. This research study was supported by a grant from the Oticon Foundation that was awarded to the Communication

Sciences Research Center at Cincinnati Children's Hospital Medical Center. David R. Moore is partly supported by the NIHR Manchester Biomedical Research Centre.

Preregistration: https://clinicaltrials.gov/ct2/show/NCT03771287

Data sharing: Deidentified behavioural and audiological data and dictionaries can be downloaded indefinitely from https://osf.io/d72n8/. We certify that this repository includes sufficient information to reproduce the reported results.

Material sharing: Tasks, stimuli, data extraction and analysis scripts can be downloaded indefinitely from https://github.com/stewarthannahj/OtiS.git. We certify that this repository includes sufficient information to reproduce the methodology.

Acknowledgments: This research was generously supported by grant 17-2441 from The Oticon Foundation. The authors would like to thank the large team involved in completing this study: Jasmin Martinez and Sarah Ferguson for behavioural data collection; Audrey Perdew for database management; Andrew Wagner for programming support; Boys Town National Research Hospital for supplying the word repetition in noise test stimuli; Elena Plante for providing the novel grammar learning task and stimuli; and Elaine Ng and Thomas Behrens from Oticon for help with funding, providing hearing aids, and other general support. David Moore was supported in part by the NIHR Manchester Biomedical Research Centre.

Author contributions: HJS designed the protocol, registered the study, performed experiments, analyzed data and wrote the manuscript. EKC and JP performed experiments, analyzed data and wrote the manuscript. JP also registered the study. LL designed and executed the analysis and commented on the manuscript. CNvM designed the protocol and commented on the manuscript. CCHMC Division of Audiology designed the protocol and performed experiments. LLH designed the protocol and commented on the manuscript. DRM designed the protocol and wrote the manuscript. All authors discussed the results and implications. The CCHMC study audiologists were Tommy Evans, Tim Nejman, Jeanie Hamilton, Anne-Marie Wollet and Erin Stewart.

Abbreviations

BKB-SIN - Bamford-Kowal-Bench Sentence in noise test

BTE - behind-the-ear

CCHMC - Cincinnati Children's Hospital Medical Center

CV – consonant-vowel

DSL - Desired Sensation Level

GCBI - Glasgow Children's Benefit Inventory

HAB – hearing aid benefit

MRI - magnetic resonance imaging

OMNI – Omnidirectional, control algorithm

OPN – Oticon OPN hearing aid

OSN – OpenSound Navigator algorithm

PROMIS - Patient-Reported Outcomes Measurement Information System

PTA – audiometric pure tone average

REDCap - Research Electronic Data Capture

RV1, RV2 – research visit 1, 2

SNR – signal/noise ratio

SPL - sound pressure level

SRT - speech reception threshold

SSN – steady state noise

SSQ - Speech, Spatial and Qualities of Hearing Scale

TTM – two-talker masker

ABSTRACT

Objectives: We completed a registered double-blind randomized control trial to compare acclimatization to two hearing aid fitting algorithms by experienced pediatric hearing aid users with mild to moderate hearing loss. We hypothesized that extended use (up to 13 months) of an adaptive algorithm with integrated directionality and noise reduction, OpenSound Navigator (OSN), would result in improved performance on auditory, cognitive, academic, and caregiver- or self-report measures compared to a control, omnidirectional algorithm (OMNI).

Design: Forty children aged 6 - 13 years with mild to moderate/severe symmetric sensorineural hearing loss completed this study. They were all experienced hearing aid users and were recruited through the Cincinnati Children's Hospital Medical Center Division of Audiology. The children were divided into 20 pairs based on similarity of age (within one year) and hearing loss (level and configuration). Individuals from each pair were randomly assigned to either an OSN (experimental) or OMNI (control) fitting algorithm group. Each child completed an audiology evaluation, hearing aid fitting using physically identical Oticon OPN hearing aids, follow-up audiological appointment, and two research visits up to 13 months apart. Research visit outcome measures covered speech perception (in quiet and in noise), novel grammar and word learning, cognition, academic ability, and caregiver report of listening behaviors. Analysis of outcome differences between visits, groups, ages, conditions and their interactions used linear mixed models. Between 22 - 39 children provided useable data for each task.

Results: Children using the experimental (OSN) algorithm did not show any significant performance differences on the outcome measures compared to those using the control (OMNI) algorithm. Overall performance of all children in the study increased across the duration of the trial on word repetition in noise, sentence repetition in quiet, and caregivers' assessment of hearing ability. There was a significant negative relationship between age at first hearing aid use, final Reading and Mathematical ability, and caregiver rated speech hearing. A significant positive relationship was found between daily hearing aid use and study-long change in performance on the Flanker test of inhibitory control and attention. Logged daily use of hearing aids related to caregiver rated spatial hearing. All results controlled for age at testing/ evaluation and false discovery rate.

Conclusions: Use of the experimental (OSN) algorithm neither enhanced nor reduced performance on auditory, cognitive, academic or caregiver report measures compared to the control (OMNI) algorithm. However, prolonged hearing aid use led to benefits in hearing, academic skills, attention, and caregiver evaluation.

INTRODUCTION

About 15% of children aged 6-19 years in the US have pure tone average (PTA) hearing thresholds > 15 dB HL (Niskar et al., 1998; Su & Chan, 2017), the level at which hearing loss in children is considered functionally significant. However, only a small minority of these children have bilateral PTA ≥ 26 dB HL, the threshold used in Butcher and colleagues' (2019) meta-analysis to calculate a 0.11% prevalence of bilateral hearing loss at newborn screening and, hence, to be routinely referred for hearing aid assessment. This seems a missed opportunity, since hearing aid use provides benefit for even the most mild hearing loss (McCreery et al., 2020).

Both children (Tomblin et al., 2014) and adults (Gatehouse, 1992) have also been found to obtain additional benefit from hearing aid use across time following initial fitting (acclimatization). While such acclimatization is controversial in adults, with several studies finding no greater, long-term benefit (Bender et al., 1993; Dawes et al., 2014; Humes et al., 2002; Humes & Wilson, 2003; Taylor, 1993), the weight of evidence in children favors increased benefit (Moeller & Tomblin, 2015). This issue is clinically significant because some of the same evidence shows the importance of earliest possible intervention for hearing loss (Moeller & Tomblin, 2015). The greater improvement in language and academic skills obtained by early hearing aid use presumably reflects improved speech perception through a process of auditory perceptual learning (Walker et al., 2020).

Intervention for hearing impairment is especially important in children who are learning language through everyday activities and in educational settings (McCreery et al., 2019; Walker et al., 2015a, 2020). Hearing aids are ideally used extensively during waking hours (Muñoz & Hill, 2015), and provide enhanced speech intelligibility (Ertmer, 2011), enabling long duration exposure to socially- and educationally-relevant stimulation. However, children using hearing aids exhibit greater listening effort, evidenced by slower reaction times to verbal instructions, especially at more adverse signal to noise ratios, which may negatively impact learning and academic achievement in the classroom (McGarrigle et al., 2019). Both duration and salience of stimulation are considered to be key elements of perceptual learning (Wildt et al., 2019). Children also learn more easily than adults (Lucas et al., 2014), possibly due to enhanced brain plasticity and/or greater motivation with the task regimen.

It seems plausible that improving the features of hearing aids that are most useful for children, such as directional microphones (Chung & Zeng, 2009; Gravel et al., 1999), may further improve acclimatization. For example, children may be more dependent than adults on spatial localization of sound, due to reduced selective attention (Wightman et al., 2006), and the need to learn in noisy classrooms (Picard & Bradley, 2001) and home environments (Erickson & Newman, 2017). In response to these learning and environmental considerations, OpenSound Navigator (OSN) was developed to provide adaptive, integrated directionality and noise reduction. Briefly, OSN utilizes a dual microphone array to analyze the acoustic environment. Subsequent directionality and noise reduction processing attenuate noise sources and diffuse noise respectively. The system updates 500 times/second across 16 independent frequency channels, enabling a rapid, spatial-based adaptive system with high selectivity to speech sounds (Le Goff et al., 2016).

In a previous, exploratory study (Pinkl et al., 2021), we showed that 6-12 year old children fitted with OSN-enabled hearing aids for two months received enhanced caregiver assessments of their speech and sound perception, spatial sound awareness and ability to participate in conversations. However, their measured speech perception, language, cognitive and academic skills were unaffected by using the hearing aids. As there was no control group in that study, it is possible that the enhanced assessments were affected by reporter bias. In addition, the acclimatization period was short and the sample size limited. Finally, it is possible that the range of measures used, although wide, missed other important aspects of listening and learning.

Browning et al. (2019) assessed speech in noise perception in children using OSN or omnidirectional (OMNI) algorithms after no acclimatization. They found that speech in steady state noise (SSN) thresholds were better for the OSN than for the OMNI algorithm, regardless of whether the target talker was facing the participant or not. No algorithm difference was found for a speech in two-talker speech condition. However, they used a within-subject design; each child was tested with both algorithm options rather than using a separate OMNI control group.

To address these gaps in evidence, we designed a new, expanded, registered double-blind randomized control trial that compared extended (up to 13 months) use of the OSN algorithm with a control, OMNI algorithm, programmed into the physically identical Oticon OPN hearing aids. We added outcome measures for speech in noise (BKB-SiN), statistical grammar learning and participant fatigue to our previous behavioural testing battery of speech perception, cognition, academic and caregiver report outcomes. We hypothesized that extended use of the OSN algorithm would result in improved performance on the range of skills previously examined and on other skills (statistical learning, and self-report effort measures) newly introduced in this study.

MATERIALS AND METHODS

< Table 1 about here >

Participants

Recruitment: Eligible participants were initially identified through a CCHMC Division of Audiology medical record search. Study staff pre-screened each eligible child and mailed paper recruitment materials. After 2 weeks, families were contacted by phone and/or email regarding their interest in the study. If interested, the parent/guardian of the potential participant completed a series of online questionnaires. If all eligibility criteria (see below) were met, study staff consented the parent/guardian over the phone using the eConsent Framework in the Research Electronic Data Capture (REDCap; Harris et al., 2009) platform, also used for collecting and managing all study data. REDCap is a secure, web-based application providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources. Participants were assented if they were 11 years of age or older.

Inclusion: Forty-two children were enrolled, but two were withdrawn due to noncompliance or discomfort. The 40 remaining were experienced pediatric HA users, ages 6,7 to 13,2 (years,months; mean = 9,9). Age of hearing loss diagnosis ranged from birth to 8 years (mean = 3,8; Table 1), while age at receipt of first hearing aids ranged from 3 months to 9 years (mean = 4,1). Inclusion criteria were a) native English speakers, b) no history of ear surgeries, c) symmetrical sensorineural hearing loss in the mild to moderately-severe range from 500-4000 Hz (see “Audiological evaluation” below), d) no history of developmental delays, e) no medical diagnoses of neurologic/psychiatric disorders or attention deficits, f) history of consistent binaural HA use, and g) no prior experience with Oticon’s OSN algorithm. A further criterion was magnetic resonance imaging (MRI) compatibility; MRI results will be reported elsewhere. Three participants who were not MRI-compatible were enrolled to provide matches for other participants (one of whom was withdrawn).

Ethics and preregistration: Ethical approval for all clinical tests, services and data collection procedures was obtained from the CCHMC Institutional Review Board. The study was preregistered with clinicaltrials.gov (NCT03771287).

Blinding and Randomization: Participants were recruited in age (within one year) and hearing loss (level and configuration) matched pairs, irrespective of sex. The individuals in each pair were randomly assigned to a group; one child used the omnidirectional setting (control), and the other used the OpenSound Navigator (experimental). This randomization was assigned via a coin flip by a research coordinator who was not involved in recruitment, testing, or analysis for this study. Each child's enabled hearing aid algorithm was saved in REDCap and was accessible only to this external research coordinator and to the study audiologists. When research study staff scheduled each participant's hearing aid fitting appointment, the external research coordinator notified the fitting audiologist. As part of standard care, the hearing aid algorithm was enabled (OSN or omnidirectional) and added in the participants' medical record notes. Participants, their caregivers, and study research staff were unaware of each participant's enabled algorithm until both participants in any pair had completed the study.

Incentives: Participants received a payment of $30 for completing the first research and MRI visits and $60 for completing the second research and MRI visits. At the end of the study they received a $40 bonus if they wore their hearing aids for at least 8 hours a day (as per their logging data). The children were allowed to keep their study hearing aids and accessories at the end of the study.

< Figure 1 about here >

Procedures

The study schedule of participant visits is shown in Fig. 1. Many of the procedures have been described in detail previously (Pinkl et al., 2021) and will be presented only briefly here. All audiological visits followed a checklist of procedures to ensure that all study audiologists followed the same evaluation and fit protocols.

Audiologic evaluation: A clinical evaluation which included pure-tone thresholds, obtained using the modified Hughson-Westlake procedure (Hughson and Westlake,1944), word recognition and immittance testing, was performed on all participants by licensed audiologists at the Division of Audiology, CCHMC. Air and bone conducted pure-tone signals were presented through EARTone 3A insert earphones and a MelMedtronics B71 adult bone oscillator, respectively, using a GSI 61 Clinical Audiometer (Grason-Stadler, Eden Prairie, MN). Pure-tone thresholds were obtained from 250-8000 Hz at half-octave increments via air-conduction (Fig. 2) and 500-4000 Hz at octave increments via bone-conduction.

< Figure 2 about here >

Monaural and binaural speech testing was completed via air-conduction and included speech reception thresholds (SRT, left, right ears separately) and monosyllabic word recognition. Speech recognition testing was performed using recorded male voices with open-set lists (NU-6, W-22 or PKB; Auditec, Inc.) presented in quiet at 40 dB SL (sensation level; based on SRT) or the participant's most-comfortable level. Standard 226-Hz tympanometry (Hunter & Blankenship, 2017) was completed using a GSI Tympstar Middle Ear Analyzer or Titan/IMP440. Previous audiometric results were used for participants who received an audiologic evaluation at CCHMC within six months of study enrollment. Earmold impressions were also taken at this visit; participants were permitted to continue use of prior earmolds if the audiologists thought they provided adequate fit and comfort.

Hearing aid fitting: Within one month of the audiologic evaluation, bilateral Oticon OPN 1 PP behind-the-ear (BTE) HAs with standard tubing, custom ear molds and a compatible wireless microphone and Bluetooth streamer (ConnectClip), were dispensed for each participant. Ear molds were either skeleton or canal style and made from silicone material with venting size selected by the fitting audiologist. Manual controls on the hearing aids were disabled. The OSN algorithm was enabled, where used (see Randomization, above), with all features set to manufacturer default settings. This included noise reduction with the levels set to 0 dB for simple environments and -7 dB for complex listening environments. The directionality mode was set to “Open Automatic” with the transition handles set to medium. For the control, OMNI group, the directionality mode was set to “Pinna Omni” with the noise reduction disabled.

HA verifications were performed using Verifit 2.0 (Audioscan, Dorchester, ON, Canada). To measure the Speech Intelligibility Index (SII) real-ear speech mapping was completed using a standard recorded passage presented at 55, 65 and 75 dB (low, average and loud, respectively) sound pressure level (SPL). Prescriptive targets were set using Desired Sensation Level (DSL) v5.0 based on each participant’s audiometric thresholds and individual real ear-to-coupler differences. Real-ear probe microphone measurements were used to ensure that HA gain met prescriptive targets. Fine-tuning gain adjustments were made so that HA output was within 5 dB at 250, 500, 1000 and 2000 Hz and 8 dB at 3000 and 4000 Hz of the prescriptive targets (Bagatto et al., 2016; Bagatto et al., 2011). Participants and accompanying caregivers were instructed on HA use and care by the fitting audiologist. They were asked to ensure hearing aids were used whenever possible, with a minimum of 8 hours per day. Participants who utilized hearing aid compatible assistive listening devices and Bluetooth streaming for classroom learning were instructed to continue using those same devices with adapters and audio boots provided by Oticon when necessary. Usage data on these listening devices was not collected.

Research visit 1 (RV1): Participants completed the first round of experimental behavioral testing, as well as an initial MRI, 2-7 days after hearing aid fitting. The test battery included 6 different assessments: a free-field word repetition task, sentence repetition tasks in quiet and in noise, a novel word learning task, a set of standardized cognition tasks, and an academic achievement task. Parents of participants were asked to complete caregiver report questionnaires that assessed their child's speech and hearing abilities.

Audiology follow-up: One month after hearing aid fitting, the fitting audiologist checked the hearing aid fit and advised on the child's average daily usage. Validation measures included aided free-field narrow-band thresholds (center frequencies of 500, 1000, 2000 and 4000 Hz) and speech recognition in quiet testing (NU-6, W-22 or PBK lists, consistent with the initial audiologic test). Word recognition was completed using recorded male voices presented at 35 and 55 dB HL presented at 0° azimuth.

Research visit 2 (RV2): Participants returned 6-13 months after their hearing aid fitting. The initial battery of testing was repeated, with different versions to reduce learning effects, in addition to a novel statistical grammar learning task. Caregivers completed the same questionnaires, this time in reference to their child's speech and language abilities with their study hearing aids. A novel pediatric fatigue questionnaire, the PROMIS, was administered both as a self-report by the participant and, independently, as a caregiver report.

Optional Algorithm Change: Upon completion of the study, participants were unblinded and had the option of scheduling a final study follow-up appointment with their fitting audiologist for programming changes. At this appointment, they could change the hearing aid programming following discussion with the audiologist. Most participants chose to continue use of the OSN program, or to enable OSN if they had been in the OMNI group. Each participant's daily hearing aid usage (Fig. 2B) was reviewed and discussed with caregivers.

Unblinding: Research staff involved in testing and/or analysis were unblinded after all participants had completed the study.

Behavioral testing

All tests were performed in a double-walled sound-treated booth (IAC) with an observation window. Participants wore their study hearing aids in the experimental setting. Unless otherwise stated, auditory test stimuli were generated in Matlab and presented through an M-Track Eight interface (M-Audio, Cumberland, RI), Servo 120a power amplifiers (Samson Technologies, Hicksville, NY), and JBL Control 1Xtreme 4" loudspeakers in manufacturer's enclosures (Harman International Industries, Stamford, CT), atop microphone stands at a height of 96 cm. The tester was outside the booth and communicated with the participant visually and via intercom.

Order of testing was randomized in a Latin square design across participants. All tests were completed twice: within one week of hearing aid fitting (RV1; Fig. 1) and 6-13 months post-fitting (RV2). Test sessions of sentence repetition and novel word learning were recorded using Audacity (v. 2.3.2). Recordings were re-scored by a second researcher. If there was agreement between the two researchers, the initial score was used. Disagreements were reconciled by a third researcher.

Word repetition in noise: This test (Figure 3A) was adapted from custom software supplied by Boys Town National Research Hospital (Browning et al., 2019) and measured frontally-presented monosyllabic target word thresholds in rearward masking noise. Target words were from English reading lists suitable for 5 and 6 year olds (Corbin et al., 2016), normalized to a fixed intensity of 65 dBrms SPL and spoken by adult males in an American English dialect. Two-talker masker (TTM) speech streams were created from separate passages from *Jack and the Beanstalk.* Speech-shaped noise (SSN) was based on the spectral envelope of the TTM speech streams. Maskers were set at an initial level of 55 dBrms SPL. Three test conditions varied either the target direction or the noise type (Fig. 3A).

Participants used a chin rest for head and hearing aid microphone positioning to remain in the center of a speaker ring, diameter 2m. They were instructed to remain still, look straight ahead, listen closely for the target word, and repeat it back. If they were unsure what word they heard, they were instructed to guess. Correct/incorrect identification of Target increased/decreased masker level by 4 dB on next trial. After the second reversal, step size was reduced to 2 dB. Speech reception threshold (SRT) was the mean of the last 6 reversals, following 8 total reversals.

Sentence repetition in quiet: Speech perception, verbal working memory and grammatical knowledge were assessed by listening to and repeating recorded sentences (after Moll et al., 2015). The recorded sentences were provided by an online research platform from Uppsala University (*Audio Research*, 2018; Ranjbar & Nakeva von Mentzer, 2020) and presented by a native female speaker at 50 dBrms SPL in Standard American English. Two loudspeakers were situated either side of a computer monitor (±5° azimuth) on a table 84 cm in front of the seated participant, who was instructed to listen closely to the stimuli, and repeat back the sentence. If the participant forgot or did not hear the sentence, they were encouraged to guess. Each sentence was scored based on the accuracy of word content and order. Sentences were categorized for memory by length (involving use or non-use of adjectives) and for grammar by complexity (use or nonuse of a ditransitive passive sentence structure where subject and object were expressed by two prepositions). Sentences could thus be short with low complexity (e.g. “The mom baked her daughter a pie.”), short with high complexity (e.g. “A piano was delivered by the dad to his son.”), long with low complexity (e.g. “The kind man ordered the tired woman a hot coffee.”), or long with high complexity (e.g. “A purple pencil was offered by a friendly girl to the new boy.”). Different sets of sixteen sentences comprised the task list at each research visit. Each set had four sentences from each of the four categories.

Sentence repetition in noise: The Bamford-Kowal-Bench Sentences in Noise (BKB-SIN) test contains 18 list pairs of recorded sentences in four-talker babble noise (Bench et al., 1979; Etymōtic Research, 2005). In this study, participants were presented with one list pair (2 x 10 sentences) at a constant 70 dBrms SPL from a frontal speaker. Sentences were preceded by the verbal cue “Ready” and spoken by a male voice in Standard American English. The concurrent babble noise level increased by 3 dB after each sentence, progressively reducing signal-to-noise ratio (SNR) from +21 to -6 dB across each 10 sentence sublist (Etymōtic Research, 2005; Holder et al., 2018). Participants were asked to repeat each sentence verbally and were scored by the number of correctly identified keywords, averaged across the two sublists. Reported scores are the SNR-50, the SNR at which 50% of target words were correctly identified, normalized as SNR loss relative to normal hearing listeners.

Novel word learning in quiet: The NEPSY-II, a standardized neuropsychological test battery, assesses multiple neurocognitive domains (Brooks et al., 2009; Korkman et al., 1997). The Repetition of Nonsense Words subtest gauges novel word-learning ability. Successful completion of the task requires participants to decode phonological stimuli and articulate novel words. Thirteen recorded nonsensical words in a male voice were presented at 65 dB SPL through the same speaker configuration as the sentence repetition in quiet task. Words varied from two to five syllables in length, where one correctly-pronounced syllable amounted to one point. Points from each word were summed to formulate a composite score, which was then scaled by age.

Novel grammar learning in noise: The children were exposed to a novel grammar (von Koss Torkildsen et al., 2013) in the form of a game where they were teaching an alien a new language. During a single exposure at RV2, children were introduced to an aX and Yb grammar, with a and b components consisting of a single syllable nonword (CV), X representing one-syllable nonwords (CVC), and Y representing one-syllable nonwords (CVCCV). Examples of the aX ‘poe’ grammar are “poe zek” and “poe zug”, whereas Yb ‘koo’ grammar is exemplified by “wagso koo” and “zikvoe koo”. After four examples, to introduce the participant to the format of the task, the grammar exposure was split into 12 blocks, each 4 trials long ending with 2 two-alternative forced choice trials to assess their learning. This totaled 48 exposures to the grammar with 24 test trials to assess grammar knowledge. Target words were presented at 0° azimuth and speech shaped background noise, created from the spectral envelope of the nonwords used in the task, was presented at 180° azimuth throughout the task. This test was administered only at RV2.

Cognition: The NIH Toolbox Cognition Battery assesses the brain’s higher-level functions language, perception, planning and execution of behavior, and memory (Weintraub et al., 2013). This battery was administered because of the previously demonstrated close relationship between hearing loss, listening difficulty and cognitive ability (Moore et al., 2020; Petley et al., 2021). Four subtests from this battery, each lasting 5-15 minutes, were administered to participants on an iPad with the tester seated next to them. Individual subtests produced a raw score as well as an age-standardized score. Together the results of each standardized subtest comprise the ‘Early Childhood Composite’ score.

- The Picture Vocabulary Test assesses receptive vocabulary. Four pictures were presented together on the screen, and a single word was spoken by a female voice in Standard American English. Participants were instructed to select the picture that best matched the meaning of the spoken word.
- The Flanker Inhibitory Control and Attention Test measures attention and inhibitory control, an element of executive function. Five shapes (fish or arrows) were presented in a row on the screen, each pointing either left or right. Participants selected the direction (left or right) of the middle stimulus, while ignoring the distractors, as quickly as possible. Accuracy and response time were factored into scoring.
- The Dimensional Change Card Sort Test assesses the planning and execution of behavior (attention shifting, executive function). In each trial, participants were presented one of two shapes (e.g. a ball or a boat) that could be either yellow or blue. The word “shape” or “color” appeared on the screen, instructing the participant how to choose quickly the appropriate matching item.
- The Picture Sequence Memory Test assesses episodic memory. Illustrations were presented in a specific order with verbal narration, after which they were randomly shuffled. Participants were asked to place the illustrations back in the order in which they were presented.

Academic achievement: Four standardized subtests were selected from The Woodcock-Johnson IV (Schrank et al., 2014) to test reading and mathematical ability. Questions progressively increased in difficulty throughout each subtest with scoring procedures following basal and ceiling rules. Participants were guided through the assessment by the tester, who was seated next to them.

- The *Letter-Word Identification* test measures word identification ability and pronunciation accuracy. Single words were listed vertically on a screen, and the participant was instructed to read each word out loud from top to bottom.
- The *Passage Comprehension* test measures reading comprehension. Younger children are initially presented with pictures and must point to a specific one upon tester instruction. These transition to sentences with a word missing wherein the participant must silently read the sentence and tell the tester what the missing word is. Older children were tasked with lengthier passages that required the use of context clues to determine the missing word.
- The *Applied Problems* test presented math problems in word or picture form. Participants were given a pencil and scrap paper to use at their discretion.
- The *Calculation* test assesses straightforward mathematical knowledge. Participants were given a worksheet that began with simple numerical calculations (addition, subtraction, etc.) and eventually geometric, logarithmic, and calculus-based problems.

Caregiver- and self-report scales

Speech, Spatial and Qualities of Hearing Scale (SSQ): The SSQ (Gatehouse & Noble, 2004) is a self-report measure of hearing ability in varying contexts. This study used a version adapted for children (Galvin & Noble, 2013) in which caregivers rated their child's listening ability in everyday scenarios on a scale from 0-100. Twenty-seven items were included across four categories: speech hearing, spatial hearing, qualities of hearing, and conversational uses of hearing. A higher average score for each category indicated better ability. This questionnaire was completed at both RV1 and RV2 (Fig. 4).

Glasgow Hearing Aid Benefit Profile: The GHABP (Gatehouse, 1999) is a self-report measure of the effectiveness of assistive technology in hearing-impaired adults. In this study, a modified pediatric version based on caregiver report was used (Kubba et al., 2004). The Glasgow Children's Benefit Inventory (GCBI) consists of 24 questions on the effectiveness of recent or present hearing aids covering hearing disability, handicap, hearing aid usage, hearing aid benefit, hearing aid satisfaction, and residual disability relative to the benefit of prior hearing aids. This questionnaire was administered at both RV1, when it asked about effectiveness of the pre-study hearing aids, and RV2, when it asked about the study (OPN) hearing aids. All questions utilized a five-point Likert scale with each scalepoint corresponding to a numerical score (100 = much better, 50 = a little better, 0 = no change, -50 = a little worse, -100 = much worse). The 24 responses were averaged to create a composite wherein higher/more positive values reflected greater intervention benefit.

Pediatric Fatigue Questionnaire: The Patient-Reported Outcomes Measurement Information System (PROMIS) short form includes measures of physical, mental, and social health (Irwin et al., 2010; Lai et al., 2013; Varni et al., 2014). The shortened form of the Pediatric Fatigue measure includes 10 items assessing feelings of tiredness on a five-point Likert scale (1 = Never, 2 = Almost Never, 3 = Sometimes, 4 = Often, 5 = Almost Always; Lai et al., 2013). Recent research suggests that hearing impairment may increase risk for fatigue due to the increased need for deliberate listening and attention (Hornsby et al., 2016). Though much of this evidence is based on subjective accounts, the use of patient-reported outcomes is arguably the most direct source of information on health outcomes (Gerhardt et al., 2018). Participants completed a self-report version with the assistance of study staff, and parents completed a proxy version. Scores were averaged to create a composite for each version; higher scores indicate more frequent feelings of fatigue. The PROMIS Pediatric Fatigue measure was administered only at RV2.

Statistical analysis

Linear mixed effects models with correlated errors (Raudenbush & Bryk, 2002) were applied to study the main outcome differences between research visits 1 and 2 and the algorithm groups (OSN, OMNI; Table 2). The interaction effect between visit and group was also explored. Differences between test conditions were tested using a PROC Mixed model, controlling for test age and family Benjamini-Hochberg adjustment for false discovery rate. Where available, standardized test scores were used. Novel grammar learning in noise was administered only at RV2. A repeated measures ANOVA compared accuracy on the aX and Yb grammars. Planned follow up analysis used one sample t-tests to compare the accuracies of these grammars to chance (50%). The PROMIS questionnaire was also only administered at RV2. ANCOVA was used separately to study the group difference for the self-report and the care-giver report scores, controlling for age. A planned paired samples t-test was used to compare if the self- and care-giver reports were significantly different from one another. Linear mixed models were also applied to study effects of hearing aid use (age at first use, mean daily use during study). Participants who showed a clear button preference (i.e. clicking the same button for the majority of the trials: OSN n = 3, OMNI n = 4), or who logged less than 6 hours of daily HA use (OSN n = 3, OMNI n = 3) were excluded from the analysis of that task, or of the whole study, respectively. An OSN participant with HA technical difficulties was also excluded. Between 22 – 39 participants provided usable data for each task. Data were analyzed employing SAS statistical software, version 9.4 (SAS Institute, Cary, N.C.). A two-sided significance level was set at 0.05

RESULTS

Audiometry, hearing aid use and aided audibility dose

The two groups (OSN and OMNI) were well matched for age, mean hearing loss and interaural symmetry (Table 1; Fig. 2A) with no mean PTA group difference (left $t_{38} = 0.77$, $p = 0.44$; right $t_{38} = 0.37$, $p = 0.72$). However, there was a wide range of individual audiogram configurations and hearing loss severity. Hearing aid use varied considerably within both groups (Fig. 2B). Nevertheless, both the median and the variability of hearing aid use between the two groups was near identical (Wilcoxon rank-sum exact test, $p = 0.98$), suggesting that the dose of aided audibility (Walker et al., 2020) would also be near identical in the two groups. To confirm this, we calculated a weighted hearing aid benefit (HAB) by multiplying the aided SII for each ear by the average use hours for the same ear, divided by the maximal possible benefit (SII=1.0, use = 24 hours). Thus, the maximum benefit would be equal to 1.0 for each ear using this modified formula. The average was 0.37 for the right ear and 0.38 in the left ear for the OMNI group. For the OSN group, the average HAB was also 0.37 in the right and 0.38 in the left ear. There was no difference between groups for either ear in the HAB ($p=0.984$ for right, and $p=0.914$ for left ears, student t-test).

Behavioural testing

Linear mixed models showed no significant interaction between algorithm group and research visit for any outcome (Table 2). The hypothesis that extended use of the OSN algorithm would result in improved performance was therefore rejected. The following presentation provides further detail of performance by all participants on each outcome measure across the study and in relation to hearing aid use.

< Figure 3 about here >

Word repetition in noise (Fig. 3A): Performance (threshold SNR) differed between the three conditions ($F_{2, 78} = 7.45$, $p = 0.001$). Word repetition was poorer in the two- talker speech masker (Condition 3) compared with the speech-shaped noise masker (Conditions 1; $t_{71} = -3.71$, $p = 0.001$; and 2; $t_{78} = -2.81$, $p = 0.017$) but was unaffected by whether the target was in front (Condition 1) or 60^o left (Condition 2) of the listener ($t_{78} = -0.72$, $p = 0.75$). Overall, performance improved across time between RV 1 and 2 ($F_{1, 29}$ = 11.72, p = 0.002), but only in Conditions 2 and 3 (Table 2). No significant interactions were found between Condition and RV.

Sentence repetition in quiet (Fig. 3B): Short sentences were reproduced more accurately than long sentences ($F_{1,28}$ = 15.61, p < 0.001, ηp^2= 0.36) and sentences of low complexity were reproduced more accurately than those of high complexity ($F_{1,28}$ = 22.80, p < 0.001, ηp^2= 0.45). However, following Tukey-Kramer adjustment, no significant difference was found between Long & low and Long & high sentences. Accuracy on Long sentences of low complexity decreased significantly between visits, and younger children performed more poorly on the Long sentence/high complexity task than did older children (Table 2).

Sentence repetition in noise (Fig. 3C): SNR loss scores were mostly in the range 3-7 dB, described by the test manual as a “Mild SNR loss”, where 0-3 dB is “Normal” and 7-15 dB is “Moderate”. No SNR differences were observed based on group, visit or age (Table 2).

Novel word learning in quiet (Fig. 3D): No significant change in scores across RV was observed.

Novel grammar learning in noise (Fig. 3E): This was assessed at RV2 only, as the stimuli and paradigm needed to be novel. The participants learnt the Yb grammar significantly better than the aX grammar ($F_{1,20}$ = 8.96, p = 0.007, ηp^2= 0.31). Further analysis showed that participants did not learn the aX grammar, where the grammar key ‘poe’ was at the start of the word, significantly more than chance (50%). But they did learn the Yb grammar ($F_{1, 20}$ = 14.63, p = 0.001; r = 0.66, p = 0.0006), where the grammar key ‘koo’ was at the end of the word. Accuracy for the Yb grammar increased significantly with the child’s age ($F_{1,20}$ = 14.63, p = 0.001; r = 0.66, p = 0.0006), by 6.72% for each year. There was no significant difference between OSN and OMNI groups on either 'poe’ ($F_{1, 20}$ = 1.94, p = 0.18) or ‘koo’ ($F_{1, 20}$ = 0.04, p = 0.84) accuracy.

Cognition (Fig. 3F): Age-standardized scores have a mean of 100 and a SD of 15. Participant scores were in the range 80-120 for all of the four subtests. Across all participants there was a significant difference between subtest scores ($F_{3,93} = 5.29$, $p = 0.005$, $\eta p^2 = 0.15$ Greenhouse-Geisser corrected), with picture sequence memory test performance significantly better than that for the flanker task ($t = -3.97$, $p < 0.001$, $d = -0.69$).

Academic achievement (Fig. 3G): Age-standardized scores ranged between 90 and 120 on both outcomes and, compared with most other tests and excepting a few outliers, were remarkably stable for each individual across time. There was no significant change between RV1 and RV2 (Reading: $F_{1, 61.9} = 3.08$, $p = 0.08$; Math: $F_{1, 58.7} = 0.05$, $p = 0.83$) and no significant difference between OSN and OMNI groups (Reading: $F_{1, 37.8} = 3.50$, $p = 0.07$; Math: $F(1, 37.2) = 3.15$, $p = 0.08$) overall or between research visits.

Questionnaires

Speech, Spatial and Qualities of Hearing Scale (SSQ; Fig. 4A): Between RV1 and RV2, all scores either increased significantly (Speech, Spatial, Quality) or showed a trend to do so (Conversation) (Table 2). The Spatial score also showed a trend towards improved performance between RV1 and RV2, specifically in the OMNI group. This may suggest that caregivers perceived an increased benefit of the OMNI strategy as a result of prolonged use of that strategy. However, note that OMNI scores were lower than OSN scores at RV1. Finally, for these non-standardized scores there was an overall trend for increasing score with age.

< Figure 4 about here >

Glasgow Children's Benefit Profile (GCBP; Fig. 4B): Significantly higher overall scores for the OSN group than for the OMNI group were found, but significantly lower overall scores were found at RV2 than at RV1 (Table 2). The difference between groups at RV2 was not significant (Two sample t-test: $t_{31} = 1.61$, $p = 0.12$). This pattern of results is difficult to interpret but does not clearly support or challenge the primary hypothesis.

Pediatric Fatigue Questionnaire (PROMIS; Fig. 4C): The PROMIS was administered at RV2 only and showed no significant difference between OSN and OMNI groups on either the participant self-report scores ($F_{1, 28} = 4.01$, $p = 0.06$) or those by their caregivers ($F_{1, 29} = 3.45$, $p = 0.07$). There was no significant difference between participants' self-report and their caregiver scores (Paired t-test: $t_{29} = 0.12$, $p = 0.90$).

Hearing aid use and academic achievement

Given the similarity between OSN and OMNI performance, we combined both groups to evaluate the effect of hearing aid fitting and use patterns with increased power. Participants' age at first hearing aid fit (Table 1) correlated significantly with their WJ-IV Reading ($r = -0.42$, $p = 0.016$) and Mathematics ($r = -0.43$, $p = 0.012$) outcomes at RV2, but not at RV1 (Reading: $r = -0.36$, $p = 0.037$; Mathematics: $r = -0.34$, $p = 0.052$; Fig. 5A), following test age and family-wise correction. The improvement in the Flanker task between RV1 and RV2 correlated significantly with their average daily hearing aid usage during the trial ($r = 0.47$, $p = 0.006$; Fig. 5B). These comparisons at RV2 remained significant after family correction and indicated that the overall duration of hearing aid use was associated with better academic and cognitive outcomes. However, no other significant relationships between test performance and age at first fit or duration of hearing aid use were found. A significant negative relationship was found between age at first hearing aid use and caregiver rated speech hearing. A significant positive relationship was found between daily hearing aid use and caregiver rated spatial hearing (data not shown).

< Figure 5 about here >

DISCUSSION

Long term use of a hearing aid processing algorithm (OSN) designed to improve speech hearing in situations of competing speech neither enhanced nor reduced performance on a variety of age-standardized auditory, cognitive, and academic tasks, relative to a control, omnidirectional processor. Some improvements in performance, both objective (perceptual, linguistic, cognitive, academic) and subjective (SSQ, GCBP), were observed between RV1 and RV2 in both treatment groups. These were likely due to practice (Taylor et al, 2020) or caregiver expectation (SSQ). However, some exceptions to these otherwise unsurprising results were observed. One was the emergence of a significant relationship between age at first hearing aid use and both Reading and Mathematics scores that became significant at RV2 and another was the improvement of inhibitory control and attention with daily hearing aid use during the trial.

Enhanced target word repetition in SSN, compared with a two-talker masker (TTM), has previously been reported in children with mild/moderate hearing loss using OPN hearing aids, regardless of whether the target talker was directly in front of the child (Browning et al., 2019). Other studies using headphone stimulation have also shown more TTM in younger children, but not in older youth or adults (McCreery et al, 2020; Buss et al, 2021). In the study reported here, we confirmed these findings of greater TTM, and extended them by showing improved performance in the off-center target test (Condition 2) after 7-13 months of OPN hearing aid use. Browning and colleagues (2019) additionally found that performance using the OSN processing algorithm was superior to that of the same children using the OMNI algorithm in SSN masking, but not in TTM. Here, different groups of children used OSN and OMNI algorithms, and we did not find enhanced benefit of OSN during either SSN masking or TTM. It is possible that the difference between these studies was due to the reduced variance associated with retesting the same participants in two test conditions.

Previously, we found that word repetition in noise did not improve after 2-3 months of using OSN in a new fitting, but caregiver reports of ability improved over the study period (Pinkl et al., 2021). Here, we report improved word repetition in noise after 7-13 months of continuous use of OSN and OMNI. In addition, we found a general improvement in attention and caregiver-reported (SSQ) speech abilities. Together, these results provide evidence for longer-term acclimatization. Acclimatization may reflect enhanced training, motivation and/or expectation, along different timelines. For example, there may be a general and rapid, positive motivation and caregiver expectation after a new intervention (Gilliver et al., 2013), followed by a more gradual perceptual learning.

There is a growing literature on the effects of complex processing enhancements on speech perception using hearing aids (Cox et al., 2014; directional microphones - Magnusson et al., 2013; noise reduction - Brons et al., 2014; Magnusson et al., 2013; frequency compression - Hopkins et al., 2014; Picou et al., 2015). In a recent review, Lesica (2018) summed up these enhancements as showing only “modest additional benefit” (p. 175) beyond amplification. By showing no significant difference in a highly controlled study of the auditory and related cognitive and academic performance of children using omni or directional microphones with noise reduction processing strategies, the results reported here appear to support Lesica’s thesis. However, his optimistic suggestion was that alternative strategies based on recent developments in physiology and artificial intelligence should improve matters in the near future. In particular, an emphasis on the importance of cognitive and environmental factors to determine optimal, individualized processing strategies seems particularly suited to a pediatric population.

Moeller & Tomblin (2015), Ching et al (2018), and Nakeva von Mentzer et al. (2020) have shown the importance of early and prompt fitting of hearing aids for children with hearing loss. The average age of our participants was 9 years 9 months and the length of hearing aid usage prior to the study was 4 years 1 month. We showed that, regardless of processing strategy, the age of hearing aid fit continued to have an effect on academic achievement during development with the relationship between age of first hearing aid use and reading/mathematics scores showing significance at RV2, and trending at RV1.

The end of our study was conducted during the COVID-19 pandemic with eight of our participants (OSN = 5, OMNI = 3) completing RV2 during periods of school closures (Table 1). It is of note that four of these children (OSN = 2, OMNI = 2) were not included in the final analysis as they stopped wearing their hearing aids consistently with their daily usage below 6 hours. Consistent daily use of HAs has been highlighted as a strong predictor of cognitive and academic scores (Walker et al., 2015b). We echo this with the relationship between daily HA use and change in performance on the selective attention (flanker) task. Adding a covariate of whether the participants completed testing during the pandemic did not affect our findings. Significant relationships between hearing aid use and caregiver reports must be interpreted with caution because of possible bias related to caregiver knowledge of hearing aid use.

We have provided a broad spectrum of test approaches including auditory sensitivity, several forms of speech listening, learning in noise and in quiet, cognitive and academic performance, and caregiver- and self-report. On none of these tests was any significant performance difference found between the OMNI and the OSN processing strategies. A potential exception to this was the GCBP, where significantly enhanced scores for the OSN group were found across research visits. However, the questions asked in each visit differed, and neither the interaction between group and visit, nor the difference between groups at RV2, was significant. Nevertheless, it remains a possibility that in more realistic test environments differences would be found between these hearing aid strategies, for example, where distracting sounds were presented simultaneously from several different directions or where visual information is combined with auditory. While a within-subject design reduces individual differences, our study had multiple levels of quality checks built in. These included groups well matched in hearing level, audiogram shape and age, double-blinding throughout the study until the last participant completed their RV2, checklists to ensure all audiologists followed the same rigorous fitting protocols, and study pre-registration. We found no evidence that children learned how to use the additional features provided by OSN.

References

Audio Research. (2018). http://www.sentencerepetition.com/

Auditec, Inc. (2021, June 14). *Speech Discrimination for Children*. https://auditec.com/2015/11/09/speech-discrimination-for-children/

Bagatto, M., Moodie, S., Brown, C., Malandrino, A., Richert, F., Clench, D., & Scollie, S. (2016). Prescribing and verifying hearing aids applying the American Academy of Audiology pediatric amplification guideline: Protocols and outcomes from the Ontario infant hearing program. *Journal of the American Academy of Audiology*, *27*(3), 188–203.

Bagatto, M. P., Moodie, S. T., Malandrino, A. C., Richert, F. M., Clench, D. A., & Scollie, S. D. (2011). The University of Western Ontario pediatric audiological monitoring protocol (UWO PedAMP). *Trends in Amplification*, *15*(1), 57–76.

Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108–112. https://doi.org/10.3109/03005367909078884

Bender, R. A., Niebuhr, D. P., Getta, J. P., & Anderson, C. V. (1993). Longitudinal study of hearing aid effectiveness. I: Objective measures. *Journal of Speech, Language, and Hearing Research*, *36*(4), 808–819.

Brons, I., Houben, R., & Dreschler, W. A. (2014). Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. *Trends in Hearing*, *18*, 2331216514553924.

Brooks, B. L., Sherman, E. M. S., & Strauss, E. (2009). NEPSY-II: A Developmental Neuropsychological Assessment, Second Edition. *Child Neuropsychology*, *16*(1), 80–101. https://doi.org/10.1080/09297040903146966

Browning, J. M., Buss, E., Flaherty, M., Vallier, T., & Leibold, L. J. (2019). Effects of adaptive hearing aid directionality and noise reduction on masked speech recognition for children who are hard of hearing. *American Journal of Audiology*, *28*(1), 101–113.

Browning Jenna M., Buss Emily, Flaherty Mary, Vallier Tim, & Leibold Lori J. (2019). Effects of Adaptive Hearing Aid Directionality and Noise Reduction on Masked Speech Recognition for Children Who Are Hard of Hearing. *American Journal of Audiology*, *28*(1), 101–113.

Buss, E., Calandruccio, L., Oleson, J., & Leibold, L.J. (2021). Contribution of stimulus variability to word recognition in noise versus two-talker speech for school-age children and adults. *Ear and Hearing*, *42 (2)*, 313–322.

Butcher, E., Dezateux, C., Cortina-Borja, M., & Knowles, R. L. (2019). Prevalence of permanent childhood hearing loss detected at the universal newborn hearing screen: Systematic review and meta-analysis. *PloS One*, *14*(7), e0219600.

Ching, T. Y. C., Dillon, H., Leigh, G., & Cupples, L. (2018) Learning from the Longitudinal Outcomes of Children with Hearing Impairment (LOCHI) study: summary of 5-year findings and implications. International Journal of Audiology, 57(suppl. 2), S105-S111.

Chung, K., & Zeng, F.-G. (2009). Using hearing aid adaptive directional microphones to enhance cochlear implant performance. *Hearing Research*, *250*(1–2), 27–37.

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of open-set word recognition in children: Speech-shaped noise and two-talker speech maskers. *Ear and Hearing*, *37*(1), 55.

Cox, R. M., Johnson, J. A., & Xu, J. (2014). Impact of advanced hearing aid technology on speech understanding for older listeners with mild to moderate, adult-onset, sensorineural hearing loss. *Gerontology*, *60*(6), 557–568.

Dawes, P., Munro, K. J., Kalluri, S., & Edwards, B. (2014). Acclimatization to hearing aids. *Ear and Hearing*, *35*(2), 203–212.

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, *26*(5), 451–457.

Ertmer, D. J. (2011). Assessing speech intelligibility in children with hearing loss: Toward revitalizing a valuable clinical tool. *Language, Speech, and Hearing Services in Schools*.

Etymōtic Research. (2005). *Bamford-Kowal-Bench Speech-in-Noise Test*.

Galvin, K. L., & Noble, W. (2013). Adaptation of the speech, spatial, and qualities of hearing scale for use with children, parents, and teachers. *Cochlear Implants International*, *14*(3), 135–141. https://doi.org/10.1179/1754762812Y.0000000014

Gatehouse, S. (1992). The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids. *The Journal of the Acoustical Society of America*, *92*(3), 1258–1268.

Gatehouse, S. (1999). Glasgow hearing aid benefit profile: Derivation and validation of. *J Am Acad Audiol*, *10*(80), 103.

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology*, *43*(2), 85–99.

Gerhardt, W. E., Mara, C. A., Kudel, I., Morgan, E. M., Schoettker, P. J., Napora, J., Britto, M. T., & Alessandrini, E. A. (2018). Systemwide Implementation of Patient-Reported Outcomes in Routine Clinical Care at a Children's Hospital. *The Joint Commission Journal on Quality and Patient Safety*, *44*(8), 441–453. https://doi.org/10.1016/j.jcjq.2018.01.002

Gilliver, M., Ching, T. Y., & Sjahalam-King, J. (2013). When expectation meets experience: Parents' recollections of and experiences with a child diagnosed with hearing loss soon after birth. *International Journal of Audiology*, *52*(sup2), S10–S16.

Gravel, J. S., Fausel, N., Liskow, C., & Chobot, J. (1999). Children's speech recognition in noise using omni-directional and dual-microphone hearing aid technology. *Ear and Hearing*, *20*(1), 1–11.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J. G., & others. (2009). A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*, *42*(2), 377–381.

Holder, J. T., Levin, L. M., & Gifford, R. H. (2018). Speech Recognition in Noise for Adults With Normal Hearing: Age-Normative Performance for AzBio, BKB-SIN, and QuickSIN. *Otology & Neurotology*, *39*(10), e972. https://doi.org/10.1097/MAO.0000000000002003

Hopkins, K., Khanom, M., Dickinson, A.-M., & Munro, K. J. (2014). Benefit from non-linear frequency compression hearing aids in a clinical setting: The effects of duration of experience and severity of high-frequency hearing loss. *International Journal of Audiology*, *53*(4), 219–228.

Hornsby, B. W. Y., Naylor, G., & Bess, F. H. (2016). A Taxonomy of Fatigue Concepts and Their Relation to Hearing Loss. *Ear and Hearing*, *37*, 136S. https://doi.org/10.1097/AUD.0000000000000289

Hughson, W., Westlake, H., & others. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Trans Am Acad Ophthalmol Otolaryngol*, *48*(Suppl), 1–15.

Humes, L. E., & Wilson, D. L. (2003). An examination of changes in hearing-aid performance and benefit in the elderly over a 3-year period of hearing-aid use. *Journal of Speech, Language, and Hearing Research*.

Humes, L. E., Wilson, D. L., Barlow, N. N., & Garner, C. (2002). Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *Age (Years)*, *72*, 73–0.

Hunter, L., L., & Blankenship, C. (2017). Immittance Measures in Children. In *Comprehensive Handbook of Pediatric Audiology* (2nd Ed. R Seewald and AM Tharpe). Plural Publishing.

Irwin, D. E., Stucky, B. D., Thissen, D., DeWitt, E. M., Lai, J. S., Yeatts, K., Varni, J. W., & DeWalt, D. A. (2010). Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. *Quality of Life Research*, *19*(4), 585–594.

Korkman, M., Kemp, S., & Kirk, U. (1997). *NEPSY: A developmental neuropsychological assessment*.

Kubba, H., Swan, I. R., & Gatehouse, S. (2004). The Glasgow Children's Benefit Inventory: A new instrument for assessing health-related benefit after an intervention. *Annals of Otology, Rhinology & Laryngology*, *113*(12), 980–986.

Lai, J.-S., Stucky, B. D., Thissen, D., Varni, J. W., DeWitt, E. M., Irwin, D. E., Yeatts, K. B., & DeWalt, D. A. (2013). Development and psychometric properties of the PROMIS® pediatric fatigue item banks. *Quality of Life Research*, *22*(9), 2417–2427.

Le Goff, N., Jensen, J., Pedersen, M. S., & Callaway, S. L. (2016). An introduction to OpenSound Navigator™. *Oticon A/S*.

Lesica, N. A. (2018). Why do hearing aids fail to restore normal auditory perception? *Trends in Neurosciences*, *41*(4), 174–185.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

Magnusson, L., Claesson, A., Persson, M., & Tengstrand, T. (2013). Speech recognition in noise using bilateral open-fit hearing aids: The limited benefit of directional microphones and noise reduction. *International Journal of Audiology*, *52*(1), 29–36.

McCreery, R. W., Miller, M. K., Buss, E., & Leibold, L. J. (2020). Cognitive and Linguistic Contributions to Masked Speech Recognition in Children. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3525–3538.

McCreery, R. W., Walker, E. A., Spratford, M., Lewis, D., & Brennan, M. (2019). Auditory, cognitive, and linguistic factors predict speech recognition in adverse listening conditions for children with hearing loss. *Frontiers in Neuroscience*, *13*, 1093.

McCreery, R. W., Walker, E. A., Stiles, D.J., Spratford, M., Oleson, J.J., & Lewis, D.E. (2020).

Audibility-based hearing aid fitting criteria for children with mild bilateral hearing loss. *Language, Speech and Hearing Services in Schools, 51*(1), 55-67.

McGarrigle, R., Gustafson, S. J., Hornsby, B. W., & Bess, F. H. (2019). Behavioral measures of listening effort in school-age children: Examining the effects of signal-to-noise ratio, hearing loss, and amplification. *Ear and Hearing*, *40*(2), 381–392.

Moeller, M. P., & Tomblin, J. B. (2015). An introduction to the outcomes of children with hearing loss study. *Ear and Hearing*, *36*(0 1), 4S.

Moll, K., Hulme, C., Nag, S., & Snowling, M. J. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Applied Psycholinguistics*, *36*(2), 203.

Muñoz, K., & Hill, M. M. (2015). Hearing aid use for children with hearing loss: A literature review. *Perspectives on Hearing and Hearing Disorders in Childhood*, *25*(1), 4–14.

Nakeva von Mentzer, C., Wallfelt, S., Engström, E., Wass, M., Sahlén, B., Pfändtner, K., Lyxell, B., Kallioinen, P., Lindgren, M., Ors, M., & others. (2020). Reading Ability and Working Memory in School-Age Children Who Are Deaf and Hard of Hearing Using Cochlear Implants and/or Hearing Aids: A 3-Year Follow-Up on Computer-Based Phonics Training. *Perspectives of the ASHA Special Interest Groups*, *5*(6), 1388–1399.

Niskar, A. S., Kieszak, S. M., Holmes, A., Esteban, E., Rubin, C., & Brody, D. J. (1998). Prevalence of hearing loss among children 6 to 19 years of age: The Third National Health and Nutrition Examination Survey. *Jama*, *279*(14), 1071–1075.

Picard, M., & Bradley, J. S. (2001). Revisiting Speech Interference in Classrooms: Revisando la interferencia en el habla dentro del salón de clases. *Audiology*, *40*(5), 221–244.

Picou, E. M., Marcrum, S. C., & Ricketts, T. A. (2015). Evaluation of the effects of nonlinear frequency compression on speech recognition and sound quality for adults with mild to moderate hearing loss. *International Journal of Audiology*, *54*(3), 162–169.

Pinkl, J., Cash, E. K., Evans, T. C., Neijman, T., Hamilton, J. W., Ferguson, S. D., Martinez, J. L., Rumley, J., Hunter, L. L., Moore, D. R., & others. (2021). Short-Term Pediatric

Acclimatization to Adaptive Hearing Aid Technology. *American Journal of Audiology*, *30*(1), 76–92.

Raudenbush, S.W., & Bryk, A.S. (2001). *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed.* In: Advanced Quantitative Techniques in the Social Sciences, Vol. 1. Thousand Oaks: SAGE Publications.

Ranjbar, P., & Nakeva von Mentzer, C. (2020). *The Web-based Sentence Repetition Test to diagnose developmental language disorder (DLD): User manual.* http://oru.diva-portal.org/smash/get/diva2:1511789/FULLTEXT01.pdf

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Riverside.

Su, B. M., & Chan, D. K. (2017). Prevalence of hearing loss in US children and adolescents: Findings from NHANES 1988-2010. *JAMA Otolaryngology–Head & Neck Surgery*, *143*(9), 920–927.

Taylor, K. S. (1993). Self-perceived and audiometric evaluations of hearing aid benefit in the elderly. *Ear and Hearing*, *14*(6), 390–394.

Tomblin, J. B., Oleson, J. J., Ambrose, S. E., Walker, E., & Moeller, M. P. (2014). The influence of hearing aids on the speech and language development of children with hearing loss. *JAMA Otolaryngology–Head & Neck Surgery*, *140*(5), 403–409.

Varni, J. W., Magnus, B., Stucky, B. D., Liu, Y., Quinn, H., Thissen, D., Gross, H. E., Huang, I.-C., & DeWalt, D. A. (2014). Psychometric properties of the PROMIS® pediatric scales: Precision, stability, and comparison of different scoring and administration options. *Quality of Life Research*, *23*(4), 1233–1243.

von Koss Torkildsen, J., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). *Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability*.

Walker, E. A., Holte, L., McCreery, R. W., Spratford, M., Page, T., & Moeller, M. P. (2015a). The influence of hearing aid use on outcomes of children with mild hearing loss. *Journal of Speech, Language, and Hearing Research*, *58*(5), 1611–1625.

Walker, E. A., Holte, L., McCreery, R. W., Spratford, M., Page, T., & Moeller, M. P. (2015b). The influence of hearing aid use on outcomes of children with mild hearing loss. *Journal of Speech, Language, and Hearing Research*, *58*(5), 1611–1625.

Walker, E. A., Sapp, C., Dallapiazza, M., Spratford, M., McCreery, R. W., & Oleson, J. J. (2020). Language and reading outcomes in fourth-grade children with mild hearing loss compared to age-matched hearing peers. *Language, Speech, and Hearing Services in Schools*, *51*(1), 17–28.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11 Suppl 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America*, *119*(6), 3940–3949.

Wildt, E., Rohlfing, K. J., & Scharlau, I. (2019). The Role of Saliency in Learning First Words. *Frontiers in Psychology*, *10*, 1150.

Table legends

Table 1: Participant demographics and summary of hearing aid models used prior to the study along with how long they wore their study hearing aids (time from HA fitting to RV2).

Table 2: Linear mixed model main outcome p-values. Values in black were non-significant, those in red were significant, following Benjamini-Hochberg (B-H) adjustment, and those in blue were no longer significant following (B-H) adjustment. Refer to Figures 3 and 4 for data. Note that Novel grammar learning in noise, and the PROMIS fatigue questionnaire were not included in the model.

Figure legends

Figure 1: Study timeline.

Figure 2: Hearing loss and hearing aid use. (A) Audiograms, showing individual thresholds in light grey and group means in color (OSN dark, OMNI light). PTAs calculated from 500, 1000 and 2000 Hz. (B) Average daily hearing aid use during the study, as measured by the hearing aid logging feature. Points are individual data. Boxplots show the groups' median and quartiles, with the whiskers indicating maximum and minimum number of hours.

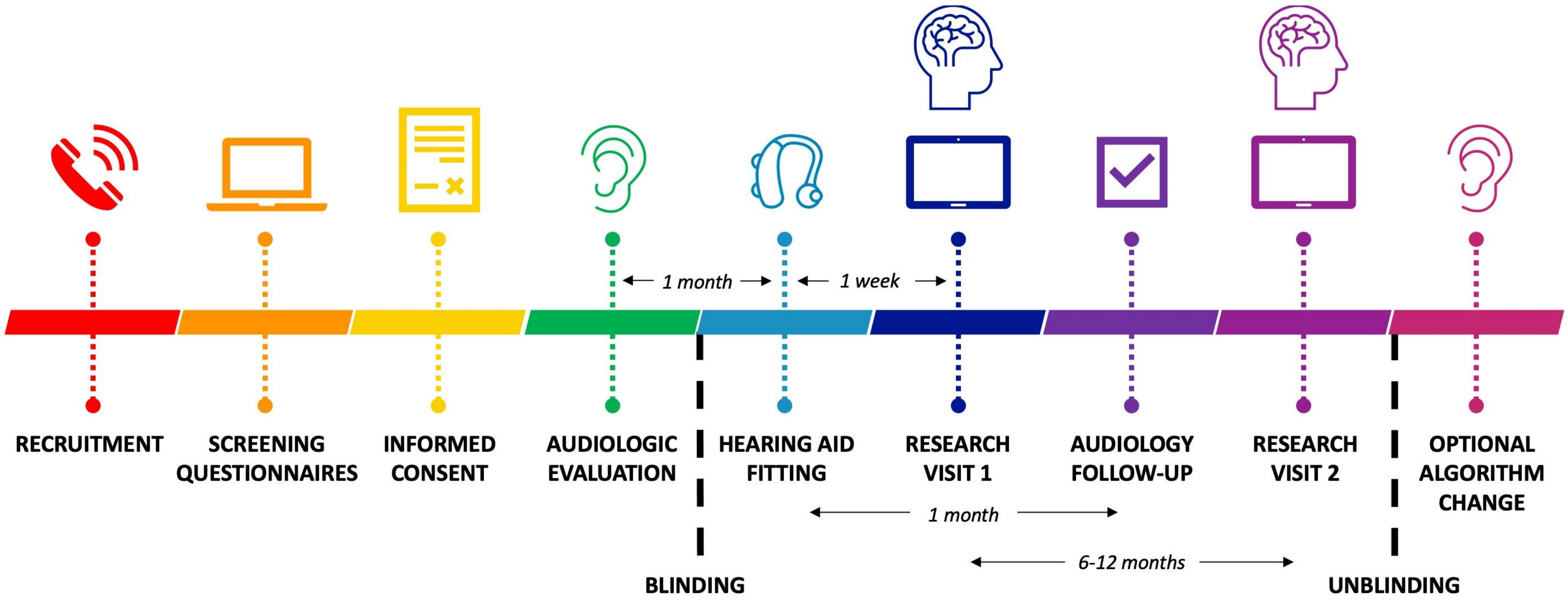
Figure 3: Behavioral test performance. No difference was found between the OMNI and OSN group on any task. (A) Word repetition in noise. Signal-to-noise ratio (SNR) thresholds for words in noise for three Conditions: 1: Target speech at front with speech-shaped noise (SSN) maskers behind; 2: Target on left with SSN maskers behind; 3: Target speech at front with two-talker speech masking (TTM) behind. (B) Sentence repetition in quiet scores for the four sentence types. (C) Sentence repetition in noise. (D) Novel word learning in quiet showing age standardized scores from the NEPSY. (E) Novel grammar learning in noise, measured only at

RV2. The artificial grammar was presented at front with SSN presented at rear. The horizontal dashed line on the graph marks vocabulary learning at chance, 50%. Error bars show SEM. (F) Cognition measured by age standardized scores on 4 subtasks from the NIH Toolbox (Weintraub et al., 2013): picture vocabulary; attention – flanker task; executive function – dimensional change card sorting task; episodic memory – picture sequence memory test. (G) Academic achievement measured by composite reading and mathematics age standardized scores (Woodcock-Johnson IV Tests of Achievement). For all figures, colored dots and joining show individual score change between RV1 and 2. Boxplots show the 25th, 50th and 75th percentile of each group; whiskers indicate maximum/minimum scores.

Figure 4: Questionnaires. (A) Speech, spatial and qualities of hearing scale (SSQ; Galvin & Noble, 2013). (B) Glasgow hearing aid benefit profile (Kubba et al., 2004). For figures (A) and (B), the colored dots show individual scores with the lines linking each individual's performance between RV1 and RV2. (C) PROMIS pediatric fatigue questionnaire showing standardized T-scores from the self- and caregiver-reports (Varni et al., 2014). Lower scores indicate more frequently reported fatigue. Colored dots show individual scores and lines link self- with caregiver-report, both measured at RV 2 only. In all graphs the boxplots show group 25th, 50th and 75th percentiles and the whiskers indicate maximum and minimum scores.

Figure 5: Hearing aid (HA) use predicted academic performance and attention. (A) Negative correlations between the children's age at first hearing aid fit with WJ-IV Reading and Mathematics outcomes at RV2. (B) Positive correlation between daily hearing aid usage and change in performance on the NIH toolbox attention task (flanker task).

Figure 1



1 month
1 week
RECRUITMENT
SCREENING QUESTIONNAIRES
INFORMED CONSENT
AUDIOLOGIC EVALUATION
HEARING AID FITTING
RESEARCH VISIT 1
AUDIOLOGY FOLLOW-UP
RESEARCH VISIT 2
OPTIONAL ALGORITHM CHANGE
1 month
6-12 months
BLINDING
UNBLINDING

Figure 2

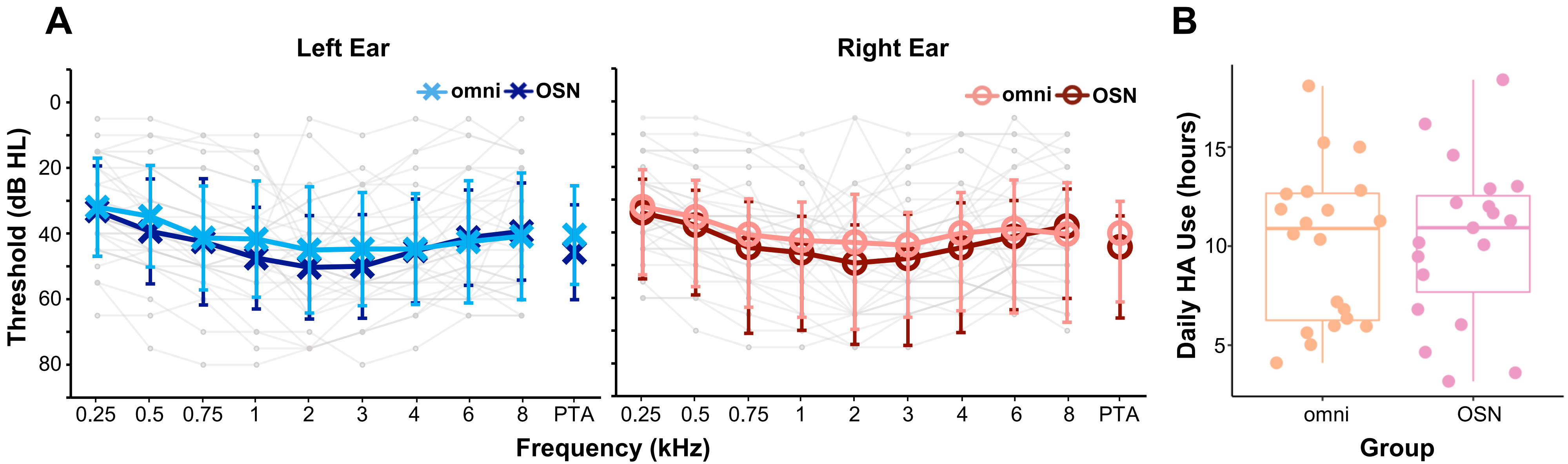

A
Left Ear
omni
OSN
Right Ear
omni
OSN
Threshold (dB HL)
0
20
40
60
80
0.25
0.5
0.75
1
2
3
4
6
8
PTA
Frequency (kHz)
B
Daily HA Use (hours)
5
10
15
omni
OSN
Group

Figure 3

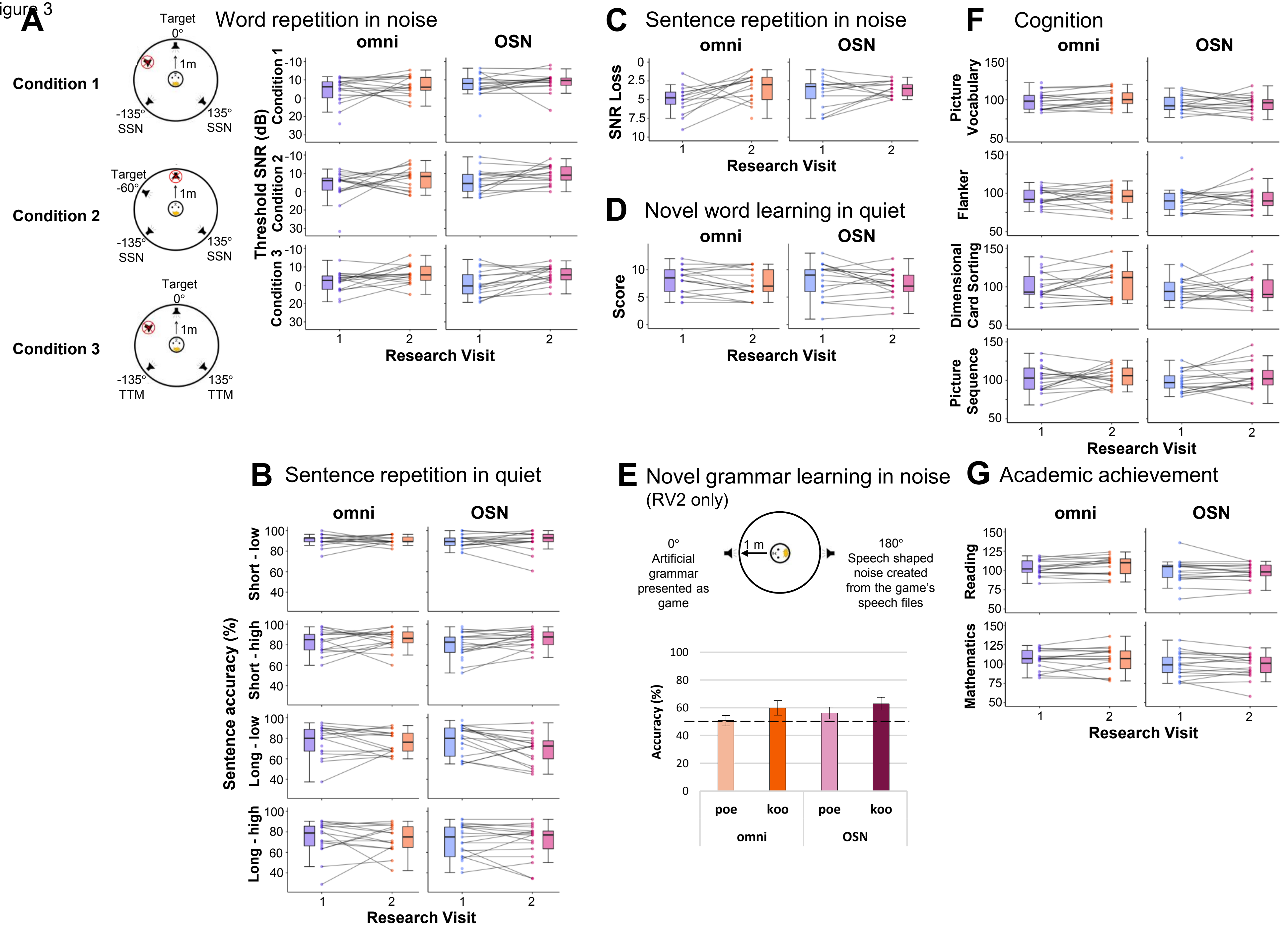

A
Word repetition in noise
Condition 1
Condition 2
Condition 3
Target
0°
1m
-135°
SSN
135°
SSN
Target
-60°
-135°
TTM
135°
TTM
omni
OSN
Threshold SNR (dB)
Research Visit
B
Sentence repetition in quiet
Sentence accuracy (%)
Short - low
Short - high
Long - low
Long - high
C
Sentence repetition in noise
SNR Loss
D
Novel word learning in quiet
Score
E
Novel grammar learning in noise
(RV2 only)
0°
Artificial grammar presented as game
1 m
180°
Speech shaped noise created from the game's speech files
Accuracy (%)
poe
koo
F
Cognition
Picture Vocabulary
Flanker
Dimensional Card Sorting
Picture Sequence
G
Academic achievement
Reading
Mathematics

Figure 4

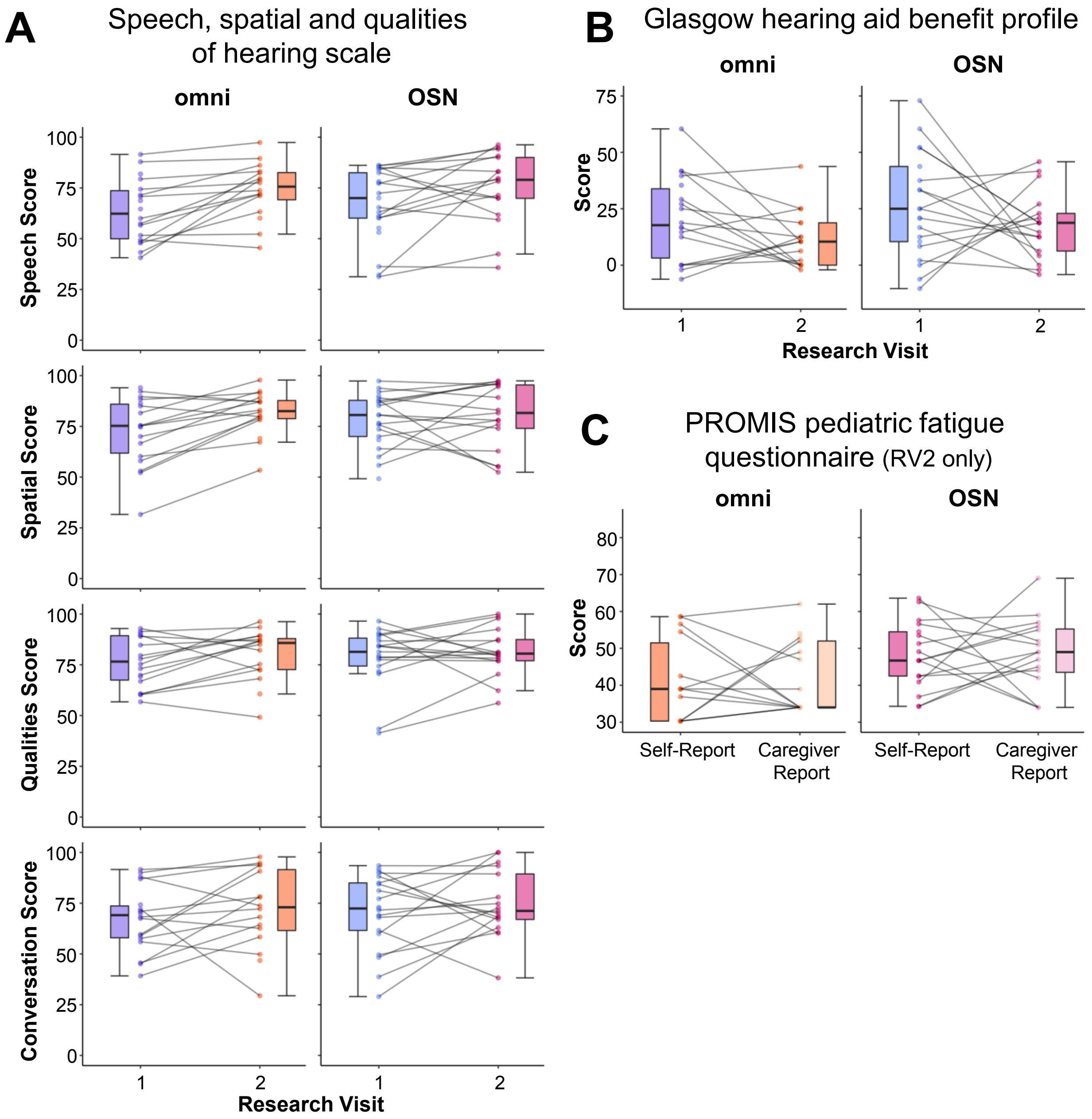
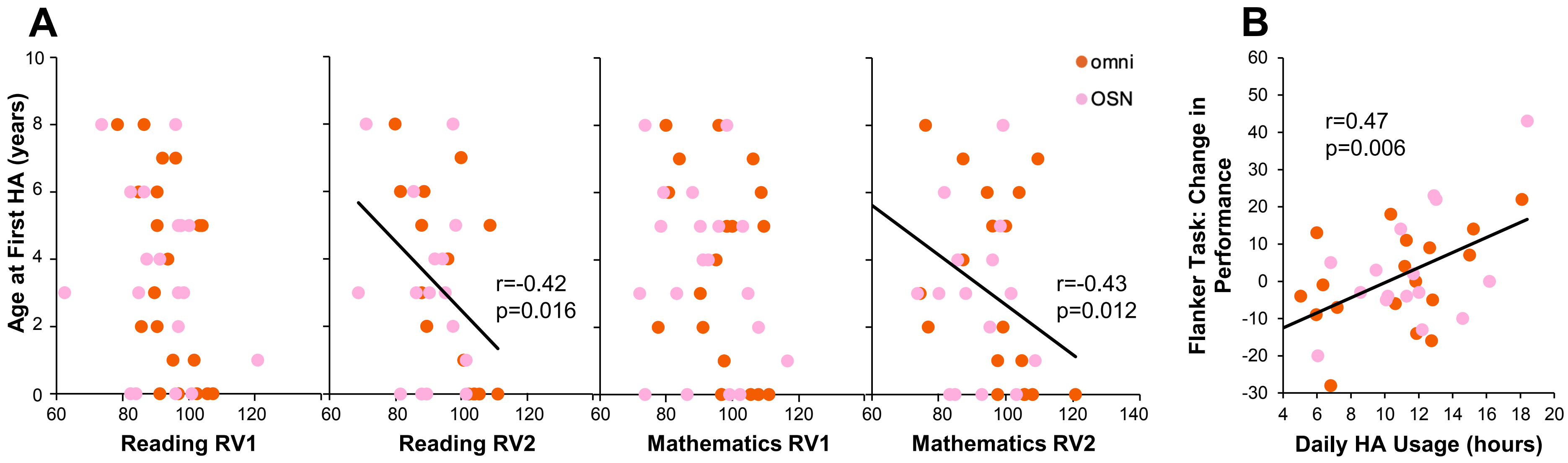

A
Speech, spatial and qualities of hearing scale
omni
OSN
Speech Score
Spatial Score
Qualities Score
Conversation Score
Research Visit
B
Glasgow hearing aid benefit profile
omni
OSN
Score
Research Visit
C
PROMIS pediatric fatigue questionnaire (RV2 only)
omni
OSN
Score
Self-Report
Caregiver Report
Self-Report
Caregiver Report

Figure 5


A
B
Age at First HA (years)
10
8
6
4
2
0
60
80
100
120
140
Reading RV1
Reading RV2
Mathematics RV1
Mathematics RV2
r=-0.42
p=0.016
r=-0.43
p=0.012
omni
OSN
Flanker Task: Change in Performance
60
50
40
30
20
10
-10
-20
-30
r=0.47
p=0.006
4
6
8
10
12
14
16
18
20
Daily HA Usage (hours)

Table

Table 1

Pair	Group	Shape of HL	R PTA	L PTA	Maternal education	Age at RV1 (y, m)	Age of diagnosis (y)	Age at first HA (y)	Previous HA Type Make	Previous HA Type Model	Time with previous HA (y, m)	Time with study HA (m)	RV2 during COVID
1	OSN	Bowl/notch	40.0	53.3	No high school	8, 7	6	6	Oticon	Sensei Pro	2, 0	8	no
	omni	Bowl/notch	43.3	43.3	College degree	9, 1	1	5	Widex	Dream 220	3, 2	8	no
2	OSN	Bowl/notch	71.7	76.7	College degree	11, 6	1	1	Phonak	Sky Q-70 M13	4, 4	11	no
	omni	Bowl/notch	58.3	66.7	College degree	13, 4	2	2	Phonak	Sky V50 P	2, 9	6	no
3	OSN	Falling	45.0	36.7	College degree	7, 7	birth	9 months	Oticon	Sensei Pro	1, 8	8	no
	omni	Falling	61.7	63.3	College degree	8, 5	5	5	Oticon	Sensei Pro SP	1, 7	8	no
4	OSN	Falling	20.0	20.0	College degree	11, 5	3	3	Oticon	Safari 600P	8, 4	8	no
	omni	Falling	23.3	25.0	College degree	11, 7	1	1	Phonak	Sky Q50 M13	4, 5	9	no
5	OSN	Falling	23.3	26.7	Some college	11, 10	8	8	Phonak	Sky V90 M	2, 2	7	no
	omni	Bowl/notch	50.0	46.7	No high school	11, 8	7	8	Phonak	Sky Q-50 SP	3, 11	8	no
6	OSN	Flat	68.3	68.3	College degree	11, 0	3	3	Phonak	Sky Q50-SP	4, 7	9	no
	omni	Flat	65.0	61.7	Post-graduate degree	10, 0	birth	2 months	Phonak	Bolero V 90-P	3, 3	11	no
7	OSN	Rising	35.0	36.7	Some college	10, 6	3	3	Oticon	Safari 600P	6, 2	10	no
	omni	Rising	31.7	31.7	Post-graduate degree	9, 6	5	6	Phonak	Sky Q50-M13	2, 7	9	no
8	OSN	Bowl/notch	28.3	36.7	Some college	9, 9	5	5	Phonak	Sky 50 Q M 13	4, 6	12	no
	omni	Falling	30.0	23.3	College degree	10, 0	1	1	Phonak	Nios S H20 III	5, 5	8	no
9	OSN	Bowl/notch	40.0	41.7	College degree	6, 8	4	4	Oticon	Sensei Pro	2, 0	11	no
	omni	Bowl/notch	43.3	45.0	College degree	6, 2	4 months	9 months	Oticon	Safari 600P	5, 6	10	no
10	OSN	Falling	41.7	41.7	College degree	11, 7	0	3	Widex	Inteo BTE	7, 11	8	no
	omni	Bowl/notch	31.7	38.3	Some college	12, 4	3	3	Oticon	Safari	8, 0	9	no
11	OSN	Rising	26.7	30.0	College degree	8, 10	6	6	Oticon	Sensei Pro	2, 2	15	yes
	omni	Falling	40.0	23.3	College degree	9, 8	7	7	Oticon	Sensei Pro	2, 6	9	no

12	OSN	Bowl/notch	51.7	51.7	College degree	11, 11	3-6 months	3-6 months	Oticon	Sensei Pro	2, 9	8	yes
	omni	Bowl/notch	56.7	45.0	Some college	10, 10	8	8	Phonak	Sky Q 50 SP	2, 7	12	yes
13	OSN	Flat	45.0	45.0	Some college	7, 6	5	5	Oticon	Sensei Pro	1, 11	9	yes
	omni	Rising	38.3	43.3	Some college	7, 11	1 month	2 months	Phonak	Sky Q50 SP	3, 8	10	no
14	OSN	Flat	48.3	48.3	College degree	8, 6	1 month	5 months	Oticon	Safari 600 P	8, 1	13	yes
	omni	Flat	48.3	48.3	Some college	8, 7	4	4	Widex	Dream 440	4, 7	13	yes
15	OSN	Falling	51.7	50.0	Some college	10, 5	4 months	5 months	Oticon	Vigo Pro	9, 11	9	no
	omni	Falling	10.0	11.7	College degree	9, 4	5	6	Phonak	Sky Q 50 M13	3, 3	10	no
16	OSN	Bowl/notch	58.3	58.3	College degree	9, 5	4	5	Phonak	Sky Q70 SP	4, 5	10	no
	omni	Flat	30.0	30.0	Post-graduate degree	9, 7	Unknown	Unknown	Unknown	Unknown	9, 4	9	no
17	OSN	Flat	36.7	38.3	Some high school	10, 11	8	9	Oticon	Sensei Pro	1, 10	9	no
	omni	Rising	43.3	50.0	Some college	11, 11	7	7	Resound	LiNX3D977	1, 10	10	yes
18	OSN	Flat	35.0	35.0	Post-graduate degree	10, 4	4	4	Oticon	Safari 600	5, 10	13	yes
	omni	Falling	35.0	36.7	Post-graduate degree	10, 5	5	5	Oticon	Sensei Pro	1, 1	11	no
19	OSN	Flat	45.0	48.3	Post-graduate degree	6, 10	2	2	Phonak	Sky Q50 m13	4, 2	9	no
	omni	Flat	45.0	45.0	Some college	7, 5	1 month	3 months	Phonak	SkyQ 50 m13	3, 5	7	no
20	OSN	Falling	36.7	35.0	Post-graduate degree	9, 7	5	5	Oticon	Sensei Pro	5, 0	11	no
	omni	Falling	31.7	31.7	College degree	10, 2	2	2	Phonak	Sky V90-P	2, 2	8	no

Table 2

Outcome measure	Algorithm Group	Research Visit	Group* Visit	Age
Word repetition in noise				
Condition 1 (dB SNR)	0.061	0.127	0.796	0.137
Condition 2	0.306	0.019	0.655	0.826
Condition 3	0.843	<0.001	0.674	0.458
Sentence repetition in quiet				
Short and low accuracy (%)	0.321	0.998	0.839	0.869
Short and high accuracy	0.869	0.113	0.583	0.328
Long and low accuracy	0.404	0.016	0.323	0.196
Long and high accuracy	0.525	0.143	0.881	0.007
Sentence repetition in noise				
SNR loss (dB)	0.297	0.117	0.391	0.321
Novel word learning in quiet				
Standard score	0.746	0.409	0.869	0.063
NIH Cognition Toolbox				
Picture vocabulary (Standard score)	0.162	0.371	0.574	0.240
Flanker inhibition and attention	0.622	0.308	0.728	0.048
Dimensional change card sort	0.225	0.164	0.203	0.286
Picture sequence memory	0.327	0.036	0.720	0.055
SSQ hearing evaluation scale				
Speech (score)	0.883	<0.001	0.475	0.277
Spatial	0.931	0.008	0.083	0.057
Quality	0.670	0.024	0.628	0.104
Conversation	0.732	0.086	0.984	0.083
Glasgow Children's Benefit				
Score	0.041	0.021	0.642	0.148

Academic achievement				
Reading (score)	0.125	0.097	0.088	0.045
Math (score)	0.058	0.778	0.230	0.698