



Published in final edited form as:

Ear Hear. 2017 ; 38(3): 344–356. doi:10.1097/AUD.0000000000000393.

Some Neurocognitive Correlates of Noise-Vocoded Speech Perception in Children with Normal Hearing: A Replication and Extension of Eisenberg et al., 2002

Adrienne S. Roman¹, David B. Pisoni², William G. Kronenberger³, and Kathleen F. Faulkner²

¹Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA

²Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

³Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA

Abstract

Objectives—Noise-vocoded speech is a valuable research tool for testing experimental hypotheses about the effects of spectral-degradation on speech recognition in adults with normal hearing (NH). However, very little research has utilized noise-vocoded speech with children with NH. Earlier studies with children with NH focused primarily on the amount of spectral information needed for speech recognition without assessing the contribution of neurocognitive processes to speech perception and spoken word recognition. In this study, we first replicated the seminal findings reported by Eisenberg et al. (2002) who investigated effects of lexical density and word frequency on noise-vocoded speech perception in a small group of children with NH. We then extended the research to investigate relations between noise-vocoded speech recognition abilities and five neurocognitive measures: auditory attention and response set, talker discrimination and verbal and nonverbal short-term working memory.

Design—Thirty-one children with NH between 5 and 13 years of age were assessed on their ability to perceive lexically controlled words in isolation and in sentences that were noise-vocoded to four spectral channels. Children were also administered vocabulary assessments (PPVT-4 and EVT-2) and measures of auditory attention (NEPSY Auditory Attention (AA) and Response Set (RS) and a talker discrimination task (TD)) and short-term memory (visual digit and symbol spans).

Results—Consistent with the findings reported in the original Eisenberg et al. (2002) study, we found that children perceived noise-vocoded lexically easy words better than lexically hard words. Words in sentences were also recognized better than the same words presented in isolation. No significant correlations were observed between noise-vocoded speech recognition scores and the PPVT-4 using language quotients to control for age effects. However, children who scored higher

Address correspondence to Adrienne S. Roman, DHSS Graduate Studies and Research, Medical Center East, 1215 21st Avenue South, Suite 8310, Nashville, TN 37232, USA.: adrienne.s.roman@vanderbilt.edu.

Financial Disclosures/Conflicts of Interests: This research was funded by grants from the NIH- National Institute on Deafness and Other Communication Disorders: T32 DC000012, R01 DC000111, and R01 DC009581 We have no conflicts of interest to report

on the EVT-2 recognized lexically easy words better than lexically hard words in sentences. Older children perceived noise-vocoded speech better than younger children. Finally, we found that measures of auditory attention and short-term memory capacity were significantly correlated with a child's ability to perceive noise-vocoded isolated words and sentences.

Conclusions—First, we successfully replicated the major findings from the Eisenberg et al. (2002) study. Because familiarity, phonological distinctiveness and lexical competition affect word recognition, these findings provide additional support for the proposal that several foundational elementary neurocognitive processes underlie the perception of spectrally-degraded speech. Second, we found strong and significant correlations between performance on neurocognitive measures and children's ability to recognize words and sentences noise-vocoded to four spectral channels. These findings extend earlier research suggesting that perception of spectrally-degraded speech reflects early peripheral auditory processes as well as additional contributions of executive function, specifically, selective attention and short-term memory processes in spoken word recognition. The present findings suggest that auditory attention and short-term memory support robust spoken word recognition in children with NH even under compromised and challenging listening conditions. These results are relevant to research carried out with listeners who have hearing loss, since they are routinely required to encode, process and understand spectrally-degraded acoustic signals.

Introduction

Researchers have learned a great deal about the development of speech perception and spoken language processing by studying listeners with hearing loss, especially children who are born deaf and later receive cochlear implants (CIs) (Niparko, 2009; Waltzman & Roland, 2006; Zeng, Popper, & Fay, 2004). While the use of CIs has provided substantial benefits to profoundly deaf listeners, a period of early auditory deprivation can also be detrimental to cognitive and linguistic development (Kral & Eggermont, 2007; Niparko, 2009; Nittrouer, 2010). Because the brain is plastic, a period of auditory deprivation followed by compromised acoustic input, especially during the critical periods for language development, can result in neural reorganization from lack of normal sensory input during early stages of development (Gilley, Sharma, & Dorman, 2008; Kral, Kronenberger, Pisoni, & O'Donoghue, 2016).

Cognitive Development and Hearing Loss

Auditory deprivation and language delays have been shown to negatively impact cognitive development and executive functioning in deaf children with CIs. For example, children with CIs often have significant delays and deficits in verbal working memory dynamics (Burkholder & Pisoni, 2003; Pisoni, Kronenberger, Roman, & Geers, 2011), verbal short-term memory capacity (Harris et al., 2013), language development and reading (Johnson & Goswami, 2010), implicit sequence learning (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011), visual attention (Horn, Davis, Pisoni, & Miyamoto, 2005; Quittner, Smith, Osberger, Mitchell, & Katz, 1994) and theory-of-mind (Peterson, 2004).

Noise-Vocoded Speech and Spoken Word Recognition

The use of noise-vocoded speech has been a valuable research tool because it allows researchers to test specific hypotheses about the effects of spectral-degradation on listeners with normal hearing (NH) with typically-developed auditory and cognitive systems and to investigate how neurocognitive functioning affects speech recognition abilities (Conway, Deocampo, Walk, Anaya, & Pisoni, 2014; Dorman, Loizou, Kemp, & Kirk, 2000; Eisenberg, Martinez, Holowecky, & Pogorelsky, 2002; Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Newman & Chatterjee, 2013; Warner-Czyz, Houston, & Hynan, 2014). Noise-vocoded speech refers to speech signals that have been processed to preserve gross temporal and amplitude information but have degraded fine spectral information in the signal (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). This type of signal processing strategy was originally created by Shannon et al. (1995) to model the way a CI processes speech and to investigate perception of degraded speech in listeners with NH.

CIs and acoustic simulations of CIs create difficulty for spoken word recognition processes because they produce spectrally-degraded underspecified acoustic-phonetic information. While the temporal information is persevered, this specific alteration to the signal makes perception of vowels and some consonants more difficult because there is insufficient information available to support speech recognition. According to the principles of the Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998) underspecified acoustic-phonetic information results in coarsely-coded lexical representations of words that become organized into larger lexical neighborhoods with increased densities and greater competition among words for recognition (Bell & Wilson, 2001; Charles-Luce & Luce, 1990; Dirks, Takayana, & Moshfegh, 2001; Dirks, Takayanagi, Moshfegh, Noffsinger, & Fausti, 2001). Because of increased competition among similar-sounding words, spoken word recognition becomes significantly more challenging often requiring downstream predictive coding and contextual support from cognitive and linguistic processes. The term “downstream” refers to neurocognitive processes that do not only rely solely on sensory input for interpretation of the stimulus, but also integrate cognitive and linguistic knowledge and context to more accurately predict and perceive what is being heard based on context and knowledge of prior probabilities. These processes are especially important when the stimulus is compromised or challenging. Noise-vocoded speech has become an important research tool for studying the effects of spectral degradation on speech perception and spoken word recognition because it is possible to manipulate experimental variables and test specific hypotheses with listeners with NH, which is often not possible with a clinical population of CI users.

More than 10 years ago, Eisenberg et al. (2002) reported the results of a study that investigated spoken word recognition in children with NH using noise-vocoded speech. In one of their experiments, Eisenberg et al. (2002) assessed the effects of word frequency and neighborhood density on noise-vocoded word recognition and sentence perception using a small group of children 5–14 years in age. Eisenberg and colleagues created sets of “easy” and “hard” words based on manipulations of lexical density and word frequency (see Kirk, Pisoni, & Osberger, 1995). “Easy” words were high-frequency words selected from sparse lexical neighborhoods; “hard” words were low-frequency words from dense lexical neighborhoods. Eisenberg et al. (2002) found that children with NH recognized lexically

“easy” words better than lexically “hard” words when the same words were presented in both isolation and in sentences under noise-vocoded conditions. They also reported that accuracy for words in sentences was much better than for the same words presented in isolation showing a substantial context benefit gain (Miller, Heise, & Lichten, 1951). Eisenberg et al.’s (2002) study was important because it demonstrated the robustness of word frequency and neighborhood density in open-set word recognition tasks when listeners received spectrally-degraded speech signals. Successful word recognition is not only dependent on the initial quality and precision of sensory information encoded in the speech signal, but also reflects the contribution of underlying neurocognitive components that contribute to facilitating the recognition process as well. Since the time that Eisenberg et al.’s (2002) pioneering study was carried out, to the best of our knowledge, no attempts at replication have been published, although this study is widely cited in the literature. One goal of the present study was to carry out a replication of the original Eisenberg et al. (2002) findings.

Noise-Vocoded Speech and Children with Normal Hearing

The existing literature on noise-vocoded speech with children with NH has focused primarily on the early sensory aspects of speech perception (Dorman, Loizou, Kemp, & Kirk, 2000; Eisenberg et al., 2000; Newman & Chatterjee, 2013; van Heugten et al., 2013; Warner-Czyz, Houston, & Hynan, 2014). Because successful speech recognition relies on the integration and functioning of the whole information processing system working together, it is also important to investigate the effects of neurocognitive functioning on the perception of spectrally-degraded speech. Previous research has shown that speech recognition is not only dependent on the amount and quality of acoustic-phonetic information available in the speech signal (Eisenberg et al., 2002; van Heugten et al., 2013) but also reflects the contribution of powerful downstream predictive neurocognitive processes that contribute to the observed variability in performance. Thus, a second goal of this research was to assess relations between vocoded speech perception abilities in children with NH and several core aspects of cognition and executive functioning including auditory attention, inhibition, and short-term memory. These cognitive factors have been shown to play an important compensatory supporting role in pediatric CI users spoken word recognition performance (Cleary, Pisoni, & Kirk, 2005; Pisoni & Geers, 2000; Pisoni et al., 2011). If the same core underlying information processing operations are also used by children with NH listening to spectrally-degraded speech, we would also expect these same information processing factors to contribute to word recognition performance using isolated words and sentences.

Cognitive Factors Involved in Spoken Word Recognition

Attention is a broad theoretical construct in the field of cognitive science that refers to the properties of cognition that involve the control and effortful allocation of limited processing resources and information processing capacity (Cowan, 1995). Most theories assume that attention operates to keep information active in working memory during capacity-demanding information processing tasks. Auditory attention specifically refers to maintaining modality-specific auditory information active in immediate memory. Although auditory attention skills are critical to speech perception and the development of spoken

language processing, almost all of the research on attention in deaf children with CIs has focused on visual attention (Horn et al., 2005; Quittner et al., 1994; Smith, Quittner, Osberger, & Miyamoto, 1998; Tharpe, Ashmead, & Rothpletz, 2002) or preference to sound over silence (Houston, 2009). The existing literature on auditory attention in CI users has been concerned with the child's ability to attend to a stream of speech in the presence of noise or other distractors used in informational masking studies (Wightman & Kistler, 2005). Although this research is important because of its ecological validity, the presence of noise or competing distracting tasks creates an additional cognitive load and increased information capacity demands. Additional cognitive load and processing demands are often the result of greater mental effort and processing resources required to inhibit competition from irrelevant information while the listener actively directs attention to the critical target information in the signal (Zekveld, Kramer, & Festen, 2011).

Models of language processing such as the Ease of Language Understanding Model (Rönnberg et al., 2013), Interactive Compensatory Model (Stanovich, 1980), Two Process Theory of Expectancy (Posner & Snyder, 2004; Posner, Snyder, & Davidson, 1980), and Single Resource Model (Hula & McNeil, 2008) posit that phonological and semantic processing of language reflect the combined results of two processes. The first process is an early implicit automatic activation process that places few demands on active conscious neurocognitive processing. Second is an explicit, conscious, attention-demanding process that is effortful and places demands on active neurocognitive processing. Although these theories differ somewhat in their details, they all share in common the assumption that speech-language processing under challenging conditions (whether because of the linguistic complexity of the information or limitations in sensory abilities of the individual) requires the activation and use of compensatory effortful neurocognitive processes because fast automatic language processes are not sufficient to support perception and recognition under these adverse conditions. As a result, individual differences in core neurocognitive processes such as working memory capacity (Rönnberg et al., 2013), controlled focused selective attention (Posner & Snyder, 2004; Posner et al., 1980), and inhibitory control (Norman & Shallice, 1986) contribute to differences in spoken language processing under challenging conditions. Hence, processing of degraded speech signals may be influenced by higher order neurocognitive functions, and in particular the executive functions (which include working memory, controlled selective attention and inhibitory control), which are activated automatically to provide compensatory support in adverse spoken language situations (Kronenberger, Colson, Henning, & Pisoni, 2014). In the current study, we sought to identify core neurocognitive functions that underlie processing of spectrally-degraded speech signals. Based on our earlier research, we hypothesized that several components of executive functioning, especially controlled auditory selective attention and working memory capacity, would be positively related to better speech perception performance under challenging conditions (see Kronenberger, Colson, Henning, & Pisoni, 2014).

One novel approach to studying the relations between controlled auditory attention and speech perception has been to measure a listener's talker discrimination (TD) skills. Cleary and Pisoni (2002) developed a simple TD task for use with young children. Their procedure required the child to attend to the indexical properties of speech - the talker's voice - and actively inhibit processing of the linguistic content and meaning of short sentences.

Indexical properties of speech provide personal information about the talker such as gender, dialect, and emotional or physical states whereas linguistic properties of speech refer to the phonological and lexical symbolic content of the talker's intended utterance (Pisoni, 1997). Discriminating between talkers requires the listener to be able to perceive and encode the indexical properties of speech that are specific to each talker's voice. Cleary and Pisoni (2002) found that the TD was quite difficult for pediatric CI users because of the reduced spectral detail provided by the CI (see also Cleary et al., 2005). In order to successfully complete the TD task, the child has to selectively ignore and inhibit the allocation of attention and processing resources to the linguistic content of the sentences and instead consciously focus attention on the indexical characteristics of the vocal source -- the talker's voice.

In addition to controlled auditory attention, working memory is considered to be a central component of executive functioning that is activated as a compensatory neurocognitive process under challenging listening conditions (Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008). In this study, we also investigated relations between short-term/working memory and noise-vocoded speech perception. There is now strong agreement among cognitive scientists that short-term memory is the memory subsystem that stores and processes limited amounts of information for brief periods of time (Baddeley, 2012; Cowan, 2008; Unsworth & Engle, 2007) and plays a central foundational role in speech perception and language acquisition (e.g. Baddeley, Gathercole, & Papagno, 1998; Frankish, 1996; Gathercole, Service, Hitch, Adams, & Martin, 1999; Jacquemot & Scott, 2006; Jusczyk, 1997; Pisoni, 1975). Earlier studies have found that CI users show delays and deficits in verbal short-term memory compared to age-matched NH peers (Dawson, Busby, McKay, & Clark, 2002; Harris et al., 2013; Pisoni et al., 2011), which ultimately affects their speech and language processing skills as well as other aspects of cognitive development.

Short-term memory capacity is typically assessed with measures of immediate memory span such as digit or word span tests (Richardson, 2007; Wechsler, 2003). These methods require an individual to retain item and order information over a short period of time before recall of the information is required. Digit spans are frequently used as one index of short-term memory capacity and have consistently revealed strong relations with performance on a wide range of speech and language outcome measures in deaf children with CIs (Pisoni & Geers, 2000). Numerous other studies have found significant correlations between digit spans and speech and language scores showing that children with higher digit span scores display better performance in both open-set and closed-set spoken word recognition (Burkholder & Pisoni, 2003; Pisoni & Cleary, 2003; Pisoni & Geers, 2000; Pisoni et al., 2011), speech perception (Pisoni et al., 2011), speech intelligibility (Pisoni & Geers, 2000; Pisoni et al., 2011), vocabulary (Fagan, Pisoni, Horn, & Dillon, 2007; Pisoni et al., 2011), language comprehension (Pisoni & Geers, 2000), reading (Fagan et al., 2007; Pisoni & Geers, 2000; Pisoni et al., 2011), verbal rehearsal speed (Burkholder & Pisoni, 2003), and nonword repetition (Pisoni et al., 2011). These findings are not limited to clinical populations with hearing loss, however. Recently, Osman and Sullivan (2014) reported that digit spans of children with NH were also significantly correlated with performance on a speech perception in noise task. Taken together, these earlier findings suggest a central role for verbal short-term memory in the perception of degraded speech.

As noted earlier, research with children who use CIs has documented the importance of auditory attention and short-term memory in speech and language processing measures, but very little research has investigated these areas of cognition in children with NH listening to spectrally-degraded noise-vocoded speech. Only one study to date has used any cognitive processing measures with noise-vocoded speech. Eisenberg et al. (2000) had two groups of children with NH (5–7 and 10–12 years in age) and a group of adults with NH complete digit span tasks under varying amounts of acoustic degradation and then correlated their performance on a noise-vocoded digit span task with performance on several different speech perception measures. They found a significant positive correlation between digit span under 8-channel simulation and a language quotient derived from the Peabody Picture Vocabulary Test (PPVT) that controlled for chronological age. We extended these earlier findings by investigating associations between several additional neurocognitive measures and noise-vocoded speech perception in a larger group of children with NH 5–13 years of age.

In summary, the present study was designed to achieve two goals: first, we sought to replicate the findings of Eisenberg et al. (2002) on vocoded speech perception by children with NH. Our hypothesis was that the Eisenberg et al. (2002) findings would replicate in a larger sample, adding support to their original findings. Second, we sought to identify neurocognitive factors related to vocoded speech perception in children with NH. Consistent with existing theories of speech perception and our earlier research with deaf children who use CIs, we hypothesized that two components of executive function, controlled auditory attention and short-term/working memory capacity, would be positively related to better vocoded speech perception because these processes serve as effortful compensatory processes that support speech perception under adverse and challenging conditions.

Materials and Methods

Participants

Thirty-one children between 5;9 years and 13;3 years in age ($M = 10;0$ years, $SD = 2;4$ years; 12 females, 19 males) were recruited for this study. The majority of the sample was Caucasian ($n = 27$), with the remaining identified as either Native Hawaiian/Pacific Islander ($n = 2$) or multiracial ($n = 2$). Thirty-seven typically-developing monolingual English-speaking children from 5;2 years (years; months) to 13;3 years in age were originally recruited for this study, but six children had to be excluded for the following reasons: technical problems ($n = 2$), noncompliance ($n = 3$), and reported speech delays ($n = 1$). Based on parent-report, all children had normal hearing and vision and no diagnosed cognitive/developmental delays. All children were recruited through an IRB approved departmental subject database at Indiana University in Bloomington. The majority (58%) of the children who participated in this study were from a families with incomes in the \$50,000 to \$100,000 range; 19% reported incomes less than \$50,000; 6% reported incomes within the \$100,000 – \$150,000 range; 10% reported incomes within the \$150,000 – \$200,000 range; and 6% reported incomes greater than \$200,000. All children included in the final data analyses passed a pure-tone hearing screening at 15 dB HL between 250 – 4000 Hz to verify that their hearing was within normal limits.

Equipment

All speech perception testing was carried out in an IAC sound attenuated booth in the Speech Research Laboratory at Indiana University in Bloomington. A high-quality Advent AV570 loudspeaker was located approximately two feet from the listener. A Radio Shack Digital Sound Level Meter was used to verify stimulus presentation levels over loudspeaker at 65 dB HL using C-Weighting. All stimuli used in the speech perception, digit span, and symbol span tasks were presented using programs running on a Power Mac G4 Apple computer with a Mac OS 9.2 using Psyscript (Bates & D'Oliveiro, 2003). A 12" Keytec LCD Touch Monitor was used to present visual stimuli. The colored touchscreen presented stimuli at a brightness level of 150 cd/m² and a contrast ratio of 100:1. During presentation of the visual stimuli, participants were seated at a table directly in front of the touchscreen. The touchscreen's presentation angle was 120°.

Noise-Vocoded Speech Stimuli

Noise-vocoded speech signals were created using the techniques described in Shannon et al. (1995) and Eisenberg et al. (2002). Original audio recordings of the unprocessed speech stimuli were obtained from Dr. Laurie Eisenberg for the replication of Experiment 2 reported in the Eisenberg et al. (2002) paper. AngelSim (TigerCIS), an online speech-processing program, was used to generate all of the noise-vocoded speech stimuli. These signal processing algorithms preserved the temporal cues and amplitude of the speech signal by creating digitally filtered noise bands that were modulated by the original speech amplitude envelope from the same spectral band (Shannon et al., 1995). The original speech signals were processed to four spectral channels with bandwidth frequencies set at 300, 722, 1528, 3066, and 6000 Hz using a noise-vocoded setting with white noise as the carrier type. We consulted with Drs. Laurie Eisenberg and Mr. John Galvin directly when vocoding the original stimuli so we could verify that our vocoded stimuli were identical to the spectrally-degraded stimuli used by Eisenberg et al. They provided us with copies of the parameter values that they used in their studies.

Performance Measures

Eisenberg Word Familiarity Rating Scale—An Eisenberg word familiarity rating scale was used to assess each child's familiarity with the test words. Parents were asked to rate their child's familiarity of each of the 150 Eisenberg lexically controlled words on a Likert scale ranging from 1 (not at all familiar) to 7 (very familiar) (Kirk, Sehgal, & Hay-McCutcheon, 2000; Lewellen, Goldinger, Pisoni, & Greene, 1993; Nusbaum, Pisoni, & Davis, 1984).

Peabody Picture Vocabulary Test- 4th Edition (PPVT-4)—The PPVT-4 was used to obtain a measure of the child's receptive vocabulary. This test is a standardized vocabulary assessment that can be administered to participants ranging in age from 2.5 – 90+ years. During administration of the PPVT-4, the experimenter showed the child a page with four different colored illustrations displayed in a 2x2 box format. The experimenter said the stimulus word out loud and instructed the child to either point to or say the number associated with the picture that best illustrated the meaning of the word. Guessing was

encouraged if the child was uncertain. The child began the PPVT-4 at a predetermined point based on the child's chronological age and continued until he or she reached ceiling on the assessment (to reach ceiling the child had to miss at least eight items in one set). The PPVT-4 contains a total of 228 test items divided into 19 groups of 12 items. Items increased in difficulty as the test proceeded. The PPVT took approximately 15 minutes to complete. Raw and standard scores (normative mean=100, $SD=15$) were obtained for each child. All children were administered Form A of the PPVT-4 (Dunn & Dunn, 2007).

Expressive Vocabulary Test- 2nd Edition (EVT-2)—The EVT-2 was used to obtain a measure of each child's expressive vocabulary knowledge. This standardized vocabulary assessment can be administered to participants ranging in age from 2.6 – 90+ years. During administration, the examiner presented the child with an illustration and then read a stimulus question asking the child to either label the illustration or provide a synonym. The participant began at a predetermined point based on chronological age and continued until the participant reached ceiling on the assessment (the child reached ceiling when he/she made five consecutive errors). There were a total of 190 items arranged in order of increasing difficulty. The EVT-2 took approximately 15 minutes to complete. Raw and standard scores (normative mean=100, $SD=15$) were obtained for each child. All children were administered Form A of the EVT-2 (Williams, 2007).

Noise-Vocoded Word Intelligibility by Picture Identification- 2nd Edition (WIPI)—The WIPI test was a closed-set spoken word recognition test (Ross, Lerman, & Cienkowski, 2004) that was used to familiarize the children with the noise-vocoded speech stimuli. Each child listened to List A of the WIPI test, which consisted of 25 isolated noise-vocoded words. All children responded by pointing to one of six pictures that matched the word he or she heard. This assessment was scored for accuracy but was not included in any final data analyses reported below.

Noise-Vocoded Lexically Controlled Words and Sentences—The stimulus lists of lexically controlled words and sentences originally developed by Eisenberg et al. (2002) were used to investigate the effects of word frequency and neighborhood density on word recognition in isolation and in sentences. Eisenberg et al. (2002) created two lists of words based on lexical competition: one lexically “easy” list and one lexically “hard” list. Each list consisted of 15 practice words followed by 60 test words produced by one female speaker. The easy and hard word lists were then combined to create one set of 30 practice words and 120 test words that were noise-vocoded and presented in a randomized order. Practice trials always preceded test trials (see PDF, Supplemental Digital Content 1). For the word recognition task, children were instructed to repeat what they heard out loud to the experimenter. No feedback was provided regarding response accuracy.

Eisenberg et al. (2002) also created two lists of 25 low-predictability sentences (5 practice and 20 test sentences) using the same easy and hard words that were produced by one female speaker and previously presented in isolation (see Bell & Wilson, 2001). Each sentence was five to seven words in length and contained three “key” words from either the easy or hard list. The two sentence lists were combined to create one larger set of 10 practice sentences and 40 test sentences that were also noise-vocoded and presented in a randomized

order to each child. Once again, practice trials always preceded the test trials (see PDF, Supplemental Digital Content 2). For the sentence recognition task, children were instructed to repeat what they heard out loud to the experimenter. The noise-vocoded sentences were scored for number of key words correct. Children did not receive any feedback regarding their response accuracy.

Auditory Attention and Response Set Subtests (NEPSY-2)—The Auditory Attention (AA) subtest of the NEPSY-2 was used to obtain a measure of the child's selective auditory attention and ability to sustain attention. During administration of AA, the child listened to a three-minute series of 180 prerecorded spoken words presented over a loudspeaker and was instructed to touch the appropriate colored circle when the target color word (i.e., "red") was presented randomly on 30 of the 180 trials.

The Response Set (RS) subtest of the NEPSY-2 was used to assess the child's ability to shift to and maintain a new set of complex instructions while also inhibiting previously learned responses by correctly attending and responding to matching or contrasting stimuli. During administration of the RS, the participant listened to a three-minute series of 180 prerecorded spoken words presented over a loudspeaker and either touched the color, a contrasting color, or did nothing when a target color word was presented on 36 of the trials.

Raw and scaled scores were obtained from on both tests for each child. These two subtests can be administered to participants ranging in age from 5–16 years (AA) and 7–16 years (RS). Three children (1 five-year-old female, 1 six-year-old female, and 1 six-year-old male) were not administered the RS because of age restrictions of the assessment. Both subtests are part of the Attention and Executive Functioning domain of the NEPSY-2 and have a mean scaled score of 10 and *SD* of 3 (Korkman, Kirk, & Kemp, 2007). Following the test manual, the spoken words on the AA and RS subtests were presented in the clear and were not noise-vocoded. Only the raw scores from AA and RS subtests were used for statistical analyses reported here.

Talker Discrimination—A TD task was used to assess selective auditory attention using noise-vocoded speech. This task was modeled after the TD task originally created by Cleary and Pisoni (2002) to investigate the ability of deaf children with CIs to discriminate differences between talkers. In this task, the child heard pairs of short meaningful English sentences and was asked to make a judgment as to whether the talker who produced the first sentence in each pair was the "same" or "different" from the talker who produced the second sentence. Responses were recorded using a touchscreen monitor. We modified this task from the original procedure developed by Cleary and Pisoni (2002) to use noise-vocoded speech with children with NH. The sentences were noise-vocoded to four channels following the parameters described previously in the Eisenberg et al. (2002) paper. The key methodological design feature of the TD task is that the child is required to consciously ignore and inhibit the lexical-symbolic linguistic information in the pair of sentences and focus his or her attention on the indexical properties of the signal to make a same or different judgment based on the vocal source information.

Following Cleary & Pisoni (2002) two presentation conditions were used: a fixed-sentence condition and a varied-sentence condition. During the fixed-sentence condition, the same sentence was used for all trials. Participants completed eight practice trials (four pairs) per condition with experimenter feedback using unprocessed speech prior to listening to the noise-vocoded speech stimuli in order to verify that the subjects understood the instructions and task requirements. The practice trials consisted of two male talkers (Talker 1 and Talker 21) from the Indiana Multi-Talker Sentence Database (IMTSD) developed by Karl and Pisoni (1994). Two practice trials used the same talker and two practice trials used different talkers. The test trials followed the practice trials. The test trials consisted of three female talkers (Talkers 6, 7, and 23 from the IMTSD). There were a total of 24 sentences (12 pairs) used as test trials. Six trials used the same talker (each talker paired with herself twice) and six trials used pairs of different talkers (each talker paired with the other twice). The varied-sentence condition followed the same presentation format as the fixed-sentence condition except that each test pair contained a unique combination of two different sentences. Supplemental Digital Content 3 (PDF) shows the list of sentence pairings used in both conditions. To respond correctly on each of the varied-sentence trials, the child must actively inhibit attention and processing resources to the linguistic content of the sentences and focus his or her attention on the talker's voice. All sentence and child.

Visual Digit Span—A forward visual digit span task was used to obtain a measure of each child's verbal short-term memory capacity. This task had three types of trials: familiarity, practice, and test. Trials were administered via a touchscreen computer display. During the familiarity trials, the child saw a single digit (between one and nine) randomly presented on the touchscreen monitor. The visual digits appeared as a black number encased in a black box on a white backdrop in the center of the screen for one second. When the digit disappeared, the response screen appeared and displayed the digits, one through nine, in a 3x3 fixed grid format. The child was instructed to touch the number previously displayed on the screen. After the familiarity trials, the child began the practice trials. During the practice trials, the child saw a set of two and then a set of three single digits presented sequentially on the screen one after another, each digit displayed for a period of one second. The response screen then appeared and the child was instructed to reproduce the numbers previously seen in the order in which they were presented by touching the digits on the computer screen. The child had a window of five seconds after each presentation to respond before the experimental program advanced to the next trial. Each child had to successfully complete the practice trials to proceed to the test trials. The test trials began with a set size of two digits (i.e. list length of two). Each list length was presented twice and had to be correctly reproduced during one of the two trials before the list length increased by one digit on the next trial. When the child failed to correctly reproduce both trials at a given list length, the assessment was automatically terminated. The digit span task was scored for points correct. One point was awarded for each digit correctly reproduced in its correct serial order. For example, if a child was presented with the digits "1...3...5" and responded with "1...4...5" the child was awarded two points because the "1" and "5" were reproduced in their correct serial order. Use of a manual touchscreen response in the digit span task was used to eliminate any confounding issues related to verbal output and response organization in speech motor control.

Symbol Span—A forward symbol span task was created to assess each child's nonverbal short-term memory capacity. This assessment was administered using the same procedures and format as the visual digit span test except that nine black and white abstract visual symbols were used as stimuli in place of the familiar digits. The set of visual symbols is shown in Supplemental Digital Content 4 (PDF).

Statistical Analyses

In replicating the original Eisenberg et al. (2002) study, we carried out a series of analyses examining relations between lexical competition and recognition of noise-vocoded speech. Following Eisenberg et al. (2002), the percent-correct word recognition scores were first normalized by subjecting them to an arcsine transformation and then entered into a repeated measure ANOVA to test for main effects of lexical competition (easy and hard) and stimulus type (words and sentences). Next, we computed language quotients for the vocabulary scores, the PPVT-4 and EVT-2, and then calculated Spearman correlations between these two language quotients and the noise-vocoded speech perception measures. Language quotients were derived by taking the PPVT-4 age-equivalency scores and dividing them by the child's chronological age. This method provided an index of language development that controlled for the child's chronological age. We also examined the effects of chronological age on speech recognition performance through Spearman correlations.

In our initial assessment of the relations between neurocognitive measures and noise-vocoded speech recognition performance, we did not include separate analyses of the easy and hard lexical variable in order to increase the statistical power of analyses. Correlations were also carried out separately for the easy and hard words after the initial analyses averaging over both types of words. Because no consistent differences in the pattern of correlations among any of the neurocognitive measures were observed between easy and hard words, those analyses are not reported. Spearman correlations and t-tests were computed to assess relations between speech recognition and neurocognitive measures.

Procedures

All children were tested individually by the first author (ASR). The study was completed in one test session lasting about two hours. Parental consents and child assents, when applicable, were obtained prior to testing as per the guidelines of Indiana University's Institutional Review Board (IRB study #1203008368). All assessments, with the exception of the two vocabulary tests, were administered in an IAC sound booth in the Speech Research Laboratory at Indiana University in Bloomington. All children received monetary compensation, two books, and numerous stickers that were distributed throughout the testing session to maintain motivation and interest.

Results

Part I: Replication of Eisenberg et al. (2002)

Results from averaging Eisenberg's word familiarity rating scale data indicated that all children were highly familiar with the test words used (see Table 1). Figure 1 displays recognition performance for each individual child for words in isolation (top panel) and

words in sentences (bottom panel). Three horizontal lines were included to show mean performance for all words combined (total), easy words, and hard words. On average, children were able to accurately recognize fewer words presented in isolation compared to words presented in sentences with poorer performance in both conditions for lexically hard words.

Repeated measures ANOVA analyses revealed that words in sentences were recognized better than words in isolation, $F(1,30)=37.53$, $p<.001$, and that lexically easy words were recognized better than lexically hard words, $F(1,30)=5.82$, $p=.022$. The interaction between lexical competition and context was not significant.

As in the original study by Eisenberg et al. (2002), no significant correlations were observed for performance on speech recognition and the PPVT-4 (see Table 2). Only one significant correlation was found between the EVT-2 language quotient and easy sentences ($r=.37$, $p=.043$).

We also examined the effects of chronological age on speech recognition performance. Chronological age was significantly correlated with the noise-vocoded sentence perception measures in the original Eisenberg et al. (2002) study, and this pattern was also replicated with all of the noise-vocoded speech perception measures in the current study. Chronological age was significantly correlated with performance on words in isolation ($r=.56$, $p<.001$), lexically easy words in isolation ($r=.60$, $p<.001$), and lexically hard words in isolation ($r=.49$, $p=.006$). It was also significantly correlated with performance on sentences ($r=.71$, $p<.001$), lexically easy sentences ($r=.67$, $p<.001$), and lexically hard sentences ($r=.73$, $p<.001$).

Part II. Neurocognitive Correlates of Vocoded Speech Perception

Auditory Attention and Response Set—Raw scores from the auditory attention task and response set tests were first used in correlations with chronological age (months). Three children were removed from the AA data set because they failed to demonstrate understanding of the task. No significant correlations were observed between chronological age and auditory attention ($r=.34$, $p=.06$) or response set scores ($r=.30$, $p=.12$). Raw scores from the auditory attention and response set tests were then correlated with the noise-vocoded speech perception measures (see Table 3). Only one significant correlation was found: performance on the AA subtest was significantly correlated with performance on noise-vocoded words in isolation ($r=.40$, $p=.04$).

Talker Discrimination Task—Figure 2 shows the distribution of scores for the fixed-sentence (top panel) and varied-sentence (bottom panel) conditions in the TD task. The ordinate represents the proportion of correct responses produced out of the 12 trials. As shown in this figure, the majority of children (96.7%) performed above well above chance on the fixed-sentence condition, but only 64.5% performed above chance on the varied-sentence condition. Chronological age was strongly related to performance on both conditions (fixed-sentence: $r=.57$, $p=.001$; varied-sentence: $r=.41$, $p=.024$). As expected, performance on the fixed-sentence condition was significantly better than performance on the varied-sentence condition; $t(31)=4.76$, $p<.001$. Children's performance in both

conditions was significantly correlated with their ability to recognize words in noise-vocoded sentences (see Table 3). However, performance in the fixed-sentence condition was more strongly correlated with word recognition scores than the varied-sentence condition.

Error rates and types of errors also differed by condition. In the fixed-sentence condition, children showed a higher occurrence of “misses” responding “same” when the talkers in both sentences were actually “different” (misses: 14.79%; false alarm rate: 11.54%). In contrast, in the varied-sentence condition, children show a much higher occurrence of “false alarms,” responding “different” when the talkers in both sentences were in fact the “same” (misses: 16.91%; false alarm rate: 23.11%). The false alarm rate in the varied-sentence condition was twice the magnitude of the false alarm rate in the fixed-sentence condition, $t(30) = -.464, p < .001$.

Visual Digit Span and Symbol Span—Using a one-sample *t*-test, children did significantly better on visual digit span compared to symbol span; $t(31) = 5.91, p < .001$. Chronological age was significantly correlated with both visual digit span ($r = .61, p < .001$) and symbol span ($r = .49, p = .005$). Only visual digit span was significantly correlated with performance on noise-vocoded words in isolation ($r = .62, p < .001$). However, both visual span scores were correlated with performance on noise-vocoded sentences (see Table 3) with visual digit span being more strongly correlated with word recognition in sentences than symbol span.

Discussion

In the current study, we replicated all of the major findings reported by Eisenberg et al. (2002) in their pioneering research on the perception of noise-vocoded speech by children. They found that lexically easy noise-vocoded words were recognized better than lexically hard noise-vocoded words in both isolation and in sentences and that accuracy for words in vocoded sentences was substantially better than for the same words presented in isolation. Both sets of findings are consistent with the predictions made by the NAM model of spoken word recognition (Luce & Pisoni, 1998). According to NAM, spoken words are recognized relationally in the context of other phonetically-similar sounding words in lexical memory. The core processing operations underlying lexical selection and discrimination involve activation and competition of spoken words that are organized into lexical similarity neighborhoods in the mental lexicon. These findings also replicate the earlier results reported by Kirk et al. (1995) for spoken words in isolation with children and Bell and Wilson (2001) for spoken words in isolation and sentences with adults.

Chronological age was strongly related to spoken word recognition performance. Older children performed better than younger children on all noise-vocoded speech perception measures consistent with previous findings (e.g., Eisenberg, Shannon, Martinez, Wygonski, & Boothroyd, 2000; Vongpaisal, Trehub, Schellenberg, & van Lieshout, 2012). This finding is consistent with improved ability to manage challenging speech-language conditions with age, just as other areas of neurocognitive performance improve with age. It is likely that improvements in neurocognitive skills with age, particularly executive functions, contribute to the improvement observed in speech perception with age under challenging conditions.

Future research with larger samples and/or longitudinal methods should be carried out to better understand the contribution of age to executive functioning and speech perception under challenging conditions.

The second goal of the current study was to investigate the relations between auditory attention, short-term memory, and the perception of spectrally-degraded noise-vocoded speech in children with NH. Analyses revealed significant relations between performance on auditory attention and short-term memory tasks and a child's ability to recognize spoken words and sentences that were noise-vocoded to four spectral channels. These findings replicate earlier research demonstrating that noise-vocoded speech recognition not only reflects the contribution of early peripheral auditory processes related to sensory encoding and audibility, but also demonstrates the important contribution of downstream compensatory cognitive processes as well (Chatterjee et al., 2014; Conway, Bauernschmidt, Huang, & Pisoni, 2010; Davis et al., 2005; Eisenberg et al., 2002).

One unexpected finding was the absence of a consistent pattern of significant correlations between the NEPSY-2 AA and RS subtests and performance on the noise-vocoded speech recognition tasks. Only NEPSY-2 AA and performance on the isolated word recognition task were significantly correlated. One explanation of this finding is that both the AA and RS tasks were simply too easy for the children tested in this study because performance was found to be close to ceiling (i.e., children reached near to or maximum scores by making few errors) producing a restricted range in the test scores. The NEPSY-2 AA subtest is a simple auditory detection task that requires the child to sustain attention by attending to a stream of spoken words and respond to the target words. The NEPSY-2 RS subtest also measured sustained attention in addition to set shifting, a form of response inhibition. With the exception of the three children who were excluded from analyses in the AA task because they were unable to follow the instructions, most of the children performed very well on both attention tasks, with the majority of children performing at ceiling level on the AA task. As a result, the range of scores was restricted because of ceiling effects resulting in weak correlations with the noise-vocoded scores for words in isolation and words in sentences.

However, auditory attention as measured by the TD task did reveal significant correlations with all of the noise-vocoded speech recognition measures. These findings suggest that a child's ability to actively attend to and discriminate differences between two talkers based only on indexical vocal source properties of their speech is strongly associated with recognition of spectrally-degraded noise-vocoded speech. It is important to emphasize here that both the fixed- and varied-sentence conditions of the TD task were significantly correlated with all measures of noise-vocoded speech recognition. Although both conditions measure aspects of selective auditory attention and effortful cognitive control processes, each condition also has a substantial short-term immediate memory component. Performance in the fixed-sentence condition taps into verbal short-term memory because the children only have to detect a change or difference in the talker's voices between two sentences when the linguistic content remains constant within a test trial. Consistent with the findings from the digit span data, short-term memory was strongly related to a child's ability to recognize noise-vocoded speech, which may be one reason why the TD task was so strongly correlated with performance on all of the noise-vocoded speech recognition tasks.

In contrast, performance in the varied-sentence condition of the talker-discrimination task taps verbal working memory capacity because in this condition the children also have to actively inhibit and ignore differences in the linguistic content of the two different sentences while at the same time detecting similarities or differences between the talker's voices. This additional information processing component of actively controlling attentional focus in the face of distraction draws on limited processing resources that increases the cognitive processing load on the child, making the varied-talker test condition of the TD task much more difficult overall than the fixed-talker condition. In addition, because both the TD task and the sentence perception task were carried out under noise-vocoded conditions, the relations found between the two measures may also reflect individual differences in how the children encode, store and process degraded noise-vocoded speech (Johnsrude et al., 2013). Children who are better at recognizing words in noise-vocoded speech would also very likely display advantages on other processing tasks that use of noise-vocoded speech stimuli, such as TD (see Cleary & Pisoni, 2002; Cleary, Pisoni, & Kirk, 2005).

Results from the TD task also replicated findings from the earlier study of deaf children with CIs carried out by Cleary and Pisoni (2002). They found much better performance on the fixed-sentence condition than the varied-sentence condition and stronger correlations between performance on the fixed-sentence condition and several conventional speech recognition measures. A similar pattern was also found in the error rates for each condition in the present study with a much stronger response bias for committing false alarms observed during the more difficult varied-sentence condition. The present findings demonstrate that regardless of hearing status, reducing spectral information in the speech signal makes discriminating talkers more difficult, especially when the cognitive load is increased in the varied-sentence condition which requires active inhibitory control processes to ignore the downstream predictive semantic content of the sentences and consciously focus attentional control and processing resources on the indexical vocal source characteristics of the talker's voice.

Findings from the TD task and the NEPSY-2 AA task demonstrate the central role of cognition, specifically executive functions in controlling auditory attention, verbal short-term memory, and verbal working memory in perceiving spectrally-degraded noise-vocoded speech. The present findings on the control of auditory selective attention with noise-vocoded speech represent a significant contribution to the literature because they suggest that auditory attention and short-term memory processes are inseparable components of speech recognition under suboptimal listening conditions even in children with NH when the speech signal is compromised and significantly degraded using spectrally degraded vocoded signals. New knowledge about the critical role of selective attention and cognitive control processes may provide the foundations for the development of novel interventions for improving speech recognition under less than ideal listening environments. For example, if the speech signal is significantly degraded and cannot be processed or recoded into verbal short-term memory, then finding alternative ways of improving attentional control or increasing working memory capacity may provide new avenues for improving speech perception and spoken word recognition skills in low-functioning deaf children with CIs. The present findings with children with NH also suggest several new directions for future research on the role of selective attention and inhibitory control processes in speech

recognition adding additional new knowledge to the growing body of literature in the field (e.g., Chatterjee et al., 2014; Davis et al., 2005; Dorman et al., 2000; Eisenberg et al., 2002; Eisenberg et al., 2000; Kronenberger & Pisoni, 2009; Newman & Chatterjee, 2013; Shannon et al., 1995; Warner-Czyz et al., 2014).

Both measures of verbal short-term memory also showed significant relations with performance on noise-vocoded speech recognition measures. A child's ability to rapidly encode, store and reproduce temporal sequences of highly familiar items in serial order was associated with his or her ability to recognize noise-vocoded words in isolation and in sentences. Verbal short-term memory plays a central role in speech recognition because it is the active memory processing system that is used to retain the order of phonological and lexical information in working memory for language comprehension. Being able to rapidly encode, store, and retrieve item and order information about sequences of spoken words is a foundational building block for speech recognition and spoken language comprehension. This is especially important when only degraded phonological representations are available for spoken word recognition and speech understanding, as in spectrally-degraded noise-vocoded speech recognition tasks like those used in the present study.

Although the symbol span task was initially designed to measure of nonverbal short-term memory capacity, we found that symbol span scores were also correlated with performance with noise-vocoded sentence recognition yielding correlations close to those observed in the visual digit span task. Thus, a child's ability to recall and reproduce the serial order of abstract visual objects was also associated with his or her ability to recognize noise-vocoded sentences. This was an unexpected and incidental finding. We anticipated that the symbol span task would require less verbal coding and mediation than the conventional digit span task. Although the symbol span task was originally designed as a nonverbal measure of memory span, our informal observations suggest that older children actively used verbal coding strategies to carry out the symbol span task. To facilitate active verbal rehearsal, older children routinely applied common names and verbal labels to the visual displays to serve as verbal cues for rehearsal and recall. These observations suggest that the two memory span tasks may actually be tapping into similar aspects of verbal short-term memory. Taken together, these findings and informal observations suggest that children with more efficient and robust verbal coding strategies for encoding and rehearsing abstract visual information were able to apply more efficient processing strategies for recognizing noise-vocoded speech. Also, because sentences provide additional powerful contextual cues and downstream predictive support, it is possible that the ability to make use of sentence context efficiently is closely related to the ability to recognize degraded speech when weak or underspecified coarsely-coded sensory information is available in the speech signal.

The findings of this study should be viewed in the context of specific characteristics and limitations of the study design and methodology. One limitation of this study is that it is a correlational study and, therefore, it cannot provide causal explanations of the significant relations found in the results. Future research should use experimental manipulations to better understand the underlying causal factors including additional analyses that examine interactions between independent variables that can be used to compare variances accounted

for by each measure for a more thorough understanding of the cognitive contributions to spoken word recognition.

A second consideration in the interpretation of study results is the sample size and age range. While the current sample size was sufficient to replicate Eisenberg et al.'s (2002) findings and to uncover several significant relations between neurocognitive functions and noise-vocoded speech recognition, it did not allow for a robust evaluation of the significance of small effect sizes. Thus, some of the small but nonsignificant effect sizes such as correlations in the 0.20 to 0.30 range (e.g., see Table 2) should be interpreted with caution because the small sample size may have limited the statistical power of the study. Larger sample sizes in future research will also allow for more powerful statistics, such as multiple regression and mixed effects models, to examine the predictive role of neurocognitive functioning on vocoded speech perception in children while also accounting for demographic factors such as age and socioeconomic status.

A third consideration is the selection of the specific neurocognitive measures used, which were designed to evaluate several theoretically important areas related to noise-vocoded speech recognition (such as selective attention and short-term memory). These particular measures do not constitute a comprehensive or complete evaluation of the full range of potential neurocognitive functions that might impact on noise-vocoded speech perception (see Ruffin, Kronenberger, Colson, Henning, & Pisoni, 2013). Future research should investigate additional measures of auditory selective attention and inhibitory control processes and verbal short-term memory capacity. Additional methods and converging measures should address other possible neurocognitive influences such as information processing speed and efficiency.

Finally, we did not statistically correct for the number of correlations and other analyses in the paper. As a result, the experiment-wise error rate exceeded 0.05. We opted not to correct for the number of analyses because of a potential detrimental effect on statistical power. Furthermore, because our methods and measures were theoretically motivated and in most cases correlated moderately, a Bonferroni-type correction would have been overly conservative. Nevertheless, the possibility of alpha error should be considered when interpreting the present results.

In summary, the present results provide new knowledge about several previously unexplored domains of noise-vocoded speech recognition in children with NH. Our findings may be clinically significant for understanding processing of speech in suboptimal real-world settings for children with NH and for gaining additional new insights into the processing of degraded and underspecified speech signals in children with hearing loss who use cochlear implants.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Laurie Eisenberg and her colleagues for their generous assistance in the completion of this study along with Luis Hernandez for his assistance and help with instrumentation. We would also like to thank the NIH- National Institute on Deafness and Other Communication Disorders for funding this research (T32 DC000012, R01 DC000111, and R01 DC009581 grants).

Regarding, author contributions, ASR wrote the main paper and tested all participants included in analyses. Together, ASR, DBP, and WGK designed the experiments and analyzed the data. DBP and WGK provided extensive revisions to the main paper. KFF provided audiological guidance and feedback regarding creation of stimuli and experimental design.

References

- Baddeley A. Working memory: theories, models, and controversies. *Annual Review of Psychology*. 2012; 63:1–29. DOI: 10.1146/annurev-psych-120710-100422
- Baddeley A, Gathercole S, Papagno C. The phonological loop as a language learning device. *Psychological Review*. 1998; 105(1):158. doi: 10.1037/0033-295X.105.1.158 [PubMed: 9450375]
- Bates TC, D'Oliveiro L. PsyScript: A Macintosh application for scripting experiments. *Behavior Research Methods, Instruments, & Computers*. 2003; 35(4):565–576.
- Bell TS, Wilson RH. Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology*. 2001; 12(10):514–522. [PubMed: 11791938]
- Burkholder RA, Pisoni DB. Speech timing and working memory in profoundly deaf children after cochlear implantation. *Journal of Experimental Child Psychology*. 2003; 85(1):63–88. [PubMed: 12742763]
- Charles-Luce J, Luce PA. Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*. 1990; 17(01):205–215. [PubMed: 2312642]
- Chatterjee M, Zion D, Deroche ML, Burianek B, Limb C, Goren A, Kulkarni A, Christensen JA. Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hearing Research*. 2014; 322:151–162. [PubMed: 25448167]
- Cleary M, Pisoni DB. Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Annals of Otology, Rhinology, & Laryngology*. 2002; 111(5 Part 2):113–118.
- Cleary M, Pisoni DB, Kirk KI. Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*. 2005; 48(1):204–223.
- Conway CM, Bauernschmidt A, Huang SS, Pisoni DB. Implicit statistical learning in language processing: Word predictability is the key. *Cognition*. 2010; 114(3):356–371. [PubMed: 19922909]
- Conway CM, Pisoni DB, Anaya EM, Karpicke J, Henning SC. Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*. 2011; 14(1):69–82. [PubMed: 21159089]
- Conway CM, Deocampo J, Walk AM, Anaya EM, Pisoni DB. Deaf children with cochlear implants do not appear to use sentence context to help recognize spoken words. *Journal of Speech, Language, and Hearing Research*. 2014; 57:2174–2190.
- Cowan, N. *Attention and memory: An integrated framework*. Oxford Psychology Series; New York: 1995.
- Cowan N. What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*. 2008; 169:323–338. [PubMed: 18394484]
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*. 2005; 134(2):222. [PubMed: 15869347]
- Dawson P, Busby P, McKay C, Clark GM. Short-term auditory memory in children using cochlear implants and its relevance to receptive language. *Journal of Speech, Language, and Hearing Research*. 2002; 45(4):789–801.

- Dirks DD, Takayana S, Moshfegh A. Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*. 2001; 12(5):233–244. [PubMed: 11392435]
- Dirks DD, Takayanagi S, Moshfegh A, Noffsinger PD, Fausti SA. Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing*. 2001; 22(1):1–13. [PubMed: 11271971]
- Dorman MF, Loizou PC, Kemp LL, Kirk KI. Word recognition by children listening to speech processed into a small number of channels: data from normal-hearing children and children with cochlear implants. *Ear and Hearing*. 2000; 21(6):590–596. [PubMed: 11132785]
- Dunn, LM., Dunn, DM. Peabody Picture Vocabulary Test, Fourth Edition Manual. Minneapolis, MN: Pearson, Inc; 2007.
- Eisenberg LS, Martinez AS, Holowecky SR, Pogorelsky S. Recognition of lexically controlled words and sentences by children with normal hearing and children with cochlear implants. *Ear and Hearing*. 2002; 23(5):450–462. [PubMed: 12411778]
- Eisenberg LS, Shannon RV, Martinez AS, Wygonski J, Boothroyd A. Speech recognition with reduced spectral cues as a function of age. *The Journal of the Acoustical Society of America*. 2000; 107(5):2704–2710. [PubMed: 10830392]
- Fagan MK, Pisoni DB, Horn DL, Dillon CM. Neuropsychological correlates of vocabulary, reading, and working memory in deaf children with cochlear implants. *Journal of Deaf Studies and Deaf Education*. 2007; 12(4):461–471. [PubMed: 17556732]
- Frankish, C. Auditory short-term memory and the perception of speech. In: Gathercole, SE., editor. *Models of short-term memory*. Psychology Press; 1996. p. 179-207.
- Gathercole SE, Service E, Hitch GJ, Adams AM, Martin AJ. Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology*. 1999; 13(1):65–77.
- Gilley PM, Sharma A, Dorman MF. Cortical reorganization in children with cochlear implants. *Brain Research*. 2008; 1239:56–65. [PubMed: 18775684]
- Harris MS, Kronenberger WG, Gao S, Hoen HM, Miyamoto RT, Pisoni DB. Verbal short-term memory development and spoken language outcomes in deaf children with cochlear implants. *Ear and Hearing*. 2013; 34(2):179–192. [PubMed: 23000801]
- Horn DL, Davis RA, Pisoni DB, Miyamoto RT. Development of visual attention skills in prelingually deaf children who use cochlear implants. *Ear and Hearing*. 2005; 26(4):389–408. [PubMed: 16079634]
- Houston DM. Attention to speech sounds in normal-hearing and deaf children with cochlear implants. *The Journal of the Acoustical Society of America*. 2009; 125(4):2534–2534.
- Hula, WD., McNeil, MR. Models of attention and dual-task performance as explanatory constructs in aphasia. Paper presented at the Seminars in speech and language; 2008.
- Jacquemot C, Scott SK. What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*. 2006; 10(11):480–486. [PubMed: 16997610]
- Johnson C, Goswami U. Phonological awareness, vocabulary, and reading in deaf children with cochlear implants. *Journal of Speech, Language, and Hearing Research*. 2010; 53(2):237–261.
- Johnsrude IS, Mackey A, Hakyemez H, Alexander E, Trang HP, Carlyon RP. Swinging at a cocktail party voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*. 2013; 20:27–40.
- Jusczyk, PW. *The discovery of spoken language*. MIT Press; 1997.
- Karl JR, Pisoni DB. The role of talker-specific information in memory for spoken sentences. *The Journal of the Acoustical Society of America*. 1994; 95(5):2873–2873.
- Kirk KI, Pisoni DB, Osberger MJ. Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing*. 1995; 16(5):470–481. [PubMed: 8654902]
- Kirk KI, Sehgal ST, Hay-McCutcheon M. Comparison of children's familiarity with tokens on the PBK, LNT, and MLNT. *The Annals of otology, rhinology & laryngology Supplement*. 2000; 185:63–64.
- Korkman, M., Kirk, U., Kemp, S. *NEPSY-II Clinical and Interpretive Manual*. San Antonio, TX: PsychCorp; 2007.

- Kral A, Eggermont JJ. What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Research Reviews*. 2007; 56(1):259–269. [PubMed: 17950463]
- Kral A, Kronenberger WG, Pisoni DB, O'Donoghue GM. Neurocognitive factors in sensory restoration of early deafness: a connectome model. *The Lancet Neurology*. 2016; 15(6):610–621. [PubMed: 26976647]
- Kronenberger WG, Colson BG, Henning SC, Pisoni DB. Executive Functioning and Speech-Language Skills Following Long-Term Use of Cochlear Implants. *Journal of deaf studies and deaf education*. 2014; :456–470. DOI: 10.1093/deafed/enu011 [PubMed: 24903605]
- Kronenberger, WG., Pisoni, DB. Measuring learning-related executive functioning: Development of the LEAF scale. Paper presented at the 117th Annual Convention of the American Psychological Association; Toronto, Canada. 2009.
- Lewellen MJ, Goldinger SD, Pisoni DB, Greene BG. Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*. 1993; 122(3):316–330. [PubMed: 8371087]
- Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*. 1998; 19(1):1–36. [PubMed: 9504270]
- Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*. 1951; 41(5):329–335. [PubMed: 14861384]
- Newman R, Chatterjee M. Toddlers' recognition of noise-vocoded speech. *The Journal of the Acoustical Society of America*. 2013; 133(1):483–494. [PubMed: 23297920]
- Niparko, JK. Cochlear implants: Principles & practices. Lippincott Williams & Wilkins; 2009.
- Nittrouer, S. Early development of children with hearing loss. Plural Pub; 2010.
- Norman, DA., Shallice, T. Attention to action. Springer; 1986.
- Nusbaum HC, Pisoni DB, Davis CK. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on speech perception progress report*. 1984; 10(10):357–376.
- Osman H, Sullivan JR. Children's auditory working memory performance in degraded listening conditions. *Journal of Speech, Language, and Hearing Research*. 2014;1–9. [PubMed: 24576834]
- Peterson CC. Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry*. 2004; 45(6):1096–1106. [PubMed: 15257666]
- Pisoni DB. Auditory short-term memory and vowel perception. *Memory & Cognition*. 1975; 3(1):7–18. [PubMed: 24203819]
- Pisoni, DB. Some thoughts on “normalization” in speech perception. In: Johnson, K., Mullenix, JW., editors. *Talker variability in speech processing*. San Diego: Academic Press; 1997. p. 9–32.
- Pisoni DB, Cleary M. Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear and Hearing*. 2003; 24(1 Suppl):106S–120S. [PubMed: 12612485]
- Pisoni DB, Geers AE. Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing. *The Annals of Otolaryngology & Laryngology Supplement*. 2000; 185:92–93.
- Pisoni DB, Kronenberger WG, Roman AS, Geers AE. Measures of digit span and verbal rehearsal speed in deaf children after more than 10 years of cochlear implantation. *Ear and Hearing*. 2011; 32(1):60S–74S. DOI: 10.1097/AUD.0b013e3181ffd58e [PubMed: 21832890]
- Posner MI, Snyder CR. Attention and Cognitive Control. *Cognitive psychology: Key readings*. 2004:205–223.
- Posner MI, Snyder CR, Davidson BJ. Attention and the detection of signals. *Journal of Experimental Psychology: General*. 1980; 109(2):160–174.
- Quittner AL, Smith LB, Osberger MJ, Mitchell TV, Katz DB. The impact of audition on the development of visual attention. *Psychological Science*. 1994; 5(6):347–353.
- Richardson JT. Measures of short-term memory: a historical review. *Cortex*. 2007; 43(5):635–650. [PubMed: 17715798]
- Rönnerberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, Dahlström Ö, Signoret C, Stefan S, Pichora-Fuller K, Rudner M. The Ease of Language Understanding (ELU) model:

- theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*. 2013; 7(31):1–17. [PubMed: 23420631]
- Rönnerberg J, Rudner M, Foo C, Lunner T. Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*. 2008; 47(S2):S99–S105. [PubMed: 19012117]
- Ross, M., Lerman, J., Cienkowski, KM. *Word Intelligibility by Picture Identification-WIPI*, 2nd Edition. St. Louis, MO: Auditec; 2004.
- Ruffin CV, Kronenberger WG, Colson BG, Henning SC, Pisoni DB. Long-term speech and language outcomes in prelingually deaf children, adolescents and young adults who received cochlear implants in childhood. *Audiology and Neurotology*. 2013; 18(5):289–296. [PubMed: 23988907]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. DOI: 10.1126/science.270.5234.303 [PubMed: 7569981]
- Smith LB, Quittner AL, Osberger MJ, Miyamoto R. Audition and visual attention: the developmental trajectory in deaf and hearing populations. *Developmental Psychology*. 1998; 34(5):840–850. [PubMed: 9779732]
- Stanovich KE. Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading research quarterly*. 1980; 16(1):32–71.
- Tharpe AM, Ashmead DH, Rothpletz AM. Visual attention in children with normal hearing, children with hearing aids, and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*. 2002; 45(2):403–413.
- Unsworth N, Engle RW. On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*. 2007; 133(6): 1038–1066. DOI: 10.1037/0033-2909.133.6.1038 [PubMed: 17967093]
- van Heugten M, Volkova A, Trehub SE, Schellenberg EG. Children's recognition of spectrally degraded cartoon voices. *Ear and Hearing*. 2013; 35(1):118–125. DOI: 10.1097/AUD.0b013e3182a468d0
- Vongpaisal T, Trehub SE, Schellenberg EG, van Lieshout P. Age-related changes in talker recognition with reduced spectral cues. *The Journal of the Acoustical Society of America*. 2012; 131(1):501–508. [PubMed: 22280611]
- Waltzman, SB., Roland, JT. *Cochlear implants*. Thieme; 2006.
- Warner-Czyz AD, Houston DM, Hynan LS. Vowel discrimination by hearing infants as a function of number of spectral channels. *The Journal of the Acoustical Society of America*. 2014; 135(5): 3017–3024. [PubMed: 24815281]
- Wechsler, D. *Wechsler Intelligence Scale for Children*. 4. San Antonio, TX: The Psychological Corporation; 2003.
- Wightman FL, Kistler DJ. Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*. 2005; 118(5):3164–3176. [PubMed: 16334898]
- Williams, KT. *Expressive Vocabulary Test, Second Edition, Manual*. Bloomington, MN: Pearson, Inc; 2007.
- Zekveld AA, Kramer SE, Festen JM. Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*. 2011; 32(4):498–510. [PubMed: 21233711]
- Zeng, F.-G., Popper, AN., Fay, RR. *Cochlear implants: Auditory prostheses and electric hearing*. Vol. 20. Springer; 2004.

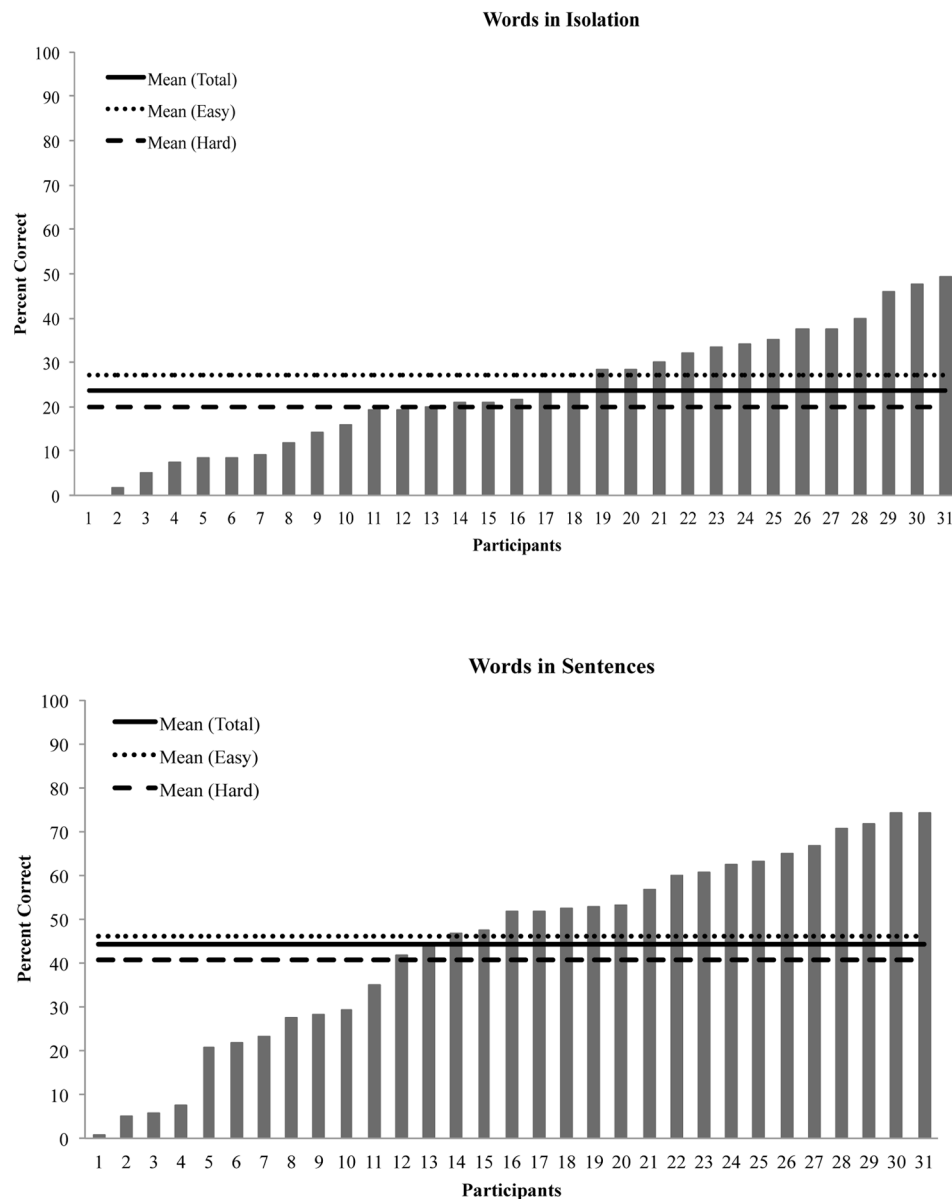
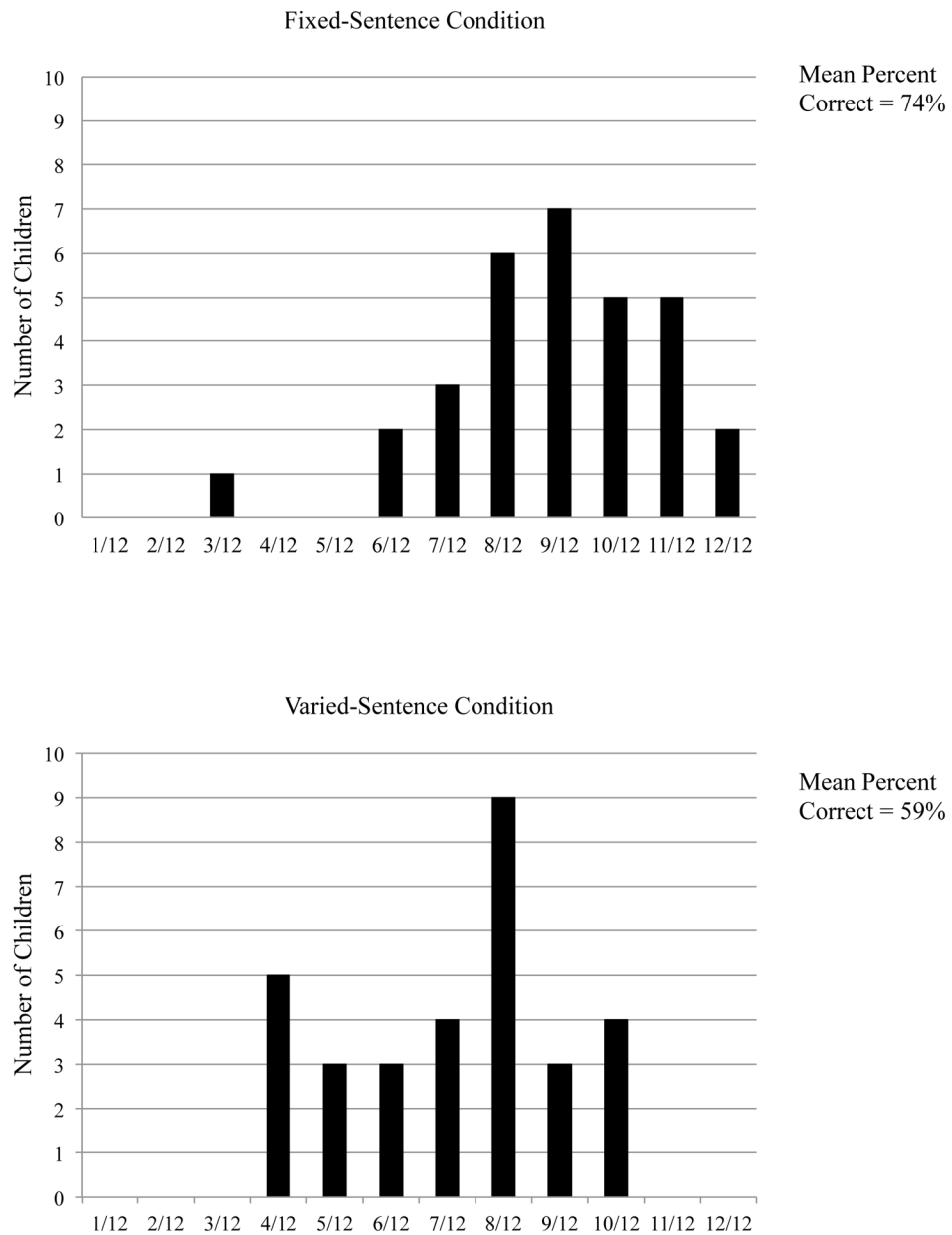


Figure 1. Individual scores for words presented in isolation (top panel) and words in sentences (bottom panel). Children are arranged in ascending order based on performance. The solid line represents the mean score of the group for all words presented (total). The dotted line displays the mean performance for lexically easy words. The dashed line represents the mean performance for lexically hard words. Performance is scored as percent correct.

**Figure 2.**

TD scores displayed as proportion of correct responses out of the 12 test trials. Top panel displays a histogram of the number of correct trials for the fixed-sentence condition; bottom panel displays a histogram of the number of correct trials for the varied-sentence condition.

Table 1Descriptive Statistics for Children's' Performance on Speech Perception and Neurocognitive Measures (*N*=31)

Measure	<i>M</i>	<i>SD</i>	Range
Word Familiarity Rating Scale	5.04	0.82	3.46 – 6.23
PPVT LQ	1.26	0.23	0.90 – 1.93
EVT LQ	1.29	0.32	0.86 – 1.94
ELCW- All words	23.50	13.60	0 – 49.17
ELCW- Easy Words	27.14	15.57	0 – 60
ELCW- Hard Words	19.83	12.24	0 – 48.3
ELCS- All Sentences	44.29	21.99	0.83 – 74.17
ELCS- Easy Sentences	46.12	23.74	0 – 80
ELCS- Hard Sentences	40.72	20.62	1.67 – 80
NEPSY-2 AA **	28.78	1.50	24 – 30
NEPSY-2 RS **	32.82	3.45	21 – 36
TD- Fixed Sentences	74.18	16.30	25 – 100
TD- Varied Sentences	59.15	16.43	33.30 – 83.33
Visual Digit Span	30.29	15.25	1 – 70
Symbol Span	14.65	11.35	0 – 44

* Note. Speech perception measures reflected percent accuracy. Language quotients (LQ) were derived by taking the age-equivalency scores and dividing them by the chronological age. NEPSY-2 AA and RS scores reflected raw scores. TD= Talker Discrimination Task. ELCW = Eisenberg Lexically Controlled Words. ELCS = Eisenberg Lexically Controlled Sentences.

**
N=28

Table 2

Correlational Coefficients for PPVT-4 and EVT-2 Language Quotients (LQ) and Noise-Vocoded Speech Perception Scores ($N=31$)

Speech Perception Measures	PPVT-4 LQ	EVT-2 LQ
Words in Isolation	-.00	.20
Easy Words	.01	.23
Hard Words	.06	.24
Sentences (Keywords)	.19	.27
Easy Sentences	.24	.37*
Hard Sentences	.13	.09

Note. Speech perception measures reflected percent accuracy. Language quotients were derived by taking the age-equivalency scores and dividing them by the chronological age

*
p<.05.

**
p<.01.

p<.001 (two-tailed)

Table 3

Correlational Coefficients for Auditory Attention and Short-Term Memory Measures and Noise-Vocoded Speech Recognition Scores

Neurocognitive Measure		Words in Isolation	Sentences
Auditory Attention	NEPSY AA	.40 *	.34
	NEPSY RS	.00	.16
	TD (Fixed)	.49 **	.51 **
	TD (Varied)	.39 *	.40 *
Short-Term Memory	Digit Span (Forward)	.62 **	.43 *
	Symbol Span (Forward)	.28	.36 *

Note. Speech perception measures reflected percent accuracy. Auditory Attention (AA) and Response Set (RS) reflect raw scores. Talker discrimination (TD) scores reflect percent correct. Span measures scores reflect points per correct response.

*
p<.05.

**
p<.01.

p<.001 (two-tailed)