Some Factors Underlying Individual Differences in Speech Recognition on PRESTO: A First Report

Terrin N. Tamati[1], **Jaimie L. Gilbert**[2], and **David B. Pisoni**[1,2,3]

[1]Department of Linguistics, Indiana University, Bloomington, Memorial Hall 322, Bloomington, IN 47405

[2]Department of Psychological and Brain Sciences, Indiana University, Bloomington, 1101 East Tenth Street, Bloomington IN 47405

[3]Department of Otolaryngology – Head and Neck Surgery, Indiana University School of Medicine, 702 Barnhill Drive, Suite 0860, Indianapolis, IN 46202

Abstract

Background—Previous studies investigating speech recognition in adverse listening conditions have found extensive variability among individual listeners. However, little is currently known about the core, underlying factors that influence speech recognition abilities.

Purpose—To investigate sensory, perceptual, and neurocognitive differences between good and poor listeners on PRESTO, a new high-variability sentence recognition test under adverse listening conditions.

Research Design—Participants who fell in the upper quartile (HiPRESTO listeners) or lower quartile (LoPRESTO listeners) on key word recognition on sentences from PRESTO in multitalker babble completed a battery of behavioral tasks and self-report questionnaires designed to investigate real-world hearing difficulties, indexical processing skills, and neurocognitive abilities.

Study Sample—Young, normal-hearing adults (N = 40) from the Indiana University community participated in the current study.

Data Collection and Analysis—Participants' assessment of their own real-world hearing difficulties was measured with a self-report questionnaire on situational hearing and hearing health history. Indexical processing skills were assessed using a talker discrimination task, a gender discrimination task, and a forced-choice regional dialect categorization task. Neurocognitive abilities were measured with the Auditory Digit Span Forward (verbal short-term memory) and Digit Span Backward (verbal working memory) tests, the Stroop Color and Word Test (attention/inhibition), the WordFam word familiarity test (vocabulary size), the BRIEF-A self-report questionnaire on executive function, and two performance subtests of the WASI Performance IQ (non-verbal intelligence). Scores on self-report questionnaires and behavioral tasks were tallied and analyzed by listener group (HiPRESTO and LoPRESTO).

Results—The extreme groups did not differ overall on self-reported hearing difficulties in real-world listening environments. However, an item-by-item analysis of questions revealed that LoPRESTO listeners reported significantly greater difficulty understanding speakers in a public place. HiPRESTO listeners were significantly more accurate than LoPRESTO listeners at gender discrimination and regional dialect categorization, but they did not differ on talker discrimination

Address Correspondence to: Terrin N. Tamati, Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405, Phone: (812) 855-6204; Fax: 812-855-1300; ttamati@indiana.edu.

accuracy or response time, or gender discrimination response time. HiPRESTO listeners also had longer forward and backward digit spans, higher word familiarity ratings on the WordFam test, and lower (better) scores for three individual items on the BRIEF-A questionnaire related to cognitive load. The two groups did not differ on the Stroop Color and Word Test or either of the WASI performance IQ subtests.

Conclusions—HiPRESTO listeners and LoPRESTO listeners differed in indexical processing abilities, short-term and working memory capacity, vocabulary size, and some domains of executive functioning. These findings suggest that individual differences in the ability to encode and maintain highly detailed episodic information in speech may underlie the variability observed in speech recognition performance in adverse listening conditions using high-variability PRESTO sentences in multitalker babble.

Suggested Keywords

Speech Recognition; Adverse Conditions; Individual Differences

---

Introduction

Listeners encounter enormous amounts of variability in speech in everyday listening conditions. Successful understanding of the intended message of an utterance entails the ability to rapidly adapt to and use variation in the talker's individual idiolectal characteristics (indexical properties), background noise and/or competition, and linguistic knowledge in long-term memory. However, individual listeners vary a great deal in their ability to perceive speech in adverse listening conditions (Mattys et al, 2012). A consistent challenge in the field of spoken word recognition is identifying the underlying sources of individual differences in speech recognition skills. Historically, research has focused on identifying properties of an idealized talker and listener, in quiet listening and speaking environments of the lab or clinic. However, this approach ignores the rich variability in real-world speech communication environments and diversity among speakers and listeners.

One of the most basic features of everyday real-world speech communication is that individuals interact with a variety of different people in different environments and contexts. Beyond the linguistic information, such as speech sounds, syllables, and words, the speech signal also encodes nonlinguistic indexical information about the talker's voice (Kreiman and Van Lancker Sidtis, 2011). Indexical variability related to talker-specific and group features, such as age, gender, regional dialect, and native language (Abercrombie, 1967), has been shown to play an important role in speech perception processes (e.g., Johnson and Mullennix, 1997; Pisoni, 1997; Cleary et al, 2005). The listener's task, then, is to simultaneously process both linguistic and indexical information in the speech signal to uncover both the intended meaning of the utterance and how the utterance was said and by whom (e.g., Nygaard, 2008). The indexical information in the speech signal is implicitly encoded into memory without conscious awareness or attention (e.g., Palmeri et al, 1993; Nygaard et al, 1995), which allows listeners to be able to make use of this additional source of information to identify a familiar talker (e.g., Van Lancker et al, 1985a; Van Lancker et al, 1985b; Kreiman and Van Lancker Sidtis, 2011), make an explicit judgment about a talker's gender (e.g., Lass et al, 1976) or age (e.g., Ptacek and Sander, 1966), or even infer the talker's region of origin (e.g., Van Bezooijen and Gooskens, 1999; Clopper and Pisoni, 2004). Furthermore, listeners benefit from learning talker-specific indexical information in speech, which has been shown to facilitate successful speech recognition under adverse listening conditions (e.g., Nygaard et al, 1994).

However, indexical variability may also present a challenge to listeners who have difficulty adjusting to this type of information or in adapting to a novel talker's voice or speaking

style. Speech from non-native talkers is less intelligible than speech from native talkers (e.g., Bradlow and Pisoni, 1999). Similarly, talkers from an unfamiliar or marked dialect region tend to be more difficult to recognize and understand (e.g., Mason, 1946; Labov and Ash, 1997; Clopper and Bradlow, 2008). Even within a particular language or regional dialect, individual talkers appear to vary substantially in their inherent intelligibility (Bradlow et al, 1996). Beyond individual talker characteristics, spoken word recognition is more difficult when the task involves recognition of speech from multiple talkers as opposed to only a single talker (Peters, 1955; Creelman, 1957; Mullennix et al, 1989). A multiple-talker environment requires the listener to repeatedly adjust and adapt to the indexical properties of each new voice, as opposed to a single repeated talker environment where the listener can more easily learn talker-contingent details, which improves word recognition in noise (Nygaard et al, 1994; Nygaard and Pisoni, 1998). Taken together, these studies demonstrate that real-world natural speech variability contributes to more challenging environments for the listener.

Background noise and/or competition from other talkers also present significant challenges to the listener. Competing speech, in particular, creates considerable difficulty for the listener (Carhart et al, 1969; Carhart et al, 1975), because they generate both energetic and informational masking (Van Engen and Bradlow, 2007; Helfer and Freyman, 2009). Additionally, previous research has shown that the type and level of background competition can interact with the indexical properties of the target signal. The voice characteristics of the target signal and competing talker influence the intelligibility of the test sentences, with more similar talkers resulting in poorer speech intelligibility (Brungart et al, 2001). In addition, the regional dialect of the talker has also been shown to reduce intelligibility to a greater extent at more challenging signal-to-noise ratios (SNRs) (Clopper and Bradlow, 2008). Thus, target variability and background competition, both independently and together, influence the listening environment and create more difficult conditions for speech recognition.

Individual listeners also vary substantially in their ability to rapidly adapt and adjust to adverse listening conditions. To study how these factors influence speech recognition performance, Gilbert et al (in press) recently demonstrated that young, normal-hearing listeners showed a large range of performance on a sentence recognition task using high-variability sentence materials from PRESTO, a new sentence recognition test. PRESTO (Perceptually Robust English Sentence Test Open-set) was developed to investigate the effects of variability in speech and individual listener differences in speech recognition. PRESTO contains variability in talker, gender, and regional dialect, as well as syntactic structure and lexical properties (e.g., frequency). PRESTO test materials consist of sentence-length utterances from the TIMIT (Texas Instruments/MIT Speech Corpus) database (Garofolo et al, 1993). Originally, Felty (2008) used sentences from the TIMIT database to construct PRESTO sentences lists, each of which contained 18 unique utterances produced by 18 different talkers, equal numbers of key words (76 key words per list), male and female talkers (9 male and 9 female talkers per list), average word familiarity ($M = 6.9$), and average log word frequency ($M = 2.5$) (Nusbaum et al, 1984). Additionally, talkers were selected from various regions of the United States and each test list included talkers from at least five different dialect regions of the United States. No sentence was ever repeated across lists.

Gilbert et al (in press) assessed the speech recognition abilities of 121 normal-hearing, native speakers of American English using 180 PRESTO sentences (10 lists) in multitalker babble at four different SNRs, +3 dB, 0 dB, −3 dB, and −5 dB SNR. They found a wide range of performance on the PRESTO sentences, on average and across all four SNRs. Performance on the initial test and a re-test with new sentences (N = 40) was strongly

correlated, displaying excellent test/re-test reliability. Normal-hearing, young adult listeners' overall accuracy on the high-variability PRESTO ranged from approximately 40% to 76% accuracy (N = 121). Clearly, some listeners performed better on PRESTO lists than others, despite hearing within normal limits.

Other studies examining speech recognition in different types of listening conditions have also reported individual variability in performance. Recent studies have shown that some listeners are more susceptible to competing acoustic signals than other listeners (Neff and Dethlefs, 1995; Richards and Zeng, 2001; Wightman et al, 2010). Additionally, although not generally a topic of primary interest to hearing scientists, individual listener variability is also present in studies on speech recognition of unfamiliar regional dialects (e.g., Mason, 1946) or foreign accented speech (e.g., Bent and Bradlow, 2003).

Taken together, the recent findings of Gilbert et al (in press), and others, demonstrate the important contribution of adverse listening conditions in the study of speech perception. These studies also suggest a pressing need to investigate how and why speech recognition performance varies across individual listeners. Indeed, adverse listening environments are ideal for studying individual variability in speech perception processes, because most individuals do not tend to vary as greatly under idealized speech recognition conditions with little talker and linguistic variability (e.g., Clopper and Bradlow, 2008). These dissociations have been previously observed in the field of clinical audiology, where one foundational problem in assessing speech recognition ability is that different speech recognition tests are not equally sensitive to individual differences (Wilson et al, 2001). In particular, some conventional low-variability speech recognition tests have been shown to be less sensitive to individual variability in listener performance than high-variability tests (Gifford et al, 2008; Park et al, 2010).

However, little is currently known about the core, underlying factors contributing to the individual variability observed in speech perception tasks. Some earlier research focusing on group characteristics has demonstrated that a listener's unique linguistic background and development influences his or her ability to understand speech in adverse conditions. Previous research has found that listeners who have had more experience communicating with deaf talkers are better at understanding speech from deaf children than inexperienced listeners (e.g., McGarr, 1983). Greater familiarity with regional dialects results in better speech recognition performance, such that familiar local or supraregional dialect varieties are more intelligible (e.g., Mason, 1946; Labov and Ash, 1997; Clopper and Bradlow, 2008), although the effect of talker dialect on the perception of regional dialects appears to be modulated by the presence and level of background noise (Clopper and Bradlow, 2008). Similarly, in the perception of foreign-accented speech, several studies have suggested that sharing a native language with the talker benefits the listener in speech intelligibility tasks even when the talker is speaking in the second language (Bent and Bradlow, 2003). Taken together, these results suggest that some aspects of a listener's background and linguistic experience may result in better individual performance in speech recognition tasks including accented speech (i.e., target variability).

Beyond group differences, differences in elementary neurocognitive abilities may also influence variability in speech perception in adverse listening conditions. Research has suggested that cognitive operations involving cognitive control and executive function are central to speech perception processes, and may underlie differences in speech perception, especially in noise and under other adverse listening conditions (see for example, Pisoni, 2000; Arlinger et al, 2009; Stenfelt and Rönnberg, 2009; Mattys et al, 2012). Indeed, several current models of speech perception have focused on the interaction of speech perception and neurocognitive processes to explain and predict speech recognition in hearing-impaired

listeners (e.g., Akeroyd, 2008), in the aging population (e.g., Schneider et al, 2005; Pichora-Fuller and Singh, 2006; Arlinger et al, 2009), and in normal-hearing young adults (e.g., Francis and Nusbaum, 2009).

Although few studies have focused specifically on individual differences in speech perception in adverse listening conditions, several studies have investigated the contributions of neurocognitive abilities to variability in speech recognition performance. Research in child language development has provided new insights about individual differences in speech perception and language processing. Recent findings have shown that spoken word recognition skills of pediatric cochlear implant (CI) users are related to individual differences in working memory, specifically, working memory capacity and verbal rehearsal (Pisoni and Geers, 2000; Cleary et al, 2002). Pisoni and Cleary (2003) found spoken word recognition scores of pediatric CI users were related to forward and backward digit span and speaking rate. These measures were also strongly related to speech and language outcome measures of the same children ten years later in life as teenagers (Pisoni et al, 2011). Beer et al (2011) also found that the working memory domain of executive function (as measured by the BRIEF, a parent report questionnaire of executive function in children) was significantly correlated with the CI children's speech perception in noise scores, suggesting links between spoken language processing and executive control.

Some studies on adults have also suggested that individual differences in neurocognitive abilities affect speech recognition performance. Examining the influence of attentional abilities on speech recognition in an adverse condition, Jesse and Janse (2012) measured the benefit older and younger listeners received when they were able to see the target speaker in a speech recognition task in single-talker noise. Although listeners benefited overall from seeing the speaker, the authors also found that the size of the response time benefit in a phoneme-monitoring task was predicted by differences in attentional abilities in both the older and younger listener groups. Conway et al (2010) investigated the relationship between implicit learning and speech perception and showed that the implicit learning abilities of individual listeners in learning visual sequences was related to the listeners' ability to use knowledge of word predictability to understand sentences under degraded listening conditions. Several recent studies with older hearing-impaired adults have also shown that neurocognitive abilities play a crucial factor in speech recognition difficulties in the aging population, especially when auditory factors have been taken into account (e.g., van Rooij and Plomp, 1990; Gordon-Salant and Fitzgibbons, 1997; Humes, 2007; Akeroyd, 2008). Taken together, studies on individual differences in hearing-impaired children and adult speech recognition abilities establish a close link between speech perception and cognition, and also provide converging evidence that individual differences in speech perception in adverse listening conditions may be related to these underlying, core neurocognitive processes.

The Current Study

Despite the recent studies investigating speech perception in adverse listening conditions, little is currently known about the core, underlying factors that contribute to individual differences in speech recognition in highly variable listening conditions. In order to fill this gap in our basic knowledge, the current study used an extreme groups design to investigate differences in speech recognition between good and poor listeners in adverse listening conditions.

Listeners from Gilbert et al (in press) who fell in the upper or lower quartile of the original distribution of 121 normal-hearing, native listeners' performance on the PRESTO sentences in multitalker babble were invited to return to the laboratory for an additional battery of

tests. Performance on the PRESTO sentences was used to assign listeners to the extreme groups in the study because the PRESTO sentences contained several sources of linguistic and indexical variability and showed a wide range of scores when tested in multitalker babble. Although the use of an extreme groups design has some limitations, such as the loss of information (and individual variability) from removing a large portion of the distribution of data, or the potential to overestimate the relationships between variables (see, for example, Humphreys, 1978; Cohen, 1983; Preacher et al, 2005), this methodology can be a useful heuristic for establishing that individual differences are real and meaningful, and for determining the existence of relationships between experimental variables (for some discussion of the appropriate uses of the extreme groups approach see, for example, Conway et al, 2005; Preacher et al, 2005).

The current study was carried out to uncover relationships between speech recognition in adverse listening conditions and several other measures. One benefit of using an extreme groups design is that it gives the experimenter control over performance on the issue under investigation in the study, namely, the ability to understand speech in adverse listening conditions. It also provides an efficient and effective methodology to begin exploring the associated skills or abilities related to speech perception in adverse listening conditions. The two extreme groups used in this study were formed uniquely based on their speech recognition abilities on PRESTO in multitalker babble. Evaluating the two groups on various tasks allows for significance testing on differences in performance between the two groups, in addition to an evaluation of the association of performance between tasks. Good and poor listeners completed several self-report questionnaires on everyday activities and behaviors, indexical processing tasks, and neurocognitive assessments. These tests were selected for the current study to help uncover some of the factors that may underlie individual differences in speech recognition in high-variability listening conditions. In particular, the present study sought to evaluate: 1) perceived hearing difficulties in real-world environments, 2) associated indexical processing abilities, and 3) neurocognitive abilities of listeners who are good (HiPRESTO) or poor (LoPRESTO) at speech understanding in adverse listening conditions.

Real-World Hearing

Differences in hearing difficulties of the two groups in everyday real-world listening environments were assessed with a self-report questionnaire on situational hearing and hearing health history. The Hearing Health Quick Test (HHQT) developed by the American Academy of Audiology, which for the purposes of the current study was later divided into a Situational Hearing Scale and Hearing Health Scale, was used to obtain a measure of the participants' subjective perception of any hearing difficulties in everyday listening environments (Situational Hearing Scale) and any health-related risk factors for hearing problems (Hearing Health Scale). Although all participants from both groups passed a hearing screener, the LoPRESTO group performed more poorly on the PRESTO test. If the ability to understand the sentences on PRESTO is representative of real-world everyday listening conditions, the LoPRESTO group may also display difficulties in daily life. These problems would emerge in the responses on the Situational Hearing Scale, which assesses perceived hearing difficulties in different day-to-day situations. Thus, we predicted that the LoPRESTO group would report more difficulties on the Situational Hearing Scale, but would not show any specific hearing health related risks on the Hearing Health Scale.

Indexical Processing Abilities

Indexical processing abilities were assessed with three indexical processing tasks: a gender discrimination task, a talker discrimination task, and a forced-choice regional dialect categorization task. Each of these tasks focuses on a key component of the PRESTO

sentences, which contain multiple male and female talkers from different regional dialect backgrounds. Because successful sentence recognition requires rapidly adjusting to these sources of variability, we predicted that listeners who are able to encode more detailed indexical information in speech would show an advantage in recognizing words in the PRESTO sentences in multitalker babble. This ability would also be expected to emerge on the three indexical processing tasks in the form of more accurate and/or faster gender discrimination and talker discrimination, and in more accurate identification of the region of origin of unfamiliar talkers. Although none of the listeners were expected to have difficulty perceiving acoustic-phonetic differences among different talkers of different genders or regions of origin, listeners who are able to encode more detailed indexical information (related to talker, gender, or regional dialect) were expected to display an advantage at processing the indexical information in these tasks. However, given that discrimination should not be very difficult for any of the participants (LoPRESTO or HiPRESTO), it is possible that the better indexical processing abilities would either surface in accuracy or speed, and not both. In other words, a speed-accuracy trade-off would be greater for those listeners who displayed more difficulty processing indexical information. Thus, better indexical processing abilities would benefit the listeners in recognizing the PRESTO sentences. Therefore, we expected HiPRESTO listeners to perform more accurately, and faster, on the indexical processing tasks, especially the dialect categorization task.

Neurocognitive Processing Skills

The current study also assessed the neurocognitive functioning of both groups of listeners. Short-term and working memory capacity, attention/inhibition, vocabulary size, executive functioning, and non-verbal IQ were measured for both groups of listeners. Specific predictions about the contributions of these processes to speech perception in adverse listening conditions are as described below.

The ability to understand speech in highly variable listening conditions requires the listener to devote considerable processing resources to encode highly detailed information in the speech signal (e.g., Martin et al, 1989; Mullennix et al, 1989; Mullennix and Pisoni, 1990). Additionally, energetic and informational masking from competing background talkers could result in substantial acoustic-phonetic ambiguity, which would require holding and manipulating the available speech information in verbal short-term memory (e.g., Cowan, 1988, 1997; Gathercole and Baddeley, 1993; Nusbaum and Magnuson, 1997; Baddeley et al, 1998; Baddeley, 2003). Listeners who are able to allocate more resources to this highly demanding process may show better abilities to use the available acoustic-phonetic and lexical information in the signal, along with context cues in adverse listening conditions. In the current study, the contributions of short-term and working memory were assessed using the auditory digit span forward and backward tests. Because HiPRESTO listeners showed better performance in recognizing words in the PRESTO sentences in babble, we predicted they would have greater short-term and working memory capacity than LoPRESTO listeners.

Understanding the target talkers against competing talkers at various signal-to-babble levels requires listeners to be able to simultaneously focus and attend to the target talker and content, while inhibiting competition from the background talkers and content (e.g., Cherry, 1953; Alain and Arnott, 2000; Brungart et al, 2001). More flexible attentional and inhibitory abilities may help listeners in this difficult task. The current study assessed attention and inhibition skills in both groups of listeners. Given that the PRESTO sentences in multitalker babble over several SNRs presents a challenging listening environment to the listeners, we predicted that better attentional and inhibitory skills would enable listeners to reliably attend selectively to the linguistic and indexical information in the target utterance and quickly adjust to new talkers and SNRs from trial to trial in a test list. Given that HiPRESTO

listeners were better at recognizing speech in adverse listening conditions than LoPRESTO listeners, we expected that HiPRESTO listeners would display better attentional and inhibitory abilities.

Vocabulary size may also influence the ability to understand speech in highly variable, adverse listening conditions. Lexical knowledge has been found to play an important role in speech perception and spoken word recognition (e.g., Ganong, 1980; Samuel, 1986), and any individual differences in lexical knowledge may result in variability in speech recognition performance. Listeners with larger vocabularies may be better able to use top-down lexical information due to greater lexical connectivity among words (e.g., Pisoni et al, 1985; Altieri et al, 2010), especially in adverse listening conditions when available linguistic information in the signal is poor. Additionally, the PRESTO sentences contain words varying in lexical frequency and familiarity, which may be difficult to recognize in such conditions. To measure vocabulary size, we used a self-report word familiarity questionnaire that contained high-, medium-, and low-familiarity English words. Since all participants were normal-hearing, young adults with no reported history of speech or hearing difficulties, they would be expected to be familiar with all the high-familiarity items. However, HiPRESTO listeners may have larger vocabularies, which would have aided them with the PRESTO sentences, and, as such, would know more words in the language. Therefore, we expected that HiPRESTO listeners would be more familiar with less familiar items, especially the low-familiarity items, on the questionnaire.

Executive functioning was also measured in both groups of listeners. As mentioned above, individual differences in executive functioning and cognitive control have been found to be related to speech perception skills (Beer et al, 2011). Executive functioning is a system of cognitive control that manages other cognitive processes, including attention/inhibition, working memory, and regulating behaviors and actions (Barkley, 1997, 2012). Thus, executive function is involved in a wide range of daily activities. Individual differences in executive functioning, among other things, would influence a listener's ability to rapidly adapt and adjust to new or changing environments, to start and stop activities or behaviors, focus or change attention, and allocate processing resources. PRESTO sentences in multitalker babble present sentences for recognition that are constantly changing from trial to trial, and at times the babble may mask indexical and linguistic information. Successful adaptation to such materials draws heavily on executive functions and, as such, listeners with better executive functioning may be better at adapting to such conditions. Thus, we predicted that the HiPRESTO listeners would show better executive functioning than LoPRESTO listeners, especially in domains related to changes in attention under variable or unpredictable environments.

To test the hypothesis that the extreme groups did not differ on all neurocognitive abilities, which would indicate that speech recognition in adverse listening conditions is simply a global domain-general ability, a measure of non-verbal intelligence was also obtained for all listeners. Non-verbal intelligence is related to the ability of participants to perform different types of tasks that do not rely on verbal coding and the use of language mediation. We predicted that the ability to recognize speech in highly variable conditions is specific to processing operations related to speech perception, and not to domain-general, non-verbal intelligence. In other words, we predicted that the extreme groups would not be inherently different on all abilities, PRESTO included, but rather they would be different on measures of neurocognitive processes specific to speech perception in adverse listening conditions. Thus, although the extreme groups were expected to differ on several perceptual and neurocognitive tasks described above, they were not expected to differ on measures of non-verbal intelligence.

To summarize, the goal of the current study was to investigate differences in real-world hearing, the processing of indexical information in speech, and neurocognitive abilities in good (HiPRESTO) and poor (LoPRESTO) listeners. HiPRESTO listeners were expected to show fewer real-world hearing difficulties, more accurate and/or faster processing of indexical information in speech, and better neurocognitive functioning than LoPRESTO listeners. These basic differences in perceptual and neurocognitive abilities were predicted to underlie differences in speech recognition in highly variable, adverse listening conditions.

Methods

Participants

Forty young adults were recruited to participate in the current study. All of the participants had completed Phase I of the study described in Gilbert et al (in press). Of these 40, 19 participants scored within the upper quartile of the Phase I distribution of the group of 121 (HiPRESTO) listeners and 21 scored within the lower quartile of the Phase I distribution (LoPRESTO) listeners. The original distribution of 121 participants on PRESTO, with HiPRESTO and LoPRESTO ranges indicated by grey dashed vertical bars, is shown in Figure 1. The 19 HiPRESTO listeners included 16 females and 3 males, aged 18.3–31.0 years old ($M = 22.7$; $SD = 3.3$ years) and had a mean score of 69.4% across all SNR conditions of the PRESTO test in Phase I. The 21 LoPRESTO listeners included 14 females and 7 males, aged 18.7–24.8 years ($M = 21.3$; $SD = 1.6$ years). They had a mean score of 55.5% across all SNR conditions of the PRESTO test in Phase I. The difference in overall performance between the HiPRESTO and LoPRESTO groups in Phase I was highly significant by a t-test [$t(38) = -15.46$, $p < .001$]. As reported in Gilbert et al (in press), the HiPRESTO group was consistently more accurate across all SNR conditions on a retest of PRESTO containing novel sentences [$t(38) = -9.06$, $p < .001$] and on sentences from the Hearing In Noise Test (HINT) [$t(38) = -3.00$, $p = .005$]. A summary of the HiPRESTO and LoPRESTO groups' performance on PRESTO (Phase I and Phase II), as described in Gilbert et al (in press), is provided in Table 1.

All participants in the extreme groups were self-reported normal-hearing native speakers of American English with no significant history of hearing or speech disorders at the time of testing. As reported in Gilbert et al (in press), participants also passed a hearing screener, which consisted of responding to pure tones presented at 25 dB HL at 250–8000 Hz (octave intervals) in the right ear and in the left ear. Missed responses to the first presentation of a tone were excluded to take into consideration any possible misunderstanding of the instructions. Participants were paid a total of $35 for participating in the second phase of testing. All listeners received $35 for 120 minutes of participation, which included a $15 completion bonus.

Materials and Procedures

Participants were tested individually in a quiet testing room, where they completed a series of computer-based perceptual tasks, neurocognitive tasks, and self-report questionnaires. For the computer-based tasks, each participant was seated in an enclosed testing carrel in front of a PowerMacG4 computer running MacOS 9.2 with diotic output to Beyer Dynamic DT-100 circum-aural headphones. Computer-based experimental tasks were controlled by custom PsyScript 5.1d3 scripts. Output levels of the target sentences for all computer-based perception tasks were calibrated to be approximately 64 dB SPL. Procedures for individual tasks are described below.

Hearing Questionnaire

Hearing Health Quick-Test (HHQT)—To obtain a measure of the participant's perception of his/her own hearing difficulties, the Hearing Health Quick Test (HHQT) from the American Academy of Audiology (available at www.howsyourhearing.org/ FSQuickTest.pdf) was also administered to all listeners. This assessment asks about perceived hearing difficulties and risk factors for hearing impairment, to which the respondent answers *yes*, *no*, or *sometimes*. For the purposes of the current experiment, the HHQT was divided into two different scales: one to evaluate the perceived presence of hearing difficulties in different situations (Situational Hearing Scale) with a maximum of 16 points, and one to identify significant hearing health history (Hearing Health Scale) with a maximum of 8 points. An analysis of the individual items was also carried out to compare responses across the extreme groups and assess perceived hearing difficulties in specific environments or situations.

Indexical Processing Tasks

Gender Discrimination—Participants completed a gender discrimination task. This task assessed the participants' ability to perceive and process gender-specific information in isolated sentences. Four talkers from the Indiana Multitalker Sentence Database (Karl and Pisoni, 1994; Bradlow et al, 1996) were selected for this task. Materials consisted of 32 unique utterances, 8 for each of 4 talkers (2 female and 2 male). On each trial, listeners were presented with two sentences, separated by 1000 ms of silence. Participants were asked to decide if the talkers in each pair of sentences were of the same gender or different genders. Overall, all talkers were paired with themselves once (a single trial) and with the other three talkers twice (6 trials), for 8 presentations per talker (over 7 trials), with no utterance repeated. Thus, the entire task included a total of 16 trials, of which 8 pairs were 'Same Gender' talkers and 8 pairs were 'Different Gender' talkers. Participants were instructed to respond as quickly as possible without compromising accuracy. They responded by pressing one of two buttons on a button box. For all participants, 'Same Gender' responses were always represented by the button farthest to the right, and 'Different Gender' responses always the farthest button to the left. Response accuracy (correctly responding 'Same' to a 'Same Gender' trial or 'Different' to a 'Different Gender' trial) and response times (RT) were collected and analyzed separately.

Talker Discrimination—A talker discrimination task was also included to assess the participants' ability to perceive and process within-gender talker-specific information. The talker discrimination task was based on the methodology developed by Cleary and Pisoni (2002). Six different talkers from the Indiana Multitalker Sentence Database (Karl and Pisoni, 1994; Bradlow et al, 1996) were selected for this task. Materials consisted of 48 unique utterances, 8 per each of 6 talkers (3 female and 3 male). On each trial, listeners were presented with two sentences, separated by 1000 ms of silence, that were produced by a single talker or two different talkers. Listeners were asked to decide if the two sentences were produced by the same talker or by two different talkers. Sentences produced by male talkers were always paired with another sentence produced by the same or a different male talker; similarly, female talkers were always paired together in trials. Overall, all talkers were paired with themselves two times (2 trials) and with the other two talkers of the same gender twice (4 trials), for 8 presentations per talker (over 6 trials). No utterance was repeated. The talker discrimination task consisted of a total of 24 trials, of which 12 were 'Same Talker' trials and 12 were 'Different Talker' trials.

Participants were instructed to respond as quickly as possible without compromising accuracy. They responded by pressing one of two buttons on a button box. For all participants, 'Same Talker' responses were always represented by the button farthest to the

right, and 'Different Talker' responses always the farthest button to the left. Trial accuracy (correctly responding 'Same' to a 'Same Talker' trial or responding 'Different' to a 'Different Talker' trial) and response time (RT) were collected and analyzed.

Regional Dialect Categorization—Participants completed a regional dialect categorization task. This task assessed a listener's ability to perceive dialect-specific information in the acoustic signal and use stored knowledge of such information to identify the region of origin of unfamiliar talkers. The forced-choice dialect categorization task used in this study was based on the methodology developed by Clopper and Pisoni (2004). The talkers and test sentences were selected from the TIMIT acoustic-phonetic speech corpus (Garofolo et al, 1993). Forty-eight talkers, 24 female and 24 male, were used in this task. Eight talkers (4 female and 4 male) were from each of the following six dialect regions: New England, North Midland, South Midland, North, South, and West. Two sentences were used for each talker for the task. One sentence, "She had your suit in greasy wash water all year," was the same for all talkers. This sentence is one of the baseline calibration sentences collected from all talkers in the TIMIT database and was designed to obtain dialectal differences. The other sentence was unique to each talker and was selected to contain phonological/phonetic variation representative for each dialect region (Clopper and Pisoni, 2004).

The dialect categorization task was divided into two blocks. In the first block, participants heard the talkers saying the same baseline sentence. In the second block, participants heard the talkers produce the unique sentences. Each talker was presented one time per block, for a total of 96 trials (48 trials during the first block and 48 trials during the second block). Participants were permitted to take a break after each block. For each trial, participants heard a single talker and were asked to select the region where the talker was from using a closed set of response alternatives. Participants were able to choose from the six regions (New England, North Midland, South Midland, North, South, and West) represented on a map of the United States displayed on a computer monitor and responded by clicking on a labeled box located within a particular region. Participants could take as long as they wanted to respond. Once they responded, the next trial began. Participants' responses were collected and coded for the region selected and overall accuracy. Both incorrect and correct responses were analyzed.

Neurocognitive Tests

Auditory Digit Span—Participants completed Digit Span Forward and Digit Span Backward tasks (Wechsler Memory Scale 3rd Edition, Wechsler, 1997). For the current study, Digit Span Forward was used as a measure of verbal short-term memory capacity, and Digit Span Backward was used as a measure of verbal working memory capacity. An experimenter, who was a native speaker of American English, read a list of digits out loud, ranging from two to nine digits in length, to the participant. The participant was asked to repeat back the list of numbers, either in the same order as dictated by the experimenter (Digit Span Forward) or in a backward order (Digit Span Backward). Both digit span tasks started with a list of two numbers. Each list length was presented twice and the task continued until the participant missed both trials for the same list length or until the lists were exhausted (Forward: length of 8 digit sequences, total 16 possible trials; Backward: length of 7 digit sequences, total 14 possible trials). No participant was able to correctly repeat all lists in either task. Responses were scored on-line by the experimenter. The participant had to repeat all numbers back in the correct order in order for that trial to be counted as correct. Trials were counted as either correct or incorrect; partial credit for partially correct answers was not given.

Stroop Color and Word Test—Participants completed the Stroop Color and Word Test (Golden, 1978; Golden and Freshwater, 2002), which provided a measure of attention/ inhibition. The test is divided into three parts. In the first part (the Word condition), participants were instructed to read a list of color words ('blue', 'green', 'red') typed with black ink aloud as quickly as possible, but to correct any mistakes that they made by returning to the word and reading it over again. Participants had 45 seconds to read as many words aloud as possible. Similarly, in the second part (the Color condition), participants were instructed to say the color of three *XXX*s written with blue, green, or red ink in a list aloud, going through as many as possible in 45 seconds. In the third part (the Color-Word condition), participants had to say the color of the printed ink out loud for the color words ('blue' 'green', 'red') in a list. The color of the ink did not match the color word (i.e., the word 'blue' was not written with blue ink). Again, participants had 45 seconds to go through as many trials in the list as possible, while correcting any errors made by repeating the item. Each part of the test was scored by counting the number of items that the participant was able to complete in the 45 seconds. Color, Word, and Color-Word T-scores were calculated from the difference between raw scores and predicted scores (based on age and education) for each condition. Raw Interference scores were then calculated by subtracting a predicted Interference score (based on raw color and word scores) from the raw Color-Word score. These scores were transformed into T-scores.

WordFam—Participants completed the WordFam test (Pisoni, 2007), a self-report word familiarity rating questionnaire originally developed by Lewellen et al (1993). Responses on the WordFam test provide a measure of vocabulary size. In this task, participants were instructed to rate how familiar they were with a set of English words using a seven-point scale, ranging from 1 ('You have never seen or heard the word before') to 7 ('You recognize the word and are confident that you know the meaning of the word.'). Participants were asked to follow the instructions and rating guidelines provided on the first page of the questionnaire. The WordFam questionnaire contains a total of 150 English words, including 50 low familiarity, 50 mid familiarity, and 50 high familiarity items based on ratings from Nusbaum et al (1984). Participants responded by marking the number corresponding to the familiarity rating for a given item. Responses for all 150 items were recorded and averaged for each of the three lexical frequency conditions.

BRIEF-A—Participants also completed the Behavioral Rating Inventory of Executive Function – Adult Version (BRIEF-A; Roth et al, 2005; Psychological Assessment Resources). The BRIEF-A is a self-report questionnaire designed to assess a participant's own rating of his/her executive functions in everyday life. The BRIEF-A consists of 75 statements to which participants must respond if their behavior is a problem: never (1 point), sometimes (2 points), or often (3 points). The BRIEF-A is used to evaluate nine clinical domains of executive functions: Shift, Inhibit, Emotional Control, Self-Monitor, Initiate, Working Memory, Plan/Organize, Task Monitor, and Organization of Materials. These domains are grouped into two aggregate indexes, a Behavioral Regulation Index (BRI) and a Metacognitive Index (MI), and an overall global General Executive Composite (GEC) score. The Behavior Regulation Index (BRI) is comprised of Shift, Inhibit, and Emotional Control scales. The Shift scale is related to the ability to move or change from one situation, topic, or task to another. The Inhibit scale is related to the ability to inhibit or stop oneself from acting on impulse in different situations. Emotional Control refers to one's ability to control or modulate emotional behavior. Overall, the Behavioral Regulation Index (BRI) is related to one's ability to appropriately change or adapt their emotional behavior demonstrating good inhibitory control.

The Metacognitive Index (MI) is comprised of the Self-Monitor, Initiate, Working Memory, Plan/Organize, Task Monitor, and Organization of Materials scales. Self-Monitor refers to

one's ability to evaluate how his/her actions or behaviors affect others. Initiate is related to the ability to start a new task or generate new, independent thoughts or ideas. The Working Memory scale assesses the participant's ability to hold information in memory for completing, or taking the necessary steps to complete, a task or goal. Plan/Organize refers to the ability to manage information to complete current tasks or in anticipation of future tasks. Task Monitor refers to one's ability to evaluate his/her actions or behaviors during or after a task or activity. The scale of Organization of Materials is related to the orderliness or organization of one's belongings and space or actions. Overall, the Metacognitive Index (MI) is related to the ability to appropriately manage and complete different tasks and actions.

The Global Executive Composite (GEC) of the BRIEF includes all individual scales and provides an overall assessment of executive functioning in everyday life. Scores on all nine domains of executive functions, the two aggregate indexes, and the GEC composite are transformed into T-scores. Responses were compared across groups for T-score differences in GEC, BRI, MI, and each of the nine individual domains. An analysis of the individual items was also carried out to compare responses across groups regarding specific behaviors.

WASI Performance IQ—To obtain a measure of non-verbal IQ, participants completed two performance subtests of the WASI (Wechsler Abbreviated Scale of Intelligence, Wechsler, 1999), Block Design and Matrix Reasoning. The Block Design task is used to assess the participant's visuospatial processing and problem solving abilities. For this task, participants arranged two-tone blocks in figures to match a model or picture of a design. Participants were timed, and points were given for accuracy and amount of time needed to complete the design. The Block Design task continued until participants had three consecutive incorrect responses or until all trials were completed. Points were tallied and converted into T-scores following conventional procedures.

Similarly to the Block Design task, the Matrix Reasoning task assesses the participant's ability to carry out abstract visuospatial reasoning and problem solving. Participants were presented with an incomplete picture or pattern (on the top of the test booklet) and were instructed to choose one of five options (given on the bottom of the test booklet) that they thought best completed the visual pattern. They were asked to pick the best option, even if they were unsure if any of possible answers entirely fit the pattern. Participants could take as long as they desired to give their response. The task continued until participants had given at least four incorrect responses in the previous five test items or until all trials were completed. Responses were recorded as either correct or incorrect, and no partial credit was given. Points were tallied and converted into T-scores. Scores from both tasks were combined to obtain a general measure of Performance IQ.

Results

Hearing Questionnaire

Hearing Health Quick-Test (HHQT)—Overall HHQT scores, the Situation Hearing Scale scores, and the Hearing Health Scale scores were calculated for both the extreme groups. Independent samples t-tests were carried out between groups on all three measures. Although the LoPRESTO group reported having slightly more overall hearing difficulty and hearing health history risk factors on the HHQT than the HiPRESTO group (LoPRESTO: M = 8.29, SD = 4.80; HiPRESTO: M = 6.84, SD = 4.12), overall HHQT scores were not significantly different between the two groups [$t(38) = -1.02$, $p = .317$]. A similar non-significant trend occurred for the Situational Hearing Scale, with the LoPRESTO group averaging a score of 6.81 (SD = 3.44) compared to the HiPRESTO group average of 5.26 (SD = 3.19) [$t(38) = -1.47$, $p = .150$]. The difference between groups on the Hearing Health

Scale was also not significant. An analysis of individual questions revealed group differences in response to one individual item. The LoPRESTO group reported significantly higher responses to the question, *Do you sometimes find it difficult to understand a speaker at a public meeting or a religious service?* [t(38) = −2.30, p = .027], suggesting that the LoPRESTO group might be aware of having more difficulties understanding speech in public places, which do not provide an ideal, quiet listening environment.

Indexical Processing Tasks

Gender Discrimination—Mean accuracy measures were calculated for both listener groups for overall accuracy, as well as 'Same Gender' and 'Different Gender' conditions. Independent sample t-tests were carried out between groups on all three measures. The t-test revealed that the HiPRESTO group was significantly more accurate on the Gender Discrimination task [t(38) = 2.09, p = .043] than the LoPRESTO group. The overall significant difference, however, was likely due to more accurate performance on 'Same Gender' trials, as HiPRESTO participants were significantly more accurate than LoPRESTO participants on 'Same Gender' trials [t(38) = 2.14, p = .039] but not 'Different Gender' trials [t(38) = .290, p = .773]. Table 2 shows overall accuracy for 'Same Gender' and 'Different Gender' trials for both groups.

Mean RT measures were also calculated for all correct trials, for all correct trials within the 'Same Gender' conditions, and within the 'Different Gender' condition. A series of t-tests were also carried out between groups on these three measures. The HiPRESTO and LoPRESTO groups were not significantly different on any RT measure. This suggests that while HiPRESTO and LoPRESTO groups responded with roughly the same speed, the LoPRESTO group's accuracy was compromised.

Talker Discrimination—Similarly to the Gender Discrimination task, mean accuracy measures were calculated for both groups for overall accuracy, as well as 'Same Talker' and 'Different Talker' conditions. Independent sample t-tests between groups on all three measures revealed no significant difference on any accuracy measure between the HiPRESTO and LoPRESTO groups. Mean RT measures were also calculated for all correct trials, for all correct trials within the 'Same Talker' conditions, and for all correct trials within the 'Different Talker' condition. A series of t-tests again revealed no significant differences between the groups on any RT measure. Thus, HiPRESTO and LoPRESTO groups performed similarly in terms of response accuracy and speed on the Talker Discrimination task.

Regional Dialect Categorization—Participants' responses were collected and analyzed for overall accuracy. Figure 2 shows the overall accuracy and accuracy for each talker dialect by group. Because the focus of this study is on extreme group differences, a series of t-tests with listener group (HiPRESTO or LoPRESTO) as the between-subject factor was carried out on overall accuracy and accuracy on each talker dialect. The t-tests revealed that the HiPRESTO group was significantly better than the LoPRESTO group on overall accuracy [t(38) = 2.11, p = .041] and for the Southern talkers [t(38) = 2.55, p = .015]. This pattern suggests that the overall difference is driven by better performance on the Southern talkers, and indicates that HiPRESTO listeners may have greater dialect-specific representations for the non-standard (marked) dialects than the LoPRESTO group.

Neurocognitive Tests

Auditory Digit Span—Responses were tallied and scored for the longest list length that a participant could correctly recall forward (Max Forward) and the longest list that a participant could correctly recall in the reverse order (Max Backward). A series of t-tests

were carried out on Max Forward and Max Backward scores between the two listeners groups. HiPRESTO participants had significantly longer Max Forward spans [$t(38) = 3.60$, $p = .001$] and Max Backward spans [$t(38) = 2.28$, $p = .029$]. Figure 3 displays the average Max Forward and Max Backward spans for HiPRESTO and LoPRESTO groups.

Stroop Color and Word Test—Color, Word, Color-Word, and Interference T-scores were calculated for each participant. A series of t-tests were then carried out between groups for all scores. No significant differences were found for any of the scores on the Stroop.

WordFam—Responses were collected and tallied for all responses and for each familiarity condition (high, medium, and low familiarity). A series of t-tests were carried out between groups on average responses for all items and for each familiarity condition. Overall, the HiPRESTO group displayed significantly higher familiarity ratings with all words on the WordFam questionnaire than the LoPRESTO group [$t(38) = 5.01$, $p < .001$]. However, HiPRESTO participants were also significantly more familiar with a greater number of low familiarity items [$t(38) = 4.78$, $p < .001$] and medium familiarity items [$t(38) = 5.47$, $p < .001$], but not with high familiarity items [$t(38) = 1.99$, $p = .053$], which suggests that the overall significant difference between the two listener groups was driven by the low and medium familiarity words. Figure 4 shows the average familiarity rating to all words, as well as for low, medium, and high familiarity words, for both participant groups.

BRIEF-A—Using the standardized domains, aggregate indexes, and global GEC scores, no differences were observed between HiPRESTO and LoPRESTO groups on self-reports of executive function. Overall, both groups' GECs were within normal limits (HiPRESTO: $M = 46.47$, $SD = 9.12$; LoPRESTO: $M = 49.19$, $SD = 11.66$). However, an analysis of individual questions revealed several consistent differences in the groups' responses to individual items. The LoPRESTO group demonstrated significantly elevated responses to three statements compared to the HiPRESTO group: Item 35, Working Memory domain, *I have a short attention span* [$t(38) = -2.18$, $p = .036$]; Item 46, Working Memory domain, *I forget instructions easily* [$t(38) = -2.79$, $p = .009$]; and Item 47, Planning and Organization domain, *I have good ideas but cannot get them on paper* [$t(38) = -2.85$, $p = .008$], suggesting awareness of cognitive workload difficulties.

WASI Performance IQ—Block Design T-scores, Matrix Reasoning T-scores, and the Performance IQ score were calculated for each participant. A series of t-tests were carried out on average scores between HiPRESTO and LoPRESTO groups. No significant differences between listener groups were found on either task or the overall Performance IQ score.

Discussion

The current study investigated the real-world hearing difficulties, indexical processing skills, and neurocognitive abilities of two extreme groups of listeners who were selected from the 121 listeners originally tested by Gilbert et al (in press). To assess real-world hearing, listeners completed a self-report questionnaire on situational hearing difficulties and hearing health history. While no significant differences were found on either subscale or overall, LoPRESTO listeners did report having slightly more difficulty hearing in different environments or situations. Indeed, one question that was significantly different between groups assessed perceived hearing ability in public places, where listening conditions might be less than ideal. Although preliminary in nature, these results suggest that performance on PRESTO may be related to real-world hearing experiences. One limitation of the Hearing Health Quick-Test used in this study is that it was self-report, making it subject to individual

participants' subjective perception of their ease of hearing. Moreover, the response alternatives were limited to three (*yes*, *no*, or *sometimes*), and perhaps a more gradient rating scale would better capture individual differences in self-perceived hearing difficulties. Nevertheless, the link between real-world hearing and speech recognition skills based on self-reports should be further explored.

HiPRESTO and LoPRESTO listeners also completed several perceptual tasks designed to investigate whether those listeners who are better at recognizing the high-variability PRESTO sentences would also be better at indexical processing tasks. For the talker discrimination task, the HiPRESTO and LoPRESTO groups showed no significant difference in performance. Thus, the poor performance on PRESTO by the LoPRESTO group is not related to audibility or sensory factors associated with their ability to quickly process talker-specific information in spoken sentences in the quiet. The HiPRESTO and LoPRESTO groups, however, did significantly differ on the Gender Discrimination Task and Dialect Categorization Task. Although performance was at ceiling for the HiPRESTO group and nearly at ceiling for the LoPRESTO group on the gender discrimination task, suggesting that neither group had much difficulty at distinguishing the gender-specific information, the LoPRESTO group was significantly less accurate both overall and on trials with two talkers of the same gender, but not on trials with talkers of different genders. The two listener groups also showed no significant differences in response times for the task, suggesting that the LoPRESTO group made their judgments as rapidly as the HiPRESTO group but sacrificed some degree of accuracy. While the overall mean difference between the two groups was small (Difference for all trials = 3.84%; Difference for "Same Gender" trials = 7.14%), the significant tendency for the LoPRESTO group to be slightly less accurate on the task suggests that they may have some difficulty processing gender-specific information even under quiet presentation conditions. While this cannot be because of lack of exposure to talkers of different genders, the listeners may have trouble perceiving and encoding gender-specific episodic context information. In any case, given this finding, the LoPRESTO group overall may have more difficulty rapidly adjusting to talkers of different genders, a foundational processing skill that would be necessary for success on the PRESTO sentences in multitalker babble.

HiPRESTO listeners were also significantly more accurate overall on the forced-choice dialect categorization task than LoPRESTO listeners, and for reliably categorizing the Southern regional dialect in particular. This pattern suggests that HiPRESTO listeners may have more robust dialect-specific representations and memory codes in long-term memory for the American English dialects. HiPRESTO listeners, who were significantly more accurate on the dialect categorization task, may benefit from a better ability to encode more detailed episodic information in speech. This skill would help these listeners to perceive, encode, and store more robust dialect-specific information in day-to-day life, which they would be able to draw upon in both the dialect categorization task and the high-variability sentence recognition test. One argument against this account is that, by chance, the HiPRESTO listeners might simply be more familiar with the different regional dialects of American English, and perhaps the Southern regional dialect in particular, through previous exposure from residential history or casual contact. To assess this explanation, we examined the residential histories of the listeners and found that the two groups did not differ in terms of region of origin or geographic mobility (HiPRESTO: 9 General American (Midland, West, parts of New England), 2 Mid-Atlantic, 4 Mobile, 4 North; LoPRESTO: 12 General American, 6 Mobile, 3 North). Therefore, previous residential history is not likely to have affected performance on the task. If HiPRESTO listeners are indeed better at encoding fine context-sensitive episodic information in speech, one would predict that the HiPRESTO listeners would also be better at processing other types of indexical variability, like foreign accent, voice quality, and emotion. In any case, listeners showing better processing of

acoustic-phonetic regional dialect information appear to benefit from this ability on a speech recognition task containing speech samples of talkers from multiple dialects.

Taken together, the results obtained from the perceptual tasks suggest that better indexical processing skills may be related to better performance on a high variability sentence recognition task, like PRESTO. Although the difference between HiPRESTO and LoPRESTO listeners was not large on either discrimination task because of ceiling effects under quiet presentation, HiPRESTO listeners were significantly more accurate on gender discrimination and dialect categorization accuracy, suggesting that LoPRESTO listeners may have difficulty perceiving, encoding, and storing fine acoustic-phonetic details of speech and episodic context information. Based on the present findings, one might expect that if a listener has difficulties processing indexical variability, he/she will also have problems understanding the PRESTO sentences in babble. These findings support previous research on indexical information and word recognition, which suggests that speech perception is a talker-contingent process. Earlier studies have shown that familiar talkers, after explicit voice identification training, are more intelligible in noise than novel talkers (Nygaard et al, 1994; Nygaard and Pisoni, 1998). However, these studies also found that listeners vary greatly in their ability to explicitly learn the talker voices in the laboratory during training and that those listeners who were particularly poor at learning the voices did not show a recognition benefit for familiar talkers. Similarly, research on deaf children with CIs has shown that the ability to discriminate talker-specific voice information is related to other speech and language outcome measures (Cleary and Pisoni, 2002; Cleary et al, 2005). Children with CIs who were better able to process talker-specific indexical information also scored better on measures of spoken word recognition (Cleary and Pisoni, 2002; Cleary et al, 2005). Taken together, the findings of the current study support previous research, suggesting that a listener's ability to perceive and encode indexical information in speech is related to his/her ability to rapidly adjust to this type of information for successful word recognition in highly variable, adverse listening conditions.

Listeners also completed several neurocognitive tasks designed to assess how individual differences in speech perception may be related to core, underlying neurocognitive processes. Given the close links established between several neurocognitive processes and speech perception skills (e.g., Pisoni, 2000; Arlinger et al, 2009; Stenfelt and Rönnberg, 2009), our expectation was that the HiPRESTO listeners would also display greater short-term and working memory capacity, better inhibition/attention, and better executive functioning, all of which previous research has suggested are important underlying components of speech perception in adverse listening conditions. To examine this issue, both HiPRESTO and LoPRESTO groups completed forward and backward Digit Span tasks, the Stroop Color and Word Test, and the BRIEF-A questionnaire on executive function. As expected HiPRESTO listeners had significantly larger forward (short-term) and backward (working memory) digit spans. This finding suggests that short-term and working memory capacity play a foundational role in speech perception in adverse listening conditions, and individual differences in such abilities may contribute to observed variability in speech perception performance on a variety of tasks (Pisoni and Cleary, 2003; Pisoni et al, 2011). As such, the current findings add to other reports in the literature demonstrating evidence for speech perception models that incorporate short-term and working memory processes (e.g., Gathercole and Baddeley, 1993; Baddeley et al, 1998; Goldinger, 1998; Vaughan et al, 2008; Conway et al, 2010).

Contrary to our predictions, however, HiPRESTO and LoPRESTO listeners did not differ in terms of inhibition/attention abilities, as measured by the Stroop Color and Word Test. This finding suggests that aspects of attention/inhibition measured by Stroop do not play a key role in adapting to the highly variable PRESTO sentences in babble. However, it is also

quite possible that the Stroop Color and Word Test did not provide enough of a challenge to the young, normal-hearing listeners and as such, there was not enough meaningful variability in performance on the test for any relationship to emerge. One other potential problem with using the Stroop Color and Word Test for the current study is that the task is a predominantly visual processing task. Attention/inhibition tasks specific to the auditory modality, such as the auditory Stroop task (Jerger et al, 1988; Jerger et al, 1993; Jerger et al, 1994) or a speeded-classification task (Garner, 1974; Mullennix and Pisoni, 1990), may provide better measures for investigating skills that may be more heavily influenced by modality-specific auditory attention/inhibition processes.

The extreme groups were significantly different on the WordFam test, a measure of lexicon size/vocabulary. As expected, we found that HiPRESTO listeners were significantly more familiar with the set of 150 words than the LoPRESTO group. In particular, the groups did not significantly differ on the high familiarity words, but the HiPRESTO listeners were more familiar with the low and medium familiarity words than the LoPRESTO listeners. This pattern indicates that HiPRESTO listeners benefited from knowing more words and greater lexical connectivity that helped them to recognize words in the PRESTO sentences under degraded listening conditions. These findings are consistent with previous research on the role of the lexicon in spoken word recognition (e.g., Pisoni et al, 1985; Altieri et al, 2010), and converge with findings on the importance of lexicon size and connectivity in speech perception in child language development (e.g., Munson, 2001; Edwards et al, 2004; Munson et al, 2005a; Munson et al, 2005b) and second language speech perception (e.g., Bradlow and Pisoni, 1999; Ezzatian et al, 2010; Bundgaard-Nielsen et al, 2011). Thus, lexicon size/vocabulary provides a direction for future research on individual differences in speech recognition in adverse listening conditions.

Similarly to the Stroop Color and Word Test, no differences were found between the HiPRESTO and LoPRESTO groups on any of the nine clinical scales, the two aggregate indexes, or the global score on the BRIEF-A. For individual questions, however, significant differences emerged between groups on three statements. On those three items, which were related to self-reports of cognitive load, the LoPRESTO listeners showed significantly higher scores (i.e., greater self-awareness of difficulty under cognitive load) than the HiPRESTO group, as would be predicted. The significant group difference on these three items of self-reported executive function suggest that the two groups differed in several aspects of executive functioning and that different tests, including performance tests assessing executive functioning, and specifically cognitive load under dual task conditions, should be included to further assess individual differences in spoken word recognition.

The BRIEF-A was designed to measure the difficulties arising from executive function that individuals with Attention Deficit Hyperactivity Disorder (ADHD) have in everyday situations (Roth et al, 2005). Furthermore, the BRIEF was designed to identify individuals from an ADHD population who are having clinically significant difficulties. The BRIEF-A is not a general-purpose instrument for use in identifying individuals with executive function problems in the general population who are functioning within normal limits. These design issues may limit the BRIEF-A's sensitivity to identify individual differences in executive functioning in young, normal-hearing, highly educated participants. Still, the three questions separating the two extreme groups suggest that LoPRESTO listeners have specific difficulties in everyday situations involving cognitive load and the allocation of cognitive resources under competition. Other types of performance tests should be included in future studies to further evaluate the role of executive function in speech perception in adverse listening conditions.

As predicted, the HiPRESTO and LoPRESTO groups did not differ on the WASI Performance IQ or any of its subtests, which was used to measure non-verbal intelligence. This finding supports the prediction that individual differences in speech perception in adverse listening conditions are not related to some global overall difference in IQ. These results are in contrast with the lexicon size/vocabulary measurement of the WordFam test, which may be considered a proxy for verbal intelligence. The distinction between non-verbal and verbal intelligence is perhaps unsurprising, given that the groups were formed based on spoken word recognition performance. However, in this initial exploration into the underlying factors associated with individual differences on PRESTO, it was imperative to document that there was no difference in non-verbal intelligence between the two groups.

The results of the neurocognitive tasks suggest that the ability to rapidly adjust to the target sentence variability of PRESTO against background competition from other competing talkers may be related to individual differences in several underlying, core neurocognitive processes. In particular, our findings suggest that short-term and working memory capacity, along with lexical knowledge are underlying factors that are associated with speech recognition in adverse listening conditions. Inhibition/attention and selected aspects of executive function should be further explored with more sensitive performance-based tests and, perhaps, more diverse listener groups. However, the presence of such a wide range of performance on speech perception tasks in normal-hearing listeners is indicative of basic, underlying information processing differences, which we have only begun to elucidate through the current investigation using a very limited test battery. Individual differences in speech perception reflect the contribution of multiple foundational neurocognitive processes. In addition, this study of normal-hearing listeners provides the first evidence that indexical/perceptual processes may also contribute to the wide range of individual performance on speech perception in adverse listening conditions.

Although the findings of the present study suggest that several types of indexical processing skills and neurocognitive abilities may underlie speech recognition in adverse conditions, our investigation of individual differences was limited by the number of component processes investigated and the number and types of tasks used. In this preliminary investigation, the types of tasks used were necessarily constrained by our attempts to sample several perceptual and neurocognitive domains, i.e., the goal of the study was to examine multiple potential factors that might be associated with speech recognition on the PRESTO sentences in multitalker babble. Additionally, the current study examined only three types of indexical processing, talker and gender discrimination and regional dialect categorization. Certainly, other associated perceptual skills related to the ability to understand speech in high-variability conditions may involve the processing of other types of indexical variability, such as foreign accent discrimination/categorization, identification of emotion in speech, novel talker learning and talker identification, the listener's perceptual vowel space size or shape, and environmental sound perception.

Similarly, other neurocognitive abilities should also be investigated, including processing speed, rapid phonological coding, multi-tasking, cognitive effort or mental workload. Also, the modality- and domain-specificity of the specific neurocognitive tests (auditory or visual), especially for tasks assessing attention/inhibition, should be carefully considered in future studies. Multiple tasks and the use of converging methods of assessments should be used to obtain additional measures of the different perceptual and neurocognitive abilities involved in the current study, as well as those that have not yet been tested. While performance on PRESTO sentences was chosen to form the extreme groups in the design of this study, other types of sentence tests and speech perception tasks could be used to examine the underlying factors that play a role in speech perception in different types of conditions. For example, other types of tests could include different types of target

variability (e.g., foreign-accented speech) or different degrees of background competition (e.g., noise or babble). To gain a better understanding of the contribution of basic sensory, perceptual, and neurocognitive factors to speech perception in adverse listening conditions and how these factors influence individual differences in speech perception abilities, additional studies using diverse methods and multiple perspectives are necessary.

Conclusions

The goal of the current study was to examine and identify the strengths and weaknesses of good and poor listeners who were selected based on their performance in recognizing speech in highly variable, adverse listening conditions. In order to carry out this study, young, normal-hearing adults who fell within the upper and lower quartiles of performance on PRESTO completed several neurocognitive and perceptual tasks designed to uncover the relations between speech perception and neurocognitive processes, and any associated skills in the perception of indexical properties of speech. Our results showed that the two extreme groups differed on tests of indexical processing abilities, short-term and working memory abilities, and vocabulary size. However, the two groups did not differ on measures of real-world hearing, talker discrimination, attention/inhibition, and non-verbal intelligence. While the groups did not differ on the standardized clinical scales of the BRIEF, they did show differences on specific test items related to cognitive load and the use of mental processing resources. These results suggest short-term and working memory capacity, vocabulary size, and the ability to encode highly detailed episodic information in speech play important roles in speech perception in adverse listening conditions and may contribute to individual differences observed in the ability to recognize speech in such conditions. Although no causal relationship between these abilities and speech recognition performance on PRESTO can be identified from this study, this preliminary investigation of individual differences provides several new directions for future research using a variety of other methodologies. Further investigation of the sensory, perceptual, and neurocognitive abilities should be carried out using a variety of different methods and approaches to uncover the potential sources of individual differences in speech recognition in adverse listening conditions.

Acknowledgments

Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
BRI	Behavioral Regulation Index
BRIEF	Behavioral Rating Inventory of Executive Function
CI	Cochlear Implant
GEC	Global Executive Composite
HHQT	Hearing Health Quick Test

HiPRESTO	Group of participants in the upper quartile of performance distribution in Phase I
IQ	Intelligence Quotient
LoPRESTO	Group of participants in the lower quartile of performance distribution in Phase I
MI	Metacognitive Index
PRESTO	Perceptually Robust English Sentence Test Open-set
RT	Response Time
SNR	Signal-to-Noise Ratio
TIMIT	Texas Instruments/Massachusetts Institute of Technology
WASI	Wechsler Abbreviated Scale of Intelligence

References

Abercrombie, D. Elements of general phonetics. Chicago, IL: Aldine; 1967.

Akeroyd MA. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. Int J Audiol. 2008; 47(Suppl 2):S53–S71. [PubMed: 19012113]

Alain C, Arnott SR. Selectively attending to auditory objects. Front Biosci. 2000; 5:D202–D212. [PubMed: 10702369]

Altieri N, Gruenenfelder T, Pisoni DB. Clustering coefficients of lexical neighborhoods: Does neighborhood structure matter in spoken word recognition? Men Lex. 2010; 5:1–21.

Arlinger S, Lunner T, Lyxell B, Pichora-Fuller MK. The emergence of Cognitive Hearing Science. Scand J Psychol. 2009; 50:371–384. [PubMed: 19778385]

Baddeley A, Gathercole S, Papagno C. The phonological loop as a language learning device. Psychol Rev. 1998; 105:158–173. [PubMed: 9450375]

Baddeley A. Working memory and language: an overview. J Commun Disord. 2003; 36:189–208. [PubMed: 12742667]

Barkley RA. Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. Psychol Bull. 1997; 121:65–94. [PubMed: 9000892]

Barkley, RA. Executive Functions: What They Are, How They Work, and Why They Evolved. New York: The Guilford Press; 2012.

Beer J, Kronenberger W, Pisoni DB. Executive function in everyday life: Implications for young cochlear implant users. Cochlear Implants Int. 2011; 12(Suppl 1):S89–S91. [PubMed: 21756484]

Bent T, Bradlow AR. The interlanguage speech intelligibility benefit. J Acoust Soc Am. 2003; 114:1600–1610. [PubMed: 14514213]

Bradlow AR, Pisoni DB. Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. J Acoust Soc Am. 1999; 106:2074–2085. [PubMed: 10530030]

Bradlow AR, Toretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine- grained acoustic-phonetic talker characteristics. Speech Commun. 1996; 20:255–272. [PubMed: 21461127]

Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J Acoust Soc Am. 2001; 110:2527–2538. [PubMed: 11757942]

Bundgaard-Nielsen RL, Best CT, Tyler MD. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowels categories. Appl Psycholinguist. 2011; 32:51–67.

Carhart R, Johnson C, Goodman J. Perceptual masking of spondees by combinations of talkers. J Acoust Soc Am. 1975; 58(Suppl 1):S35.

Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. J Acoust Soc Am. 1969; 45:694–703. [PubMed: 5776931]

Cherry EC. Some experiments on the recognition of speech, with one and two ears. J Acoust Soc Am. 1953; 25:975–979.

Cleary M, Pisoni DB. Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. Ann Otol Rhinol Laryngol Suppl. 2002; 189:113–118. [PubMed: 12018337]

Cleary M, Pisoni DB, Kirk KI. Working memory spans as predictors of spoken word recognition and receptive vocabulary in children with cochlear implants. Volta Rev. 2002; 102:259–280. [PubMed: 21666765]

Cleary M, Pisoni DB, Kirk KI. Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. J Speech Lang Hear Res. 2005; 48:204–223. [PubMed: 15938065]

Clopper CG, Bradlow AR. Perception of dialect variation in noise: Intelligibility and classification. Lang Speech. 2008; 51:175–198. [PubMed: 19626923]

Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. J Phon. 2004; 32:111–140. [PubMed: 21451736]

Cohen J. The cost of dichotomization. Appl Psychol Meas. 1983; 7:249–253.

Conway AR, Kane MJ, Bunting MF, Hambrick DZ, Wilhelm O, Engle RW. Working memory span tasks: A methodological review and user's guide. Psychon Bull Rev. 2005; 12:769–786. [PubMed: 16523997]

Conway CM, Bauernschmidt A, Huang SS, Pisoni DB. Implicit statistical learning in language processing: Word predictability is the key. Cognition. 2010; 114:356–371. [PubMed: 19922909]

Cowan N. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychol Bull. 1988; 104:163–191. [PubMed: 3054993]

Cowan, N. Attention and memory: An integrated framework. New York: Oxford University Press; 1997.

Creelman CD. Case of the unknown talker. J Acoust Soc Am. 1957; 67:253–361.

Edwards J, Beckman ME, Munson B. The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. J Speech Lang Hear Res. 2004; 47:421–436. [PubMed: 15157141]

Ezzatian P, Avivi M, Schneider BA. Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking? Speech Commun. 2010; 52:919–929.

Felty, R. Unpublished manuscript. Indiana University; Bloomington: 2008. Perceptually Robust English Sentence Test (Open-Set).

Francis AL, Nusbaum HC. Effects of intelligibility on working memory demand for speech perception. Atten Percept Psychophys. 2009; 71:1360–1374. [PubMed: 19633351]

Ganong WF. Phonetic categorization in auditory word perception. J Exp Psychol Hum Percept Perform. 1980; 6:110–125. [PubMed: 6444985]

Garner, WR. The Processing of Information and Structure. Potomac, MD: Lawrence Erlbaum; 1974.

Garofolo, JS.; Lamel, LF.; Fisher, WM.; Fiscus, JG.; Pallett, DS.; Dahlgren, NL. The DARPA TIMIT acoustic-phonetic continuous speech corpus. Linguistic Data Consortium; Philadelphia: 1993.

Gathercole, SE.; Baddeley, AD. Working memory and language. Mahway, NJ: Lawrence Erlbaum Associates; 1993.

Gifford RH, Shallop JK, Peterson AM. Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. Audiol Neurootol. 2008; 13:193–205. [PubMed: 18212519]

Gilbert JL, Tamati TN, Pisoni DB. Development, Reliability and Validity of PRESTO: A New High-Variability Sentence Recognition Test. J Am Acad Audiol. in press.

Golden, CJ. Stroop Color and Word Test: A manual for clinical and experimental uses. Wood Date, IL: Stoelting Co; 1978.

Golden, CJ.; Freshwater, SM. The Stroop Color and Word Test: Revised examiner's manual. Wood Dale, IL: Stoelting Co; 2002.

Goldinger SD. Echoes of echoes: An episodic theory of lexical access. Psychol Rev. 1998; 105:251–279. [PubMed: 9577239]

Gordon-Salant S, Fitzgibbons PJ. Selected cognitive factors and speech recognition performance among young and elderly listeners. J Speech Lang Hear Res. 1997; 40:423–431. [PubMed: 9130210]

Helfer KS, Freyman RL. Lexical and indexical cues in masking by competing speech. J Acoust Soc Am. 2009; 125:447–456. [PubMed: 19173430]

Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. J Am Acad Audiol. 2007; 18:590–603. [PubMed: 18236646]

Hearing Health Quick Test [PDF document]. American Academy of Audiology; n.d. Retrieved from < http://www.howsyourhearing.org/FSQuickTest.pdf> July 9, 2012

Humphreys LG. Research on individual differences requires correlational analysis, not ANOVA. Intelligence. 1978; 2:1–5.

Jerger S, Elizondo R, Dinh T, Sanchez P, Chavira E. Linguistic influences on the auditory processing of speech by children with normal hearing or hearing impairment. Ear Hear. 1994; 15:138–160. [PubMed: 8020647]

Jerger S, Martin R, Pirozzolo F. A developmental study of the auditory stroop effect. Brain Lang. 1988; 35:86–104. [PubMed: 3179704]

Jerger S, Stout G, Kent M, Albritton E, Loiselle L, Blondeau R, Jorgenson S. Auditory stroop effects in children with hearing impairment. J Speech Hear Res. 1993; 36:1083–1096. [PubMed: 8246474]

Jesse A, Janse E. Audiovisual benefit for recognition of speech presented with single- talker noise in older listeners. Lang Cogn Process. 2012; 27:1167–1191.

Johnson, K.; Mullennix, JW. Talker Variability in Speech Processing. San Diego, CA: Academic Press; 1997.

Karl J, Pisoni DB. The role of talker-specific information in memory for spoken sentences. J Acoust Soc Am. 1994; 95:2873.

Kreiman, J.; Van Lancker Sidtis, D. Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Boston: Wiley-Blackwell; 2011.

Labov, W.; Ash, S. Understanding Birmingham. In: Bernstein, T.; Nunnally, T.; Sabino, R., editors. Language Variety in the South Revisited. Tuscaloosa, AL: Alabama University Press; 1997. p. 508-573.

Lass NJ, Hughes KR, Bowyer MD, Waters LT, Bourne VT. Speaker sex identification from voiced, whispered, and filtered isolated vowels. J Acoust Soc Am. 1976; 59:675–678. [PubMed: 1254794]

Lewellen MJ, Goldinger SD, Pisoni DB, Greene BG. Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. J Exp Psychol Gen. 1993; 122:316–330. [PubMed: 8371087]

Martin CS, Mullennix JW, Pisoni DB, Summers WV. Effects of talker variability on recall of spoken word lists. J Exp Psychol Learn Mem Cogn. 1989; 15:676–681. [PubMed: 2526857]

Mason HM. Understandability of speech in noise as affected by region of origin of speaker and listener. Speech Monographs. 1946; 13:54–68.

Mattys S, Davis MH, Bradlow AR, Scott SK. Speech recognition in adverse conditions: A review. Lang Cogn Process. 2012; 27:953–978.

McGarr NS. The intelligibility of deaf speech to experience and inexperienced listeners. J Speech Hear Res. 1983; 26:451–458. [PubMed: 6645470]

Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. J Acoust Soc Am. 1989; 85:365–378. [PubMed: 2921419]

Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. Percept Psychophys. 1990; 47:379–390. [PubMed: 2345691]

Munson B. Relationships between vocabulary size and spoken word recognition in children aged 3–7. Contemp Issues Commun Sci Disord. 2001; 28:20–29.

Munson B, Edwards J, Beckman ME. Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. J Speech Lang Hear Res. 2005a; 48:61–78. [PubMed: 15938060]

Munson B, Kurtz BA, Windsor J. The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. J Speech Lang Hear Res. 2005b; 48:1033–1047. [PubMed: 16411794]

Neff DL, Dethlefs TM. Individual differences in simultaneous masking with random- frequency, multicomponent maskers. J Acoust Soc Am. 1995; 98:125–134. [PubMed: 7608391]

Nusbaum, HC.; Magnuson, JS. Phonetic constancy as a cognitive process. In: Johnson, K.; Mullennix, JW., editors. Talker Variability in Speech Processing. San Diego: Academic Press; 1997. p. 109-132.

Nusbaum, HC.; Pisoni, DB.; Davis, CK. Research on Speech Perception Progress Report No 10. Speech Research Laboratory, Indiana University; Bloomington: 1984. Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words; p. 357-376.

Nygaard, LC. Perceptual Integration of Linguistic and Nonlinguistic Properties of Speech. In: Pisoni, DB.; Remez, RE., editors. The Handbook of Speech Perception. Oxford, UK: Blackwell Publishing Ltd; 2008. p. 390-413.

Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. Percept Psychophys. 1998; 60:355–376. [PubMed: 9599989]

Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. Psychol Sci. 1994; 5:42–46. [PubMed: 21526138]

Nygaard LC, Sommers MS, Pisoni DB. Effects of stimulus variability on perception and representation of spoken words in memory. Percept Psychophys. 1995; 57:989–1001. [PubMed: 8532502]

Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. J Exp Psychol Learn Mem Cogn. 1993; 19:309–328. [PubMed: 8454963]

Park, H-Y.; Felty, R.; Lormore, K.; Pisoni, D. PRESTO: Perceptually Robust English Sentence Test: Open-set design, philosophy, and preliminary findings. Poster presented at the 159th Acoustical Society of America Meeting; Baltimore, Maryland. 2010 Apr.

Peters, RW. Joint Project Report No 56. Vol. 56. US Naval School of Aviation Medicine; Pensacola, FL: 1955. The relative intelligibility of single-voice and multiple-voice messages under various conditions of noise; p. 1-9.

Pichora-Fuller MK, Singh G. Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. Trends Amplif. 2006; 10:29–59. [PubMed: 16528429]

Pisoni, DB. Some thoughts on "normalization" in speech perception. In: Johnson, K.; Mullennix, JW., editors. Talker Variability in Speech Processing. San Diego, CA: Academic Press; 1997. p. 9-32.

Pisoni DB. Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. Ear Hear. 2000; 21:70–78. [PubMed: 10708075]

Pisoni, DB. WordFam: Rating Word Familiarity in English. Indiana University; Bloomington: 2007.

Pisoni DB, Cleary M. Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. Ear Hear. 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. Ann Otol Rhinol Laryngol Suppl. 2000; 185:92–93. [PubMed: 11141023]

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 Hear. 2011; 32(1 Suppl):60S–74S. [PubMed: 21832890]

Pisoni DB, Nusbaum HC, Luce PA, Slowiaczek LM. Speech perception, word recognition and the structure of the lexicon. Speech Commun. 1985; 4:75–95. [PubMed: 23226910]

Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of extreme groups approach: A critical reexamination and new recommendations. Psychol Methods. 2005; 10:178–192. [PubMed: 15998176]

Ptacek PH, Sander EK. Age recognition from voice. J Speech Hear Res. 1966; 9:273–277. [PubMed: 5925532]

Richards VM, Zeng T. Informational masking in profile analysis: Comparing ideal and human observers. J Assoc Res Otolaryngol. 2001; 02:189–198. [PubMed: 11669393]

Roth, R.; Isquith, P.; Gioia, G. BRIEF-A: Behavioral Rating Inventory of Executive Function – Adult Version: Professional Manual. Lutz, FL: Psychological Assessment Resources; 2005.

Samuel, AG. The role of the lexicon in speech perception. In: Schwab, EC.; Nusbaum, HC., editors. Perception of Speech and Visual Form: Theoretical Issues, Models, and Research. New York: Academic Press; 1986.

Schneider BA, Daneman M, Murphy DR. Speech comprehension difficulties in older adults: Cognitive slowing or age-related changes in hearing? Psychol Aging. 2005; 20:261–271. [PubMed: 16029090]

Stenfelt S, Rönnberg J. The Signal-Cognition interface: Interactions between degraded auditory signals and cognitive processes. Scand J Psychol. 2009; 50:385–393. [PubMed: 19778386]

Van Bezooijen R, Gooskens C. Identification of language varieties: The contribution of different linguistic levels. J Lang Soc Psychol. 1999; 18:31–48.

Van Engen KJ, Bradlow AR. Sentence recognition in native- and foreign-language multi- talker background noise. J Acoust Soc Am. 2007; 121:519–526. [PubMed: 17297805]

Van Lancker D, Kreiman J, Emmorey K. Familiar voice recognition: Patterns and parameters: Part I. Recognition of backward voices. J Phon. 1985a; 13:19–38.

Van Lancker D, Kreiman J, Wickens T. Familiar voice recognition: Patterns and parameters: Part II. Recognition of rate-altered voices. J Phon. 1985b; 13:39–52.

van Rooij JCGM, Plomp R. Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. J Acoust Soc Am. 1990; 88:2611–2624. [PubMed: 2283434]

Vaughan N, Storzbach D, Furukawa I. Investigation of potential cognitive tests for use with older adults in audiology clinics. J Am Acad Audiol. 2008; 19:533–541. [PubMed: 19248729]

Wechsler, D. Wechsler Memory Scale—Third Edition (WMS-III) administration and scoring manual. San Antonio, TX: Harcourt Assessment; 1997.

Wechsler, D. Wechsler: Abbreviated Scale of intelligence (WASI). San Antonio, TX: Harcourt Assessment; 1999.

Wightman FL, Kistler DJ, O'Bryan A. Individual differences and age effects in a dichotic informational masking paradigm. J Acoust Soc Am. 2010; 128:270–279. [PubMed: 20649222]

Wilson RH, McArdle RA, Smith SL. An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. J Speech Lang Hear Res. 2007; 50:844–856. [PubMed: 17675590]
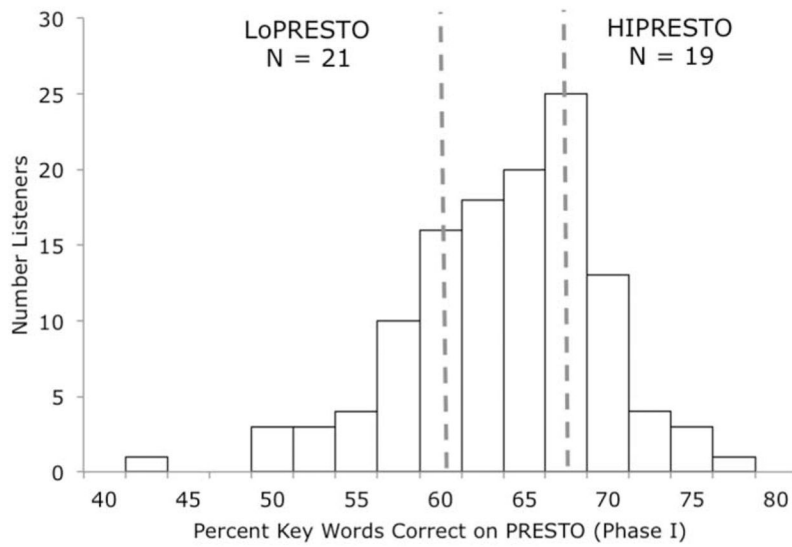
Figure 1.
Histogram of the original frequency distribution of mean keyword accuracy scores in PRESTO (Phase I) from 121 participants, reported by Gilbert et al (in press). Gray dashed vertical bars indicate divide for LoPRESTO listeners (lower quartile) and HiPRESTO listeners (upper quartile).
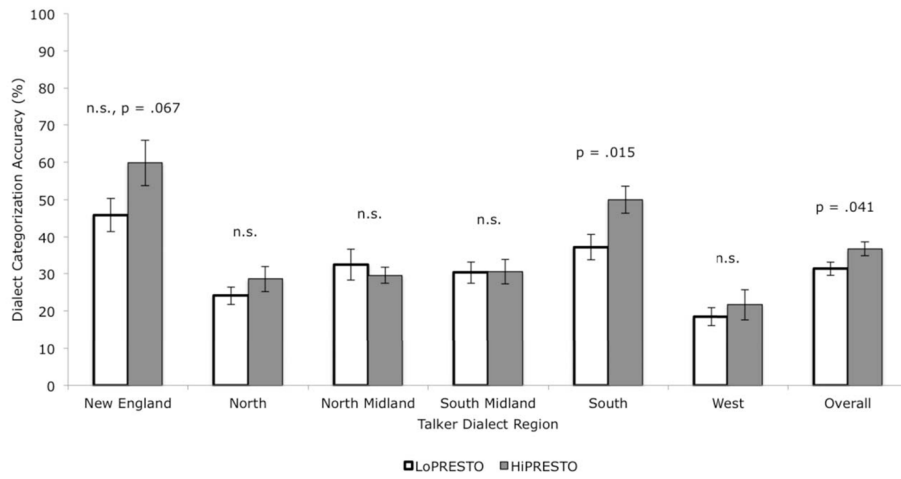
Figure 2.
Mean accuracy of regional dialect categorization for LoPRESTO and HiPRESTO groups for the six regional dialects presented and overall combined. Error bars represent standard error of the mean.
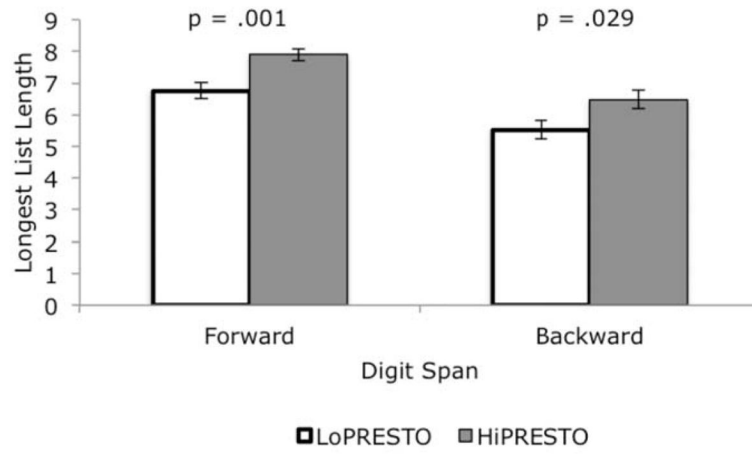
Figure 3.
Mean maximum (longest list) digit span recalled correctly in the forward and backward conditions for the LoPRESTO and HiPRESTO groups. Error bars represent standard error of the mean.
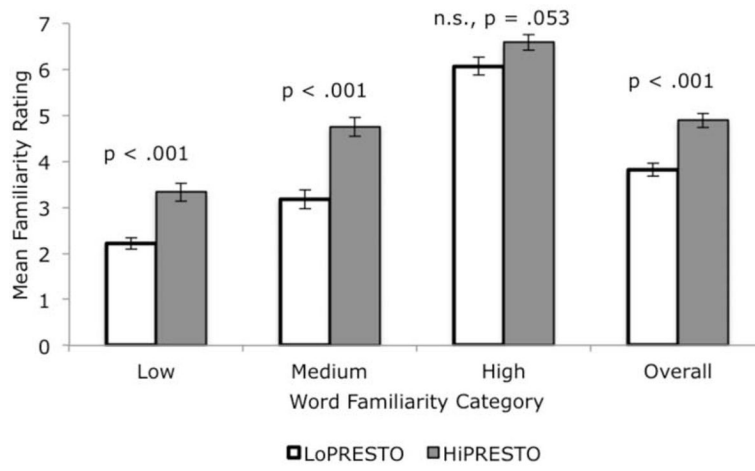
Figure 4.
Mean familiarity rating of low, medium, and high familiarity words and overall for LoPRESTO and HiPRESTO groups. Error bars represent standard error of the mean.

Table 1

Means and Standard Deviations of key words correct on PRESTO (Phase I & Phase II) in Multi-Talker Babble for 19 HiPRESTO Listeners and 21 LoPRESTO Listeners, as reported in Gilbert et al. (in press).

Signal-to-Noise Ratio	PRESTO Phase I Mean (SD)		PRESTO Phase II Mean (SD)3	
	HiPRESTO	**LoPRESTO**	**HiPRESTO**	**LoPRESTO**
Overall Mean	69.45 (2.71)	55.55 (2.95)	70.93 (3.91)	57.75 (5.13)
+3 dB	91.69 (2.76)	80.51 (6.11)	83.24 (3.50)	74.87 (5.99)
0 dB	77.56 (3.31)	64.45 (3.98)	79.36 (5.30)	66.85 (7.43)
−3 dB	62.83 (4.06)	46.60 (3.41)	69.60 (7.23)	51.44 (7.48)
−5 dB	44.94 (4.54)	30.64 (3.43)	51.52 (6.14)	37.84 (8.21)

Table 2

Means and standard deviations of gender discrimination accuracy (% correct) for LoPRESTO and HiPRESTO groups for "same" trials, "different" trials, and all trials combined.

	Same Trials	Different Trials	All Combined
LoPRESTO	92.9 (14.56)	98.2 (4.48)	95.5 (7.43)
HiPRESTO	100 (0)	98.8 (5.74)	99.4 (2.87)