Attention, Perception, & Psychophysics (2018) 80:1559–1570
https://doi.org/10.3758/s13414-018-1537-4

Coping with adversity: Individual differences in the perception of noisy and accented speech

Drew J. McLaughlin[1,2] · Melissa M. Baese-Berk[1] · Tessa Bent[3] · Stephanie A. Borrie[4] · Kristin J. Van Engen[5]

Published online: 8 May 2018

Abstract

During speech communication, both environmental noise and nonnative accents can create adverse conditions for the listener. Individuals recruit additional cognitive, linguistic, and/or perceptual resources when faced with such challenges. Furthermore, listeners vary in their ability to understand speech in adverse conditions. In the present study, we compared individuals' receptive vocabulary, inhibition, rhythm perception, and working memory with transcription accuracy (i.e., intelligibility scores) for four adverse listening conditions: native speech in speech-shaped noise, native speech with a single-talker masker, nonnative-accented speech in quiet, and nonnative-accented speech in speech-shaped noise. The results showed that intelligibility scores for similar types of adverse listening conditions (i.e., with the same environmental noise or nonnative-accented speech) significantly correlated with one another. Furthermore, receptive vocabulary positively predicted performance globally across adverse listening conditions, and working memory positively predicted performance for the nonnative-accented speech conditions. Taken together, these results indicate that some cognitive resources may be recruited for all adverse listening conditions, while specific additional resources may be engaged when people are faced with certain types of listening challenges.

Keywords Speech perception · Working memory · Inhibition

Much spoken communication appears effortless and occurs without error. However, factors such as noisy environments or speakers with unfamiliar accents can adversely affect the speech perception process. These adverse conditions may vary in their origin, but the effects for listeners are often similar: They may not understand individual words or entire utterances, and/or they may need more time than usual to accurately decode what was heard (Van Engen & Peelle, 2014). Previous research has suggested that there are vast individual differences in listeners' abilities to understand speech under adverse listening conditions (e.g., Benichov, Cox, Tun, & Wingfield, 2012; Bent, Baese-Berk, Borrie, & McKee, 2016; Wightman, Kistler, & O'Bryan, 2010). In the present study, we first compared individuals' performance in multiple types of adverse listening conditions, to determine whether listening ability in one condition was related to listening ability in all, or only in specific, other types of adverse listening conditions. Next, we examined whether several cognitive, linguistic, and perceptual skills predict individuals' proficiency in speech perception under adverse conditions, and whether particular skills are linked to aptitude with specific types of degraded and/or nonnative-accented speech.

✉ Drew J. McLaughlin
drewjmclaughlin@wustl.edu

[1] Department of Linguistics, University of Oregon, Eugene, OR, USA

[2] Present address: Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, MO 63130, USA

[3] Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN, USA

[4] Department of Communicative Disorders and Deaf Education, Utah State University, Logan, UT, USA

[5] Department of Psychological and Brain Sciences, Washington University, St. Louis, MO, USA

Types of adverse listening conditions

A review of adverse listening conditions by Mattys, Davis, Bradlow, and Scott (2012) characterizes difficult listening conditions as belonging to two main categories: environmental degradations and source degradations. Environmental degradations affect the speech signal during transmission from the speaker to the listener. Common examples regularly employed in lab research include speech in noise and speech in babble [i.e., with competing talker(s) in the background]. Both noise

and babble cause perceptual, or energetic, interference for the listener due to the physical overlap of the target signal and a nontarget signal. When there is a competing talker, informational masking can pose an additional challenge to the listener (Cooke, Garcia Lecumberri, & Barker, 2008). That is, listeners must segregate the target and competing signals, suppress nontarget signals, and manage interference caused by higher-level lexical activation when the speaker or speakers in the background are producing language that the listener can understand (Bregman, 1990). Some of the consequences of informational masking, therefore, may dissociate from those of energetic masking due to differences in the processing demands caused by the presence of higher-level interference. That is, whereas both energetic and informational masking create challenges for speech processing, the stages of language processing at which the masking interferes with speech recognition may differ.

Source degradations to the speech signal, according to the classification of Mattys et al. (2012), are caused by speaker deviations, as in disordered (e.g., dysarthric) speech or nonnative-accented speech. These deviations result in mismatches between the signal and the listener's long-term linguistic representations, making the unfamiliar speech more difficult to process (Van Engen & Peelle, 2014). In nonnative-accented speech, such mismatches may arise because of systematic segmental and/or suprasegmental differences between the talker's speech patterns and the listener's speech patterns (Adank, Evans, Stuart-Smith, & Scott, 2009; Davidson, 2011; Munro & Derwing, 1995).

For the purposes of the present study, the term "environmentally degraded speech" will be used to refer to degradations such as energetic and informational masking, whereas "nonnative-accented speech" will be referred to as such to distinguish it from other types of source degradation (e.g., speech affected by neurogenic disorders).

Cognitive, linguistic, and perceptual resources

Previous research indicated that listeners use additional cognitive, linguistic, and/or perceptual resources for speech perception under adverse conditions (Heinrich, Schneider, & Craik, 2008; Pichora-Fuller, Schneider, & Daneman, 1995; Rabbitt, 1968), and that individuals vary substantially in their ability to perceive speech in adverse listening conditions (e.g., Benichov et al., 2012; Bent et al., 2016; Wightman et al., 2010). Skills such as auditory working memory, receptive vocabulary, inhibition, and rhythm perception have been investigated as indicators of aptitude in the perception of environmentally degraded or accented speech. Below, we review previous findings on each of these measures.

Working memory The Ease of Language Understanding (ELU) model is one attempt to explain the relationship between working memory capacity and speech perception for both individuals with normal hearing and individuals with hearing loss (Rönnberg, 2003; Rönnberg, Rudner, Foo, & Lunner, 2008). One assumption of the model is that conditions such as hearing loss and noisy environments cause "mismatches" between the input and stored representations due to signal degradation or loss, and that resolving these mismatches requires explicit processing supported by working memory (Rönnberg et al., 2008). Although the model focuses on mismatches caused by signal degradations and loss, nonnative-accented speech also deviates from native listeners' representations. Given this relationship, the ELU would predict that working memory capacity is related to performance in both environmentally degraded and nonnative-accented speech conditions.

Receptive vocabulary Behavioral studies examining multiple types of challenging listening conditions have found relationships between receptive vocabulary and the intelligibility of disordered speech as well as of speech with unfamiliar accents (Banks, Gowen, Munro, & Adank, 2015; Bent et al., 2016; Janse & Adank, 2012). For example, Janse and Adank investigated listeners' perceptual adaptation to a novel constructed accent in both auditory-only and audiovisual presentations, and compared adaptation with the listeners' cognitive abilities. Vocabulary size, as well as inhibition, predicted improvement in listening accuracy over the experiment, and auditory short-term memory and working memory predicted overall listening accuracy. For speech in noise, Tamati, Gilbert, and Pisoni (2013) found that participants who performed better on speech recognition tasks in multitalker babble had significantly greater word familiarity ratings on a vocabulary questionnaire.

Inhibition In addition to working memory and receptive vocabulary, Banks et al. (2015) assessed the role of inhibition in perceptual adaptation to a novel constructed accent. Their results revealed that better inhibition scores on a Stroop test correlated with faster perceptual adaptation to the constructed accent; however, Stroop scores did not significantly correlate with overall perception of the constructed accent. The authors suggest that listeners with better Stroop scores are better able to resolve ambiguities in speech signals, meaning they are faster to correctly identify items and therefore to learn to match unfamiliar patterns in the accented speech to their existing representations. However, it is unclear why this advantage would only contribute to the rate of adaptation and not to overall performance. For environmentally degraded speech, Janse (2012) found that inhibition scores predicted the impact of a competing talker on speech-processing performance for older adults on a phoneme-monitoring task. However, another study of speech perception in multitalker babble examining young adults (Gilbert, Tamati, & Pisoni, 2013) did not find a significant relationship

between inhibition and speech perception performance. Inhibition has yet to be investigated in the context of young adults in especially difficult adverse listening conditions, such as those that combine nonnative accents and environmental degradations.

Rhythm perception Rhythm perception has been shown to predict listener performance for both source and environmental speech degradations. Slater and Kraus (2015) found that the ability to differentiate rhythms was positively related to perception scores for sentences in four-talker babble. This significant finding did not extend to the perception of individual target words in four-talker babble, suggesting that strong rhythm perception skills provide a greater advantage in sentence formats in which the temporal pattern can be identified and then bootstrapped for segmenting the speech signal. For source degradations, rhythm perception abilities do not predict initial intelligibility of dysarthric speech (a neurogenic speech disorder characterized by segmental and suprasegmental distortion), although they do predict the magnitude of intelligibility improvements following familiarization with the disordered speech signal (Borrie, Lansford, & Barrett, 2017). The authors surmise that an opportunity to learn about the disordered rhythm cues of dysarthric speech is required for the benefit of rhythm perception to be realized.

Differences between types of adverse listening conditions

The cognitive, linguistic, and perceptual skills discussed above have been shown to predict individuals' abilities when perceiving environmentally degraded and accented speech; however, some of these studies have shown relationships between the skills in question and only specific types of adverse listening conditions. This set of findings raises the question of whether there are differences in how specific types of adverse listening conditions are processed by the listener.

Studies using brain-imaging and other physiological measures have revealed some differences in how each type of adverse listening condition is processed (Adank, Davis, & Hagoort, 2012; Francis, MacPherson, Chandrasekaran, & Alvar, 2016; Miettinen, Alku, Salminen, May, & Tiitinen, 2010). For example, using fMRI, Adank et al. (2012) demonstrated that the neural systems used to process speech may differ depending on whether the speech signal has an unfamiliar constructed accent or has a familiar accent with environmental degradation. Similar results indicating differences between source and environmentally degraded speech were found using magnetoencephalography to compare brain activity during the perception of speech with reduced amplitude resolution and of speech in noise (Miettinen et al., 2010). Francis et al. (2016) found evidence of differences in physiological responses between multiple types of degraded speech; their research used unmasked natural speech (as a control condition), a speech-shaped noise masker (an energetic environmental degradation), a two-talker babble masker (an informational environmental degradation), and unmasked synthetic speech (a source degradation). Physiological measures (e.g., skin conductance as measured by electrodes and blood pulse) along with intelligibility measures (i.e., keywords recalled correctly) suggested the presence of additional or different processing demands for each of the environmental degradation conditions, in contrast with the source degradation condition; however, it could not be determined whether the difference resulted from the use of additional cognitive mechanisms or the presence of an "emotional stress-like response" (p. 13) due to the masker being more challenging and/or frustrating in one condition than in another.

Within each adverse listening condition category, listeners may also vary in their ability to accurately perceive subtypes (e.g., regional-accented speech vs. nonnative-accented speech within the category of accented speech). Dysarthric, regional-accented, and nonnative-accented speech were investigated by Bent et al. (2016), whose results showed that perception of Irish talkers (regional dialect) and Spanish-accented talkers (nonnative accent) significantly correlated, and that the scores for the Spanish-accented condition significantly correlated with those for dysarthric speech; however, there was not a significant relationship between performance in the Irish and dysarthric conditions. These results indicated that, rather than possessing a general skill for processing speech deviations, listeners may be adept at processing specific types of phonetic and prosodic deviations. All three of the investigated speech varieties deviated from standard American English both segmentally and suprasegmentally. However, of the three varieties, dysarthric and Spanish-accented speech may be more prosodically dissimilar from standard American English than the Irish dialect, whereas the segmental features of the Spanish-accented speech and Irish dialect are more dissimilar from standard American English than is dysarthric speech. If this explanation is correct, it would indicate that some listeners may be more skilled at recovering from suprasegmental deviations than from segmental deviations, or visa versa.

Similar to the perception of unfamiliar accents, differences in processing abilities within the category of environmental degradations have also been observed, specifically between energetic and informational masking. Taitelbaum-Swead and Fostick (2016) examined younger and older listener groups in three background noise conditions (speech-shaped noise, babble noise, and white noise) at multiple signal-to-noise ratios (SNRs). The results showed that an increase in noise caused a significantly greater reduction in participant accuracy in the babble condition than in the other noise conditions, and that there was an interaction with age (i.e., a greater decrease in accuracy for older than for younger adults in the babble condition). Furthermore, a speech recognition training study

in speech-shaped noise, Mandarin babble, and English babble showed that intelligibility improved after training in the babble conditions, but not in the speech-shaped noise condition (Van Engen, 2012). These results suggested that speech recognition training for informational masking may be more effective than training for energetic masking.

Research examining listener performance in multiple types of adverse listening conditions has in some cases shown relationships between conditions (e.g., Borrie, Baese-Berk, Van Engen, & Bent, 2017), but in other cases has shown disassociations (e.g., Bent et al., 2016). The aim of the present study was to examine multiple types of adverse listening conditions and the potential cognitive, linguistic, and perceptual skills that support success in each condition. In the case of working memory, the ELU model (Rönnberg, 2003; Rönnberg et al., 2008) posits that mismatches between the incoming signal and stored representations require explicit processing supported by working memory. Success with both nonnative-accented speech and speech in noise, then, was predicted to be related to working memory capacity.

Rhythm perception was expected to positively predict performance for speech in noise (e.g., Parbery-Clark, Skoe, Lam, & Kraus, 2009; Slater & Kraus, 2015). For nonnative-accented speech, we expected that rhythm perception would not positively predict performance, given the unfamiliar rhythmic qualities that often characterize nonnative-accented speech. Receptive vocabulary, on the other hand, has been correlated with both disordered and accented speech (Bent et al., 2016; Janse & Adank, 2012; McAuliffe, Gibson, Kerr, Anderson, & LaShell, 2013). However, vocabulary knowledge has not been compared with both nonnative-accented speech and environmentally degraded speech within the same experiment. Perhaps more importantly, including this measure in our analyses allows us to statistically control for vocabulary knowledge. Given the previous findings, we expect receptive vocabulary to predict performance for multiple adverse listening conditions.

Inhibition has been related to listening performance for environmentally degraded speech in older adults (Janse, 2012), but not in young adults (Gilbert et al., 2013). However, inhibition has not been investigated, to our knowledge, for young adult populations' perception of especially difficult combinations of adverse listening conditions. Thus, we expected that if inhibition were related to listening performance in young adults, this relationship would only hold for conditions of comparably low intelligibility (i.e., nonnative-accented speech with environmental degradation).

By examining the adverse listening conditions and cognitive, linguistic, and perceptual skills discussed above, we aimed to address two key research questions: first, whether performance on one adverse listening condition is related to performance on all, or only on specific, different types of adverse conditions, and, second, whether specific types of adverse listening conditions are linked to particular cognitive, linguistic, and perceptual skills.

Method

Participants

Participants (n = 65) were recruited using the University of Oregon's Psychology and Linguistics Human Subjects Pool with approval from University of Oregon's Institutional Review Board. The target sample size of 50 participants was chosen on the basis of comparable studies (Bent et al., 2016; Borrie, Lansford, et al., 2017). Participants were compensated for 2 h of participation, either with $20 or with class participation credit. Advertisements for the study indicated that participants must be monolingual speakers of English with normal hearing in order to participate. However, in total, 14 participants had to be excluded from the analyses: Seven of the participants were excluded because they were bilingual or had extensive exposure to Spanish-accented speech; three because they were not native speakers of American-English; and four because they did not pass the hearing screening. The majority of the unqualified participants were recruited via the Human Subjects Pool for class participation credit, and, thus, were not turned away. These participants were replaced, resulting in data from 51 monolingual American-English participants with normal hearing (see below for details) for our analysis. Of the 51 participants included in the analyses, 36 self-identified as female and 15 self-identified as male. The age range of the participants was 18–31 years old (M = 20.61, SD = 2.36).

Experimental procedure

Participants completed a series of tasks including: a hearing test, a phrase recognition task, the Peabody Picture Vocabulary Test (PPVT-4; Dunn, Dunn, & Pearson Assessment, 2007), the color Stroop test (Stroop, 1935), the rhythm perception subtest of the Musical Ear Test (MET; Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), and the Word Auditory Recognition and Recall Measure (WARRM; S. L. Smith, Pichora-Fuller, & Alexander, 2016). All of the tasks were administered on a Mac OS X computer in a quiet room, and all auditory stimuli were played through Sennheiser HD 202 headphones at the same predetermined comfortable listening level. Before beginning, the participants also filled out a questionnaire regarding their language experience and background. With the exception of the hearing test and the phrase recognition test (which were administered first), the order of the tasks was randomized for each participant. The raw data and program code are available on Github (https://github.com/drewjmclaughlin/Multi-ACs-Study) for public access.

The machine-learning hearing test The online hearing test ML Audiogram (Song, Garnett, & Barbour, 2017; Song et al., 2015) was used to estimate the hearing thresholds of participants and determine whether they had normal hearing. ML Audiogram is a machine-learning hearing test, which is designed to adapt to listeners' responses in order to quickly and accurately estimate a hearing threshold. The main computer's volume setting was calibrated using a sound pressure level meter. Participants were instructed to listen for a sequence of three short beeps and press the spacebar on the keyboard whenever they heard them. Before the test began, an example of the three short beeps was played for the participant at an intensity of 50 dB SPL and a frequency of 2000 Hz. During the test, the thresholds at octave frequencies within a range of 250 to 8000 Hz were assessed.

The phrase recognition task The phrase recognition test included stimuli in four conditions of environmentally degraded and/or nonnative-accented speech: a native speaker masked in speech-shaped noise (environmental degradation via energetic masking), a native speaker with a single-talker masker (environmental degradation via informational masking), a nonnative speaker in quiet (source variation), and a nonnative speaker masked in speech-shaped noise (source variation and environmental degradation via energetic masking). These four conditions will be abbreviated NE, NI, NNQ, and NNE, respectively.

The stimuli were created using semantically anomalous phrases, developed by Liss, Spitzer, Caviness, Adler, and Edwards (1998) and modeled on similar phrases used by Cutler and Butterfield (1992). These phrases contained real English words in normal syntactic frames; however, the words lacked meaning and context (e.g., "account for who could knock" and "cheap control in paper"). As in previous studies, these types of phrases were used because they reduce the use of top-down processing—thus preventing the listener from inferring misperceived words on the basis of the context. A male, native standard American English speaker was recorded reading 80 semantically anomalous phrases in a quiet environment for the NE and NI conditions. For the NNQ and NNE conditions, a male speaker with Spanish-accented English (i.e., a speaker whose native language is Spanish) was recorded reading the same 80 phrases.

To create the informational masking condition, a second male native English speaker was recorded reading a different set of 80 semantically anomalous phrases in a quiet environment; these phrases were then edited into one continuous sound file to create a single-talker masker. A speech-shaped noise file was created in Praat to match the spectral properties of the target sentences. Both energetic masking conditions (NE and NNE) used the same speech-shaped noise file. A Python program was written such that each masking condition was created by combining the target phrases (i.e., the phrases that the listener is asked to transcribe) with randomly selected sections of the masker files. The masker files began 500 ms before the onset of the target phrase and continued 500 ms after the target phrase ended. This procedure ensured that each participant had a unique combination of target phrase and masking noise, and any behavior on a particular item across listeners could not be attributed to specific qualities of the masker.

Each masking condition was mixed at a specific SNR determined by the results from pilot testing of the stimuli on Amazon Mechanical Turk (a crowdsourcing website through which human subjects can be recruited for research). The average intelligibility of the NNQ condition ($M = .63$, $SD = .06$) was known from results of a previous study in which the same speaker recordings were used (Bent et al., 2016), and the SNRs of the NE (– 2 dB) and NI (– 5 dB) conditions were chosen to approximate this intelligibility level. Because adding noise to the NNQ condition would reduce its intelligibility, the NNE condition was not expected to match the others for intelligibility. After piloting a range of SNRs, the NNE condition was presented at 0 dB SNR in order to avoid floor effects.

There were four practice trials (one for each condition) before the actual experimental trials began. Each adverse listening condition contained 20 trials, for a total of 80 trials. The order of the trials was randomized for each participant, as was the assignment of each phrase to each adverse listening condition. Thus, the phrase "account for who could knock" may have appeared in the NE condition for one participant and in the NNQ condition for another participant.

Participants were provided with both verbal and on-screen written instructions before beginning the phrase recognition task. They were instructed to pay close attention to each phrase and to try to determine what had been said. They were also instructed to take their best guess if they were unsure of what they had heard. After each phrase was played, a box appeared on the screen for the participant to type a response in before the next trial began. For the NI condition, participants were told to pay attention to the talker who began speaking half a second after the first talker. Participants were not able to replay the stimuli and were not provided with any feedback.

The Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) The PPVT-4 is a standardized test that measures receptive vocabulary (Dunn et al., 2007). Although the PPVT is often used to assess younger age groups, it can also be used to measure vocabulary for adults up to 90 years of age. In the present study, an online version of the test was administered via Q-global (https://qglobal.pearsonclinical.com). For each trial, a single word would play over the headphones, and the participant would choose the one of four illustrations that best represented the word. The PPVT-4 has 19 sets of vocabulary

items with 12 items per set, which are presented to the participant with increasing difficulty. If the participant makes one or fewer errors within a set, then it is considered a *basal set*, and the participant proceeds on to a more difficult set. However, if participants make more than one error, then they return to a less difficult set so that a basal set can be determined. Testing is concluded when a participant responds incorrectly for eight or more items in a set, which is considered their *ceiling set*. Participants could replay the word as many times as needed.

The color Stroop test The color version of the Stroop test from the PEBL Test Battery (Mueller & Piper, 2014) was used to measure inhibition (Stroop, 1935). In each trial, a word appeared in the middle of the screen, and participants used the horizontal numbers on the keyboard to indicate the word's font color. Four colors appeared in the task, each corresponding to a number (e.g., 1 = red), and a reference key was displayed at the bottom of the screen throughout the test. Three conditions were present in the test: congruent, incongruent, and neutral. *Congruent* conditions were those in which the word on the screen appeared in the matching color (e.g., the word "red" written in the color red); incongruent conditions were those in which the word on the screen appeared in a different color (e.g., the word "red" written in the color green); and neutral conditions were those in which the word on the screen did not correspond to any particular color (e.g., the word "when" written in any color). Response times were averaged for each condition, and then the difference between the incongruent and congruent conditions was calculated as a measure of participants' inhibition.

The Rhythm Subsection of the Musical Ear Test (R-MET) The R-MET was used to determine individuals' rhythm perception abilities in the musical domain (Wallentin et al., 2010). In each of the 52 trials, participants listened to two sets of prerecorded beats played on a wood block and, using a forced choice paradigm, had to decide whether the beats composed the same or different rhythms. Participants marked their responses on a paper answer sheet. Before the scored portion of the test began, a recording of verbal instructions was played for the participants over headphones, and then two practice rounds were given with correct answers provided. Participants were not allowed to repeat trials.

The Word Auditory Recognition and Recall Measure (WARRM) WARRM is a working memory task developed for rehabilitative audiology (S. L. Smith et al., 2016). In the present study, recall measures from the task were used to estimate individuals' working memory. Participants were randomly assigned to one of three versions of the WARRM test, in which auditory stimuli are played in different orders. Before beginning the task, participants were given instructions via a short PowerPoint presentation. Auditory stimuli were played over headphones at a comfortable level, and participant responses were recorded during the experiment by an experimenter. Participants were given two practice trials prior to the start of the test trials. Target words were presented to the listener in the carrier phrase "You will cite ___." Following this sentence, the listener was instructed to first repeat the target word out loud, and then to make a judgment as to whether the first letter of the target word was from the first or the second half of the alphabet (i.e., the listener would say "first" if the letter was between *A* and *M*, and "second" if the letter was between *N* and *Z*). At the end of the trial, a beep played, indicating to listeners that they should recall the target words from the set. If listeners could not remember all of the words in the set, they were instructed to take a guess or to move on to the next trial. There were five trials for each set size, beginning with two words per trial and ending with six words per trial. Participants were given two practice trials prior to the start of the test trials and allowed to take a short break between trials, if needed.

Analysis

Participant transcriptions from the phrase recognition test were scored in order to derive intelligibility scores. In addition to words that matched the intended target precisely, homophones or obvious misspellings of the target word were scored as correct, as were differences in tense, plurality, and substitutions between "a" and "the" (e.g., "account for who could knock" would be worth five points total). This scoring procedure was chosen to match previous literature that had used the same semantically anomalous phrases (Borrie et al., 2012; Liss et al., 1998). For each participant, a measure of intelligibility was calculated on the basis of the proportion of words correctly transcribed in each of the four listening conditions (Table 1). Measures from the PPVT-4 (receptive vocabulary), the Stroop test (inhibition), the R-MET (rhythm perception), and the WARRM test (working memory) were either automatically scored by the testing software or manually scored using the standard protocols (Table 2). For the PPVT-4, a measure of percentile rank was used; for the color Stroop test, a measure of the difference in reaction times between the congruent and incongruent conditions was used; for the WARRM, a measure of auditory word span (i.e., working memory capacity) was used; and for the R-MET, a measure of the proportion of answers correct was used. For the color Stroop test, an additional analysis in which outliers were removed from each participants' response set was also conducted and returned similar results; thus, only the initial analysis with all participants' responses included is reported below.

The measures of participants' cognitive, linguistic, and perceptual skills were analyzed in a logistic mixed-effects model with the intelligibility of each keyword from the phrase recognition task as the dependent variable. Fixed factors for the model included the scores for receptive vocabulary, inhibition,

Table 1 Descriptive statistics of intelligibility scores from the phrase recognition task for each environmentally degraded and/or nonnative-accented speech condition

Comparison	NE	NI	NNE	NNQ
Mean	.57	.63	.25	.62
Standard dev.	.07	.13	.07	.08
Max	.74	.88	.40	.79
Min	.41	.30	.06	.44

Intelligibility scores are calculations of proportions of words correct. Abbreviations are as follows: Native speaker in energetic masking (NE), native speaker in informational masking (NI), nonnative Spanish-accented speaker in energetic masking (NNE), and nonnative Spanish-accented speaker in quiet (NNQ)

working memory, and rhythm perception; type of adverse condition (i.e., NE, NI, NNQ, and NNE); and the interactions between each cognitive, linguistic, or perceptual measure and each adverse condition. Scores from the four cognitive, linguistic, and perceptual measures were all centered and scaled prior to entering them into the model. Although ideally the models would also include target words as random effects, the maximal random-effects structure that would allow the models to converge included participants as random intercepts only.

A series of model comparisons was used to determine the significance of each fixed factor. On the basis of these comparisons, it was determined that the model of best fit (Table 3) included the fixed factors of adverse listening condition (i.e., NE, NI, NNQ, and NNE), receptive vocabulary, working memory, the interaction between rhythm perception and adverse listening condition, and the interaction between working memory and adverse listening condition. Both the measure of inhibition and the interaction between inhibition and adverse listening condition were excluded from the model of best fit because they did not significantly improve the fit of the model [$\chi^2(1) = 0.7669$, $p = .381$, and $\chi^2(3) = 6.394$, $p = .094$, respectively]. The measures of cognitive, linguistic, and perceptual skills that were significant predictors and those interactions that significantly improved the model fit are discussed further below.

Table 2 Descriptive statistics of scores from the cognitive, linguistic, and perceptual skill tasks

Comparison	Receptive vocabulary (percentile ranking)	Rhythm perception (proportions of answers correct)	Inhibition (ms)	Working memory (auditory word span)
Mean	65.66	.69	138.08	4.01
Standard dev.	21.40	.10	69.85	0.91
Max	97	.90	330.21	6
Min	19	.42	19.85	2.67

Table 3 Logistic mixed-effects model of best fit

Predictor	Estimate	Standard error	*z* value
(Intercept: NE)	0.26661	0.04167	6.398*
NI	0.23223	0.04574	5.077*
NNE	– 1.36319	0.04864	– 28.024*
NNQ	0.20713	0.04567	4.535*
PPVT-4	0.11040	0.03437	3.212*
R-MET	0.05061	0.05026	1.007
WARRM	– 0.03283	0.04856	– 0.676
NI : R-MET	– 0.02000	0.05326	– 0.376
NNE : R-MET	– 0.06187	0.05674	– 1.091
NNQ : R-MET	– 0.18325	0.05304	– 3.455*
NI : WARRM	0.10793	0.05354	2.016*
NNE : WARRM	0.19517	0.05563	3.509*
NNQ : WARRM	0.16727	0.05349	3.127*

Asterisks reflect a $p < .05$

Abbreviations are as follows: Native speaker in energetic masking (NE), native speaker in informational masking (NI), nonnative Spanish-accented speaker in energetic masking (NNE), nonnative Spanish-accented speaker in quiet (NNQ), the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), the rhythm subsection of the Musical Ear Test (R-MET), and the Word Auditory Recognition and Recall Measure (WARRM)

Results

Intelligibility scores for each of the four listening conditions are reported in Table 1. The results from the pairwise correlations among the adverse listening conditions showed two significant correlations (the reported significant results reflect a Bonferroni-corrected *p* value of .008): the scores for the two conditions with energetic masking, NE and NNE ($r = .45$, $p = .001$; Fig. 1) and the scores for the two conditions with nonnative-accented speech, NNQ and NNE ($r = .43$, $p = .002$; Fig. 1). We found no significant correlation between the NE and NI conditions ($r = .32$, $p = .023$), the NE and NNQ conditions ($r = .14$, $p = .315$), the NI and NNQ conditions ($r = .25$, $p = .081$), or the NI and NNE conditions ($r = .13$, $p = .351$). These results indicate that performance in one type of adverse listening condition does not predict performance in all of the other types of adverse listening conditions. The lack of a significant correlation between the NE and NI conditions is particularly notable, since listeners were presented with the same native talker in those conditions. Furthermore, the relationships between these adverse listening conditions suggest that listeners may be adept at perceiving speech under similar types of adverse listening conditions. That is, there were significant correlations between the NNE

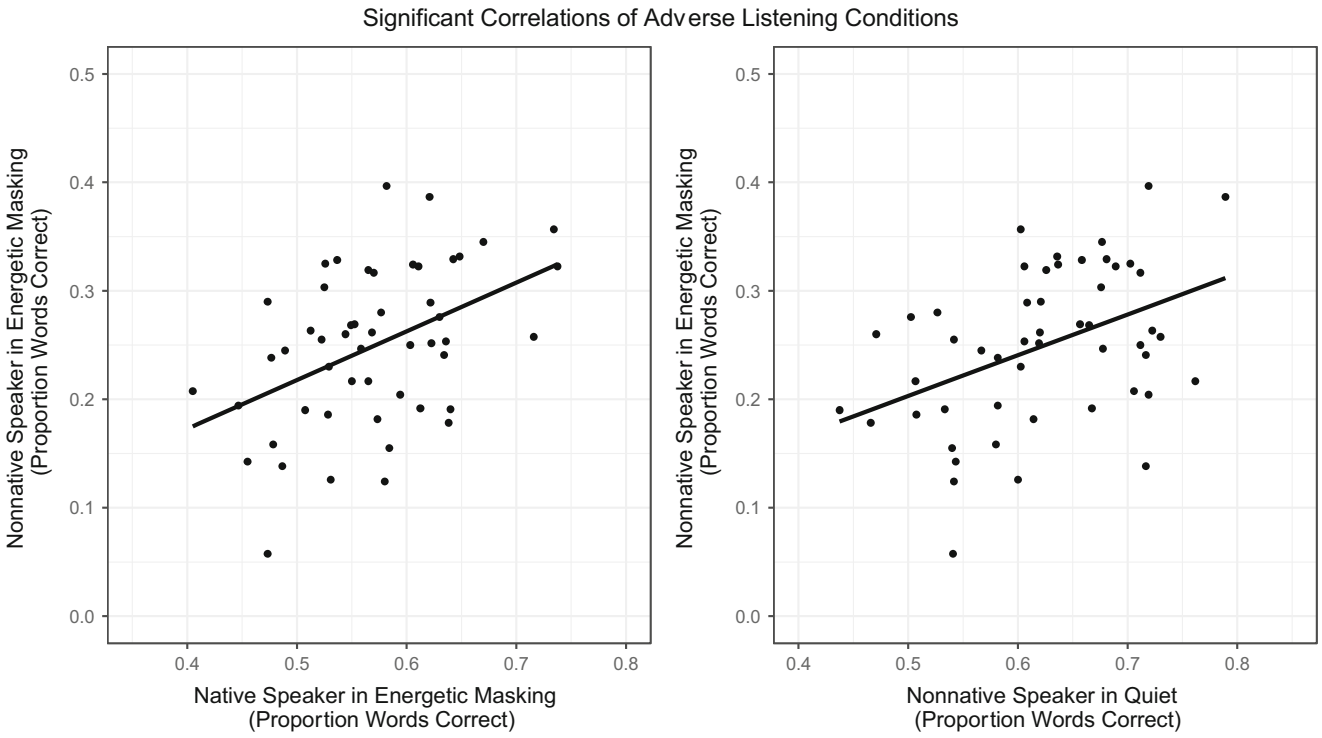

Fig. 1 Significant correlations between the conditions of a native speaker in energetic masking and the nonnative speaker in energetic masking ($r = .45$, $p = .001$; left) and of the nonnative speaker in quiet and the nonnative speaker in energetic masking ($r = .43$, $p = .002$; right)

and NE conditions (both of which have energetic masking) and the NNE and NNQ conditions (both of which have nonnative-accented speech; Fig. 1).

A correlation matrix of performance on the four cognitive, linguistic, and perceptual skills is reported in Table 4 (for statistical significance, the critical p value is .008, as determined by Bonferroni corrections). Rhythm perception significantly correlated with working memory capacity ($r = .50$, $p < .008$), which was anticipated, because the R-MET task requires listeners to remember two rhythmic sequences and determine whether they are the same or different. R-MET was also correlated with receptive vocabulary ($r = .40$, $p < .008$). All variance inflation factors were less than 3, suggesting that collinearity does not pose a problem in the logistic mixed-effects model (Miles, 2005).

Table 4 Correlation matrix of cognitive, linguistic, and perceptual skills

	PPVT-4	R-MET	Stroop
R-MET	.40*		
Stroop	.16	– .01	
WARRM	.24	.50*	.06

Abbreviations are as follows: the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), the rhythm subsection of the Musical Ear Test (R-MET), and the Word Auditory Recognition and Recall Measure (WARRM). * Bonferroni corrections were applied such that $p < .008$ for significant values

Within the logistic mixed-effects model, receptive vocabulary significantly improved the model fit [$\chi^2(1) = 9.403$, $p = .002$], but the interaction between receptive vocabulary and adverse listening condition was not significant [$\chi^2(3) = 3.450$, $p = .327$]. These results indicate that receptive vocabulary may be a positive global predictor of perceptual expertise during adverse listening conditions. Both the measure of working memory and the interaction between working memory and adverse listening condition significantly improved the model fit [$\chi^2(1) = 4.790$, $p = .029$, and $\chi^2(3) = 14.767$, $p = .002$, respectively]. The measure of rhythm perception did not significantly improve the model fit [$\chi^2(1) = 0.171$, $p = .679$]; however, the interaction between rhythm perception and adverse listening condition did significantly improve the model fit [$\chi^2(3) = 13.192$, $p = .004$].

To explore the interactions of working memory and rhythm perception, additional versions of the model of best fit with alternative intercepts (i.e., NI, NNQ, and NNE in place of NE) were constructed (Appendix 1). This analysis allowed us to observe the direct relationships between working memory and rhythm perception with each adverse listening condition. We recommend that these additional models be interpreted with some caution, but note that they confirm what the interactions in the original model indicated: Working memory positively predicted performance in the NNQ and NNE conditions, and rhythm perception negatively predicted performance in the NNQ condition. These findings suggest that working memory capacity may play a larger role in the perception of nonnative-

accented speech than of native speech in noise, and that better rhythm perception may actually hinder a listener's ability to understand a nonnative accent.

Discussion

Although listeners, in general, may experience difficulty understanding both environmentally degraded and nonnative-accented speech, their ability to understand speech depends upon the type of adverse listening condition they are faced with. By examining intelligibility performance in four adverse listening conditions, we have shown that performance in one type of adverse listening condition is not related to performance on all other types of adverse listening conditions; rather, such relationships exist within specific types of adverse listening conditions, such as those that had the same accented speaker or same type of environmental degradation. When examining the roles of cognitive, linguistic, and perceptual skills in these various listening conditions, we found the following: Receptive vocabulary positively predicts performance across adverse listening conditions; working memory positively predicts performance for conditions with nonnative-accented speech; and rhythm perception negatively predicts performance for nonnative-accented speech under quiet conditions. Inhibition (as measured with the Stroop task) did not significantly predict performance for any of the tested conditions.

The pairwise correlations of adverse listening conditions revealed that conditions with similar sources of difficulty were significantly related. Specifically, the two conditions with energetic masking, NE and NNE, were significantly correlated, as were the two nonnative-speaker conditions, NNQ and NNE. It should be noted that the NE and NI conditions, which included recordings from the same native speaker with different masker types, did not significantly correlate with one another. Thus, whereas the relationship between the NNQ and NNE conditions could be explained by the fact that the stimuli were recorded by the same speaker, this factor is unlikely to be the sole explanation of the result, given the lack of a significant correlation between conditions using a native speaker. Rather, the results of the present study suggest that performance on similar types of adverse listening conditions may be correlated because they recruit the same cognitive, linguistic, and/or perceptual skills.

Inhibition was not found to be a significant predictor of performance for the four adverse listening conditions, nor was it correlated with any of the other predictor variables. This result is in line with previous research on environmental speech degradations that had suggested that inhibition was predictive for older adults (Janse, 2012) but not younger adults (Gilbert et al., 2013). Similarly, for the NNQ condition, the lack of a correlation accords with the results of Banks et al. (2015), which revealed a relationship between inhibition and perceptual adaptation to an unfamiliar constructed accent, but not between inhibition and overall recognition scores. However, for the NNE condition it was predicted that inhibition scores might be related to listener performance, because of the especially difficult combination of nonnative accent and environmental degradation. Since this result was not found, the present results, in combination with the results of previous research, suggest that in the case of phrase-level speech recognition, inhibition may only be predictive of listening in adverse conditions for older adult populations.

In addition to the relationships demonstrated in the present study with environmentally degraded and/or nonnative-accented speech types, receptive vocabulary has been found to positively predict listener performance for Irish English (i.e., an unfamiliar dialect) and dysarthric speech (i.e., a source degradation; Bent et al., 2016), indicating a robust relationship between receptive vocabulary and multiple types of adverse listening conditions. Although in the present study semantically anomalous phrases were employed, vocabulary has been associated with the ability to anticipate future words (Borovsky, Elman, & Fernald, 2012), and for accented speech, lexical knowledge may be crucial for perceptual adaptation (Norris, McQueen, & Cutler, 2003). Thus, it may be the case that listeners with greater receptive vocabularies perform better in adverse listening conditions because they have stronger lexical mappings that allow them to access semantic representations from input even when it is environmentally degraded or accented in an unfamiliar way, or, conversely, that listeners who perform well in adverse listening conditions are able to acquire larger vocabularies because they are better able to understand speech from a variety of talkers and in a variety of situations.

Rhythm perception predicted poorer listening performance for nonnative-accented speech in quiet and did not positively predict native speech in noise, as had been shown in previous research (Parbery-Clark et al., 2009; Slater & Kraus, 2015). For nonnative-accented speech in noise, rhythm perception scores were not significantly predictive of intelligibility scores, although there was a negative trend. These results indicate that expertise in rhythm perception may in fact be *disadvantageous* for the perception of nonnative-accented speech. One interpretation of this finding is that the differing rhythmic structure of non-native speech (in this case, Spanish-accented English) may "mislead" rhythmically skilled English listeners—an interpretation rooted in the previously advanced notion that listeners with expertise in rhythm perception may be better equipped to exploit the rhythm cues of English to segment speech in adverse listening conditions (Borrie, Lansford, et al., 2017). In English, strong syllables (those receiving relative stress through longer

duration, fundamental frequency change, increased loudness, and a relatively full vowel) can be used to identify the onset of a new word (Cutler & Norris, 1988). Exploiting this statistical structure has been shown to be particularly useful in adverse listening conditions such as speech in noise (M. R. Smith, Cutler, Butterfield, & Nimmo-Smith, 1989) and dysarthric speech (Borrie, Baese-Berk, et al., 2017), although large individual variation in the degree to which listeners exploit this strategy has been observed (Borrie, Baese-Berk, et al., 2017). In languages such as Spanish, however, speakers do not produce large differences in syllable stress; rather, syllables are relatively isochronous (White & Mattys, 2007). Thus, English listeners with expertise in rhythm, likely the same listeners who rely heavily on syllabic stress contrast cues to segment speech, may be increasingly challenged by Spanish-accented English, in which less robust stress cues are available in the speech signal to exploit than in native English speech. This speculation warrants further exploration, particularly because the rhythmic disadvantage was not significant in the case of nonnative-accented speech in noise.

For working memory capacity, greater capacity was predictive of better performance in adverse listening conditions with nonnative-accented speech. Previous research has suggested a similar relationship between working memory and the perception of constructed unfamiliar accents (Banks et al., 2015; Janse & Adank, 2012), indicating that working memory may be recruited during the perception of nonnative-accented speech, so that the signal can be stored and reanalyzed when it deviates from the listener's long-term representations. On the basis of the ELU model (Rönnberg et al., 2008), we also predicted that working memory would significantly predict performance in the speech-in-noise conditions. This prediction was not supported by our data, although the null relationship found between working memory and the two environmentally degraded native speaker conditions is in accord with Füllgrabe and Rosen (2016). Thus, working memory may play an important role in processing nonnative-accented speech in young adult listeners, but only play a role for processing speech in noise in later adulthood.

One of the most notable findings of the present study was that performance under one type of adverse listening condition did not predict performance for all other adverse listening conditions. Specifically, good performance in noisy conditions did not entail good performance on nonnative-accented speech. This result indicates that listeners employ different cognitive and perceptual strategies to cope with various types of environmentally degraded and/or accented speech. Additionally, some skills, such as receptive vocabulary, may be important for all types of adverse listening conditions, whereas others, such as working memory, may be more important for specific types of adverse listening conditions.

Author note This work was partially funded by an undergraduate fellowship awarded to D.J.M. by the Office of the Vice President for Research and Innovation at the University of Oregon, and by a University of Oregon Faculty Research Award given to M.M.B.-B.

Appendix 1

Table 5 Alternative intercepts for the logistic mixed-effects model of best fit

Predictor	Estimate	Standard error	*z* value
1. NI intercept			
(Intercept: NI)	0.49884	0.04214	11.84*
NE	– 0.23223	0.04574	– 5.08*
NNE	– 1.59541	0.04907	– 32.52*
NNQ	0.02510	0.04609	– 0.54
PPVT-4	0.11039	0.03437	3.21*
R-MET	0.03060	0.05054	0.61
WARRM	0.07511	0.04914	1.53
NE : R-MET	– 0.02001	0.05326	0.38
NNE : R-MET	– 0.04186	0.05700	– 0.73
NNQ : R-MET	– 0.16325	0.05328	– 3.06*
NE : WARRM	– 0.10793	0.05353	– 2.02*
NNE : WARRM	0.08723	0.05615	1.55
NNQ : WARRM	0.05935	0.05401	1.10
2. NNQ intercept			
(Intercept: NNQ)	0.47374	0.04206	11.26*
NE	– 0.20712	0.04567	– 4.53*
NI	0.02511	0.04609	0.54
NNE	– 1.57031	0.04900	– 32.05*
PPVT-4	0.11039	0.03437	3.21*
R-MET	– 0.13265	0.05032	– 2.64*
WARRM	0.13444	0.04908	2.74*
NE : R-MET	0.18326	0.05305	3.45*
NI : R-MET	0.16326	0.05329	3.06*
NNE : R-MET	0.12138	0.05680	2.14*
NE : WARRM	– 0.16728	0.05349	– 3.13*
NI : WARRM	– 0.05933	0.05402	– 1.10
NNE : WARRM	– 0.16728	0.05349	– 3.13*
3. NNE intercept			
(Intercept: NNE)	– 1.09657	0.04526	– 24.23*
NE	1.36319	0.04865	28.02*
NI	1.59541	0.04907	32.51*
NNQ	1.57031	0.04900	32.05*
PPVT-4	0.11039	0.03437	3.21*
R-MET	– 0.01125	0.05418	– 0.21
WARRM	0.16233	0.05136	3.16*
NE : R-MET	0.06186	0.05674	1.09
NI : R-MET	0.04185	0.05701	0.73
NNQ : R-MET	– 0.12138	0.05679	– 2.14*
NE : WARRM	– 0.19517	0.05562	– 3.51*
NI : WARRM	– 0.08722	0.05616	– 1.55
NNQ : WARRM	– 0.02790	0.05608	– 0.50

Asterisks reflect a $p < .05$

Abbreviations are as follows: Native speaker in energetic masking (NE), native speaker in informational masking (NI), nonnative Spanish-accented speaker in energetic masking (NNE), nonnative Spanish-accented speaker in quiet (NNQ), the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), the rhythm subsection of the Musical Ear Test (R-MET), and the Word Auditory Recognition and Recall Measure (WARRM)

References

Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, *50*, 77–84. doi:https://doi.org/10.1016/j.neuropsychologia.2011.10.024

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 520–529. doi:https://doi.org/10.1037/a0013552

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *Journal of the Acoustical Society of America*, *137*, 2015–2024. doi:https://doi.org/10.1121/1.4916265

Benichov, J., Cox, L. C., Tun, P. A., & Wingfield, A. (2012). Word recognition within a linguistic context: Effects of age, hearing acuity, verbal ability, and cognitive function. *Ear and Hearing*, *33*, 262–268. doi:https://doi.org/10.1097/Aud.0b013e31822f680f

Bent, T., Baese-Berk, M., Borrie, S., & McKee, M. (2016). Individual differences in the perception of unfamiliar regional, nonnative, and disordered speech varieties. *Journal of the Acoustical Society of America*, *140*, 3775–3786. doi:https://doi.org/10.1121/1.4966677

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*, 417–436. doi:https://doi.org/10.1016/j.jecp.2012.01.005

Borrie, S. A., Baese-Berk, M., Van Engen, K., & Bent, T. (2017). A relationship between processing speech in noise and dysarthric speech. *Journal of the Acoustical Society of America*, *141*, 4660–4667. doi:https://doi.org/10.1121/1.4986746

Borrie, S. A., Lansford, K. L., & Barrett, T. S. (2017). Rhythm perception and its role in perception and learning of dysrhythmic speech. *Journal of Speech, Language, and Hearing Research*, *60*, 561–570. doi:https://doi.org/10.1044/2016_JSLHR-S-16-0094

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes*, *27*, 1039–1055. doi:https://doi.org/10.1080/01690965.2011.610596

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Cooke, M. P., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail effect party problem: Energetic and informational masking effects in non-native speech perception. *Journal of the Acoustical Society of America*, *123*, 414–427.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121. doi:https://doi.org/10.1037/0096-1523.14.1.113

Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 270–282.

Dunn, L. M., Dunn, D. M., & Pearson Assessment. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Minneapolis, MN: Pearson Assessment.

Francis, A. L., MacPherson, M. K., Chandrasekaran, B., & Alvar, A. M. (2016). Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in Psychology*, *7*, 263. doi:https://doi.org/10.3389/fpsyg.2016.00263

Füllgrabe, C., & Rosen, S. (2016). Investigating the role of working memory in speech-in-noise identification for listeners with normal hearing. In *Physiology, psychoacoustics and cognition in normal and impaired hearing* (pp. 29–36). New York, NY: Springer. doi:https://doi.org/10.1007/978-3-319-25474-6_4

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology*, *24*, 26–36. doi:https://doi.org/10.3766/jaaa.24.1.4

Heinrich, A., Schneider, B. A., & Craik, F. I. M. (2008). Investigating the influence of continuous babble on auditory short-term memory performance. *Quarterly Journal of Experimental Psychology*, *65*, 735–751. doi:https://doi.org/10.1080/17470210701402372

Janse, E. (2012). A non-auditory measure of interference predicts distraction by competing speech in older adults. *Aging, Neuropsychology, and Cognition*, *19*, 741–758. doi:https://doi.org/10.1080/13825585.2011.652590

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*, 1563–1585. doi:https://doi.org/10.1080/17470218.2012.658822

Liss, J., Spitzer, S., Caviness, J., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America*, *104*, 2457–2466. doi:10.1121/1.423753

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953–978. doi:https://doi.org/10.1080/01690965.2012.705006

McAuliffe, M. J., Gibson, E. M. R., Kerr, S. E., Anderson, T., & LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *Journal of the Acoustical Society of America*, *134*, 1358–1368. doi:https://doi.org/10.1121/1.4812764

Miettinen, I., Alku, P., Salminen, N., May, P. J. C., & Tiitinen, H. (2010). Responsiveness of the human auditory cortex to degraded speech sounds: Reduction of amplitude resolution vs. additive noise. *Brain Research*, *1367*, 298–309. doi:https://doi.org/10.1016/j.brainres.2010.10.037

Miles, J. (2005). Tolerance and variance inflation factor. In B. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science*. Hoboken, NJ: Wiley. doi:https://doi.org/10.1002/0470013192.bsa683

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. doi:https://doi.org/10.1016/j.jneumeth.2013.10.024

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97. doi:https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238. doi:10.1016/S0010-0285(03)00006-9

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*, 653–661. doi:https://doi.org/10.1097/AUD.0b013e3181b412e9

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America*, *97*, 593–608. doi:https://doi.org/10.1121/1.412282

Rabbitt, P. M. A. (1968). Channel capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology*, *20*, 241–248. doi:https://doi.org/10.1080/14640746808400158

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, *42*, S68–S76.

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(Suppl. 2), S99–S105. doi:https://doi.org/10.1080/14992020802301167

Slater, J., & Kraus, N. (2015). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing*, *17*, 79–87. doi:https://doi.org/10.1007/s10339-015-0740-7

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech, Language, and Hearing Research*, *32*, 912–920.

Smith, S. L., Pichora-Fuller, M. K., & Alexander, G. (2016). Development of the word auditory recognition and recall measure: A working memory test for use in rehabilitative audiology. *Ear and Hearing*, *37*, e360–e376. doi:https://doi.org/10.1097/AUD.0000000000000329

Song, X. D., Garnett, R., & Barbour, D. L. (2017). Psychometric function estimation by probabilistic classification. *Journal of the Acoustical Society of America*, *141*, 2513–2525. doi:10.1121/1.4979594

Song, X. D., Wallace, B. M., Gardner, J. R., Ledbetter, N. M., Weinberger, K. Q., & Barbour, D. L. (2015). Fast, continuous audiogram estimation using machine learning. *Ear and Hearing*, *36*, e326–e335. doi:https://doi.org/10.1097/AUD.0000000000000186

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:10.1037/h0054651

Taitelbaum-Swead, R., & Fostick, L. (2016). The effect of age and type of noise on speech perception under conditions of changing context and noise levels. *Folia Phoniatrica et Logopaedica*, *68*, 16–21. doi:https://doi.org/10.1159/000444749

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology*, *24*, 616–634. doi:https://doi.org/10.3766/jaaa.24.7.10

Van Engen, K. J. (2012). Speech-in-speech recognition: A training study. *Language and Cognitive Processes*, *27*, 1089–1107. doi:https://doi.org/10.1080/01690965.2012.654644

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, 577. doi:https://doi.org/10.3389/fnhum.2014.00577

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*, 188–196. doi:https://doi.org/10.1016/j.lindif.2010.02.004

White, L., & Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. *Amsterdam Studies in the Theory and History of Linguistic Science (Series 4)*, *282*, 237.

Wightman, F. L., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *Journal of the Acoustical Society of America*, *128*, 270–279. doi:https://doi.org/10.1121/1.3436536