# Individual Variation in the Perception of Speech in Multiple Types of Adverse Listening Conditions

Drew J. McLaughlin

*University of Oregon*

Abstract

During speech communication, both environmental noise and talker-related variation (e.g., accented speech) can create adverse conditions for the listener. Individuals recruit additional cognitive, linguistic, or perceptual resources when faced with such challenges, and they vary in their ability to understand degraded and/or variable speech. In the present study, we compare individuals' ability on a variety of skills—including receptive vocabulary, selective attention, 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 in single-talker babble, nonnative accented speech in quiet, and nonnative accented speech in speech-shaped noise. The results show that intelligibility scores within adverse listening conditions of the same class (i.e., either environmental or talker-related) significantly correlate. For cognitive, linguistic, and perceptual skills, receptive vocabulary significantly predicts performance on all four adverse listening conditions, while working memory only significantly predicts performance on conditions with nonnative accented speech. Rhythm perception was found to significantly predict speaker type (i.e., native versus nonnative speaker). Taken together, these results indicate that listeners may recruit similar resources when faced with adverse listening conditions in general, but specific additional resources when faced with certain types of listening challenges.

## Introduction

A number of factors, such as noisy environments and speaker accents, can either degrade or add unfamiliar variation the speech signal during communication, and therefore can adversely affect the speech perception process. These adverse conditions may vary in their cause, but the effects for the listener are often similar; the listener may not understand words or entire phrases, and need more time than usual to accurately decode what was heard. Previous research has suggested that there are vast individual differences in listeners' abilities to understand speech under adverse listening conditions (Wightman, Kistler, & O'Bryan, 2010; Benichov, Cox, Tun, & Wingfield, 2012; Bent, Baese-Berk, Borrie, & McKee, 2017). In the present study, we ask whether 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 variable (i.e., accented) speech.

### Types of Adverse Listening Conditions

A review of adverse 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 that are frequently replicated in the lab setting are speech in noise and speech in babble (i.e., speech from competing talker(s) in the background). These overlapping signals cause energetic interference and, for cases in which this is a competing talker, informational interference. Energetic and informational masking both create perceptual interference for the listener due to physical blending between the target signal and a non-

target signal. Informational masking, however, poses an additional challenge. In addition to segregating two competing signals and suppressing the non-target signal, when faced with informational masking the listener must manage interference caused by higher-level lexical activation, because the speaker or speakers in the background are producing language as well. Some of the perceptual consequences of informational masking may dissociate from those of energetic masking due to differences in processing demands caused by the presence of semantic interference. That is, while there are perceptual consequences for both energetic and informational masking, the stages of language processing at which the signals interfere may differ.

Source degradations to the speech signal, according to Mattys et al. (2012), are caused by speaker deviations, as is seen in conversational, disfluent, or accented speech. "Deviations" in this definition refers to systematic differences at the segmental and suprasegmental level in the talker's speech pattern compared to the listener's speech pattern (or their previous experience based on other speakers). These differences between the signal and the listener's expectations can make nonnative accented speech more difficult to process than native speech, and therefore create an adverse listening condition. Not all accents differ from standard speech in the same way. For example, the systematic deviations of nonnative accents (i.e., speech accented by a speaker's first language influencing their second language) have been shown to have larger processing costs than those of regional accents (i.e., speech accented due to social or geographical variation; Adank, Evans, Stuart-Smith, & Scott, 2009).

For the purposes of the present study, Mattys et al.'s (2012) definitions of adverse conditions are used with some minor changes in terminology. The term "degraded

speech" will be used with reference to environmental factors such as energetic and informational masking. The term "variable speech" will be used to refer to nonnative accented speech, thus separating accented speech from other source factors such as those caused by speech motor disorders. The goal of this change in terminology is to distinguish types of adverse conditions caused by loss and/or interference of signal from those caused by speaker and listener differences. "Adverse listening conditions" will serve as the umbrella term for degraded and variable speech.

**Cognitive, Linguistic, and Perceptual Resources**

The review of multiple types of adverse listening conditions above has illustrated similarities and differences in the source of signal degradation or variation. Previous research indicates that listeners use additional cognitive, linguistic, and/or perceptual resources for speech perception under adverse conditions (Rabbitt, 1968; Pichora-Fuller, Schneider, & Daneman, 1995; Heinrich, Schneider, & Craik, 2008), and that individuals vary substantially in their ability to perceive degraded or variable speech (Wightman et al., 2010; Benichov et al., 2012; Bent et al., 2017). Skills such as auditory working memory, receptive vocabulary, selective attention, and rhythm perception have been investigated in previous studies as indicators of aptitude in the perception of degraded or variable speech. Below, we address previous findings on these measures.

Behavioral and neuro-imaging research has indicated that working memory is related to the perception of speech under adverse conditions (Parbery-Clark, Skoe, Lam, & Kraus, 2009; Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012; Janse & Adank, 2012; Banks, Gowen, Munro, & Adank, 2015). Obleser et al. (2012) showed commonalities between the areas

of the brain which were active during the perception of degraded speech and those areas which were active during the use of auditory memory load (i.e., working memory) using magnetoencephalographic (MEG) responses; this indicated that auditory memory load was recruited to store degraded speech signals while they were processed. Using functional magnetic resonance imaging (fMRI), Eisner et al. (2010) found a similar link between phonological working memory and individuals' abilities in the perception of, and adaptation to, degraded speech. Both studies utilized speech degraded through noise-vocoding, which is a type of degradation meant to simulate the perceptual quality of cochlear implants (Eisner et al., 2010; Obleser et al., 2012). Working memory has also been correlated with individuals' ability to perceive accented speech (Janse & Adank, 2012; Banks et al., 2015), and speech in noise (Parbery-Clark et al., 2009).

Behavioral studies examining multiple types of challenging listening conditions have suggested that there is a relationship between receptive vocabulary and intelligibility of degraded and variable speech (Janse & Adank, 2012; McAuliffe, Gibson, Kerr, Anderson, & LaShell, 2013; Banks et al., 2015; Bent et al., 2017). The perception of speech produced by individuals with dysarthria, a motor speech disorder, was investigated by McAuliffe et al. (2013), and receptive vocabulary was a significant predictor of individuals' transcription accuracy for both younger and older listeners—although the effect in older listeners was dependent upon hearing thresholds. Janse and Adank (2012) investigated listeners' perceptual adaptation to a novel, constructed accent in both auditory-only and audiovisual presentations, and compared this with the listeners' cognitive strengths and abilities, as measured by a number of standardized tests. Vocabulary size, as well as selective attention, predicted improvement of listening

accuracy over the experiment, and auditory short-term memory and working memory predicted overall listening accuracy (Janse & Adank, 2012).

Participants' selective attention (sometimes also referred to as cognitive flexibility or inhibition), in addition to their working memory abilities and receptive vocabulary, were compared to perceptual adaptation to a novel accent (Banks et al., 2015). Results revealed that better selective attention scores on the standardized Stroop test correlated with faster perceptual adaptation to accented speech. However, other studies using the Perceptually Robust English Sentence Test Open-set (Gilbert, Tamati, & Pisoni, 2013) in multi-talker babble noise did not find a relationship with selective attention scores measured using the Stroop test (Tamati, Gilbert, & Pisoni, 2013).

Rhythm perception has been shown to predict listener performance for both source and environmental speech degradations. Slater and Klaus (2016) found that the ability to differentiate rhythms was positively related to perception scores for sentences in four-talker babble noise (i.e., an informational environmental degradation). This significant finding did not extend to the perception of target words in noise, suggesting that the strong rhythm perception skills provide a greater advantage in longer sentence formats in which the temporal pattern can be identified and then bootstrapped while segmenting the speech signal. For source degradations, rhythm perception has been found to predict improvement of listeners' intelligibility scores during the learning of dysarthric speech (Borrie, Lansford, & Barrett, 2017). The results of Borrie et al. (2017) suggest that there is not an initial advantage for listeners with greater rhythm perception ability, but that these listeners do show greater learning over time. This suggests a role for rhythm perception in perceptual adaptation to speech in adverse listening conditions.

**Differences Between Types of Adverse Listening Conditions**

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

Studies conducted using brain imaging and other physiological measures have indicated some differences in how each type of adverse listening condition is processed (Miettinen, Alku, Salminen, May, & Tiitinen, 2010; Adank, Davis, & Hagoort, 2012; Francis, MacPherson, Chandrasekaran, & Alvar, 2016). For example, using fMRI, Adank et al. (2012) demonstrated that the neural systems used to process speech under adverse conditions may differ depending on whether the signal includes source variation or environmental degradation. Similar results were found using MEG to compare brain activity during perception of speech with reduced amplitude resolution and speech in noise (Miettinen et al., 2010). Francis et al. (2016) found evidence of differences between variable speech and speech degraded in the environment; their research used stimuli in four conditions: unmasked natural speech (as a control condition), speech-shaped noise masker (an energetic environmental degradation), 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 that either additional or different processing demands may be present for the two environmental degradation conditions.

Within each sub-category of adverse listening conditions further differences among specific types of listening challenges have also been found. Multiple types of accented speech were investigated by Bent et al. (2017), and results indicated that listeners might be more or less adept at recovering from certain types of speech deviations (e.g., some listeners may be skilled at recovering from suprasegmental deviations but not segmental deviations, or visa-versa). Goslin, Duffy, and Floccia (2012) used event-related potentials (ERPs) to assess how unfamiliar regional accented and nonnative accented speech are processed, and their results indicated that different strategies may be used by listeners for each type of accented speech. Specifically, their results indicated that regional accented speech may be normalized during the early pre-lexical stage of language processing, while nonnative accented speech may be normalized in later stages of language processing.

Similar to variable speech, differences within the category of environmental degradations have also been observed, specifically between energetic and informational masking. Taitelbaum-Swead and Fostick (2016) conducted research on younger and older listener groups using three background noise conditions (speech-shaped noise, babble noise, and white noise) to degrade speech at signal-to-noise ratios (SNRs) of different difficulties. Results showed that the increase in SNR difficulty caused a significantly greater decrease in participant accuracy in the babble noise condition than in the other noise conditions—both in general and when comparing the two age groups.

The variation in listener comprehension when perceiving speech under adverse conditions, as well as the differences between the multiple types of degraded and variable speech, prompt further comparison between individuals' cognitive, linguistic, and perceptual skills and their ability to perceive particular types of degraded and/or variable speech. Based on previous research of speech perception during adverse conditions, the skills examined in the present study—including working memory, receptive vocabulary, selective attention, and rhythm perception—were predicted to indicate individuals' abilities to perceive speech under adverse conditions. There were two goals in the present study: first, to examine correlations of individual listeners' performance across multiple types of adverse listening conditions to determine whether performance on one adverse listening condition would be related to performance on all, or only specific, different types of adverse conditions; and second, to determine if the cognitive, linguistic, and perceptual skills discussed above would predict individuals' accuracy perceiving degraded and/or variable speech, and if so, whether specific types of adverse listening conditions would be linked to specific skills. By investigating adverse listening conditions in this way we aimed to shed light on how degraded and/or variable speech types are processed by the listener, and to determine whether different types of adverse conditions may be processed by the listener in different ways.

## Methods

### Participants

Participants of normal hearing ($n$ = 65) were recruited using the University of Oregon's Psychology and Linguistics Human Subjects Pool. Participants were compensated for two hours of participation, either with $20 in payment or with class

participation credit. In total, 14 participants were excluded from the analyses in order to control for confounding variables; 7 of the participants were excluded because they were bilingual or had extensive exposure to Spanish-accented speech, 3 because they were not native speakers of American-English, and 4 because they did not pass the hearing screening, leaving 51 participants 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.

**The Experiment**

Participants completed a series of short tasks including: a hearing test, a phrase recognition task, the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007), the color Stroop test (Stroop, 1935), the rhythm perception subtest of the Musical Ear Test (MET; Wallentin, Nielsen, Friis-Olivarius, C. Vuust, & P. Vuust, 2010), and the Word Auditory Recognition and Recall Measure (WARRM; Smith, Pichora-Fuller, Wilson, & Alexander, 2016). All of the tasks were administered on a Mac OS X computer in a quiet room, and all auditory stimuli were played for the participants through Sennheiser headphones at predetermined volumes. 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 in respective order), the order of the tasks was randomized for each participant.

**The machine learning hearing test.** The online hearing test ML Audiogram (Song, Garnett, & Barbour, 2017; Song et al., 2015) was used to estimate hearing thresholds of each participant and determine whether they had normal hearing. ML Audiogram is a

machine learning hearing test, which are designed to adapt to the listeners' responses in order to accurately estimate a hearing threshold of frequency (measured in Hz) and intensity (measured in dB). The main computer's volume setting was calibrated using a sound pressure level meter. The standard settings were used for the test itself with the exception of test type, which was set to Hughson-Westlake. Participants were instructed to listen for a sequence of three short beeps and press spacebar on the keyboard whenever they heard them. Before beginning the test, an example of the three short beeps was played for the participant at an intensity of 50 dB and frequency of 2000 Hz.

**The phrase recognition task.** The phrase recognition task was programmed in Python. The stimuli were created using recordings of semantically anomalous phrases taken from Liss, Spitzer, Caviness, Adler, and Edwards (1998) and originally modeled on similar phrases used by Cutler and Butterfield (1992; see Appendix A for the complete list of stimuli). Semantically anomalous phrases contain real English words composed into normal syntactic frames, however they lack meaning and context holistically. An example of this would be: "Account for who could knock." These types of phrases were used in the present experiment because they prevent top-down processing of the phrases—thus preventing the listener from inferring misperceived words based on the context.

The phrase recognition test included stimuli in four conditions of degraded and/or variable speech: native speaker masked in speech-shaped noise (environmental degradation via energetic masking), native speaker masked in single-talker babble (environmental degradation via informational masking), nonnative speaker in quiet (source variation), and 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.

A male, native English speaker was recorded reading 80 semantically anomalous phrases for the NE and NI conditions. For the NNQ and NNE conditions, a male speaker with Spanish-accented speech (i.e., a speaker whose native language is Spanish and second language is English) was recorded reading the same 80 phrases. In order to create the informational masking condition, a second male native English speaker was recorded reading a different set of 80 semantically anomalous phrases; these phrases were then edited into one continuous sound file to create single-talker background babble. Both energetic masking conditions (NE and NNE) used the same speech-shaped noise file, which was created using the software Praat. The Python program was written such that each masking condition was mixed by combining the target phrases (i.e., the phrases which the listener is asked to transcribe) with randomly selected sections of the masker files. This 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 signal-to-noise ratio (SNR) determined by the results from pilot testing of the stimuli. Pilot testing of the NE condition was conducted using Amazon Mechanical Turk; sound files of the target speaker were mixed with the masker sound files at a variety of SNRs and then posted to the website. Sixty-one participants took part in the pilot test of the NE condition (15 at a 0 dB SNR, 23 at a -2 dB SNR, and 23 at a -5 dB SNR). The participant responses were then scored for average words correct at each SNR. For the NI condition, it was

anticipated that a SNR equivalent to the NE condition would yield similar levels of intelligibility; however, after examining the results of pilot participants the SNR was adjusted. The average intelligibility of the NNQ condition was known from results of a previous study in which the same speaker recordings were used (Bent et al., 2017), and the SNRs of the NE and NI conditions were chosen to match this intelligibility level. The NNE condition was the exception to this, as it was expected to be much less intelligible than the other three conditions because it combined two sources of difficult listening situations within a single stimulus. The ratios selected for the experiment were as follows: NE at -2 dB SNR, NI at -5 dB SNR, and NNE at 0 dB SNR.

There were 4 practice trials (one for each condition) before the actual experiment trials began. Each adverse listening condition was presented in 20 trials for a total of 80 trials across listening conditions. The order of the trials was randomized for each participant, as was the condition that each semantically anomalous phrase appeared in. For example, this means that for one participant the phrase "Account for who could knock" may have been in the NE condition, and for another participant it may have been in the NNQ condition.

Before beginning the task, an experimenter recited a set of verbal instructions for the participant and answered any questions. A set of similar written instructions was also displayed on the computer screen before the experiment began. Participants 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 their response. 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 stimuli.

**The Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4).** The PPVT-4 is a standardized test that measures receptive vocabulary, which has been shown to correlate with perception of unfamiliar speech (Bent et al., 2017). In the present study, an online version of the test was administered. For each trial, a word would play over the headphones and the participant would choose one of four illustrations that best represented the word. Participants were able to 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 in the present study. The color Stroop test was used to measure selective attention—also commonly referred to as cognitive flexibility and 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 respond to what color the word is written in. Four colors appeared in the task, each corresponding to a number (e.g., 1 = red). The task was fast-paced, encouraging participants to respond quickly by flashing "Too Slow" if they did not respond quickly enough (i.e., approximately 2 seconds). 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 correct color (e.g., the word "red" written in the color red), incongruent conditions were those in which the word on the screen appeared in an incorrect 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). Thus, in the incongruent condition the participant had to focus on one trait and actively

suppress another; however, in the congruent condition no active suppression of traits was required. Response times were averaged for each condition, and then the difference between the incongruent and congruent conditions was calculated. This difference was used as a measure of participants' selective attention. Larger differences between the two conditions indicated that the participant had weaker selective attention, and smaller differences between the two conditions indicated that the participant had stronger selective attention.

**The Rhythm Subsection of the Musical Ear Test (R-MET).** The R-MET was used to determine individuals' rhythm perception abilities. In each trial, participants listened to two sets of beats played on a wood block and then decided whether the beats comprised the same rhythm or different rhythms. Participants marked their responses on a paper answer sheet. Before the actual test began, a recording of verbal instructions was played for the participants over headphones, and then two practice rounds were given with correct answers. 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 (Smith, Pichora-Fuller, Wilson, & Alexander, 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 the auditory stimuli are played in different orders. Before beginning the task, participants were instructed of the process using a short PowerPoint presentation. Auditory stimuli were played over headphones, and participant responses were recorded during the experiment by an experimenter. Participants were given 2 practice trials to confirm that they understood the process of

each trial before continuing into the test itself. 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 make a judgment as to whether the first letter of the target word is from the first or second half of the alphabet (i.e., the listener would say "first" if the letter is between *A* and *M*, and "second" if the letter is between *N* and *Z*). At the end of the trial, a beep played, indicating to the listener that they should recall the target words from the set. If the listener could not remember all of the words in the set, they were instructed to take a guess or move on to the next trial. There were 5 trials for each set size, beginning with 2 words per trial and ending with 6 words per trial. Participants were allowed to take a short break between trials if needed.

## Analysis

Participant transcripts from the phrase recognition test were scored for measures of intelligibility. This was done by calculating the number of words correct in each trial for each condition. Following Borrie, McAuliffe, and Liss (2012), words that were homophones or obvious misspellings of the target word were scored as correct, as were differences in tense, plurality, and substitutions between "a" and "the." Measures from the PPVT-4, the Stroop test, the R-MET, and the WARRM test were either automatically scored by the testing software itself or manually scored using the standard protocols indicated by the creators of the test. Statistical analyses are described in more detail below.

## Results

For each of the four degraded and/or variable listening conditions, measures of intelligibility were calculated based on the proportion of words correctly transcribed by

participants (Table 1A). The NNE condition had a notably lower mean intelligibility than the other three conditions ($M = .25$); this was expected due to the combination of both source variation and environmental degradation in this condition.

Following the analysis of Bent et al. (2017), we will first present correlations between each of the adverse listening conditions and then present the results of logistic mixed effects models that include the four adverse listening conditions and the four cognitive, linguistic, and perceptuals skills as fixed factors. Specifically, the correlations between adverse listening conditions address the question of whether listening performance for one type of degraded and/or variable speech is related to listening performance on other types of degraded and/or variable speech.

The results from the pairwise correlations among the adverse listening conditions showed three significant correlations: the intelligibility scores for the NE condition significantly correlated with the scores for the NI condition ($r = .32, p = .023$); the scores for the two conditions with energetic masking, NE and NNE, significantly correlated with one another ($r = .45, p = .001$; Figure 1); and, lastly, the scores for the two conditions with source variation, NNQ and NNE, significantly correlated ($r = .43, p = .002$; Figure 1). The relationships between the NE and NNQ conditions ($r = .14, p = .315$), the NI and NNQ conditions ($r = .25, p = .081$), and the NI and NNE conditions ($r = .13, p = .351$), were not significantly correlated. First, it is important to note that these results indicate that performance under one adverse listening condition does not predict performance for all other types of adverse listening conditions examined in the present study. Further, the relationships found between these adverse listening conditions suggest that listeners may be adept at perceiving speech under similar classes of adverse listening conditions (i.e.,

conditions that share a type of degradation or source variation). This is demonstrated in Figure 1 with plots of the significant correlations between the NNE and NE conditions (both of which have energetic masking), and the NNE and NNQ conditions (both of which have source variation).

Next, the measures of participants' cognitive, linguistic, and perceptual skills were analyzed in a logistic mixed effects model with the intelligibility scores from the phrase recognition task as the dependent variable. Fixed factors for the model included scores from the PPVT-4, Stroop, WARRM, and the R-MET (see Table 1B for a summary), intelligibility measures from each of the adverse conditions (i.e., NE, NI, NNQ, and NNE), and interactions between each cognitive, linguistic, or perceptual measure and each adverse condition. For the PPVT-4, a measure of percentile rank was used; for the color Stroop test, a measure of the difference in reaction times 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 proportion answers correct was used. Scores from the four tests were all centered and scaled prior to entering the model. Random effects were the maximal effects that would allow the models to converge and included participants as random intercepts. A series of model comparisons was used to determine the significance of each fixed factor. Based on these comparisons it was determined that the model of best fit (Table 2) included fixed factors of adverse listening condition (i.e., NE, NI, NNQ, and NNE), the PPVT-4, WARRM, the interaction between the R-MET and adverse listening condition, and the interaction between WARRM and adverse listening condition. Both the color Stroop test (i.e., the measure of selective attention) as well as the interaction between Stroop and adverse listening condition were excluded

from the model of best fit because they were not significant predictors of model fit ($\chi^2(1) = .7669, p = .381$, and $\chi^2(3) = 6.394, p = .094$, respectively). The measures of cognitive, linguistic, and perceptual skills with that were significant predictors and those interactions that were found to significantly improve model fit are discussed further below.

The PPVT-4 (i.e., the measure of receptive vocabulary) significantly improved model fit ($\chi^2(1) = 9.403, p = .002$), but the interaction between the PPVT-4 and adverse listening condition was not significant ($\chi^2(3) = 3.450, p = .327$). Examining the correlation of PPVT and performance on each adverse listening condition separately revealed that participants' percentile rankings on the PPVT significantly correlated with their performance on each adverse listening condition (NE: $r = .30, p = .033$; NI: $r = .36$, $p = .010$; NNQ: $r = .29, p = .042$; NNE: $r = .32, p = .024$; Figure 2). This finding is consistent with previous results in which receptive vocabulary was correlated with performance on a speech recognition task that included unfamiliar Spanish-accented English, Irish English, and disordered speech caused by ataxic dysarthria (Bent et al., 2017).

Both WARRM, the measure of working memory capacity, and the interaction between WARRM and adverse listening condition significantly improved model fit ($\chi^2(1) = 4.790, p = .029$, and ($\chi^2(3) = 14.767, p = .002$, respectively). As shown in Figure 3, working memory scores from WARRM significantly correlated with the NNQ condition ($r = .30, p = .035$), and the NNE condition ($r = .39, p = .004$). This may indicate an important role for working memory specifically in the perception of variable speech types, and perhaps a less important role in perception of speech in noise.

The measure of rhythm perception, R-MET, did not significantly improve model fit ($\chi^2(1) = .171, p = .679$); however the interaction between R-MET and adverse listening condition did significantly improve model fit ($\chi^2(3) = 13.192, p = .004$) and, thus, was included in the model. Individually, none of the adverse listening conditions were found to significantly correlate with the R-MET scores (all $p$-values > .05), however, as shown in Figure 4, the correlations between the scores from the R-MET with the NE condition ($r = .27, p = .056$) and the NI condition ($r = .27, p = .053$) approached significance. This was not true for the conditions with nonnative talkers (NNQ: $r = .09, p = .534$; NNE: $r = .09, p = .531$). This was further investigated using a logistic mixed effects models to determine if rhythm perception scores may be predictive of speaker type (i.e., native speaker or nonnative speaker). Contrast coding was used to distinguish adverse listening conditions by speaker type. The interaction between rhythm perception scores from the R-MET and speaker type significantly improved model fit ($\chi^2(1) = 9.238, p = .002$). In combination with the results of the correlation tests, this indicates that rhythm perception may play a role in the perception of adverse listening conditions in which there is a speaker with a native, or familiar, accent, but not necessarily those conditions in which there is a nonnative accented speaker.

## Discussion

While listeners may experience the same "symptom" when processing degraded and/or variable speech (that is, understanding speech is more difficult) their ability to understand speech in adverse listening conditions appears to differ depending upon the type of adverse listening condition that they are faced with. By examining performance in four adverse listening conditions we were able to determine that performance in one type

of adverse listening condition does not predict performance on all other types of adverse listening conditions, but that specific types of adverse listening conditions do significantly correlate with one another. Additionally, when examining the four cognitive, linguistic, and perceptual skills we found that receptive vocabulary was the only skill that significantly correlated with all four adverse listening conditions. Working memory and rhythm perception were both found to predict differences between adverse listening conditions, and selective attention was not significantly predictive of any condition. For the discussion of these results we will begin by addressing how various adverse listening conditions in the present experiment, and possibly in general, may be related to one another. We will then address the results from each of the cognitive, linguistic, and perceptual skills in turn before finally proposing what these results may indicate as a whole.

The pairwise correlations of adverse listening conditions revealed that conditions of the same class were significantly related. A class of adverse listening conditions in this case would be comprised of two or more conditions with the same variable or degraded speech type. In the present study, we found that listener performance for a native speaker in energetic noise significantly predicted performance for that same speaker in informational noise; in both of these conditions there is environmental degradation to the speech signal. Similarly, scores for the nonnative speaker in energetic noise significantly correlated with scores for the native speaker in energetic noise. Lastly, the two conditions in which there is source-related unfamiliar variation, the nonnative speaker in quiet and the nonnative speaker in energetic noise, significantly correlated. These results suggest that listeners perform similarly when faced with adverse listening conditions of the same

class, and this may be because the same cognitive, linguistic, and/or perceptual skills are recruited for adverse listening conditions of the same class.

No significant relationships were found between selective attention scores and intelligibility scores for the four adverse listening conditions. The color version of the Stroop test was used as a measure of selective attention, as previous research by Banks et al. (2015) had found a significant relationship between Stroop scores and adaptation to an unfamiliar constructed accent. Audiovisual research of adverse listening conditions by Janse and Adank (2012) also found a relationship between selective attention (as measured by a different test called the flanker task) and perceptual adaptation to an unfamiliar constructed accent. Thus, while it is possible that the different version of the Stroop task may account for the null result in the present study, it is more likely to be the case that selective attention plays a role in perceptual adaptation to, and not necessarily the general perception of, adverse listening conditions. Comparing selective attention to both perceptual adaptation and general perception could be an area for future research for both degraded and variable speech types.

In addition to the relationships found in the present study with degraded and/or variable speech types, the PPVT-4 has been found to predict listener performance for Irish English (i.e., a dialect) and dysarthric speech (i.e., a source degradation; Bent et al., 2017), indicating a robust relationship between receptive vocabulary and multiple types of adverse listening conditions. The measure of receptive vocabulary used in the present study, the PPVT-4, has also been positively correlated with verbal IQ (Bell, Lassiter, Matthews, & Hutchinson, 2001). Thus, it is possible that verbal IQ is also an indicator of

performance under adverse listening conditions, although this remains to be directly investigated.

The R-MET was utilized in the present study to measure rhythm perception skills, and was found to be predictive of speaker type (i.e., native versus nonnative speaker conditions). While in the present study rhythm perception scores did not significantly correlate with any adverse listening condition, the scores did approximate significance with only the two native speaker conditions. Taken together, these findings indicate that, in the present study, rhythm perception skills may have only benefitted participants for adverse listening conditions with a familiar (i.e., native) speaker. This finding complements previous research that showed a relationship between rhythm perception and the perception of sentences in environmental noise (Slater and Klaus, 2016; see Parbery-Clark et al., 2009, for evidence including energetic noise). An important finding of Slater and Klaus' (2016) work was that rhythm perception skills only provided an advantage in a sentence format (as opposed to single word format). Examining dysarthric speech, Borrie et al. (2017) found that R-MET scores predicted listener improvement scores, but not initial intelligibility. The findings of both studies indicate that the temporal pattern of speech may need to be identified before the listener is able to use it for language segmentation and processing. Additionally, the use of metrical stress cues for speech segmentation has been shown to be used by listeners for adverse listening conditions such as speech in energetic noise more than for those such as dysarthric speech (Borrie, Baese-Berk, Van Engen, & Bent, 2017). Thus, longer stimuli (i.e., sentence versus word) or longer exposure may be necessary for rhythm perception skills to be beneficial to the listener, especially in situations in which the speech signal is

rhythmically less familiar, such as with dysarthric speech and, possibly, nonnative accented speech also. If this is the case, it could explain why rhythm perception was predictive of speaker type in the present study. Because the adverse listening conditions were presented in a randomized order instead of in blocks, listeners' ability to perceptually adapt to degraded and/or variable speech using rhythm segmentation strategies may have been limited—and possibly more so for conditions with unfamiliar accented speech. Thus, while the results of the present study indicate a role for rhythm perception when there is a familiar speaker, further investigation is necessary to determine whether there may be a role for rhythm perception in the adaptation to nonnative accented speakers as well.

In the present study working memory, as measured by the WARRM, was found to significantly predict differences between the types of adverse listening conditions and to correlate with only the nonnative speaker conditions. Previous research has found a similar relationship between working memory and the perception of constructed unfamiliar accents (Janse & Adank, 2012; Banks et al., 2015), and in combination with the present results this may indicate a key role for working memory in the perception of variable speech types. However, it is also possible that for the NNE condition, which had a substantially lower mean intelligibility than the other three adverse listening conditions, there was a significant correlation with working memory scores because it was considerably more difficult.

One of the most notable findings of the present study is that performance on one type of adverse listening condition did not predict performance for all other types of adverse listening conditions. This indicates that careful scrutiny ought to be taken when

reviewing the topic of adverse listening conditions, because listener performance may vary substantially between various types of degraded and/or variable speech that have previously been grouped together in the literature. Additionally, these results prompt reflection upon the current categories of adverse listening conditions in the literature. The classifications proposed by reviews such as Mattys et al. (2012) are based heavily on the source of communication difficulty instead of the perceptual processes of the listener. This is problematic when, for example, adverse listening conditions such as neurologically disordered speech and accented speech are grouped together and labeled as source degradations, when previous evidence suggests that there are differences between how dysarthric speech and Irish accented speech are processed by the listener (Bent et al., 2017). Additionally, labeling regional and nonnative accented speech as degraded can negatively and unfairly portray the speaker. An ideal system of classification would account for the differences between the various types of adverse listening conditions without negatively characterizing speakers of less-prestigious speech types.

Taken altogether, the results of the present study indicate that not all types of adverse listening conditions are processed by the listener in the same way. When faced with degraded and/or variable speech, some skills, such as receptive vocabulary, may be recruited for all types of adverse listening conditions, while others, such as working memory and rhythm perception, may be employed for only specific types, or classes, of adverse listening conditions. Further, it is possible that different listeners employ different strategies when faced with adverse listening conditions.

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: 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: 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: 10.1121/1.4916265

Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the peabody picture vocabulary test—Third edition and Wechsler adult intelligence scale— 26 Third edition with university students. *Journal of Clinical Psychology 57*, 417-422. doi: 10.1002/jclp.1024

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: 10.1097/Aud.0b013e31822f680f

Bent, T., Baese-Berk, M., Borrie, S., & McKee, M. (2017). Individual differences in the perception of unfamiliar regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America*. Retrieved from http://dx.doi.org/10.1121/1.4966677

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric

speech: A review of experimental studies. Journal of Speech, Language, and Hearing Research, 55, 290–305. doi: 10.1044/1092-4388(2011/10-0349)

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*(3), 561-570. doi: 10.1044/2016_JSLHR-S-16-0094

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

Dunn, L. M., Dunn, D. M., & Pearson Assessments. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: Pearson Assessments.

Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior Frontal Gyrus Activation Predicts Individual Differences in Perceptual Learning of Cochlear-Implant Simulations. *The Journal of Neuroscience, 30*, 7179-7186. doi: 10.1523/JNEUROSCI.4040-09.2010

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: 10.3389/fpsyg.2016.00263

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: 10.3766/jaaa.24.1.4

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language, 122*, 92–102. doi: 10.1016/ j.bandl.2012.04.017

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

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology, 65*, 1563-1585. doi: 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. Retrieved from http://dx.doi.org.libproxy.uoregon.edu/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: 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: 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: 10.1016/j.brainres.2010.10.037

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: 10.1016/j.jneumeth.2013.10.024

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience, 32*, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician Enhancement for Speech-In-Noise. *Ear & Hearing, 30,* 653-661. doi: 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: http://dx.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: 10.1080/14640746808400158

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognitive processing*, *17*(1), 79-87. doi: 10.1007/s10339-015-0740-7

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*(6), e360-e376. doi: 10.1097/AUD.0000000000000329

Song, X. D., Garnett, R., & Barbour, D. L. (2017). Psychometric function estimation by

probabilistic classification. *The Journal of the Acoustical Society of America*, *141*(4), 2513-2525. Retrieved from http://dx.doi.org/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. Retrieved from http://doi.org/10.1097/AUD.0000000000000186

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of* Experimental Psychology, 18, 643–662. Retrieved from http://dx.doi.org/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: 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: 10.3766/jaaa.24.7.10

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: 10.1016/j.lindif.2010.02.004

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: 10.1121/1.3436536

# Tables

Table 1A

*Descriptive Statistics of Intelligibility Scores from the Phrase Recognition Task for each Degraded and/or Variable Speech Condition*

| Comparison | NE | NI | NNE | NNQ |
|---|---|---|---|---|
| Mean | 0.57 | 0.63 | 0.25 | 0.62 |
| Standard Dev. | 0.07 | 0.13 | 0.07 | 0.08 |
| Max | 0.74 | 0.88 | 0.40 | 0.79 |
| Min | 0.41 | 0.30 | 0.06 | 0.44 |

Table 1B

*Descriptive Statistics of Scores from the Cognitive, Linguistic, and Perceptual Skill Tasks*

| Comparison | PPVT-4 | R-MET | Stroop | WARRM |
|---|---|---|---|---|
| Mean | 65.66 | 0.69 | 138.08 | 4.01 |
| Standard Dev. | 21.40 | 0.10 | 69.85 | 0.91 |
| Max | 97 | 0.90 | 330.21 | 6 |
| Min | 19 | 0.42 | 19.85 | 2.67 |

Table 2

*Logistics Mixed Effects Model Summary*

| Predictor | Estimate | Standard Error | *z*-value |
|---|---|---|---|
| (Intercept) | 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 |

# Figures

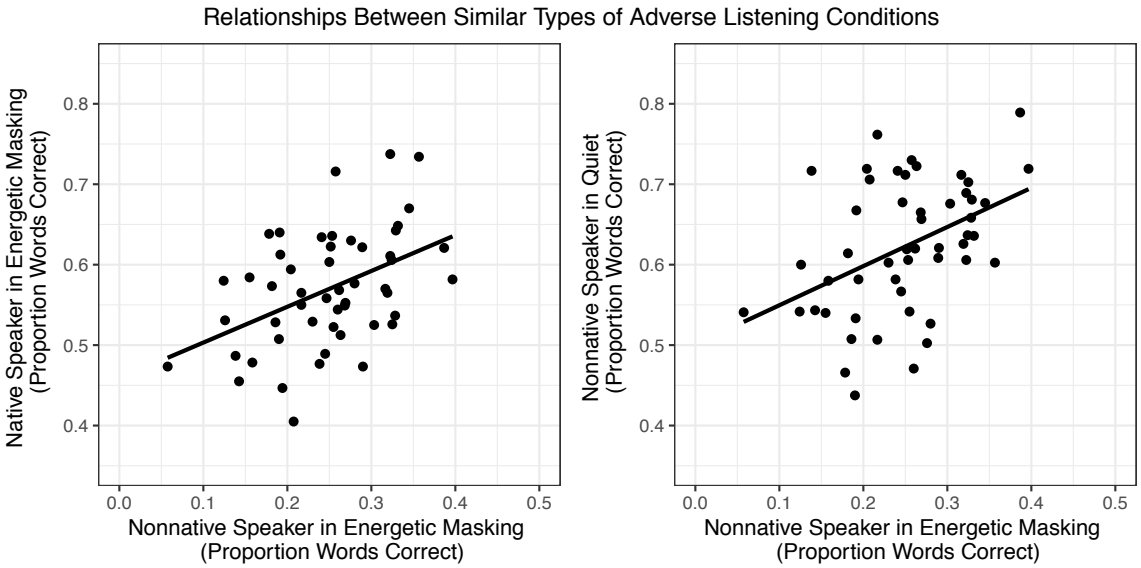


*Figure 1*. Significant correlations between the NNE and NE conditions ($r = .45, p = .001$; left) and the NNE and the NNQ conditions ($r = .43, p = .002$; right). The significant relationships found between adverse listening conditions of similar qualities indicate that listeners may be most successful at perceiving speech under types of adverse listening conditions caused by similar degradations or source variations.

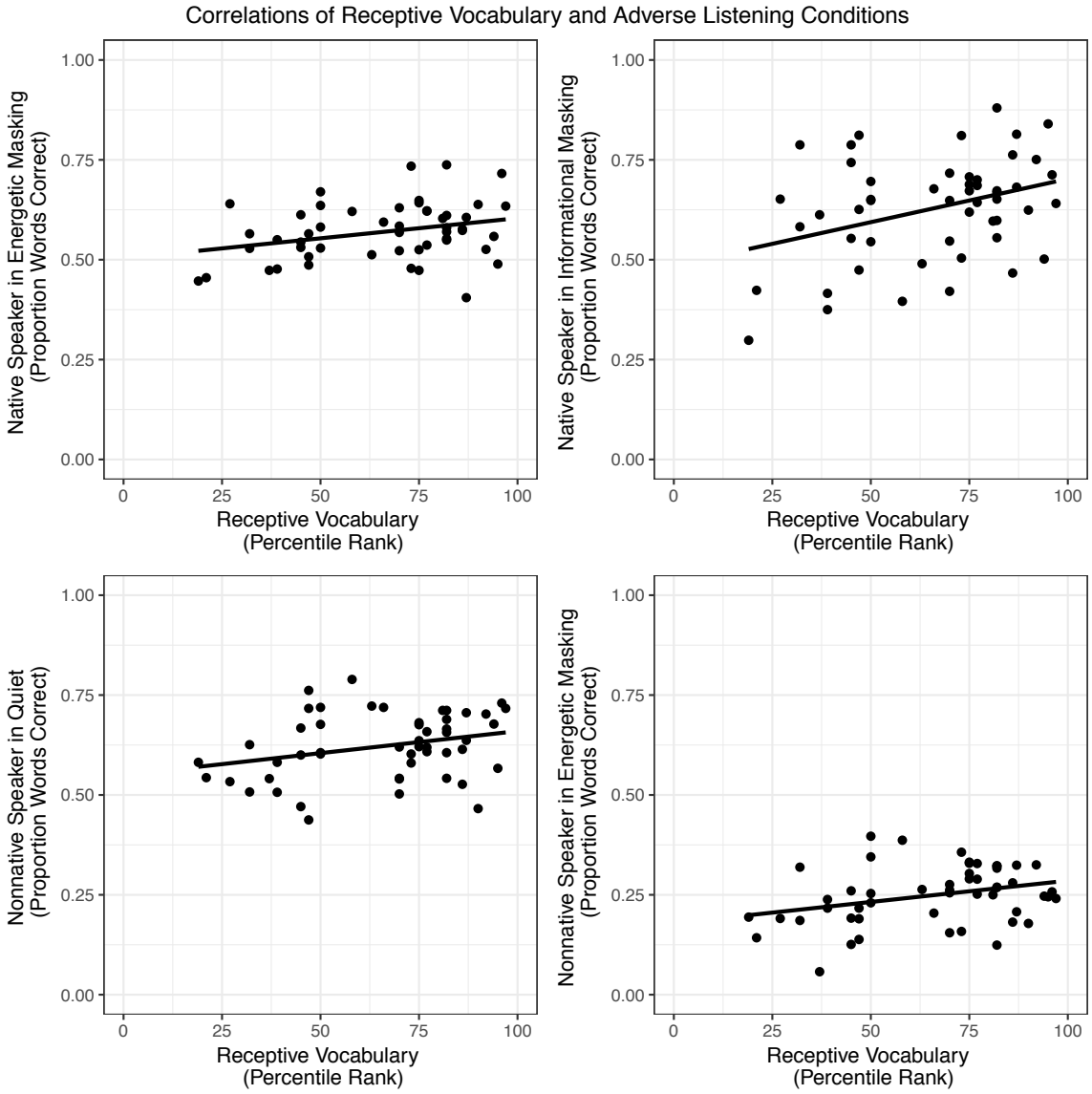


*Figure 2*. For receptive vocabulary, it was found that PPVT-4 scores significantly correlated with all four adverse listening conditions: NE ($r$ = .30 $p$ = .033), NI ($r$ = .36, $p$ = .01), NNQ ($r$ = .29, $p$ = .042), and NNE ($r$ = .32, $p$ = .024).

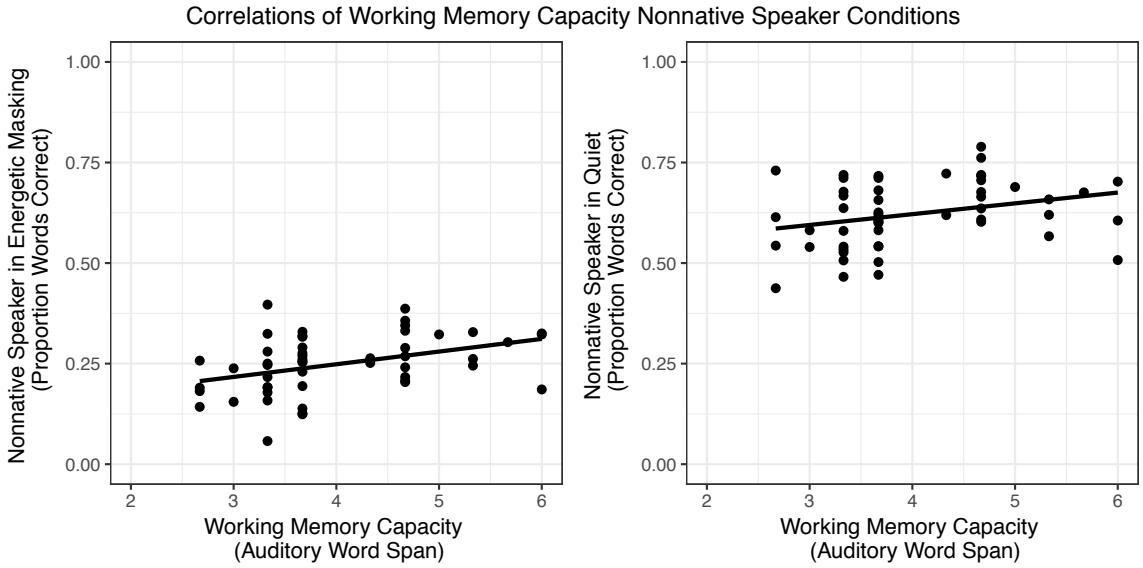


*Figure 3*. Working memory scores from WARRM significantly correlate with the NNE condition ($r = .39, p = .004$; left) and the NNQ condition ($r = .30, p = .035$; right). This may indicate that the role of working memory when listening to speech under adverse conditions is more closely related to conditions in which there is variable speech.

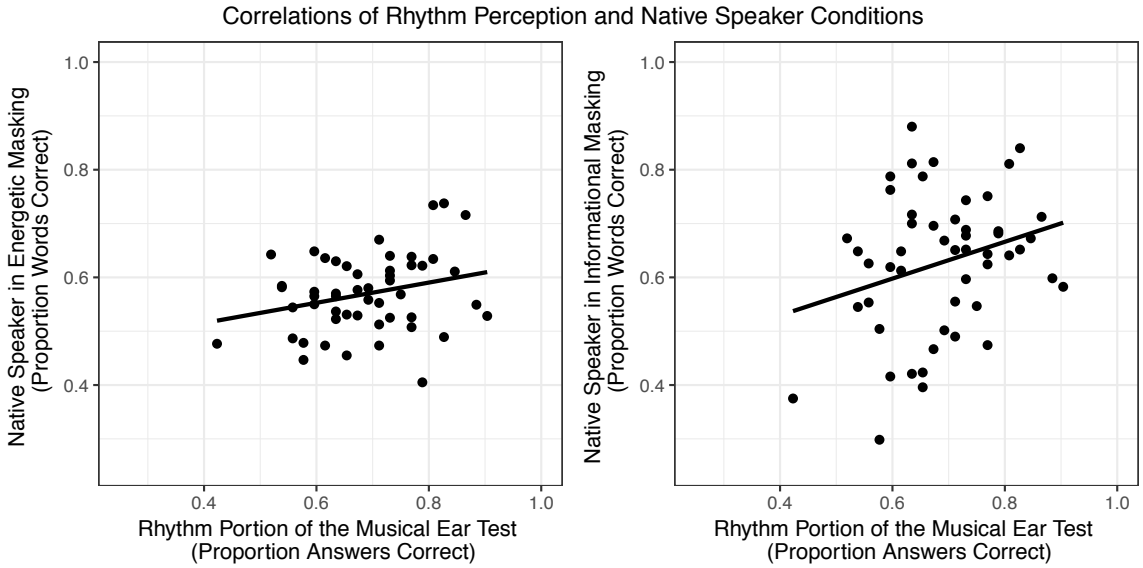


*Figure 4*. Only the two native speaker conditions, NE ($r = .27, p = .056$; left) and NI ($r = .27, p = .053$; right) were found to approach significance when correlated with rhythm perception scores.

## Appendix A

### List of Semantically Anomalous Phrases

1. account for who could knock
2. address her meeting time
3. admit the gear beyond
4. advance but sat appeal
5. afraid beneath demand
6. amend estate approach
7. and spoke behind her sin
8. appear to wait then turn
9. assume to catch control
10. attack became concerned
11. attend the trend success
12. avoid or beat command
13. award his drain away
14. balance clamp and bottle
15. beside a sunken bat
16. bolder ground from justice
17. bush is chosen after
18. butcher in the middle
19. career despite research
20. cheap control in paper
21. commit such used advice
22. confused but roared again
23. connect the beer device
24. constant willing walker
25. cool the jar in private
26. darker painted baskets
27. define respect instead
28. distant leaking basement
29. divide across retreat
30. done with finest handle
31. embark or take her sheet
32. for coke a great defeat
33. forget the joke below
34. frame her seed to answer
35. functions aim his acid
36. had eaten junk and train
37. her owners arm the phone
38. hold a page of fortune
39. increase a grade sedate
40. indeed a tax ascent
41. its harmful note abounds

42. kick a tad above them
43. listen final station
44. mark a single ladder
45. mate denotes a judgement
46. may the same pursued it
47. measure fame with legal
48. mistake delight for heat
49. mode campaign for budget
50. model sad and local
51. narrow seated member
52. or spent sincere aside
53. pain can follow agents
54. perceive sustained supplies
55. pick a chain for action
56. pooling pill or cattle
57. push her equal culture
58. rampant boasting captain
59. remove and name for stake
60. resting older earring
61. rocking modern poster
62. rode the lamp for teasing
63. round and bad for carpet
64. rowing farther matters
65. seat for locking runners
66. secure but lease apart
67. signal breakfast pilot
68. sinking rather tundra
69. sparkle enter broken
70. stable wrist and load it
71. submit his cash report
72. support with dock and cheer
73. target keeping season
74. technique but sent result
75. thinking for the hearing
76. to sort but fear inside
77. transcend almost betrayed
78. unless escape can learn
79. unseen machines agree
80. vital seats with wonder