2020

Perceptual adaptation to speech in calibrated noise

BOSTON UNIVERSITY

SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Thesis

PERCEPTUAL ADAPTATION TO SPEECH IN CALIBRATED NOISE

by

MAYA SAUPE

B.A., Wellesley College, 2017

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2020

Approved by

First Reader ______________________________

Tyler K. Perrachione, Ph.D.
Assistant Professor of Speech, Language, and Hearing Sciences

Second Reader ______________________________

Sung-Joo Lim, Ph.D.
Research Assistant Professor of Speech, Language, and Hearing Sciences

Third Reader ______________________________

Virginia A. Best, Ph.D.
Research Associate Professor of Speech, Language, and Hearing Sciences

ACKNOWLEDGMENTS

I would first like to thank my committee, Dr. Tyler Perrachione, Dr. Sung-Joo Lim, and Dr. Virginia Best. Thank you for giving me the opportunity to develop and pursue a project I was passionate about. Your support, feedback, advice, and encouragement have been invaluable, and I have learned so much from all of you. With your guidance, I have developed new skills and accomplished goals I would not have thought possible when I first started this project.

To Ja Young Choi, I want to thank you for all of the time and energy you dedicated to my project, whether it be from adding to my code, looking over my drafts, or guiding me through statistical analyses. You always found the time to provide thoughtful feedback and recommendations no matter the time of year, and the completion of my thesis would not have been possible without your support.

I would also like to extend my thanks to all the members of the Communication Neuroscience Research Laboratory. Thank you to Yaminah Carter for keeping me organized and bringing participants into the lab, and thank you to Alex Kapadia for all your advice along the way. Thank you to all of the undergraduates who assisted with running participants.

I want to thank all of the faculty and students at the Speech-Language Pathology Department at Sargent College. Thank you to Sarah Polcaro, Allison Finn, and everyone else who checked in, kept me on track, and gave me an opportunity to talk things through. I cannot imagine completing this program without you.

Finally, I would like to extend a special thank you to my family. Thank you to Sam

Buck for spending days helping put my plots together, and thank you Ellis for teaching me how to use R. You all have always given me constant and continued support in all of my endeavors, no matter where they took me. Without you, none of this would have been possible.

PERCEPTUAL ADAPTATION TO SPEECH IN CALIBRATED NOISE

MAYA SAUPE

ABSTRACT

Perceptual adaptation to a talker allows listeners to efficiently resolve inherent ambiguities present in the speech signal introduced by the lack of a one-to-one mapping between acoustic signals and intended phonemic categories across talkers. In ideal listening environments, preceding speech context has been found to enhance perceptual adaptation to a talker. However, little is known regarding how perceptual adaptation to speech occurs in more realistic listening environments with background noise. The current investigation explored how talker variability and preceding speech context affect identification of phonetically-confusable words in adverse listening conditions. Our results showed that listeners were less accurate and slower in identifying mixed-talker speech compared to single-talker speech when target words were presented in multi-talker babble, and that preceding speech context enhanced word identification performance under noise both in single- and mixed talker conditions. These results extend previous findings of perceptual adaptation to talker-specific speech in quiet environments, suggesting that the same underlying mechanisms may serve to perceptually adapt to speech both in quiet and in noise. Both cognitive and attentional mechanisms were proposed to jointly underlie perceptual adaptation to speech, including an active control process that preallocates cognitive resources to processing talker variability and auditory streaming processes that support successful feedforward allocation of attention to salient talker-specific features.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1.1 Perceptual Adaptation to Speech

One of the primary challenges human listeners must overcome during speech perception is the lack of a one-to-one mapping between acoustic signals and intended phonemic categories due to inherent variability present in speech signals. One of the prominent sources of variability in acoustic-phonemic mappings can be attributed to individual differences across talkers. Because talkers differ in vocal tract anatomy, dialect, and speech mannerisms, different acoustic signals uttered by various talkers can convey the same phoneme, or in turn, acoustically similar signals can convey different phonemes (Hillenbrand, et al., 1995; Johnson et al., 1990; Liberman et al., 1967, Miller & Dexter, 1988). Therefore, talker variability in speech hinders efficient speech processing. For example, listeners identify speech signals slower and less accurately in situations when the talker changes (mixed-talker speech) compared to when the talker remains the same (single-talker speech) (Choi, Hu & Perrachione, 2018; Choi & Perrachione, 2019; Green Tomiak, & Kuhl, 1997; Magnuson & Nusbaum, 2007; Morton, Summers & Lulich, 2015; Mullennix & Pisoni, 1990; Mullennix, Pisoni & Martin, 1989; Strange, Verbrugge, Shankweiler, & Edman, 1976). Moreover, neuroimaging studies have found that talker changes are associated with increased activity in superior temporal regions, suggesting that greater cognitive effort and listening effort are required to process mixed-talker speech (Perrachione et al., 2016; Wong, Nusbaum & Small, 2004; Zhang et al., 2016).

Both cognitive and attentional models have been proposed to explain why talker variability interferes with speech processing. One such model from a cognitive standpoint is the *active control process hypothesis* (Magnuson & Nusbaum, 2007; Heald & Nusbaum, 2014). Under this model, processing costs associated with mixed-talker speech can be attributed to the deployment of an active control mechanism. That is, when there is a change in talker, it increases the amount of ambiguity and uncertainty in the signal that listeners are perceiving. A processing cost is observed for processing mixed-talker speech because the deployment of this active control mechanism requires that certain cognitive resources be set aside or preallocated for resolving the uncertainty in mixed-talker speech, thereby reducing the resources available for speech processing.

From an attentional standpoint, an object-based model of auditory attention through auditory streaming can explain the interference from processing mixed-talker speech. This model proposes that auditory attention can be thought of similarly to visual attention, where listeners direct attention to an object in a complex scene. For example, a listener might direct attention towards one specific talker among a variety of other environmental sounds or competing speech signals. These objects are thought to be selected through auditory streaming, where acoustic stimuli from a single source are identified and linked together over time. Auditory streaming relies heavily on temporal continuity (Best et al., 2008), and successful attentional allocation via auditory streaming increases perceptual sensitivity and enhances efficiency of perceptual processing (Best, Ozmeral & Shinn-Cunningham, 2007; Kidd, Arbogast, Mason & Gallun, 2005). Under this framework, interference from processing mixed-talker speech can be attributed to attentional focus during formation of a single auditory object, in this case a talker. A change in talker requires attention to be disengaged and redirected to a new auditory object, eliminating the possibility for perceptual advantages to be afforded by auditory streaming.

Despite the increased effort that is required to process mixed-talker speech, several factors have been identified that support perceptual adaptation to mixed-talker speech, or that increase processing efficiency in mixed-talker settings. One factor that may be especially important for perceptual adaption to speech is preceding speech context. For example, Choi and Perrachione (2019) found that carrier phrases significantly reduce the interference from processing mixed-talker speech. This facilitatory effect of the carrier phrase was dependent on both the length and continuity of the signal, where the longer and more continuous the carrier phrase, the smaller the performance decrement between the single- and mixed-talker conditions. However, while longer signals reduced the interference from mixed-talker speech, the difference between mixed- and single-talker conditions was always significant, even at the longest carrier phrase tested.

These results are consistent with both active control process and auditory streaming models of speech perception, and ultimately suggest that both active control processes and auditory streaming play a role in perceptual adaptation to speech. From an active control perspective, some of the facilitatory effects of preceding speech context can be attributed to the active control mechanism being engaged at the initiation of the carrier phrase rather than at the initiation of the target word. The preceding speech context reduces the uncertainty about the upcoming target words, therefore fewer cognitive resources are required for identifying target words with preceding speech context. On the other hand, from an auditory streaming perspective, some of the facilitatory effects of preceding speech context can be explained by the successful feedforward allocation of attention afforded by the longer, uninterrupted stimulus. A longer, more continuous stimulus allows listeners to better integrate the speech signals and identify them as a single auditory object, in this case, a talker. Attention can then be better allocated to the talker, which increases perceptual sensitivity and reduces cognitive costs associated with processing mixed-talker speech. However, processing speed is still reduced in mixed-talker conditions even with long carrier phrase durations because the increased signal uncertainty is still present over longer time scales. Resources must still be allocated to resolving these uncertainties, thereby reducing the amount of cognitive resources available for speech processing. Preliminary research suggests that the maximum efficiency gains that can be afforded by preceding speech context are obtained around 600 milliseconds (Kou, 2019), indicating that both short- and long-term processes are involved in perceptual adaptation to speech.

1.2 Perceptual Adaptation to Speech in Noise

While the processing costs associated with understanding mixed-talker speech have been widely studied and consistently found across behavioral and neuroimaging studies (e.g., Nusbaum & Magnuson, 1997; Wong, Nusbaum, & Small, 2004; Zhang et al., 2016; Choi, Hu & Perrachione, 2018), very few studies have examined perceptual adaption to speech in more naturalistic listening environments. In realistic listening situations, speech typically occurs with some degree of background noise or competing stimuli, and most research investigating perceptual adaptation to speech has taken place in quiet environments with minimal distractions. This experimental design makes it difficult to determine how mixed-talker interference might affect speech processing in more realistic settings. Listening to speech in the presence of noise introduces even greater uncertainty to the signal, and requires additional cognitive resources to process (Pichora-Fuller, 2006; Zekveld, Kramer & Festen, 2011). Background noise or competing voices may also result in attentional disruptions to continuous speech signals or make the formation of an auditory object more challenging (Shinn-Cunningham, 2008). These factors may influence the processes involved perceptual adaptation to speech, as adaptation to a talker is thought to involve successful allocation of limited cognitive and attentional resources.

Two studies have investigated perceptual adaptation to speech in noise (Creelman, 1957; Mullennix, Pisoni & Martin, 1989). Both studies explored word identification in single- and mixed-talker conditions across three levels of noise. Decreased accuracy was observed in both studies in mixed-talker conditions where the talker changed compared to single-talker conditions where the talker remained consistent. However, the two studies differed in their findings regarding the relationship between talker variability and noise level. Creelman (1957) found the difference in performance between single- and mixed-talker conditions to be notably reduced at the highest noise level tested, while Mullennix, Pisoni and Martin (1989) found no relationship between talker variability and noise level. Further, limitations of both studies leave some questions unanswered as to how noise affects perceptual adaptation to speech. The percent of correct responses was the only outcome measure recorded in both studies, which does not provide any information regarding speech processing efficiency. Furthermore, both studies used fixed signal-to-noise ratios (SNRs), which may have represented quite different listening conditions depending on the individual. For instance, a 0 dB SNR could present a challenge for some listeners, whereas others may be able to identify the target word with ease (Surprenant &

Watson, 2001). Finally, talker variability was manipulated as a between-subject factor in the Mullennix, Pisoni and Martin study (1989), whereas noise level was varied within subjects, introducing additional complexity to interpreting the data. More research is clearly required to better understand how noise influences perceptual adaptation to speech.

1.3 The Current Project

The current study aimed to identify how masking noise and preceding speech context influence perceptual adaptation to speech. Participants performed a forced choice word recognition task in which they were asked to identify words spoken by a single talker or by mixed talkers. Preceding speech context was also manipulated, where participants heard target words presented both in isolation and preceded by a brief carrier phrase ("I owe you a…"). Participants performed this task in two listening environments: a *noise* condition in which target words were presented within a continuous stream of 4-talker babble, and a *quiet* condition in which words were presented without masking noise. To address limitations of previous studies investigating perceptual adaptation to speech in noise, we used an adaptive up-down procedure (Levitt, 1971) to establish the masking level at which participants achieved 70.7% accuracy on the forced choice word recognition task in the noise condition. This allowed us to identify adverse listening environments that placed similar cognitive demands across participants who may have individual differences in their ability to identify speech in noise.

In the *quiet* condition, we expected to replicate the findings from Choi and Perrachione (2019) that participants are slower to identify speech from mixed talkers and that preceding speech context facilitates processing of mixed-talker speech in ideal listening environments without competing stimuli. In the *noise* condition, we hoped to answer the following three questions:

Firstly, we asked whether talker variability (single vs. mixed talker) would influence reaction times in challenging listening environments, which we defined as SNRs within a small range around a participant's threshold SNRs. We expected to extend well-established findings of how perceptual adaptation to speech occurs in quiet listening environments (e.g., Choi & Perrachione, 2019; Mullennix & Pisoni, 1990; Nusbaum & Magnuson, 1997) to listening environments with masking noise. We anticipated that reaction times would overall be slower in the mixed-talker condition compared to the single-talker condition. Greater cognitive resources would be required to adapt to the variability in mixed-talker speech, resulting in decreased processing efficiency. We also expected SNR to influence reaction times, where lower (less favorable) SNRs would require more cognitive effort and thus result in reduced processing efficiency compared to higher (more favorable) SNRs. More adverse listening conditions are well known to result in decreased accuracy and increased cognitive effort (e.g., Pichora-Fuller, 2006; Zekveld, Kramer & Festen, 2011).

Secondly, we investigated whether talker variability (single vs. mixed talker) would influence participants' threshold SNRs where they achieved 70.7% accuracy in multi-talker babble. We expected that interference from processing mixed-talker speech would result in higher (more favorable) threshold SNRs in mixed-talker conditions compared to single-talker conditions. Greater cognitive resources would be required to process speech from mixed-talkers, which would reduce the cognitive resources available to detect the target words from masking noise in adverse listening conditions.

Finally, we explored how preceding speech context would influence speech processing in adverse listening environments in terms of both a) participants' reaction times and b) participants' threshold SNRs. We anticipated that preceding speech context would result both in faster reaction times and lower threshold SNRs in single- and mixed-talker conditions. The preceding carrier phrase would provide listeners with a longer, more continuous stimulus, allowing for additional time to identify and direct attention to relevant parts of the signal before the target word is encountered. Based on previous research regarding attention and auditory streaming, this would be expected to facilitate successful formation of an auditory object and thus reduce cognitive effort and increase perceptual sensitivity (Best, Ozmeral & Shinn-Cunningham, 2007; Kidd et al., 2005). We also expected that the facilitatory effect of preceding speech context would be greater for mixed versus single-talker conditions, as the preceding speech context would serve to reduce some of the additional uncertainty that is introduced by mixed-talker speech.

METHODS

2.1 Participants

Twenty-four native speakers of American English (20 female, 4 male, ages 18-31 years) were recruited to participate in this study. This sample size was based on power analyses detailed in previous studies investigating perceptual adaptation to speech (Choi, Hu, & Perrachione, 2018; Choi & Perrachione, 2019). All participants were self-reported to have normal speech, language, and cognition. All participants had normal hearing as assessed by pure-tone audiometry at octave frequencies from 250 Hz to 8 kHz. Participants were not previously exposed to the sound stimuli nor did they complete another study with the Cognitive Neuroscience Research Lab within the past year. None of the participants were familiar with any of the talkers used to record the auditory stimuli. All participants provided informed, written consent, which was approved and overseen by the Institutional Review Board at Boston University.

2.2 Stimuli

The target stimuli consisted of four minimal pair words, *boot*, *boat, bet,* and *bat,* spoken in standard American English. These four target stimuli were grouped into word pairs of *boot/boat* and *bet/bat*; these pairs were selected due to the substantial acoustic overlap that has been observed across talkers between /u/-/o/ and /ɛ/-/æ/ (Choi & Perrachione, 2018; Hillenbrand et al., 1995). In the task, the target words were either presented in isolation or preceded by a carrier phrase ("I owe you a [*boot/boat/bet/bat*]").

All words and carrier phrases were recorded from four native speakers of American English (2 female; 2 male). All stimuli were recorded in a sound-attenuated booth using a

Shure MX153 microphone headset, and the highest quality recordings were selected for each speaker. Carrier phrases and target words were recorded separately and later concatenated. Note that a single recording of the carrier phrase for each talker was used to ensure that listeners could not predict the upcoming target word based on differences in the carrier across trials.

The masking noise was taken from a four-talker babble recording of the QuickSIN (Killion, Niquette, Gudmundsen, Revit & Banerjee, 2004); this recording was spliced into 12, 52-second tracks. Each track was normalized to 60 dB SPL RMS amplitude using *Parselmouth* (Jadoul, Thompson, & de Boer, 2018) and *Praat* (Boersma, 2001). Throughout each block of the experiment, one randomly selected track of the babble noise was played, and repeated if the duration of the block exceeded the duration of the track. The babble noise from the QuickSIN was chosen because the QuickSIN is thought to be a reliable indicator of speech recognition in noise and is widely used clinically (Wilson, McArdle, & Smith, 2007). Multi-talker babble also affords a more realistic simulation of real-word listening conditions compared to other maskers such as speech-shaped noise.

2.3 Task Design and Procedure

Participants performed a two-alternative forced choice (2AFC) word identification task with one of the two minimal word pairs (*boot/boat* or *bet/bat*), equally assigned across participants. On each trial, participants heard a target word and provided a response to indicate which of the two words in the pair they heard using assigned keys on the number pad. Participants were instructed to respond as quickly and accurately as possible. Written instructions assigning numbers to the two target words ("boot" = 1, "boat" =2 or "bet" =

1, "bat" = 2) were provided. These instructions remained on the screen throughout the entire duration of each block. All trials had a duration of 2200ms, after which the experiment would automatically advance to the next trial regardless of whether or not the participant provided a response. No feedback was provided. Stimulus delivery was controlled using PsychoPy v.3.1.5 (Peirce, 2007). The study was completed in one 1.75-hour session. All participants were seated in a sound booth for the duration of the study. All stimuli were delivered through Sennheiser HD 380 Pro headphones

All participants performed the task across conditions manipulating *talker variability* (single vs. mixed) × *context* (isolated vs. carrier) × *presence of noise* (quiet vs. multi-talker babble), organized into eight blocks. In four consecutive blocks, words were presented without noise (i.e., quiet), and in the remaining blocks words were presented in multi-talker babble noise (i.e., noise); conditions manipulating *talker variability* (single vs. mixed) × *context* (isolated vs. carrier) alternated between the experimental blocks (Fig. 1). Order of blocks was counterbalanced using Latin Square permutations across participants. The resulting permutations were then organized so that half of participants were exposed to the quiet condition first while the other half were exposed to the noise condition first.

2.3.1 Quiet Condition

In the quiet condition, trials were divided into four blocks that varied based on *talker variability* (single talker vs. mixed talkers) and *context* (words preceded by the carrier phrase "I owe you a…" vs. words in isolation). Each block was comprised of 192 trials. Trials were presented in pseudo-randomized order so that each participant heard each target word (either *boot/boat* or *bet/bat*) with an equal probability. All stimuli were recorded, analyzed and normalized to 65 dB SPL RMS amplitude using *Praat* (Boersma, 2001).

In single-talker blocks, trials were blocked so that stimuli from each of the four talkers were presented in 48 consecutive trials, and the order of talkers was randomized. In mixed-talker blocks, stimuli from each talker were distributed throughout the block, with the restriction that no two stimuli from the same talker were presented in successive trials.

2.3.2 Noise Condition

In the noise condition, trials were divided into four blocks that varied based on *talker variability* (single talker vs. mixed talkers) and *context* (words preceded by the carrier phrase "I owe you a…" vs. words in isolation). A 1 up/2 down adaptive staircase tracking procedure (Levitt, 1971) was used during each block to establish the threshold SNR at which the participant achieved 70.7% accuracy. Within each block, four adaptive tracks were completed. In the single-talker blocks, one adaptive track was completed for words spoken by each talker, and the order of talkers was randomized. In mixed-talker blocks, talkers were randomly distributed within each adaptive track, with the restriction that same two talkers could not appear in consecutive trials.

Each adaptive track began with an initial SNR of 10 dB. The level of the multi-talker babble masker was fixed at 60 dB SPL throughout the adaptive track to reduce possible discomfort to the participant. The level of the target speech stimuli was varied using *Parselmouth* (Jadoul, Thompson, & de Boer, 2018) and *Praat* (Boersma, 2001) to manipulate the SNRs in accordance with the accuracy of the participants' responses. If the participant correctly selected the target word twice in a row, the SNR decreased by a given step size. Otherwise, the SNR increased by a given step size. The initial step size was 4 dB, and was halved after the second and fourth reversal for a final step size of 1dB. Each staircase terminated when 8 reversals were reached. The mean number of trials per staircase across participants was 37 (range 17-60). The final four reversal values were averaged to determine the threshold SNR (Fig. 2).

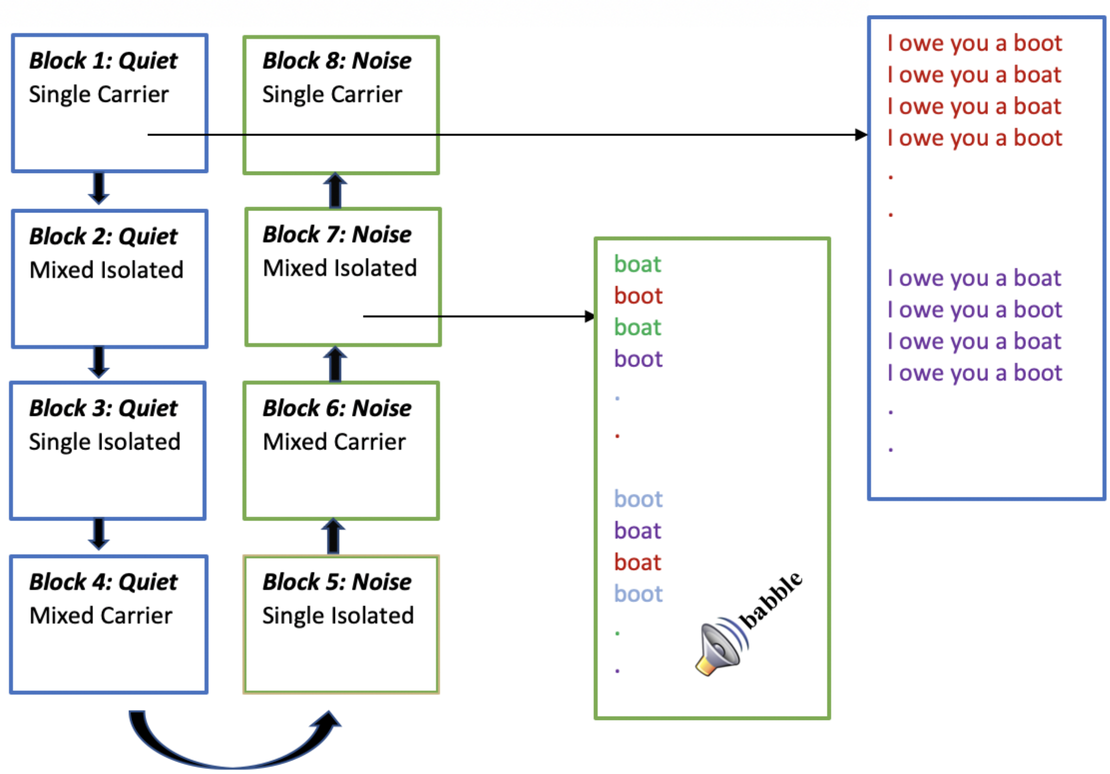

Figure 1. Task design. This figure illustrates a sample block order of the experiment for one participant and provides examples of two blocks. The *quiet* condition is represented in blue, and the *noise* condition is represented in green. Different colors represent different talkers.

2.4 Data Analysis

We assessed the effect of *talker variability* (single vs. mixed), *context* (isolated vs. carrier), and *presence of noise* (quiet vs. multi-talker babble) on participants' word identification

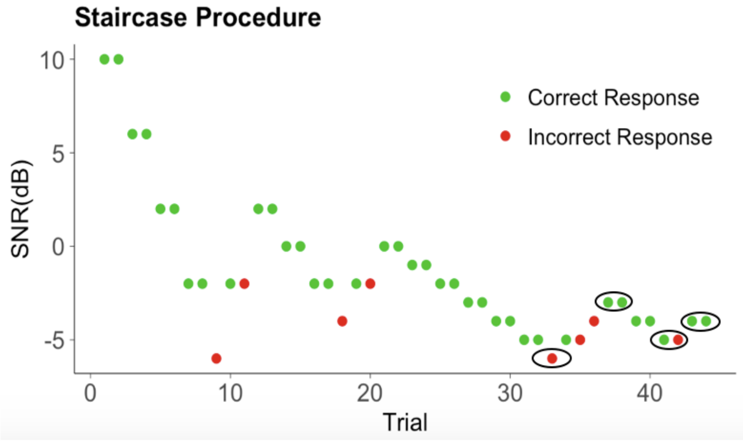

Figure 2. 1 up/2 down adaptive staircase procedure. Green circles represent correct responses and red circles represent incorrect responses. Circled values indicate reversal values that were averaged to obtain the threshold SNR for 70.7% accuracy.

performance. Participants' response times and accuracy of each trial were measured and analyzed. Reaction times were calculated from the onset of the target word and log-transformed to more closely approximate a normal distribution. Only response times from correct trials were included in the analyses. Any response time that was more than three standard deviations from each participant's mean log response time was excluded from the analysis.

For assessing performance in the noise condition, we first obtained participants' SNR thresholds for each of the four experimental conditions (*talker variability × context*). The threshold SNR for each tracking procedure was calculated by averaging the SNR values from the final four reversals of each staircase. Four staircases were completed in each single- and mixed-talker block to obtain a more accurate estimate of participants' true threshold SNRs. Each of these four threshold SNR values were included in the analyses for a total of four threshold SNR values per block. Only threshold SNRs that were more than three standard deviations from the mean threshold SNR across participants were excluded from the analysis. Based on the thresholds SNRs, we quantified participants' response times in performing the task under babble noise. Only response times from trials at SNRs within range of the final four reversal values of each staircase were included in this analysis. This range was selected so that participants' reaction times in a quiet environment could be compared to participants' reaction times in a challenging listening environment with similar cognitive demands across participants.

All analyses were completed using R (v3.6.1). Three linear mixed-effects models implemented through the *lmerTest* package (v3.1.1) were used to perform the analyses. The first model investigated the effects of *talker variability* (single vs. mixed), *context* (isolated vs. carrier), and *presence of noise* (quiet vs. multi-talker babble) on participants' response times. Thus, reaction time was included as the dependent measure and *talker variability*, *context*, and *presence of noise* were included as fixed factors. The first model also contained random effect terms of within-participant slopes for *talker variability*, *context* and *presence of noise* and random intercepts for participants, as well as slopes and random intercepts for each stimulus word spoken by all four talkers.

The second model investigated the effect of *talker variability* (single vs. mixed) and *context* (isolated vs. carrier) on participants' threshold SNRs in the noise condition. Thus, threshold SNR was included as the dependent measure and *talker variability* and *context* were included as fixed factors. The second model also contained random effect terms of within-participant slopes for *talker variability* and *context* and random intercepts for participants, as well as slopes and random intercepts for each stimulus word spoken by all four talkers.

The third model investigated the effect of SNR on participants' reaction times at individual trials in the noise condition. Trial-wise reaction time throughout each adaptive track was included as the dependent measure and *SNR*, *talker variability*, and *context* were included as fixed factors. The third model also contained random effect terms of within-participant slopes for *talker variability*, *context* and *SNR* and random intercepts for participants, as well as slopes and random intercepts for each stimulus word spoken by all four talkers.

Significance of fixed factors was determined by Type III analyses of variance (ANOVAs) for each of the three linear mixed-effects models. Post-hoc pairwise analyses were also performed to follow significant ANOVA results using differences of least-squares means tests via *difflsmeans.* A significance criterion of $\alpha = 0.05$ was adopted, with p-values based on the Satterthwaite approximation of the degrees of freedom.

RESULTS

Mean total accuracy across participants in the quiet condition was at ceiling (mean = 97.2% ± 2.4%), indicating sufficient sustained attention to the task. Results for each of the three linear mixed-effects models are summarized below. Mean reaction times are summarized in Table 1 and mean threshold SNR values are summarized in Table 2.

3.1 Reaction Times in Quiet and in Noise

	Quiet-Isolated	Quiet-Carrier	Noise-Isolated	Noise-Carrier
Single Talker	717 ± 189	695 ± 144	870 ± 221	781 ± 146
Mixed Talker	796 ± 221	734 ± 162	893 ± 218	794 ± 154
Differences	79	39	23	13

Table 1. **Reaction times in quiet and noise.** This table shows the mean ± s.d. reaction time (ms) across participants for each experimental condition in quiet and in noise. In noise, reaction times were included only from SNRs within the final four reversal points of each adaptive track. Difference in reaction times (mixed - single) are also included.

Mean reaction times for identifying words in quiet and noise conditions are shown in Figures 3-4. The first linear mixed-effects model examined the effects of *talker variability* (single vs. mixed), *context* (isolated vs. carrier), and *presence of noise* (quiet vs. multi-talker babble) on participants' reaction times. Only reaction times at SNRs near threshold (i.e., within the final four reversal points of each adaptive track) were included in this analysis. The model revealed significant main effects of *presence of noise* (F(1,23) = 53.45; $p << 0.001$), *talker variability* (F(1,25) = 41.28; $p << 0.001$) and *context* (F(1,24)

= 38.00; p << 0.001). Post-hoc analyses revealed that reaction times were slower in the noise condition than in the quiet condition (β = 0.130, s.e. = 0.018, t = 7.31, p << 0.001). Furthermore, participants were generally slower to respond in trials of mixed-talker blocks compared to single talker blocks (β = 0.051, s.e. = 0.008, t = 6.34, p << 0.001), and reaction times were faster when a carrier phrase was present compared to when words were presented in isolation (β = -074, s.e. = 0.012, t= -6.16, p << 0.001).

Importantly, the model also revealed significant interaction effects. There were significant two-way interactions of *presence of noise* × *talker variability* (F(1, 23398) = 80.53; p << 0.001), *presence of noise* × *context* (F(1, 23402) = 91.93; p << 0.001), and *talker variability* × *context* (F(1, 23401) = 33.17; p << 0.001), and a significant three-way interaction of *presence of noise* × *talker variability* × *context* (F(1, 23402) = 6.82; p << 0.01), suggesting that the effect of context and talker variability as well as the interaction between them changed based on presence of noise.

Post-hoc analyses revealed that participants' reaction times in the quiet condition were overall higher in the mixed-talker than the single-talker condition both when the target words were preceded by a carrier phrase (β = 0.053, s.e. = 0.008, t = 6.21, p << 0.001) and presented in isolation (β = 0.102, s.e.= 0.008, t= 12.16, p << 0.001). However, the difference between reaction times in single- and mixed-talker blocks was significantly reduced when speech was presented under noise; there was a significant difference in word identification speed in the mixed- vs. single-talker conditions when words were presented in isolation (β = 0.033, s.e. = 0.010, t = 3.22, p = 0.019), but there was no difference between the talker conditions when target words were preceded by a carrier phrase (β =

0.015, s.e. = 0.010, *t* = 1.41, *p* = 0.16) under noise.

Furthermore, post-hoc analyses showed that, in quiet, the effects of the carrier phrase on reaction time were driven by the mixed-talker blocks. Processing mixed-talker speech was significantly faster when there was a carrier phrase compared to when target words were presented in isolation (β = -0.070, s.e. = 0.013, *t* = -5.68, *p* << 0.001), but there was no difference in reaction times for identifying single talkers' spoken words regardless of the presence of carrier phrase (β = -0.020, s.e. = 0.012, *t* = -1.62, *p* = 0.116). On the contrary, for identifying words under noise (near threshold SNRs), participants were faster when they identified words following a carrier phrase for both single- and mixed-talker conditions (single: β = -0.093, s.e. = 0.136, *t*= -6.80, *p* << 0.001; mixed: β = -0.112, s.e. = 0.014, *t* = -8.14, *p* << 0.001).

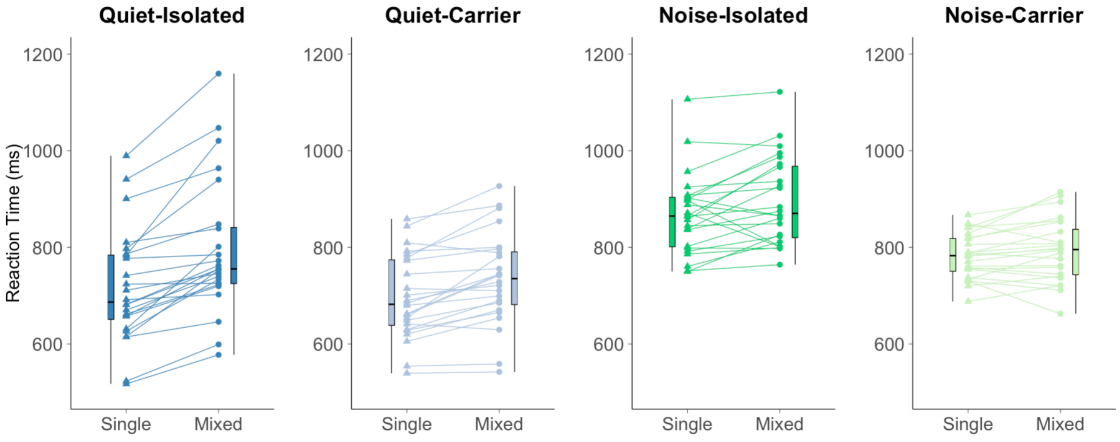

Figure 3. Effect of talker variability, context, and presence of noise on reaction time. Connected points show the difference in reaction time between single- and mixed-talker blocks for individual participants across quiet and noise conditions with and without a carrier phrase. Reaction times were taken only from SNR values within range of final four reversal points of each adaptive track in the noise condition. Box plots indicate the distribution (median, interquartile range, maximum, minimum) for each condition.

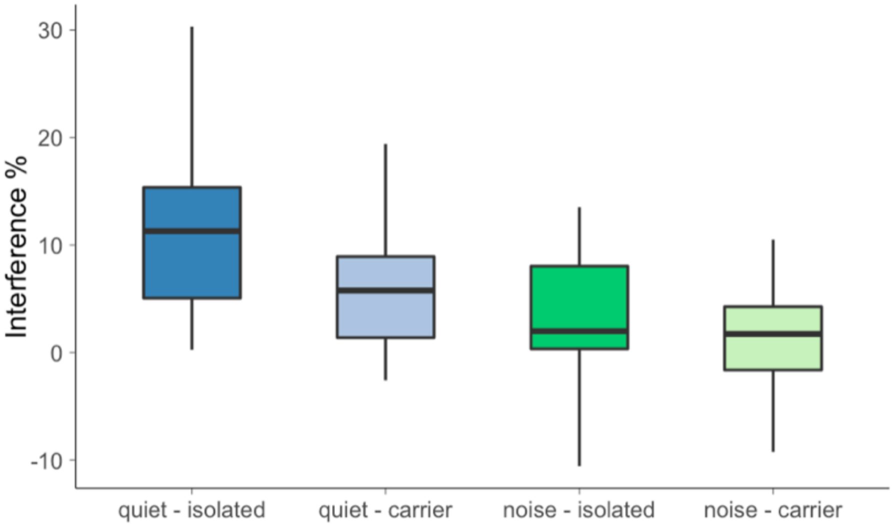

Figure 4. Interference from processing mixed-talkers' speech. Box plots show the differences in response time between the mixed- and single-talker blocks across quiet and noise conditions both with and without a carrier phrase. Differences are scaled within participants to their response time in the single-talker blocks: ((mixed-single)/single) x 100). Significant interference was observed for all but the noise-carrier condition.

3.2 Effect of Talker Variability and Context on SNR Thresholds

	Isolated	Carrier
Single Talker	-7.6 ± 3.6	-10.9 ± 2.6
Mixed Talker	-4.6 ± 2.7	-6.4 ± 2.8

Table 2. **Threshold SNRs.** This table shows the mean ± s.d. threshold SNR (dB) across participants for each experimental condition in noise.

The second linear mixed-effects model examined the effects of *talker variability* (single vs. mixed) and *context* (isolated vs. carrier), on participants' threshold SNRs in the noise condition. Figure 5 illustrates threshold SNRs and the adaptive tracking of SNRs in the four conditions. A Type III ANOVA revealed a significant main effect of *talker variability* ($F(1,22) = 122.93$; $p << 0.001$). Post-hoc analyses indicated that participants achieved 70.7% accuracy at lower (less favorable) SNRs in single-talker blocks compared to mixed-talker blocks ($\beta = 3.58$, s.e. = 0.322, $t = 11.11$, $p << 0.001$). There was also a significant main effect of *context* ($F(1,22) = 57.88$; $p << 0.001$), where participants' threshold SNRs were lower when words were preceded by a carrier phrase compared to when words were presented in isolation ($\beta = -2.53$, s.e. = 0.332, $t = -7.62$, $p << 0.001$).

Furthermore, the model revealed a significant two-way interaction of *talker variability* × *context* ($F(1,300) = 7.00$; $p = 0.009$). Post-hoc analyses indicated that there was a larger difference between the threshold SNRs achieved in the isolated and carrier conditions in the single-talker blocks ($\beta = -3.31$, s.e. = 0.445, $t = -7.45$, $p << 0.001$) compared to mixed-talker blocks ($\beta = -1.75$, s.e. = 0.445, $t = -3.92$, $p << 0.001$). As illustrated in Figure 6, there was notable variability with respect to the audibility of individual talker's speech in noise; nevertheless, lower threshold SNRs were consistently achieved for each talker's speech when a carrier phrase was present compared to when words were presented in isolation.

3.3 Influence of SNR on Reaction Time

The third linear mixed-effects model examined whether participants' reaction times in the noise condition were affected by *talker variability* (single vs. mixed), *context* (isolated vs. carrier), and *SNR* (values ranging from 26 dB to -26 dB). For this analysis, reaction times from all correct responses and trial-wise SNRs were included regardless of threshold SNR. Figure 7 illustrates how SNR affected reaction times across the four experimental conditions in noise. A significant main effect was found for *SNR* ($F(1,25) = 86.04$; $p << 0.001$), where reaction times increased as SNRs decreased (i.e., the listening condition became more adverse) ($\beta = -0.006$). Consistent with the previous analysis of reaction times at SNRs near threshold, a significant main effect of *context* ($F(1,24) = 43.85$; $p << 0.001$) was present. Participants were faster at identifying target words when the words were preceded by a carrier phrase compared to when they were presented in isolation ($\beta = -0.107$, s.e. = 0.014, $t = -7.71$, $p << 0.001$). Furthermore, there was also a significant main effect of *talker variability* ($F(1,29) = 45.99$; $p << 0.001$). Post-hoc testing revealed that participants were faster to respond in single-talker compared to mixed-talker blocks ($\beta = 0.044$, s.e. = 0.006, $t = 6.99$, $p << 0.001$).

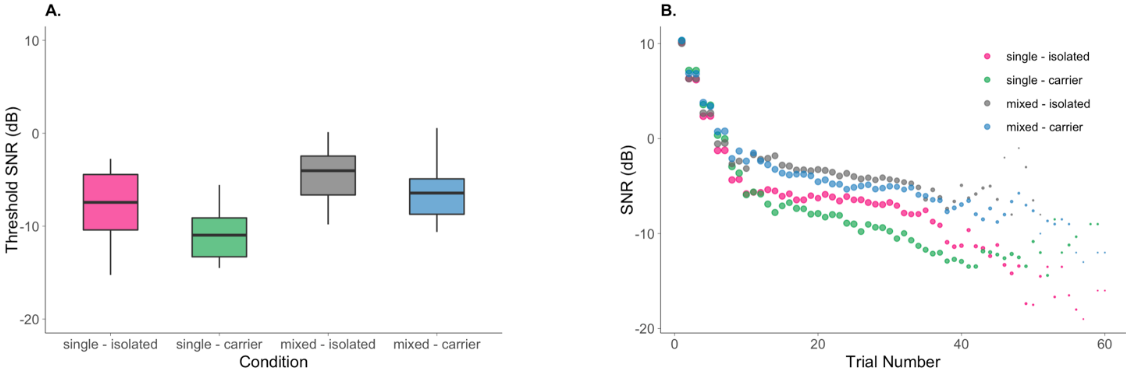

Figure 5. Effects of talker variability and context on threshold SNRs and adaptive tracking of SNRs. A) Box plots indicate the distribution (median, interquartile range, maximum, minimum) of threshold SNRs for the four experimental conditions in noise. B) Each point indicates the mean trial-wise SNR across participants. Colors represent each experimental condition in noise. Size of the points represents the number of data points included at individual trial numbers. Larger points indicate more data points, and smaller points indicate fewer data points (range is 1-96).

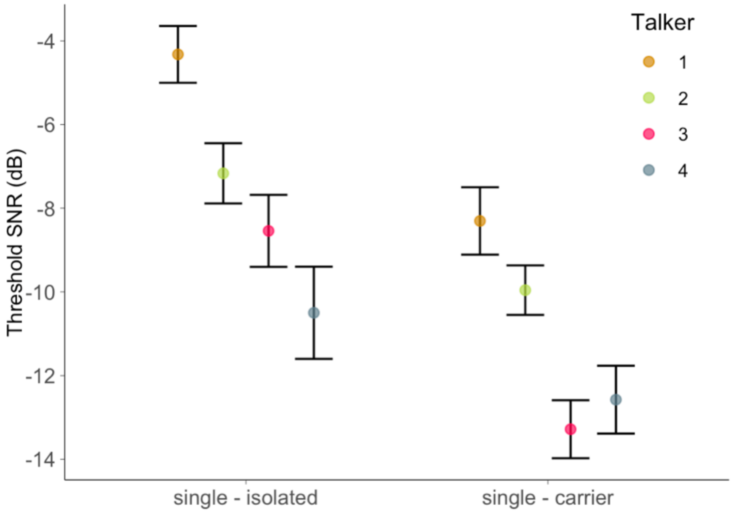

Figure 6. Threshold SNRs for individual talkers. Each point indicates the mean threshold SNR per talker across participants in single-talker blocks (single-isolated, single-carrier). Each color represents a different talker. Bars indicate standard error of the mean.

Consistent with the results from the first model, there was also a significant two-way interaction of *talker variability* × *context* ($F(1,11112) = 18.9$; $p << 0.001$). The presence of a carrier phrase had a greater effect in mixed-talker ($\beta = -0.125$, s.e. = 0.015, $t = -8.52$, $p << 0.001$) compared to single-talker blocks ($\beta = -0.090$, s.e. = 0.014, $t = -6.21$, $p << 0.001$). A significant two-way interaction of *SNR* × *context* was also revealed ($F(1,10295) = 28.23$; $p << 0.001$), where carrier phrases led to faster reaction times at less favorable SNRs than more favorable SNRs ($\beta = 0.001$). There was no significant two-way interaction between *SNR* and *talker variability* ($F(1,3508) = 0.134$; $p = 0.71$) or three-way interaction between *SNR*, *talker variability,* and c*ontext* ($F(1,10063) = 1.53$; $p = 0.22$).

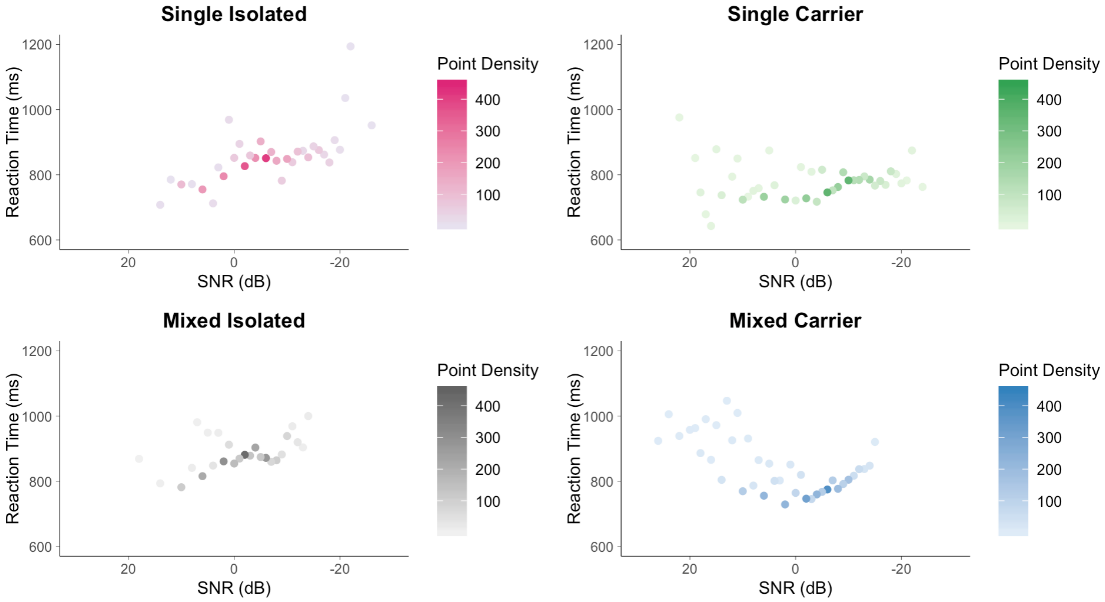

Figure 7. Effect of SNR on reaction time across four experimental conditions in noise. Each point represents the mean reaction time across participants at a given SNR for each experimental condition in noise. Colors represent different conditions. Darkness of colors represents number of data points present at given SNRs. Lighter colors indicate a smaller number of data points, and darker colors indicate a larger number of data points (range is 1-450).

DISCUSSION

4.1 Summary of Results

This study is one of the first to explore how listeners process speech with different levels of talker variability under both ideal and adverse listening conditions, as well as how listeners may utilize preceding speech context to facilitate speech processing in different listening conditions. Overall, our results suggest that talker variability and preceding speech context have similar effects on speech perception both in quiet environments and environments with masking noise, indicating that the same mechanisms may underlie perceptual adaption to speech in quiet and in noise.

Our results replicated well-established effects of talker variability and preceding speech context on processing speech without noise (e.g, Choi & Perrachione, 2019). In quiet listening environments, participants were faster at identifying speech spoken by a single consistent talker compared to speech spoken by multiple different talkers, and the performance decrement between the single- and mixed-talker conditions was smaller when target words were preceded by a brief carrier phrase. These findings reflect the additional processing costs that are incurred by accommodating mixed-talker speech, as well as demonstrate the efficiency gains that are afforded by preceding speech context in ideal listening conditions.

Our results further showed that interference from processing mixed-talker speech can also be observed under noise. Participants achieved 70.7% accuracy at significantly lower (less favorable) SNRs in single-talker conditions compared to mixed-talker conditions, indicating that participants were less able to correctly identify speech in noise when the talker changed compared to when the talker remained consistent. These findings confirm and extend previous findings (Creelman, 1957; Mullennix, Pisoni, & Martin, 1989) that talker changes have a detrimental effect on listeners' ability to understand speech in noise, and suggest that the additional cognitive effort that is required to process mixed-talker speech reduces the cognitive resources available to extract the target speech from competing speech signals. Thus, parsing mixed-talker's speech will place higher cognitive demands on the listener at relatively low noise levels compared to speech from a consistent single talker, resulting in reduced accuracy when identifying mixed-talker speech in the presence of background noise or competing stimuli.

Our findings also revealed preceding speech context to have an effect on participants' threshold SNRs. As hypothesized, participants' thresholds for achieving 70.7% accuracy occurred at significantly higher noise levels when target words were preceded by a carrier phrase compared to when target words were presented in isolation. The carrier phrase allowed participants more time to detect, isolate, and direct attention to acoustic stimuli from the target talker, thereby better enabling participants to filter out interference from the multi-talker babble and correctly identify the target word.

While participants were expected to achieve lower threshold SNRs both in single- and mixed-talker conditions when a carrier phrase was present, our results revealed the unexpected finding that carrier phrases had a more beneficial effect on participants' threshold SNRs in single- compared to mixed-talker conditions. It was hypothesized that carrier phrases would have a more beneficial effect in mixed-talker conditions, as the processing of mixed-talker speech is more cognitively demanding and thus would be expected to show greater performance differences when cognitive and attentional resources are made more available or better allocated. One possible explanation for this finding is that cognitive and attentional demands in mixed-talker conditions reached participants' maximum processing capacity due to the increased signal variability and high noise level. Thus, the potential benefit of preceding speech context was reduced because cognitive resources were not as readily available for processing speech, after resources had already been preallocated to accommodate the talker variability and to parse the signal from the noise. Another explanation might stem from the wide variation present in the threshold SNRs for each of the four individual talkers. Differences in each talker's ability to be heard in noise may have resulted in less accurate estimates of threshold SNRs in the mixed-talker staircases, as it is possible that the staircases may have been driven up or down based on whether an individual talker was more or less challenging to understand in noise. However, visual analysis of each individual adaptive track across mixed- and single-talker conditions did not indicate a failure to converge or markedly high variation in reversal points in mixed-talker conditions. Further studies may explore the effects of talker variability and preceding speech context on perceptual adaptation to speech in noise with talkers who are matched for their ability to be heard in background noise.

In contrast to the clear interference from processing mixed-talker speech observed in participants' word identification speeds in the quiet condition and in participants' threshold SNRs in the noise condition, little mixed-talker-related interference was observed in participants' word identification speeds under noise at SNRs near threshold. No differences were observed in word identification speeds between mixed- and single-talker conditions when the target words were preceded by a carrier phrase, and participants were slightly faster in single-talker conditions when target words were presented in isolation. Differences in threshold SNRs between the talker conditions could provide an explanation for these results. Threshold SNRs were significantly lower in single- compared to mixed-talker conditions. Thus, while the noise levels in mixed- and single-talker conditions presented equally challenging listening environments, single-talker response times were taken from trials with significantly higher levels of noise. Analyses of participants' word identification speeds across the entirety of the adaptive tracks found SNR to have a significant effect on processing efficiency. The more noise present in the signal, the slower participants responded, and SNR had the same deleterious effect on processing speed in single- and mixed-talker conditions. Therefore, only a small amount of mixed-talker interference was observed in this analysis because it had already been accounted for by the differences in thresholds.

Furthermore, consistent with the hypotheses, our results found preceding speech context to result in greater increases in word identification speed at noise levels near threshold compared to in quiet. In noise, word identification speeds were faster both in single- and mixed-talker conditions when target words were preceded by a carrier phrase, whereas in quiet preceding speech context facilitated processing only in mixed-talker conditions. These results can be explained by the increased cognitive and attentional demands associated with listening to speech in noise. Because identifying the target speech stream and extracting phonemically-diagnostic information are more difficult when competing speech signals are present, the additional time to lock on to the speech stream provided by the carrier phrase played a larger role in processing efficiency in noise compared to in quiet. In quiet, target speech streams are more easily formed and salient acoustic-phonemic patterns are more readily identified, thereby limiting the efficiency gains that could be demonstrated behaviorally.

4.2 Theoretical Implications

Taken together, these findings are consistent with both active control process (Magnuson & Nusbaum, 2007; Heald & Nusbaum, 2014) and auditory streaming models of speech perception (Shinn-Cunningham, 2008), and suggest that both an active control mechanism and auditory streaming play key roles in perceptual adaptation to speech. Active control process models can explain why a processing cost is incurred in mixed-talker conditions, even at long timescales (Kou, 2019). Active control models propose that mixed-talker speech requires that cognitive resources be set aside for processing talker variability, or the greater uncertainty, in the speech signal. Thus, there are always performance costs associated with mixed-talker conditions because the cognitive resources available for speech processing are limited. In line with this idea, participants in the present study were more accurate in identifying speech at low SNRs in single-talker conditions compared to mixed-talker conditions because more resources were available to extract the speech signal from the noise.

On the other hand, models of auditory streaming can provide an explanation for why carrier phrases facilitated speech processing both in single- and mixed-talker conditions in noise. Auditory streaming models propose that the length and temporal continuity of auditory signals are critical for auditory object formation and successful allocation of attention, thereby increasing perceptual sensitivity and decreasing the cognitive cost for perceptual identification (e.g., Best et al., 2008). Thus, participants were more accurate in identifying speech at low SNRs when a carrier phrase was present because the carrier phrase allowed them to better allocate attention to the talker and isolate the target speech from the background noise.

Our results not only suggest that both active control processes and auditory streaming play a role in perceptual adaptation to speech, but that the two may have additive effects. In the mixed-talker condition when words were presented in isolation, we found that participants were least able to correctly identify words in adverse listening environments. Not only could participants not benefit from auditory streaming, but also fewer resources were available for speech processing due to the preallocation of cognitive resources for resolving ambiguities in the speech signal introduced by talker variability. Conversely, participants were best able to identify speech in noise in the single-talker condition when words were preceded by a carrier phrase, as more resources remained available for speech perception and participants could benefit from the longer, more continuous stream. In the mixed-talker condition when words were preceded by a carrier phrase and in the single-talker condition when words were presented in isolation, participants' ability to identify speech in noise was in the middle, and there was only a small performance difference between mixed-talker carrier and single-talker isolated conditions. These findings can be explained by participants benefiting either from reduced talker variability or increased signal length. In the mixed-talker condition with carrier phrases, cognitive resources had to be set aside, but participants could benefit from auditory streaming. In the single-talker isolated condition, the benefits of auditory streaming were limited due to the short duration of the signal, but more resources were available for speech processing. Thus, these findings extend recent preliminary results that both active control processes and auditory streaming play a complementary role in perceptual adaptation to speech (Kapadia & Perrachione, 2019; Kou, 2019), with active control processes acting over longer timescales and auditory streaming assisting with feedforward allocation of attention in the short term.

4.3 Clinical Implications

Our findings revealed that participants were faster to identify speech when words were preceded by a carrier phrase both under noise and in ideal listening environments, and that participants were most accurate in identifying speech in noise when words were preceded by a carrier phrase and spoken by a consistent single talker. Our findings further revealed that participants' word identification speed decreased as the level of background noise increased. Not only do these findings contribute to our understanding of the underlying mechanisms involved in speech perception, but these findings also hold important clinical implications. Our findings indicate that background noise should be minimized whenever possible in order to decrease cognitive effort required to understand speech content, especially for individuals whose speech processing is already complicated due to hearing loss, language disorders, attentional disorders, or auditory processing disorders. Our results also suggest that maintaining a single consistent talker's speech while presenting information may be an effective strategy for reducing cognitive effort and increasing processing efficiency, especially in environments with high amounts of background noise. If talker changes do occur or if background noise cannot be eliminated, more processing time could be given to compensate for the increased listening demands. If possible, assessments evaluating speech, language, or cognitive skills should also be administered by the same examiner due to the increased signal variability introduced by multiple different talkers. A consistent exam administrator may reduce cognitive effort required to process the examiner's speech, allowing for more resources to be directed towards the assessment tasks.

4.4 Limitations and Future Directions

One limitation of this study is that it only included participants with normal hearing who did not have any history of language delays or disorders. While this design allowed for within-participant comparisons, our findings may not be representative of how speech perception may occur in individuals with diverse language and hearing profiles. Furthermore, none of the participants included in this study were over the age of 35. As both age and hearing status have been found to have a significant impact on speech processing and the amount of cognitive resources that are required to identify speech in noise (Pichora-Fuller, 2006), further studies could explore how talker variability and preceding speech context influence speech processing in older adults and individuals with hearing loss.

A further limitation of this study is that target stimuli were restricted to a limited set of target words and a single carrier phrase. Participants were also presented with a forced choice on each trial rather than an open response, which allowed participants to expect what they might hear next. In realistic hearing environments, incoming speech is much more variable, and listeners must select what word they heard from a large array of possibilities. Further research might explore how perceptual adaptation to speech may occur with more variable, naturalistic preceding speech context or an open set response.

Another limitation of this study was that fewer trials were included from the noise condition than the quiet condition in the first analysis comparing word identification speeds in quiet and in masking noise. This is because only correct trials with SNRs within the final four reversal points of each adaptive track were included in the analysis from the noise condition, whereas all correct trials from the quiet condition were included in the analysis. However, a separate analysis was conducted examining word identification speed throughout the entirety of each adaptive track and confirmed that the effects of talker variability and preceding speech context on word identification speed were consistent over a larger number of trials across SNRs in the noise condition.

Additionally, because the distance between the highest and lowest SNR values of the final four reversal points differed for each adaptive track, there was some variation in the range of SNRs that were defined to be “near threshold” both between and across participants depending on the shape of their adaptive tracks. However, visual analysis of the adaptive tracks did reveal a general tendency for adaptive tracks to noticeably converge around a SNR value, suggesting that the final four reversal points did effectively represent the noise levels where word identification was challenging. Further research could investigate how perceptual adaptation to speech occurs at several different noise levels. Adaptive tracking staircase procedures may be used to establish participant-specific threshold SNRs, which could then be used to define specific SNR values representing low effort, medium effort, or high effort listening conditions in noise.

Other avenues for future research that could further investigate the role that auditory steaming may play in perceptual adaptation to speech include providing listeners with spatial cues indicating where an upcoming speech signal may be presented. Spatial cues have been found to improve word identification accuracy in the presence of competing stimuli (Kidd et al., 2005), and talker changes have been found to influence the degree that spatial cues can facilitate processing (Best et al., 2008). It would be interesting to explore how spatial information and preceding speech context may interact to facilitate perceptual adaptation to speech in both single- and mixed-talker environments. Such findings could further inform how listeners process speech from multiple different talkers in more realistic environments where speech is encountered not only alongside other competing auditory stimuli, but also from varying spatial locations.

4.5 Conclusion

Our results showed that listeners are less accurate and slower to identify speech presented in multi-talker babble when the target talker changes compared to when the target talker remains consistent, and that preceding speech context enhances word identification performance under noise both in single- and mixed talker conditions. These results extend previous findings of the effects of talker variability and preceding speech context on speech processing in quiet environments to more realistic listening conditions with masking noise, suggesting that the same underlying mechanisms may serve to perceptually adapt to speech both in quiet and in noise. Overall, our findings suggest that both attentional and cognitive mechanisms may interact to explain the efficiency gains afforded by preceding speech context. An active control process may serve to preallocate cognitive resources to support processing of talker variability, and auditory streaming processes may serve to support successful feedforward allocation of attention to salient talker-specific stimuli over shorter time scales. Further research might explore how perceptual adaptation to a talker might occur when spatial information is provided in order to broaden our understanding of what information can support perceptual adaptation to speech.

BIBLIOGRAPHY

Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal for the Association for Research in Otolaryngology*, 8(2), 294-304.

Best, V., Ozmeral, E. J., Kopco, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(35), 13174–13178.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, 5, 341–334.

Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attention, Perception, & Psychophysics*, 80(3), 784-797.

Choi, J. Y. & Perrachione, T. (2019). Time and information in perceptual adaptation to speech. *Cognition*, 192.

Creelman, C. D. (1957). Case of the unknown talker. *The Journal of the Acoustical Society of America*, *29*(5), 655-655.

Green, K.P., Tomiak, G.R. & Kuhl, P.K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics,* 59, 675–692.

Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, 8, 35.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of Americ*a, 97(5), 3099-3111.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71, 1-15.

Johnson, K. (1990). The role of perceived speaker identity in F 0 normalization of vowels. *The Journal of the Acoustical Society of America*, 88(2), 642-654.

Kapadia, A.M. & Perrachione, T.K. (2019). Processing costs imposed by talker variability do not scale with number of talkers. 19th International Congress of Phonetic Sciences (Melbourne, August 2019).

Kidd Jr, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, *118*(6), 3804-3815.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America,* 116(4), 2395-2405.

Kou, R. S. N. (2019). *Time course of talker adaptation.* [Master's thesis, Boston University]. ProQuest Dissertations and Theses Global.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467-477.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 391–409.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369.

Morton, J. R., Sommers, M. S., & Lulich, S. M. (2015). The effect of exposure to a single vowel on talker normalization for vowels. *The Journal of the Acoustical Society of America*, 137(3), 1443-1451.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47(4), 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, 85(1), 365-378.

Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson and J.W. Mullenix (eds.) *Talker Variability in Speech Processing*, 109-132. San Francisco: Morgan Kaufmann Publishers Inc.

Perrachione, T. K., Del Tufo, S. N., Winter, R., Murtagh, J., Cyr, A., Chang, P., ... & Gabrieli, J. D. (2016). Dysfunction of rapid neural adaptation in dyslexia. *Neuron*, *92*(6), 1383-1397.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203.

Pichora-Fuller, M. K. (2006). Perceptual effort and apparent cognitive decline: Implications for audiologic rehabilitation. *Seminars in Hearing* 27(4), 284-293.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5), 182-186.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *The Journal of the Acoustical Society of America*, 60(1), 213-224.

Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, 110(4), 2085-2095.

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*.

Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, 16(7), 1173-1184.

Zhang, C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., Peng, G.,Wang, W. S-Y. (2016). Functionally integrated neural processing of linguistic and talker information: an event-related fMRI and ERP study. *NeuroImage*, 124, 536-549.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, 32(4), 498-510.

VITA

[redacted]