THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF COMMUNICATION SCIENCES AND DISORDERS

Effect of Speaker Voice and Background Language on Spoken Word Recall

AMIRA EL-DINARY
SPRING 2023

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree in Communication Sciences and Disorders
with honors in Communication Sciences and Disorders

Reviewed and approved* by the following:

Navin Viswanathan
Professor of Communication Sciences and Disorders
Thesis Supervisor

Carol Miller
Professor of Communication Sciences and Disorders
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Understanding spoken language is a skill that we rely on every day but think little about. The task of listening is complex and can be difficult in a variety of situations. There are differences between each of our voices that listeners must adjust to in any given conversation. Additionally, surrounding sounds compete for attention and pull it away from the conversation at hand. Previous studies have shown that recall accuracy improves when words are repeatedly spoken by a "known voice" (Mattys and Liss, 2008, Nygaard and Pisoni, 1998). It has also been found that listeners recognize speech better when the words in the background are in a different language than the words that someone is attending to (Brouwer et al., 2012, Calandruccio et al., 2013, Viswanathan et al., 2016). By adding background sentences in the exposure-test design from typical known voice studies, the current study used an online experiment which evaluated whether these effects can be measured in combination to search for possible interactive effects. Findings from this study extend to the fields of audiology, psychology, linguistics, and speech therapy.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This work was supported by NSF grants OISE PIRE#1545900 and B2 #2126888.

The contents of this paper would not be possible without the support of many professors, mentors, and friends. Thank you, Penn State's Department of Communication Sciences and Disorders. The faculty from this department have taken great interest in the success of the students. Specifically, Dr. Navin Viswanathan and Dr. Carol Miller.

Dr. Viswanathan advised me not only on the writing of this thesis but on the creation of the research project. I am grateful for his guidance in reading academic articles, forming a research question, and creating measures online to test the hypothesis. As the supervisor of this thesis, he has dedicated his time and energy to my success. Navin, it has been a pleasure working with you. Similarly, Dr. Miller has taken the time and care to work with me on finding my path. As my honors advisor, she played a large role in my academic achievements. Carol, thank you for supporting me in exploring my interests. I appreciate all the guidance that you offer.

I extend a final thank you to the faculty and friends outside of the Department of Communication Sciences and Disorders for their support of the research project discussed in this paper. Thank you, Amanda Gryskevicz for working diligently with me over the past two years. This project would not be possible without your companionship and teamwork. Dr. Janet Van Hell, I appreciate your guidance in the beginning stages of this project. And Dr. Susanne Brouwer, thank you for your support and contributions. Your work inspired this project. It was an honor to work with you in person, and I deeply appreciate all that you have done for this research study. Thank you all for making this possible.

Chapter 1

Introduction

Language is a uniquely human ability that impacts every aspect of our lives. Anatomical features of the human body, specifically the placement of our throats, make it possible to create a wide range of sounds. In fact, language is so important that despite the increased choking hazard of a higher larynx, this placement persisted through evolution; verbally communicating danger was more important than protecting our airways (Liberman, 2007). A key characteristic that distinguishes our verbal messages from other animals' systems of communication is that language is limitless. With the ability to produce a wide range of sounds, we can express infinite ideas.

Due to its complexity, many people are fascinated with language. Linguistic research pioneer, Edward Sapir, delve into the field of anthropological linguistics in the 1920s. Since then, language research has transcended to the field of speech perception which investigates the role that the listener plays in a conversation.

Any type of communication involves a sender and a receiver; in verbal communication, the receiver uses auditory processes to hear and comprehend a signal. Through listening, sound signals are transformed into meaningful messages. But since language is so intricate, how do we make meaning of these sounds especially in such a wide range of contexts? Research aims to answer this very question.

In order to understand the complexities of language, we must break down the various aspects of linguistic content. Conversations involve a collection of sentences and sentences are comprised of multiple words. These words are strings of phonemes: sounds that carry meaning (i.e., the th sound in “that” differentiates the word from “hat”). This paper focuses on language at the word level. In this focus, we will dive into the cognitive process of listening and how words are remembered. We will also explore how we attend to and process speech in a variety of settings including noisy environments. Then, we will unpack how we understand different voices and account for natural differences in sound between our conversation partners.

Making Meaning

Humans communicate using linguistic (e.g., word meanings) and non-linguistic aspects (e.g., emotions). With speech being a symbolic representation of different tangible and non-tangible ideas, each word has a referent that it connects to. The word is the symbol, and the referent is the item that is being symbolized. These symbolic representations allow us to communicate about the world around us. Researchers aim to understand how we create distinguished meanings of sounds through the process of memory-based learning.

Speech perception researchers are interested in the relationship between sounds and cognitive processes. How are sounds stored in memory? Do we recall auditory memories of an entire sound event including all noises from the environment? Or are speech sounds filtered and stripped from the background to be encoded as lexical representations? Understanding factors which influence the ability to recall words are the focus of the psycholinguistic study that this paper shares.

If these ideas of auditory memories being stored as entire acoustic events, then indexical properties such as speaker voice and background sounds would influence spoken word recall. However, if these talker variations and disruptive noises have little impact on memory task performance, that would suggest that lexical representations are created around the intrinsic content of words. Before exploring aspects of background noise and speaker variation, we will review two important factors of memory: attention and perception.

Attention and Perception

Attention is the conscious awareness that a stimulus is present. Simply acknowledging that there is a noise in your environment means that you are attending to that sound. Perception takes this identification of stimuli one step further. When we perceive stimuli, we consider more information outside of mere presence. Details such as the source, intensity, and location influence how we understand sounds; this information allows us to organize and interpret all the noises that we hear at any given time. Understanding and organizing these sounds is important for comprehending spoken language. Hearing requires the process of selective perception; through brief discernment of auditory information, listeners disregard background sounds and selectively attend to meaningful sounds. Noisy environments would be impossible to navigate without this organized categorization of sounds.

Background Noise

In any listening environment, we are constantly distinguishing between meaningful and unmeaningful sounds. The event of only a single sound being processed at a time is almost impossible to experience in a natural setting. Sounds are not isolated but, rather, combine to create an auditory event. Auditory events consist of all the sounds that you hear at a given time. This complex overlap of sounds burdens listening when attending to conversational speech in noisy public areas. Our perceptual awareness is pulled in multiple directions, and we must prioritize important information to accommodate cognitive load limits. Aside from all the other senses that we process at any given moment, co-occurring auditory inputs from the environment are particularly difficult to decipher.

When we think about the multitude of sounds in an environment, it is amazing that we can make meaning of anything that we hear. Thanks to effective auditory processing, this background input is filtered subconsciously. The common situation of deciphering competing sounds is called “the cocktail party problem” first described in detail by Cherry (1953). When people are at a cocktail party, there are many overlapping sound signals—likely, music is playing, ice cubes are hitting the sides of glasses, shoes are moving across the floor, and various conversations are converging. Not only is it loud; it is confusing.

Speaker Variations

Background noise only accounts for part of the processes of speech-perception. Once the information of our target conversation is at the forefront of our attention, we then must encode meaning from the sounds that we have filtered as important. However, each person who we listen to produces their own variation of each sound, making this process complex. Vocal qualities between every speaker, so how do we understand that "coffee" pronounced with "aw" and "coffee" said with an "ah" are referring to the same word when the sound signals are vastly different?

You might have noticed that when listening to a heavy accent for the first time, it takes time to understand the new voice. This subconscious process takes more effort than you may think. It seems that talker differences are quickly accounted for; after a few sentences, you have enough information to distinguish the sounds of your new conversation partner. This is because memory makes the process faster. Research into indexical features is useful for considering how listeners account for heavy accents and disordered speech. We want to understand what information about a speaker's voice is encoded into memory and how these memories support lexical representations.

Current Investigation

Recent studies investigate natural speech in authentic listening environments. The stimuli include unaltered words spoken by various speakers and background noise is included. Situations such as the one described by the cocktail-party-effect are important to study to understand not only how words are remembered, but also how the acoustic event in its entirety is processed. Studies that include varying talker and background features target the question: what aspects of a natural listening experience are encoded into memory?

Chapter 2

Linguistic Release from Masking

Co-occurring sound signals not only overlap with each other, they also interact with one another. Sounds are pressure waves; this means that noises which occur simultaneously in time and space create physical wave disruptions. Physicists and acoustic researchers study these wave interactions to describe the physical properties of blended signals. During his presentation for the Acoustical Society of America in 1975, Irwin Pollack coined the terms energetic masking and informational masking to describe the physical and cognitive interactions of sound. What we hear is influenced by these disturbances. Factors of time (temporal) and pitch (frequency/spectral) along with the location (spatial) of sounds influence the noises' degree of disruption both at spectral and perceptual levels.

Sound consists of both energetic and informational properties and therefore influences both physical and perceptual factors. The physical properties of competing signals manifest themselves in degrading the quality and clarity of what we hear. Through the process of hearing, we perceive the wave interactions as distracting background noise amidst our conversations. Cognitive psychologists and speech perception researchers explore the perceptual effects of these complex sensory events to understand how words are processed amidst background noise.

Since its introduction, acoustic masking has been studied in multiple contexts. Many of these experiments consist of synthetically created sounds, but more recent studies explore how we understand speech in a natural setting. Using unedited recorded speech samples, aspects of perception are explored in the presence of speech-on-speech masking. The overlap of information posed by competing speech signals is more demanding than the more automatic process of deciphering one sound source from another.

Meaningful sounds, as opposed to unidentified noise, pose a larger masking effect (Brungart, 2001). For example, when we identify a noise, a subconscious decision must be made about that sound's importance. This initial decision of importance will determine whether we continue attending to the sound or not. Physical disturbances burden the hearing system while informational disturbances burden deeper cognitive networks including the linguistic system.

Factors such as the physical location of a competing speech sounds and differences in the gender of a talker (pitch variations) have been found to reduce the effects of speech-on-speech masking (Brungart, 2001; Viswanathan et al., 2016). Talker-sex mismatch (Brungart, 2001), spatial mismatch (Viswanathan et al., 2016), and language mismatch all account for easier listening in masked conditions (Brouwer et al., 2012; Calandruccio et al., 2010, 2013; Van Engen & Bradlow, 2007; Viswanathan et al., 2016; Wang & Xu, 2021). This reduction in the effects of acoustic masking in the presence of background speech presented in a language different from the target speech is called linguistic release from masking (Viswanathan et al., 2016).

Linguistic release from masking (LRM) describes the situation in which sentence recognition in the presence of background speech is improved when there is a difference between the language of words heard in the background (masker) and the language of the words that the listener is attending to (target) (Viswanathan et al., 2016). Speech-on-speech masking burdens comprehension, but linguistic variation between the target and masker impacts the degree of masking (masker effectiveness) (Calandruccio et al., 2013). When the speech in the background occurs in a language that does not match that of the target speaker, the task of listening becomes easier—even if the phonetic qualities of the words are similar between the two languages.

In masking situations, the masker is a speaker of either an identical, similar, or dissimilar language to the target speaker. A common masking situation in America would include what are called identical target-masker conditions. In this condition, both the background speaker and the conversation partner are in the same language—English. Auditory masking studies using linguistic variance investigate the effects of background speech in languages that are a) phonetically similar (e.g., English and Dutch), meaning that many sounds are shared between the two languages, and b) masking conditions where the languages are phonetically distant (e.g., English and Mandarin) where few characteristics of the languages are shared.

One study in particular set out to investigate how variations in background language conditions in relation to target speakers influence masker effectiveness in American English listeners. Calandruccio et al. (2013) evaluated whether LRM effects are affected by the relative similarity of the target and background languages. In order to do so, they tested three background languages that varied in relative similarity to English. English, Dutch, and Mandarin made up the three sets of two-talker background babble for the study's listening task. To test the effectiveness of the masker and measure the burden that each condition posed on listening, their study utilized comprehensive evaluations. The participants of the study were asked to listen to recorded sentences in the presence of background babble (two or more talkers combined) and write what they heard. Researchers previously selected keywords for each of the sentences consisting mostly of nouns. Participants' comprehensive accuracy was scored based on how many keywords the participants accurately wrote. The greatest amount of masking release is found in linguistically distant target-masker conditions. The studies by Calandruccio et al. (2013) resulted in improved listener performance on sentence listening tasks when linguistic distance was greater.

Van Engen and Bradlow (2007) investigated variables that impact speech intelligibility. Their study used diversified sound-to-noise ratios (SNR), numbers of talkers in babble, and masker languages. Subject responses to a transcription task suggested that English and Mandarin background speech have contrastive levels of influence on speech intelligibility. When presented with two-talker babble in English and Mandarin, higher accuracy was recorded written the background language mismatched the target language (Mandarin masker; English target).

In 2012, Brouwer et al. conducted experiments to further examine the role of background language on sentence recognition. The investigation not only measured release from masking in varying language backgrounds but also looked for masking effects between meaningful and non-meaningful semantically anomalous sentences. The first experiment used English and Dutch two-talker background babble of either meaningful or semantically anomalous sentences. Monolingual English listeners heard meaningful target sentences in the presence of the four background conditions. The results of this study found that nonmeaningful English backgrounds increased target sentence recognition. Additionally, Brouwer et al. replicated findings of linguistic release from masking. English listener performance improved in the Dutch condition. The study concluded that the linguistic content of background babble in speech-in-speech listening conditions influences sentence recognition.

Chapter 3

Talker Variability

Lexical representations in memory reveal cognitive processes of attention and encoding. In speech perception, we want to understand how words are encoded in the mind and what characteristics of speech matter to the encoding process. There are properties of an acoustic event outside of the words themselves that affect perception and memory. We encode not only the bits of information that we try to pay attention to, but we also remember the noise in the environment. Background noise is not the only non-lexical component that influences perceptual processes. There are characteristics of the target speech outside of the content itself that influence our understanding of spoken words. Each of us has unique vocal qualities that distinguish us from one another. A clear example of these differences can be found in accented and disordered speech.

The previous chapter discussed how the noise in an acoustic event impacts the message that listeners must decode. Speech-on-speech asking hinders attention and perception of speech especially when the background content is meaningful and in the same language as the target. But the competing noise is only part of the acoustic event. The qualities of the verbal message itself are important for extracting meaning. Talker-specific (indexical) effects describe these qualities and their impact on the listener's language encoding.

Experimental tests provide measures to understand how words are stored in memory. When it comes to studying memory in spoken word comprehension, a two-phase test paradigm can be used to measure the recall accuracy and response times of linguistic information. Pufahl and Samuel (2014) use an exposure-test paradigm to investigate how listeners respond to words in varying conditions. The study investigates the influence of background sounds in conjunction with speaker changes. Their stimuli for the first experiment included an animate or inanimate environmental background sound (i.e., a dog barking or microwave) paired with a speaker (i.e., male or female) saying an animate or inanimate word (i.e., bird or microwave). Participants answered a question about animacy in the exposure phase when presented with trials. To test memory, listeners responded to a test phase. Trials included the same words and background sounds from exposure in different combinations of speaker voice and background sound. In this task, listeners were required to use the memory process of word recall to answer whether the words were presented in the same or different background. Results showed a decrease in accuracy when the voice and background changed between exposure and test. These findings indicate that mental lexicon is influenced by characteristics outside of the words themselves.

Nygaard, Summers, and Pisoni (1995) conducted a study on the effects that stimulus variability has on the representation of spoken words in memory and perception of spoken words. The variability of the stimuli was explored by adjusting the rate of speech, amplitude, and the person speaking. In their findings, the researchers report that words in lists presented by a single speaker were more accurately recalled than in word lists presented by multiple speakers. These results replicated findings from Martin et al. (1989) and Goldinger et al. (1991). Similarly, words presented with multiple speaker rates were less accurately recalled than same-rate lists. Amplitude did not affect word-list recall. The two variables that showed significant a significant effect on word recall both carry phonetic information while amplitude is not explained by spectral or temporal properties. The research concluded that mental processing resources are used for these phonetically relevant signals to encode and perceive words. There are variations in the demand of attention for phonetically relevant characteristics. Additionally, the processing of varying amplitudes may occur early on in the perception process and may carryout automatically, requiring fewer cognitive resources to account for the changes.

A study by Mattys and Liss (2008) uses natural speech stimuli to investigate how accurately previous spoken-word recognition models transfer/generalize to real-speech conditions. Their work describes some previous findings that, when replicated with natural speech, present contradictory conclusions. Mattys and Liss's design used eight monosyllabic words which were recorded from six different speakers. Their control group consisted of a male and female voice posing no speech disorders, a man and a woman who both had mild dysarthria, and one male and one female who presented severe dysarthric speech. By presenting participants with only one of the speaker pairs, the study measured the accuracy of transcription between subjects.

Listeners heard a total of 80 words spoken by either non-disordered, mildly disordered, or severely disordered speakers. The participants heard 60 words consecutively without repetition. Half of the words were the recorded male's voice, and the other half of the words were the female's voice. Using two blocks for the exposure-test paradigm, 60 words were played in an initial phase and 60 words were tested on. The test consisted of 40 words from the initial list of 60; these were played in either the same voice as the exposure phase or the opposite voice. Of these 40, half were male, and half were female. The remaining 20 words were not previously presented and were also balanced between the male and female speakers. The initial block asked participants only to listen and press the space bar after they heard each word. Test instructions were not presented until after the first was complete. In test, participants were asked to decide if each of the 60 words (40 old and 20 new) was played in the previous block. The task was to press either old or new. To assess the intelligibility of dysarthric speech, listeners who were in the mild or severe dysarthric groups were also given eight seconds in the test phase to type the word that they heard.

The results showed that words played in the same voice between exposure and test, regardless of the level of dysarthria, had higher test accuracy. This is to say that participants more frequently answered, and therefore remembered, that a repeated word was old when it was presented in the same voice condition between exposure and test. Errors in recall decreased in all voice conditions when words were presented in the same voice between test and exposure. Mattys and Liss mention that a possible explanation for this increase in accuracy could be explained by an "acoustic match" meaning that bypasses lexical recognition. The increased word recognition may be attributed to the fact that they remembered the sound and may not be explained by lexical recognition or encoding. However, in the trials with accurate transcriptions, the voice effect was still present. The voice effect was not significantly different between the varying voice conditions meaning that incomplete or incorrect encoding of dysarthric speech does not interfere with voice effects. It is likely that the participants who listened to dysarthric speech relied more heavily on details of the acoustic event outside of the lexical content. Listeners are encoding the entirety of the stimuli, not just the words themselves.

Previous research investigated the parts of speech situations that influence perception. Pufahl and Samuel (2014) explore the importance of background noise in an acoustic event, and how environmental sounds alter memory retention of spoken words. The main questions of this study aim to understand what listeners do with the information of variability in non-human environmental noises. Additionally, do variations in a voice that pose indexical variance in addition to lexical representations serve a pragmatic purpose? Environmental noises are non-speech sounds; in this study, these sounds occur in congruence with spoken words. Changes in environmental sounds and the voice of the talker presenting the words. The work of Pufahl and Samuel supports previous findings and presents new ideas. Results showed that there are significant drops in accuracy when words presented in the same voice undergo a change in the accompanying environmental sound compared to same-voice same-background conditions. Additionally, accuracy dropped when both the voice and background conditions were altered between exposure and test phases. Consistency in both background and talker conditions facilitates listeners' ability to identify previously heard words. Pufahl and Samuel concluded in their study that mental lexicon is a memory system that includes both linguistic and non-linguistic information. Auditory information is not separated into streams of linguistic and non-linguistic information, but rather, is processed in the entirety of the acoustic stream even when listeners are attending to either background or speech sounds.

Chapter 4

Methods

The current experiment aims to further uncover the process of understanding and remembering spoken words in natural settings. These settings consist of background noise, and multiple talkers. More specifically, the study is investigating the interaction of previously found effects of speech perception. Talker-specific effects tell us that there are characteristics of a person's voice that influence our ability to recognize and recall words. When words are played in a familiar voice, they are more easily recognized and remembered. This has been tested through the known-voice-effect studies which resulted in higher accuracy and faster response times when presented as an exposure-test design. Outside of talker-specific effects which are unique to each speaker, there are background effects. Background noise, and more specifically, background speech, also impacts how we process the spoken words that we are attending to. When there is overlapping information we are tasked with filtering the information. At the intersection of these two types of effects (background and talker-specific) lies the question: are these two perceptual influences separate, or do they interact?

Through combining talker-specific and background language variables, we can better understand the degree to which non-linguistic factors impact spoken word recognition build mental lexicon to encode words into memory. The current experiment draws on testing paradigms used to study both talker-specific and background language effects. This study uses natural speech in an online test to understand how linguistic release from masking and the known voice effect interact.

Participants

The participants of this study included 100 American-English monolinguals. Everyone who participated in the listening tasks for the study are from the United States. They are aged between 18 and 35 and have no history of hearing or language-related disorders.

Materials

The target stimuli for this experiment include 84 monosyllabic English words. These words come from Mattys and Liss' 2008 experiment. Each word was recorded in a soundproof booth by a male and a female speaker. Each word was then converted to mono and set at 65 dB using the online program PRAAT. Four of these words were left in silence to serve as practice for familiarizing participants with the voices; the rest were mixed with background babble. Pretest results from 8 English monolinguals confirmed that a –5 SNR (target at 65 dB; babble at 70 dB) was appropriate for the task.

The babble consisted of sentences from the BKB sentence list (Bench et al., 1979). These sentences were recorded by four different speakers; in Dutch by a male and a female and in English by a male and a female. Each speaker's recording was set at 70 dB and converted to mono using PRAAT. The sentences were also concatenated in Audacity to create a stream of sentences with no longer than 300 milliseconds of silence between each.

Once the sentences were edited, the male and female English speaker's files were combined and the male and female Dutch speaker's files were combined. This created a long stream of babble in Dutch and a long babble stream in English. The babble was then combined with each of the 84 monosyllabic target words. In addition to the target-babble overlap, 1000 milliseconds of babble was presented before and after each target word. The final stimuli had four practice words in silence and one target word in the middle of a stream of babble for each of the 80 trials. Half of the words were presented in a male voice and half in a female voice to counterbalance talker-sex.

Procedure

This experiment was created online using Labvanced and crowdsourced through Prolific. The experimental task consisted of 1) a headphone check, 2) a 4-word practice set, 3) an exposure phase with 60 words, 4) a test phase with 60 words, and 5) a language survey with a debrief. In this study, participants were asked to complete listening tasks. The initial headphone check ensured that listeners were using headphones at a proper volume during the experiment. Three sets of static noise were presented six times. In each set of three, one of the static sounds contained a hidden tone that was only detectable when using headphones at a comfortable listening volume. After listeners adjusted their volume, they moved on to a practice session. In this session, the task was to listen to words and type what they heard. The practice familiarized participants with the task and allowed them to hear the target voices in silence. Four words from the Mattys and Liss 2008 study were balanced between the male and female target speakers. After the participants understood the task, they moved into an exposure phase. In this phase, listeners were asked again to type what they heard. The stimuli in this section of the study were also balanced between the male and female voices but were also presented with background two-talker babble in either Dutch or English. Both the words spoken by the male and the female had an equal number of English and Dutch background. The conditions are as follows: 15 male target with English background babble, 15 female-English, 15 male-Dutch, 15 female-Dutch). Each word was played once and the order was randomized for each participant. Additionally, four groups of participants were created to rotate each word through the conditions (i.e., “penguin” was heard by Group 1 in the female-English, by Group 2 in the male-English condition, by Group 3 in the female-Dutch condition, and by Group 4 in the male-Dutch condition).

Participants prescreened and crowdsourced from Prolific were directed to the online experimental platform, Labvanced, to carry out the study. During testing, listeners went through a headphone check to ensure continuity between subjects. To check for the use of headphones, six trials were presented. In each trial, three static noises were played; one of the static sounds had a hidden tone that could only be detected through headphone use. More than two errors for tone detection resulted in failure of the headphone check.

Following the headphone check was a practice set to orient listeners to the voice of the target speakers. Four words were presented in silence. Two words were spoken by the male talker and two words were spoken by the female talker. Listeners typed the word that they heard and then moved on to the 60-word exposure set.

The exposure set was counterbalanced between speakers and background conditions. With a male and female speaker and Dutch and English babble, 15 different words were played in each of the four conditions. Given the four conditions, the study had four groups as described previously. The exposure task was the same as the practice task; participants typed the word that they heard but this time, in the presence of background noise.

To replicate previous studies, an unexpected test task was given. Participants had to draw on their memory to answer if a word played in test was an old word from the exposure phase. 40 words from exposure and 20 new words made up the 60 trials of the test. Half of the old words were presented in the opposite background condition from exposure. All conditions were counterbalanced for both old and new words. A screen with two button labels, old and new, was presented for each trial for participants to indicate their answer.

The end of the experiment included a language survey to assess participants' previous exposure to the Dutch language. Testing finished with a debrief of the study's purpose.

Chapter 5

Results and Discussion

Results from the current study replicate the presence of linguistic release from masking LRM). The findings show LRM effects at the word level through online testing. Additionally, linguistic release from masking is present not only in comprehension, but the effect was also present in the old/new memory task. Prior to this experiment, linguistic release from masking has been studied primarily at the sentence level through in-person testing to measure comprehension. The current study expands our current understanding of speech perception at the level of comprehension and adds insight into language processing at the lexical memory level in the presence of background speech.

Both the exposure phase (Figure 1.) and the test phase (Figure 2.) reflect higher response accuracy percent averages in the Dutch background conditions compared to English. There are significant differences in spoken word comprehension between the variable of background language for monolingual American-English speakers.

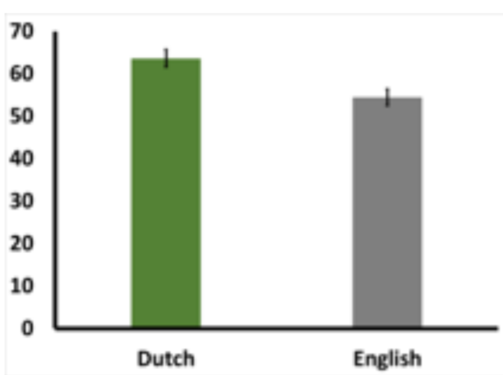

Figure 1. Exposure Phase Word Percent Accuracy Results

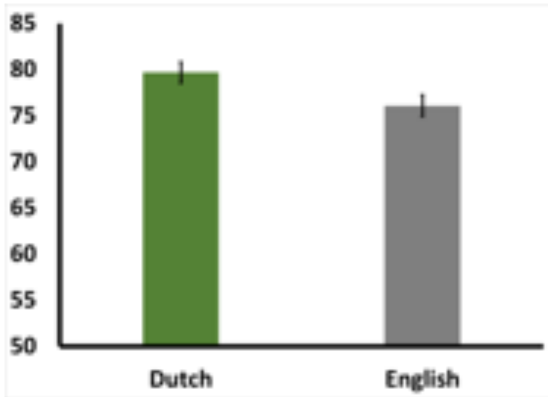

Figure 2. Test Phase Old/New Percent Accuracy Results

We see the linguistic release from masking effect at the word level in the initial recognition task (Figure 1). In the Dutch condition, percent accuracy is significantly higher for the linguistically different background babbles. The score averages for typing words in the Dutch background are 9% higher than for the English background (Dutch 63.65; English 54.49). This is a statistically significant difference in exposure task performance. Linguistic effects of mismatched background conditions were found.

We also see linguistic effects on memory tasks that measure word recall. In the test phase, performance was also higher for the Dutch background condition (Figure 2). Percent accuracy for the old/new judgment task is 3.5% higher when the background babble is in Dutch (Dutch: 79.67; English 76.08). We found that a mismatch in the background to the target language not only affects word comprehension but also influences memory judgments.

Background conditions are shown to influence both recognition and recall tasks at the word level through linguistic release from masking. This speech-in-speech study replicates findings of masking release in linguistically different background conditions and found a Dutch background language advantage for the memory task. In the current experiment, linguistic variations in background noise influence speech perception and are shown to have an effect on word memory. Implications from preliminary findings suggest that background language conditions play an important role in word comprehension and in lexical representations.

By measuring background speech influences in a paradigm used for talker-specific measures, the current investigation adds a level of understanding for how internal and external qualities of spoken words change a listener's ability to understand and remember words. Not only did we find that linguistic release from masking is present in the talker-specific testing paradigm, but we also expanded findings to online listening situations. As technology becomes more prominent, it is important that we continue exploring speech comprehension in a variety of settings. The current findings imply that word comprehension has multiple influences which impact the understanding and memory of words. This study allows for further testing using the paradigm to continue investigating these influences.

The recommended extensions of the study include testing bilingual participants, using other languages such as Mandarin to further replicate other LRM findings, using accented or disordered speech, and adding participants who have hearing loss. Other variations could include altering the speakers in the test phase of the experiment to explore the presence of interactive effects of talker-specific adjustments and linguistic release from masking. An additional research project implemented this final alteration by adding talker-variability measures to the current study's testing paradigm (Gryskevicz, 2023).

In the experiment, the same testing paradigm was used with the addition of variation of the speaker between exposure and test (mismatched) in half of the trials. Findings show an interaction between the background effects and talker effects (Figure 3).

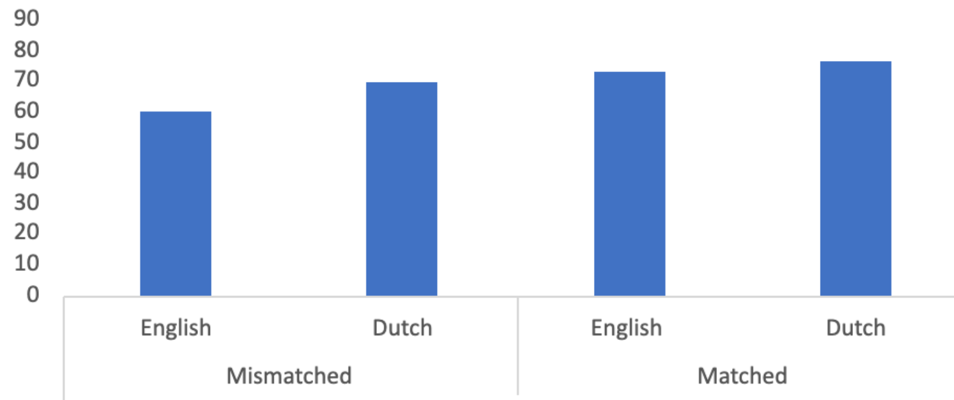

Figure 3. Test Phase Percent Accuracy Interactive Results (Gryskevicz, 2023).

The test conditions that presented a word in the same voice with a Dutch background babble have the highest accuracy in old/new judgments (Gryskevicz, 2023). These results are the average of 104 participants who were tested with the paradigm from the current study.

Natural listening environments pose challenges to speech comprehension. From current and previous research, we know that specific conditions of an auditory environment influence the extent of difficulty faced by the listener. Increased accuracy of sentence comprehension in varying background conditions has been studied in multiple ways.

The findings from the initial experiment that replicated LRM effects (Figure 1 and Figure 2) build on the overall understanding of how humans process speech in natural conditions. The presence of the release of masking effects at the word level emphasizes the level of influence of speech-in-speech listening conditions. Since the majority of our conversations occur in background noise, it is important to understand the level at which masking affects speech comprehension. Furthermore, the talker-specific influences that support or hinder comprehension are another aspect of natural listening situations that are important to understand.

The current study expanded on the benefits from linguistic variation between a target speaker and background babble through its unique paradigm. Variations between the background language and the target language improve language comprehension and recall judgments at the word level. Through an online study and a two-phase testing paradigm, the study's results bring a new level of understanding to the ongoing investigation of speech perception. Even in short language tasks such as understanding a single monosyllabic word, background speech has significant influences on understanding. With this information about background language effects on linguistic memory, models for language learning can be adapted to fit optimal listening conditions for word storage in memory.

Appendix A:

List of Stimuli

AIRPLANE	CROCODILE	MOTORCYCLE	SPARROW
ANT	DEER	OBOE	SPIDER
AXE	DRILL	OSTRICH	SPRINKLER
BAGPIPE	EEL	OTTER	SQUIRREL
BASSOON	FAN	PARAKEET	STAPLER
BEAVER	FLAMINGO	PEACOCK	STOPWATCH
BEETLE	FOX	PELICAN	SUBWAY
BLENDER	GOPHER	PENGUIN	SWAN
BROOM	GUITAR	PHEASANT	TEAPOT
BUFFALO	HAMMER	PICCOLO	THUNDER
BUGLE	HORN	PINBALL	TOASTER
BUS	HYENA	RABBIT	TOILET
BUTTERFLY	KEYS	RADIO	TOOTHBRUSH
CAMEL	LARK	ROBIN	TRIANGLE
CELLO	LAWNMOWER	SAXOPHONE	TROMBONE
CHEETAH	LLAMA	SCISSORS	TURTLE
CHIPMUNK	LOBSTER	SHOWER	VACUUM
CLARINET	MATCHES	SHREDDER	VULTURE
COBRA	MICROWAVE	SKUNK	WHALE
CRAB	MODEM	SNAIL	WORM

REFERENCES

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101–1109. https://doi.org/10.1121/1.1345696

Calandruccio, L., Brouwer, S., Van, E. K. J., Dhar, S., & Bradlow, A. R. (2013). Masking Release Due to Linguistic and Phonetic Dissimilarity Between the Target and Masker Speech. *American Journal of Audiology*, *22*(1), 157–164. https://doi.org/10.1044/1059-0889(2013/12-0072)

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*(2), 860–869. https://doi.org/10.1121/1.3458857

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152–162. https://doi.org/10.1037/0278-7393.17.1.152

Gryskevicz, A. (April 20230). *Interactive Effects on Speech Perception.* The Pennsylvania State University.

IEEE Recommended Practice for Speech Quality Measurements. (n.d.). IEEE. https://doi.org/10.1109/IEEESTD.1969.7405210

Jesse, A., McQueen, J., & Page, M. (2007). *The locus of talker-specific effects in spoken word recognition*.

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit Talker Training Improves Comprehension of Auditory Speech in Noise. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01584

Lieberman, P. (2007). The Evolution of Human Speech: Its Anatomical and Neural Bases. *Current Anthropology*, *48*(1), 39–66. https://doi.org/10.1086/509092

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(4), 676–684. https://doi.org/10.1037//0278-7393.15.4.676

Mattys, S., & Liss, J. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception & Psychophysics*, *70*, 1235–1242. https://doi.org/10.3758/PP.70.7.1235

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Perception & Psychophysics*, *57*(7), 989–1001. https://doi.org/10.3758/BF03205458

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5–S5. https://doi.org/10.1121/1.1995329

Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology*, *70*, 1–30. https://doi.org/10.1016/j.cogpsych.2014.01.001

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519–526. https://doi.org/10.1121/1.2400666

Viswanathan, N., Kokkinakis, K., & Williams, B. T. (2016). Spatially separating language masker from target results in spatial and linguistic masking release. *The Journal of the Acoustical Society of America*, *140*(6), EL465. https://doi.org/10.1121/1.4968034

Wang, X., & Xu, L. (2021). Speech perception in noise: Masking and unmasking. *Journal of Otology*, *16*(2), 109–119. https://doi.org/10.1016/j.joto.2020.12.001

ACADEMIC VITA

Amira Corine El-Dinary
ace5276@psu.edu

EDUCATION

Bachelor of Science in Communication Sciences and Disorders **May 2023**
Schreyer Honors College and College of Health and Human Development
The Pennsylvania State University, University Park, PA

Interesting Coursework: Disability Culture, Race and Ethnic Relations, Deaf Culture, American Sign Language 1 & 2, Social Demography, Community Development, Anthropology, Interpersonal Communication, Teaching English Language Learners, Intro to Human Development & Family Studies, Biology of Aging, Adult Development & Aging

RESEARCH EXPERIENCE

PENN STATE'S CENTER FOR LANGUAGE SCIENCES

Speech, Language, and Cognition Lab **2020-Present**
Dr. Navin Viswanathan

- Coding online experiments using Labvanced
- In-depth literature reading
- Running multiple experiments on Prolific
- Data cleaning and analysis

Bilingualism and AAC Project **2022**
Kassie Galley

- Qualitative data coding
- Interview study reviews
- Understand family perspectives of communication devices

Complex Communication Needs and AAC Project **Summer 2021**
Tara McCarty

- Thematic analysis
- Parent perspectives on children with Cortical Visual Impairment and Complex Communication Needs

RIVER HILL HIGH SCHOOL

Intern Mentor Course **August 2018 - May 2019**
Research Program Developed by Mary Sasser

- Shadowing of a school speech therapist
- Reading and annotating scholarly articles
- Generating and executing a lesson plan for a speech client
- Composing a research paper on factors of effective therapy

- Website design to disseminate findings

PROJECTS

Known Voice Interacts with Linguistic Release from Masking http://bitly.ws/vTkg
A Deaf Culture Parent Resource https://deafculture-parent-resource.weebly.com/
The Key to Success: Speech Therapy https://speechtherapyresearch.weebly.com/
Health and Human Development Podcast ***In Progress***

PROFESSIONAL EXPERIENCE

Peer Ally for Students with Disabilities—WorkLink **Present**
Rosemary Schwoerer

Peer Advisor—Education Abroad at Penn State Global **Present**
Casey McClellan

Alumni Mentoring Program **Present**
Natalie Hricik SLP-CC

Direct Support Professional—Strawberry Fields Inc. **February-October 2022**
Carrie Walker

American Speech and Hearing Association Mentor Program **2022**
Dr. Holly Kleiber

Youth Mental Health Counselor—Camp Attaway **Summer 2021**
Dr. Noah Weintraub

Teaching Assistant—Child Language Disorders **Fall 2021**
Dr. Carol Miller

Teaching Assistant—Intro to Communication Sciences and Disorders **Spring 2021**
Dr. Aarthi Madhavan

LEADERSHIP AND CONFERENCES

College of Health and Human Development Ambassador **Fall 2022-Present**

Schreyer Honors College Strategic Plan Working Group
Notetaker for the Diversity, Equity, Inclusion, and Belonging subgroup **Present**

Leadership Initiative Program
Penn State's College of Health and Human Development **Present**

Young Language Science Scholars Poster Presentation **January 2023**

The Psychonomic Society Annual Meeting Poster Presentation and Conference
Boston, MA **November 2022**

The International Max Planck Research School for Language Sciences Conference
Nijmegen, Netherlands **June 2022**

TOK-dag—Max Planck Insitute of Psycholinguistics
Nijmegen, Netherlands **June 2022**

The Pennsylvania State Outing Club
Served as the secretary in 2021 **Spring 2020-Present**

COMMUNITY SERIVICE

Fresh START
Penn State Student Affairs **Fall 2019**

LifeLink
State College Area School District Department of Special Education **2020-2021**

National Student Speech, Language, and Hearing Association
The Pennsylvania State University **Fall 2019-Fall 2022**

HONORS AND AWARDS

College of Health and Human Development Dean's List **Fall 2019-Present**

Judith Kroll Undergraduate Student Research Award **September 2022**

Schreyer Honors College Grant for Conference Attendance **September 2022**

Partnerships for International Research and Education Fellowship **Summer 2022**

Peer Assistant Award of Excellence **May 2019**

Student of the Month "Innovative" **December 2018**

Academic Excellence Award in Chemistry **May 2018**

River Hill High School Principal's Honor Roll **2017-2018**

INTERNATIONAL TRAVEL

The Netherlands **May-July 2022**

Study abroad research experience through the Partnerships for International Research and Education (PIRE) Fellowship funded by the National Science Foundation (NSF)

Italy **April 2016**

Family travel and cultural exposure

Japan **April 2012**

Family travel and cultural exposure

CERTIFICATES

IRB **Active**
Adult Mental Health First Aid **Active**
CPR/Frist Aid/AED **Active**
Wilderness First Aid **Active**