THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF COMMUNICATION SCIENCES AND DISORDERS

INTERACTIVE EFFECTS ON SPEECH PERCEPTION

AMANDA GRYSKEVICZ
SPRING 2023

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

Reviewed and approved* by the following:

Navin Viswanathan
Associate Professor of Communication Sciences and Disorders
Thesis Supervisor

Carol Miller
Professor of Communication Sciences and Disorders
Thesis Adviser

* Electronic approvals are on file.

ABSTRACT

Previous studies have shown that recall accuracy improves when words are repeatedly spoken by a “known voice”. It has also been found that listeners recognize speech better when the background babble is in a different language than the target speech. Using an exposure-test design, an evaluation was conducted to see if these effects interact. During exposure, English words embedded in English or Dutch background babble were presented to monolingual English listeners. During test, words from the exposure phase were repeated (old) or not repeated (new). Participants’ task is to recognize words as “old” or “new” as quickly as possible. Results include interactive effects of known voices and different maskers such that the same speaker and a target-masker mismatch yields better performance (higher accuracy and faster response times).

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank the Center for Language Science for being the principal financial contributor to this project through the PIRE program, backed with a generous grant from the National Science Foundation. This award funded a life-changing trip to the Netherlands to develop my research with international scientists in the field and positively represent the scholarship of Penn State. My research partner and I were also recipients of the Judith Kroll Undergraduate Research Award, a gift that enabled us to attend the Psychonomic Society Annual Meeting in November of 2022 to present this work to dozens of speech science professionals. Finally, I thank the Schreyer Honors College for their financial support. Each of these gifts were a vital component of my success and achievement through my Penn State career.

This research would not have been possible without the guidance of my primary mentor, Dr. Navin Viswanathan. Thank you for your endless patience while I developed as a scholar over the years. You granted me valuable opportunities to learn and grow through networking, travel, and solving challenges I otherwise would have given up on. Navin, you are a wealth of knowledge, and the Communication Sciences and Disorders Department is lucky to have you.

I extend my gratitude to my other mentors over the years. Dr. Carol Miller; your voice of reason and encouragement has kept me grounded through my last two years of college. Dr. Susanne Brouwer; your energy and enthusiasm propelled my progress and continues to support my goals, even as I exit the field of speech science. And, of course, Dr. Cathleen Hunt; the first person to guide me to my love of psychology.

My heart is full thinking of all the other friends, family, and peers who have been a part of this process. Thank you for your love and support.

Chapter 1

Introduction

As we go about our lives, we are bombarded by noise. Cars, birds, and voices of the people surrounding us interfere with our perception of the world. Then, as we age, we learn to filter out certain sounds from this bundle of noise, such as speech. With verbal communication being our primary form of interaction, humans are trained from birth to recognize speech sounds above non-speech ones (Kollmeier et al., 2008). We process speech to understand, to learn, and to interact with others. To study how speech is perceived in the natural world, different effects on speech perception can be separated and analyzed in a lab setting. This research builds on the understanding of how brains filter voices and what information is being stored. With this knowledge, we can better structure our conversations for optional speech perception and adjust to those who may be diagnosed with speech or hearing disorders.

One such effect is known as masking. Think of a verbal conversation in a busy coffee shop. There are the sounds of brewing coffee, baristas bustling around, and of course, the conversations of other patrons. To hear the conversation or person at hand, or the "target voice", these other sounds must be filtered out. The impact of these background sounds on speech perception is known as masking (Brungart, 2011). The other voices present partially cover, or "mask" the target conversation. The listener has the capability to filter out the background voices and help focus on the voice at hand, but the background makes that more difficult. Now, if those background voices were to be speaking in a language unknown to the listener of the target voice, there would not be as much difficulty processing the target conversation. This would create an effect known as linguistic release from masking (LRM) (Brouwer et al., 2012). Listeners would

have an easier time processing, remembering, and responding to the conversation at hand as they block out the unknown language. In natural conversation, LRM is only one of the speech perception effects present. One other influence happening simultaneously is the talker variability effect, otherwise known as the familiar voice effect.

The familiar voice effect is exhibited through the faster processing times and boosted memory recall when a stimulus is presented in a known voice. It could be the voices of family members, friends, significant others, or even as studies show, the voice of someone that the listener has heard or been trained on before. Research identifies that the mind is faster to process these familiar voices, or known voices, than unknown voices (Mattys & Liss, 2008). Within this paper, the phenomena will be labeled as the known voice effect.

Both these effects have been studied separately to show their results and different aspects under controlled conditions. LRM is associated with simultaneous speech sounds. This means that the effect is dissected into how it impacts monolinguals versus bilinguals (Brouwer et al., 2012) or how the two different types of masking (energetic and informational) differentiate in words-in-noise tests (Wilson et al., 2012). Meanwhile, the known voice effect describes the qualities of a voice that an individual might pay more attention to as opposed to what the voice is saying. The known voice effect, in its most notable study (Mattys & Liss, 2008), was used as a testing paradigm for individuals with dysarthric speech. What has yet to be analyzed, though, is how these two effects interact. By combining them, it can be shown whether the brain uses background language of speech, the voice of the speaker, or both to process speech and recall it. The interaction can be pictured in everyday conversation. Think of an individual talking to their loved one in a coffee shop in a foreign country or on the phone with a spouse while an international singing group is performing nearby. Both these instances showcase LRM and the

known voice. The purpose of the studies performed in this document are to control these two effects and test if an interactive effect can be formed at their combination.

Chapter 2

Linguistic Release From Masking

Linguistic release from masking is the observation that the intelligibility of target speech is improved when the background speech differs in language from the target speech (Brouwer et al., 2012). For example, an English monolingual listener experiencing English stimuli in the presence of a Mandarin background is going to have faster reaction times and higher accuracy in recall tasks than if they are experiencing English stimuli in the presence of an English background. In that case, they understand both the target and the background and must work harder to pick out the target conversation or speech. In studying LRM, scientists use a measurement called signal-to-noise ratio, or SNR. This is the ratio of the primary signal to the background. In this study it is used to measure how easy or difficult it is to pick out target speech sound from a mixed babble. For example, a higher SNR would mean greater difference between the target and background decibel (dB) levels (e.g. target 50 dB with background at 70dB, or -20 SNR) and indicate a greater challenge to pick out the target speech sound from the background. Previous studies have identified optimal results from the effect shown at -5 SNR, or when the background sound is 5 decibels higher than the target speech (Calandruccio et al., 2010). These studies provide the groundwork for ongoing LRM research.

Masking, the parent word and concept for linguistic release from masking, is split into two types: informational and energetic. Brungart's work in 2001 sought to aid in the distinction between the two, showing informational masking as conflicting semantics and energetic as conflicting levels of sound. Based on this difference, LRM studies show use of both informational and energetic masking. Both types are present in the current study. To find

linguistic differences in masking or LRM effects, Calandruccio et al. (2010) showed that the more phonetically and linguistically different target/background pairings in experimental studies yielded a greater LRM effect for monolingual speakers. For example, a target/background pair of English/Mandarin would have a greater release from masking than an English/Dutch pair, based on the linguistic differences in language alone. Because English and Dutch sound similar, they provide a more difficult listening condition than for an English/Mandarin pairing, where the languages sound drastically different. Expanding off this idea, Brouwer et al. (2012) brought bilingual listeners to the LRM discussion to discover how their LRM effect might compare to that of monolinguals. They showed that monolingual English listeners outperform their bilingual counterparts since they can only understand English. However, though English/Dutch bilinguals have a more difficult listening experience whilst understanding both the target and the background, they still show an LRM effect when the target is in a different language than the background. To find if spatial proximity impacts LRM effects Viswanathan et al. (2016) conducted a study with spatially separated sources and found correlating results to the effect. This is an indicator that regardless of variances in the distance between a listener and their target sound, LRM effects will replicate. While these studies, among others, provide a solid framework for the study of LRM, opportunities exist for further application of the concept.

A gap in LRM research so far exists in that the effect has yet to be found at the word-level. The aforementioned studies use sentences and phrases to test participants on, but the question remains as to if LRM exists with a stimulus as short as a single word. Additionally, applications and interactions of LRM with other speech effects are few. Most research has been done on LRM alone rather than combining it with other ideas to find interactions. In an effort to

remedy this, the current study aims to combine linguistic release from masking and another effect, the known voice effect.

Chapter 3

Known Voice Effect

Familiar voice studies lack the same solidified foundation as LRM. This may be attributed to the effect's multiple titles, including "familiar talker advantage" (Levi et al., 2011) and "talker-specificity effect" (Dufour et al., 2017). "Known voice effect" will be used for the purpose of this document and study at hand. When a known voice is manipulated in studies, participants exhibit faster reaction times and more accurate recall than with an unknown voice. These results provide proof of faster processing with known voices. This effect is true both with voices personally familiar to the listener and voices that the listener has been trained to recognize (Mattys & Liss, 2008).

Performing perhaps one of the most notable known voice studies, Mattys and Liss (2008) provided a valuable framework for studying the effect. They included such parameters as having an exposure/test paradigm with a phase for learning the voices at hand then a phase following with a memory task. Participants were tasked with another parameter, the old/new task, where they were presented with both words from exposure and new words then asked to label them as old or new accordingly. This supports the idea that listeners can be trained on a particular voice to make it "familiar" or more easily recognized to them. Contributing to the field, Dufour's work sought to see *when* in the speech processing procedure the known-voice effect is found. Experimentation showed that while the speaker's familiarity isn't the first aspect of a voice listeners hear, it is used to confirm recognition of a word once the lexical information has been mostly processed (2016). Based on this study, it is known when in speech recognition the known voice effect is found and how long after hearing speech do listeners recognize the voice. Levi et

al. sought to find whether language training plays a role in the known voice effect (2011). Using monolingual English listeners, the researchers trained two groups of participants. One was trained on English words while the other was tested on German for the same amount of time. They found that even though both groups had the same amount of time with the voices, the participants in the English condition outperformed their German-recognizing counterparts. This indicates that while the known voice effect is present (being that both groups improved over the training period regardless of language), language learning plays a large part in memory. Similar to LRM, while the known voice effect has been studied well on its own, it is still being explored in reference to other effects on speech perception.

Chapter 4

LRM and Known Voice Interact

To combine LRM and the known voice effect, LRM had to be found in the testing paradigms used to find the known voice effect. All previous and current research regarding LRM tested the effect at the sentence level, using full sentences as both target and background stimuli. Meanwhile, the known voice effect is typically investigated at the word level, with words as target stimuli, such as demonstrated in Mattys and Liss (2008). To simultaneously study LRM and the known voice effect, steps were taken to bridge the differences between the typical paradigms used to study each effect. In Experiment 1, LRM was tested with the exposure/test paradigms and old/new task used in Mattys and Liss' 2008 study for the known voice effect. By testing for LRM in known voice paradigms, an interaction between the effects would be able to be found later on. Another step to combine the two was to use BKB sentences as the background for the stimuli. BKB sentences are commonly used in LRM experimentation to mask target speech. With these paradigms and methods for combining LRM and known voice testing, LRM was found at the word level in Experiment 1, as seen in *Figure 1* and *Figure 2* (El-Dinary, 2023). *Figure 1* demonstrates how participants had higher accuracy when tested in the Dutch background condition than with the English background condition. Meanwhile, *Figure 2* demonstrates how participants had faster reaction times in the old/new judgment task when tested on the Dutch background condition vs the English background condition. The two components shown here are evident of an LRM effect found at the word level and within the known voice testing paradigms. These results prompted Experiment 2, where the two effects

were combined to see if an interaction was present. Methods were identical with only slight changes between Experiment 1 and Experiment 2, as detailed below.

Figure 1

Accuracy of Old/New judgments in the Test Phase

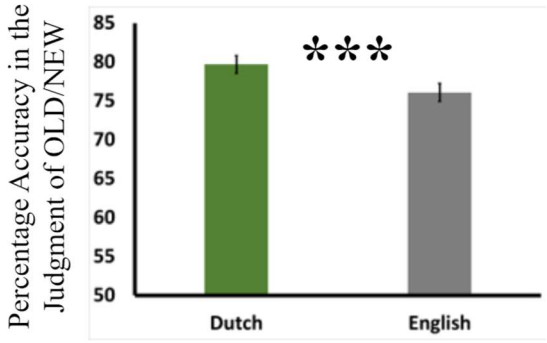
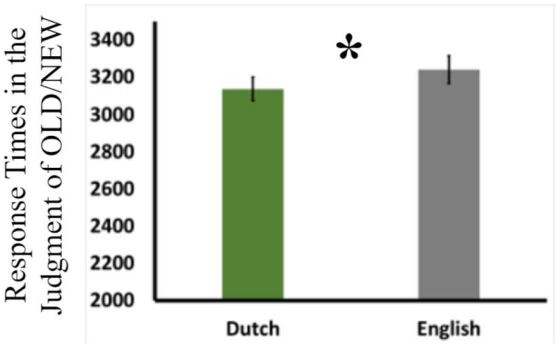


Figure 2

Response Times of Old/New judgments in the Test Phase



Methods

Participants

One hundred and eight listeners from the online research platform Prolific with American English as the first language and no exposure to Dutch took part in this second phase of the study. All participants had no history of speech or hearing disorders.

Materials

Two native English speakers, a male and female, recorded the 86 target words sourced from Pufhal and Samuels' 2014 study; nouns ideal for providing lexical information. Using Audacity, a second of silence was added to the beginning and end of the target word so that when combined with the background, a second of babble played before and after the word in addition to the background-target overlap. Viswanathan et al. (2016) used revised sentences from the Bamford-Kowal-Bench (BKB-R) Standard Sentence Test (Bench et al., 1979) together in

their study with LRM, prompting the use of those same sentences for the background stimuli of the current experiment. These sentences, ideal for their use of key words and nouns, provided the ideal conflicting stimuli for LRM to be tested at the word level. To have an LRM effect, there must be multiple languages present, with one unknown to the listener. Brouwer et al. (2012) showed successful effects of LRM when English conflicted with the Dutch language; a semantically contrasting but phonetically similar tongue. This provided the basis for using Dutch as a test background language to find LRM while the English background acted as a control. Four speakers (one male Dutch, one female Dutch, one male English, and one female English) voiced the sentence files. These were then split up into pairs, the female Dutch speaker together with the male Dutch speaker, and the female English speaker with the male English speaker. Using Praat and Audacity, each language of sentences was blended with the silence removed to create a babble that mimicked natural conversation. As supported by the work of Calandruccio et al. (2010) and a separate pre-test conducted in the present study's lab, the background stimuli was set at 70 dB with the target set at 65 dB. This created a -5 SNR, or sound-noise ratio, making a challenge for participants where the target word was slightly quieter and had to be focused on to be heard over the louder background. With an optimal SNR and stimuli ready, a study design was necessary in the search to replace LRM effects.

Procedure

Listeners were individually tested online through a Labvanced study. After confirming the use of headphones in a quiet area, the subjects were presented with 6 practice words to train them to identify the two target speakers. During the exposure phase, 60 of the target words spoken by the target speakers were randomly presented in various speaker and background conditions. Participants were asked to type the target word to test listening accuracy or type "x"

if they could not understand it. Then, 40 of those same words were repeated in the test phase while 20 new words were added; word order was randomized. In the first phase of the experiment, all "old" words were randomized and presented in the same voice as exposure. In the second phase of the experiment, half of the 40 words were presented in the same voice as exposure while the other half were presented in the opposite target voice as exposure. The accuracy of perception was measured through an old/new task during the test phase where participants indicated if they remembered hearing the target word from the exposure phase ("old") or if they believed they were hearing it for the first time ("new").

Results

Experiment 1 demonstrated the successful finding of LRM at the word level. Using the exposure/test paradigm from known voice research and the old/new task, LRM effects were replicated. This is indicated by the higher percentage accuracy and faster response times shown by *Figure 1* and *Figure 2*, respectively. Supported by Experiment 1 where the LRM effect was found at the word-level and with known voice testing paradigms, an interactive effect was found between the known voice effect and LRM in Experiment 2. Data collected from Experiment 2 identified shorter response times when the target word was mismatched in language from the babble/background but the target word was spoken in the same voice between exposure and test. Participants showed an average of 76.8% correct in the test phase with a Dutch background and an old voice (same voice as word was presented in during exposure), 73.2% correct in the English background/old voice condition, 70% correct in the Dutch background/new voice (different voice as word was presented in during exposure) condition, and 59.9% correct in the English background/new voice condition.

Chapter 5

Discussion and Conclusion

Previous studies analyzed the known voice effect and LRM separately while this is the first known record of them being studied together. We knew that LRM was present at the sentence-level and is found through similar sounding languages such as English and Dutch (Calandruccio et al., 2010). Meanwhile, Mattys and Liss (2008) showed that participants can be trained on a voice to test the known voice effect through an exposure/test paradigm and with an old/new task to assess memory. The purpose of this study was to mimic natural conversations in a lab setting and determine what components of speech humans are prone to pay attention to. Do they recall the quality of voice or the word itself? How does background speech interfere with this recall? In combining these effects and their testing paradigms, Experiment 1 showed a successful combination of testing paradigms to yield LRM effects at the word level. This builds on the known components of speech that LRM can be found and shows that the effect is not limited to longer components such as sentences or monologues. Our results indicate that LRM and the known voice effect build on each other in speech perception and produce an interactive effect, as shown in Experiment 2 when both effects were replicated simultaneously. The interactive effect shows that when hearing speech, the mind is processing both the background and the target equally. If participants were only influenced by the known voice effect, then the conditions with the same speaker in exposure and test would have the highest accuracy. If participants were only influenced by the LRM effect, then the conditions with the Dutch background would have the highest accuracy. Instead, we found the interaction of the highest

accuracy being when there was a Dutch background with the same voice in exposure/test. By testing two effects at once, we come closer to mimicking natural conversation in lab settings. It shows that both effects have the potential to be manipulated in other testing paradigms or combined with other effects on speech perception.

Future directions from these findings may be vast. All participants in this experiment were English monolinguals, however running English/Dutch bilinguals through the same paradigm would prove interesting. When participants understand both the target and background language, they may perform differently despite the known voice effect addition. It is expected that although both languages are comprehended, a mismatch between exposure and test will still yield more accurate and faster results. Running bilinguals through the study would show whether LRM can be found at the word level for participants who may understand the target and background, as opposed to only understanding the target. All participants in this study were over the age of 18. However, taking into consideration the critical periods of language development in children, participants in a younger age range may perform differently than their adult counterparts. Would the speech processing areas of their brain be developed enough to pick out a target word from background or would the conflicting noise be overwhelming for their speech processing? Through these potential applications, we may understand how different demographics and levels of language proficiency are influenced by these effects on speech processing. By replicating and controlling natural effects on speech in a lab setting, we come closer to understanding how the mind perceives language and conversation.

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., 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

El-Dinary, A. (2023). *Effect of speaker voice and background language on spoken word recall* [Unpublished undergraduate thesis]. The Pennsylvania State University.

Kollmeier, B., Brand, T., Meyer, B. (2008). Perception of Speech and Sound. In: Benesty, J., Sondhi, M.M., Huang, Y.A. (eds) *Springer Handbook of Speech Processing*. Springer Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-49127-9_4

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2011). Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible?. *The Journal of the Acoustical Society of America*, *130*(6), 4053-4062.

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception & Psychophysics, 70*(7), 1235–1242. https://doi.org/10.3758/PP.70.7.1235

Sophie Dufour, Dierdre Bolger, Stephanie Massol, Phillip J. Holcomb & Jonathan Grainger (2017) On the locus of talker-specificity effects in spoken word recognition: an ERP study with dichotic priming, Language, Cognition and Neuroscience, 32:10, 1273-1289, DOI: 10.1080/23273798.2017.1335421

Viswanathan, Navin & Kokkinakis, Kostas & Williams, Brittany. (2016). Spatially separating language masker from target results in spatial and linguistic masking release. *The Journal of the Acoustical Society of America*. 140. EL465-EL470. 10.1121/1.4968034.

Wilson, R. H., Trivette, C. P., Williams, D. A., & Watts, K. L. (2012). The effects of energetic and informational masking on The Words-in-Noise Test (WIN). *Journal of the American Academy of Audiology*, *23*(7), 522–533. https://doi.org/10.3766/jaaa.23.7.4

ACADEMIC VITA
Amanda Gryskevicz

EDUCATION

Pennsylvania State University, University Park, PA
Schreyer Honors College
Bachelor of Science, Psychology, May 2023

Study Abroad: PIRE Fellowship
Radboud University
Nijmegen, Netherlands, Summer 2022

HONORS AND AWARDS

PIRE Fellow and Grant Recipient
Schreyer Honors College Scholar
Schreyer Academic Enrichment Grant
Judith Kroll Undergraduate Research Award
Dean's List

RESEARCH EXPERIENCE

Undergraduate Honors Thesis: Interactive Effects on Speech Perception
The Speech, Language, and Cognition Lab, Penn State University
Fall 2021 - present
Thesis Supervisor: Dr. Navin Viswanathan
Thesis Advisor: Dr. Carol Miller

- Purpose - Contributing to understanding the relationship between two specific effects on speech and theories of speech processing memory.
- Reviewing relevant literature and analyzing current trends in the field.
- Developing, programming, and administering experiments to online participants.
- Presenting findings at international and domestic conferences to experienced speech scientists.

The Diversity, Equity, and Inclusion in Psychology Project, Penn State University
Spring 2022 - Spring 2023
Supervisor: Dr. Alicia Drais-Parillo

- Developed criteria for grading inclusiveness of outdated and modern educational materials, such as syllabi.
- Analyzed grades to determine growth of department DEI goals.
- Utilized findings to support proposed changes to educational materials and approaches to diversifying the Psychology Department.

The Augmentative and Alternative Communication Lab
Supervisor: Dr. Jessica Caron

- Analyzed speech therapy session recordings for behavioral patterns and use of AAC technology.
- Coded sessions based on preconceived criteria for client progress while using AAC.
- Maintained records of client sessions, AAC use, and analyzed data over time.

RELEVANT EXPERIENCE

Peer Advisor
College of the Liberal Arts, Penn State University
March 2022 - present

- Advise over 50 students in individualized settings to problem-solve academic challenges and champion campus engagement.
- Orchestrate connections between students and offices across the university to meet unique needs.
- Prioritize student goals in planning academic trajectories via classes, study abroad, research, and more.

Teaching Assistant
Psychology Department, Penn State University
January 2021 - January 2022

- Developed educational plans to structure classroom activities and fully engage students while managing remote and in-person teaching.
- Coached students on an individual level to achieve understanding and confidence with material.