# FACTORS AFFECTING THE INCIDENTAL FORMATION OF NOVEL SUPRASEGMENTAL CATEGORIES

by

JONATHAN WRIGHT

A DISSERTATION

Presented to the Department of Linguistics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2021

DISSERTATION APPROVAL PAGE

Student: Jonathan Wright

Title: Factors Affecting the Incidental Formation of Novel Suprasegmental Categories

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

| | |
|---|---|
| Melissa M. Baese-Berk | Chairperson |
| Melissa Redford | Core Member |
| Julie Sykes | Core Member |
| Caitlin Fausey | Institutional Representative |
| and | |
| Andrew Karduna | Interim Vice Provost for Graduate Studies |

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded September 2021

# DISSERTATION ABSTRACT

Jonathan Wright

Doctor of Philosophy

Department of Linguistics

September 2021

Title: Factors affecting the incidental formation of novel suprasegmental categories

Humans constantly use their senses to categorize stimuli in their environment. They develop categories for stimuli when they are young and constantly add to existing categories and learn novel categories throughout their life. A key factor when learning novel sound categories is the method a person uses to acquire the novel sound categories. Different learning methodologies interact with different neural processes and mechanisms, leading to diverse learning outcomes. However, auditory learning research has only recently begun to focus on the ways that various auditory processing structures interact with different learning methodologies. This dissertation investigates the acquisition of novel tone categories using natural tokens and an incidental learning paradigm. Throughout the experiments we demonstrated that native English participants with no prior experience with the target tone categories, from 18 to 66 years old, can use an incidental learning paradigm with natural tokens to form four novel tone categories after 30 minutes of training with very high, even perfect, accuracy. These findings confirm results from previous studies that suggest that participants can effectively learn novel sound categories through incidental learning paradigms, and we extend the investigation of factors impacting incidental learning into natural speech sound categories.

Across the four experiments we examined factors known to impact novel sound category acquisition. We demonstrated that high variability of tokens within trials resulted in greater learning than when the variability was spread out across trials. We also demonstrated that training on a single talker results in robust learning to novel tokens but a sharp decline when generalizing to novel talkers. By contrast, if participants are trained on multiple talkers during training, there is less learning, but there is little or no difference when generalizing learning to novel talkers. We also demonstrated that the presence of an unfamiliar vowel in the auditory stimuli did not impact the incidental formation of novel tone categories during perception only training. Further, we demonstrated that producing the tokens on each trial destroyed perceptual learning, and we presented multiple hypotheses regarding the nature of the disruption for future investigation. We also demonstrated that the presence of an unfamiliar vowel did not further disrupt perceptual learning over training with familiar segments. Thus, as a whole, this dissertation illustrated that incidental learning paradigms are an effective and efficient means for learning novel tone categories and investigating factors known to impact novel sound category acquisition.

CURRICULUM VITAE

NAME OF AUTHOR: Jonathan Wright

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Dallas International University, Dallas
University of Mary Hardin-Baylor, Belton

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2021 University of Oregon
Master of Arts, Linguistics, 2009, Dallas International University
Bachelor of Arts, 2002, University of Mary Hardin-Baylor

AREAS OF SPECIAL INTEREST:

Speech Perception
Speech Production
Phonetics
Second Language Acquisition

PROFESSIONAL EXPERIENCE:

Graduate Employee (researcher), Northwest Indian Language Institute, University of Oregon, 2017-2020

Graduate Employee (instructor), American English Institute, University of Oregon, 2016-2017

GRANTS, AWARDS, AND HONORS:

National Science Foundation (BCS-2017285). Doctoral Dissertation Research Improvement Grant, "Factors affecting incidental formation of novel suprasegmental categories", 2020-2021.

UO CAS Dissertation Research Fellowship, University of Oregon, 2020-2021.

General University Scholarship, University of Oregon, 2020-2021.

M. Gregg Smith Fellowship, University of Oregon, 2020-2021.

Duolingo Research Grant, Duolingo, 2020.

PUBLICATIONS:

Baese-Berk, M. M., Drake, S., Foster, K., Lee, D., Staggs, C., & Wright, J. M. (2021). Lexical Diversity, Lexical Sophistication, and Predictability for Speech in Multiple Listening Conditions. *Frontiers in Psychology*, *12*.

Wright, J. (2020). Khongso. *Journal of the International Phonetic Association,* 1-20.

# ACKNOWLEDGEMENTS

Thank you to my advisor, Melissa Baese-Berk, for your continual guidance throughout the PhD process. I am grateful for the opportunity that you gave me as I sought to expand from descriptive linguistic work to experimental linguistics and for your ongoing encouragement throughout the learning process. I have especially been grateful for your belief in me when it was difficult to believe in myself. You have provided me with a great example of what it means to be a mentor and advisor, and I hope that I will be able to pass on the same support and consideration that you have shown me throughout my time at the University of Oregon.

I have had the pleasure of regularly interacting with members of my committee through classes and events in the Department of Linguistics at the University of Oregon. Concepts developed as a student in a course with Caitlin Fausey were directly implemented in the dissertation and I am grateful for your brilliant insight and careful feedback that helped shape my work. Melissa Redford and Julie Sykes formed my advisory committee and provided important insight into the direction that my work took over the course of several years. Thank you for your guidance throughout this process.

I am grateful for my time in the Department of Linguistics at the University of Oregon. Faculty, staff, and students all worked to create an atmosphere that encourages students to expand their knowledge and explore possibilities. I have felt both encouraged and challenged to grow and learn. I am especially grateful for the other graduate students in the program. Your encouragement and cooperation have made our department a pleasure to be a part of.

Similarly, I am extremely grateful for the opportunity to be a part of the Speech Perception and Production Lab. The atmosphere of cooperation that all members worked to achieve was an ongoing source of encouragement through this process. It has been a pleasure seeing faculty, graduate students, and undergraduate students work together to investigate such a wide range of topics and factors in speech science. I have learned so much from all of you.

I am grateful for the generous funding agencies and grants that made this work possible. I am grateful to the National Science Foundation (BCS-2017285) for the DDRIG and the University of Oregon for the CAS Dissertation Research Fellowship. These grants allowed me to do work that I would not have been able to do otherwise. I am also grateful for the Duolingo

Research Grant, which also contributed greatly to my ability to do the present work. I am grateful to Cindy Blanco and the other researchers who I have had the pleasure of talking with about my research and its application to language acquisition.

Finally, my sincere appreciation goes to my family, Erin Wright and our three kids. You have worked hard to support me throughout this process and have never ceased in your encouragement. Thank you for putting up with the late nights and long hours over the last five years! This work belongs to all of us.

TABLE OF CONTENTS

# LIST OF FIGURES

Figure Page

Figure Page

Figure Page

# LIST OF TABLES

Table Page

# I. INTRODUCTION

Organisms are prone to categorize objects, sounds, and events in their environment. Categorization is fundamental for survival. Animals need to know which items are edible and which are harmful if consumed. They need to know which sounds indicate the availability of water, the presence of prey, or a potential attack from a predator. Humans use visual and auditory categorization processes and mechanisms to make split-second decisions that may be fatal if they make the wrong decision and constantly add to existing categories and learn novel categories throughout their life. The process of visual categorization involves learning to sort objects into categories and then extend that learning to novel examples of the category. Typically, the objects in a category have a feature or features that they share, but they also have features that are different. We can picture the category "dog" and imagine the shared features across the members of that category, but we can also imagine features that differ across the members of that category. Then, when we see a novel member of the category, we can identify that novel member as a dog. Similarly, auditory categorization involves sorting sounds into categories based on shared features and extending that learning to novel members of the category. We can imagine the sound of the wind even though the sounds we recognize as "wind" differ depending on the geography and vegetation of the area we are in. This categorization provides us with the ability to distinguish the sound of the wind from the sound of the waves at the beach or the sound of cars along a highway. We can then extend that knowledge when we are in a new geographical area and hear a novel sound that we immediately categorize as "wind". We also learn to sort sounds that we hear in speech into categories. For example, all languages use variations in the pitch of the voice to categorize sounds (Maddieson, 2013). In English we can tell if an utterance is a declarative statement or a question based on the pattern of the pitch across the utterance (i.e., intonation). The utterance, "We're leaving today," can be understood as a question or as an answer to a question depending on the pitch contour used to express it.[1] Different pitch categories can also be used to differentiate meaning at the word level. Pitch categories that are used to differentiate words

[1] When communicating, meaning can also be transmitted through the visual domain via gestures and facial expressions (Colin & Radeau, 2003). However, it is possible to transfer meaning via acoustic information alone, such as over the phone.

in a language are called *lexical tones*, or just *tones* for short, and languages that use tones are often called tonal languages (Maddieson, 2013). These pitch categories are typically differentiated based on the height of the pitch and/or the contour of the pitch. For example, a language may have three tone categories where the pitch remains even across the word and one category is identified by its low pitch, another by its high pitch, and another by its medium level pitch. When the pitch remains at a fairly constant level across the word, the tone category is called a *level tone* category. Therefore, this language would contain three level tone categories. Another possibility would be for the pitch in one or more categories to change by rising or falling across the word. The pitch could also rise and then fall or fall and then rise across the word. Tone categories where the pitch changes across the word are called *contour tones*. When discussing tones, we often use numbers to describe the pitch level. Thus, T11 would refer to a low level tone, T44 would refer to a high tone, T41 would refer to a high falling contour tone, and T13 could refer to a low rising contour tone. The tone categories in the current study come from Thai, which includes T45, T241, T315, T33 and T21.[2] The Thai tone categories used in the current study are discussed in detail in Chapter 2.

When communicating using a tonal language, the accurate perception and production of tones is vital. Mispronunciations can result in an inability to communicate, which is especially important to adults attempting to learn tonal languages. As discussed below, it is well noted that tone categories are difficult for adults to learn, especially if they do not have experience with tone categories in their first language. Resulting miscommunications from mispronunciations or misperceptions of tones can end in discouragement and eventual failure to learn the language. Despite the central role that tone categories have in language acquisition, relatively little research has been done regarding factors that impact novel tone acquisition. This lack of progress is in part due to the difficulty researchers face during experimentation in the lab. Traditionally, studying novel tone category formation requires learners to return to the lab over several days or weeks for training sessions (Chandrasekaran 2010; Francis et al. 2008). These training sessions typically include explicit instruction regarding the target categories and feedback on performance. Considering the effort required for such experimentation, findings have mostly been limited to inherent factors within listener and within L1 group that affect novel tone category formation. A wider range of factors such as number of talkers in the

[2] Tone notation from Chao (1930) is used throughout.

stimulus set, segmental composition of the stimuli, and the effect of production during learning are understudied. However, more recent methodologies have arisen that permit examinations of a wider range of factors that impact novel tone category acquisition and the learning mechanisms that contribute to novel sound category formation in adults.

In the present chapter, we provide an overview of tone perception research in light of the wider literature on category acquisition. In section 1.1 we discuss work on novel tone discrimination and novel tone category acquisition. In Section 1.2 we discuss novel sound category learning in the field of auditory perceptual learning by examining methodologies used to study stimuli categorization. Specifically, we discuss categorization, focusing on the categorization of speech sounds and then methodological approaches to novel sound category learning. In Section 1.3 we discuss the potential contributions of the current research and the hypotheses examined in this dissertation.

## 1.1 NOVEL TONE PERCEPTION

As a child learns their first language, they progressively develop speech sound categories specific to that language (Eimas et al. 1971; Kuhl 1987; Werker 1989; Kuhl et al. 1992). Speech sound categories that contrast with other speech sound categories to differentiate lexical meaning in a language are often referred to as *phonemes*. Children learn to differentiate between phonemes based on features such as voice onset time, phonation, F1, F2, F0 (pitch) height, and F0 contour. Further, as the child learns their first language, they learn to differentiate phonemic categories using only features present in their first language (L1). Consequentially, as they gain experience with the L1, the ability to discriminate between phonemic contrasts not present in their L1 is reduced. A commonly used example is the difficulty Japanese speakers face when discriminating between the English “r” and “l” (Goto 1971; MacKain et al. 1981, Sheldon and Strange 1982). For Japanese speakers this difficulty arises as English differentiates between two phonemes, but in Japanese a single sound category utilizes the same perceptual space (Sheldon and Strange 1982). When a person attempts to learn a second language, they often attempt to map the perceptual space of L2 phonemes onto the available perceptual spaces of their L1 phonemic categories. This mapping occurs in various ways, and several models account for possible mappings.

The Perceptual Assimilation Model (PAM) was originally designed for naïve learners (e.g., Best 1995), but was later extended to account for L2 learners (Best and Tyler 2007). PAM, the Speech Learning Model (SLM) (e.g., Flege, 1995), and the Second Language Linguistic Perception Model (L2LP) (e.g., Escudero 2005) all predict that L2 learners will adapt L2 phonemic categories to L1 categories that are closest to them in native perceptual space. PAM also provides predictions for several possibilities depending on the target phonemes in the L1 and L2, predicting that two target L2 phonemes may be mapped onto a single L1 category, as in the Japanese example above. Two target L2 phonemes may also map well onto two L1 categories. Similarly, an individual target L2 phoneme may map well onto an L1 category or it could fall in between two L2 categories. Another possibility is that there simply is no L1 category for the L2 phoneme to map onto. In this case discrimination ability can vary widely from poor to excellent. Acquiring L2 tone categories provides an example of mapping.

Some languages, such as Mandarin and Thai, are tonal languages and have tone categories, where pitch height and/or pitch contour differentiates speech sound categories. L2 learners with tonal L1s are able to map L2 tone categories onto L1 tone categories (Reid et al. 2015; Chen et al. 2018; Chen et al. 2019). Chen et al. (2019) had Mandarin listeners match Thai tones to Mandarin tone categories. Thai tones that were similar to Mandarin categories were more consistently matched to Mandarin tone categories than dissimilar tones. It is suggested that experience with L1 tone benefits L2 tone discrimination ability in learners when encountering novel tones (Wayland and Guion 2004). On the other hand, L2 learners with non-tonal L1s (e.g. English) do not have L1 categories to map L2 tones onto.[3] Research suggests that novel tone discrimination is difficult for native English speakers (Kiriloff 1969; Bluhme & Burr 1971; Shen 1989; Sun 1998; Wang et al. 1999; Wayland and Guion 2004; Reid et al. 2015), but training studies have shown that they are capable of learning to discriminate between tone categories (Chen and Pederson 2017; Chen et al. 2019) forming novel tone categories (Kiriloff 1969; Wang et al. 1999; Guion and Pederson 2007) and using tone categories to learn new lexical meanings (Wong and Perrachione 2007).

[3] Non-tonal languages are not equivalent. There may be non-contrastive phonetic features that benefit learners from some L1s over other L1s. However, in the current study we limit our scope to native English speakers.

### 1.1.1 Tone Discrimination

In general, work on tone perception can be separated into three areas: discrimination, adaptation, and novel category formation. Tone discrimination results suggest that directed attention, the number of speakers in the stimulus set, and variability in the phonological context influence tone discrimination accuracy. Although the current experiment examines novel category formation, factors involved in novel phoneme discrimination can inform our hypotheses regarding category formation. For example, training on novel segmental contrasts benefits from attention directed to the target contrast (Guion and Pederson 2007; Pederson and Guion 2010). Learners improve in discrimination of contrasts that they are made aware of but do not improve on other contrasts present in the stimuli. For example, in related work, Chen and Pederson (2017) found that, when exposed to stimuli differing in both tonal and segmental contrasts, Mandarin learners improved on discriminating between novel tones when their attention was directed to the tonal contrasts, but they did not improve on tonal discrimination when their attention was directed to novel segmental contrasts. Further, due to influence from the L1, listeners may also be endogenously oriented to features in the stimuli. For example, tone perception studies show that native English listeners weigh pitch cues differently than Mandarin listeners (Guion and Pederson 2007). It may be that native English speakers, due to lack of experience with lexical pitch, are endogenously oriented to direct their attention to segmental structure during auditory perception, leading to difficulty during tone discrimination and category formation during L2 acquisition.

Tone discrimination studies have also examined the effect of speaker variability and phonotactic variability on tone perception. To avoid ceiling effects, tone perception studies have typically introduced difficulty by including tokens from multiple talkers while controlling the phonotactic structure and segmental composition of the carrier syllable. This isolates the target tones while producing enough difficulty to attain comparable results. However, Chen and colleagues (Chen et al. 2018; Chen et al. 2019) examined the effect of number of talkers in the stimulus set and segmental variability on novel tone discrimination and adaptation among native Mandarin speakers. Discrimination was easiest in conditions where tokens were from the same talker and had the same vowel (Chen et al. 2019). When tokens came from multiple talkers or when they contained different vowels, discrimination of tones became significantly more difficult.[4]

In a similar study we investigated the impact of phonotactic structure and segmental composition on novel tone discrimination with naïve English and Mandarin participants (Wright & Baese-Berk, under review). We found that native English participants' novel tone discrimination accuracy was not impacted by phonotactic structure but was negatively impacted by segmental composition. The presence of /ŋ/ onsets, which are illegal in English, significantly reduced tone discrimination accuracy. For native English participants, /ŋ/ onsets resulted in no discrimination between tones. These results suggest that the segmental composition of carrier words for tones interacts with L1 phonotactic experience in modulating novel tone perception ability. These results, along with the work of Chen and colleagues, lead to the hypotheses tested in the current study.

### 1.1.2 Novel Tone Category Formation

Tone discrimination is measured by the ability to discriminate whether the tones in auditory stimuli are the same or different. Tone category formation is typically measured by the ability to identify the tone category of an auditory token out of a set of possible tones. In general, speech categorization can be a challenging task. When categorizing speech sounds, there is no single cue that might signal category membership. Rather, there are multiple cues underlining category membership for speech sounds. Further, the realization of the multiple cues of a speech sound vary across productions of the speech sound. Therefore, the task of *speech categorization* is to generalize across acoustically variant sounds to determine which features are salient to a specific type of sound and use those salient features to classify novel sounds. The ability to classify novel sounds based on an established sound category is called *generalization* (Palmeri & Gauthier, 2004; Holt & Lotto, 2010).

Although challenging, results from tone category formation studies suggest that native English speakers can learn novel tone categories (Wang et al. 1999) and use them to contrast word meaning (Wong and Perrachione 2007). Findings from these studies also suggest that

[4] This result pertains to native Mandarin participants. A similar study with native English listeners may differ. However, we would expect native English listeners to experience greater difficulty in a similar study.

there are experimental factors and individual factors that impact novel tone category formation success.

Experimental factors contributing to the success of novel tone category formation include phonotactic variability and number of talkers in the stimulus set (Wang et al. 1999). In speech perception there have been numerous studies on the effect of number of talkers during phoneme discrimination and novel category formation. Results suggest that participants in multiple talker conditions are initially slower and less accurate (Mullennix and Pisoni 1990). However, accuracy scores in multiple talker conditions can level off to match scores in single talker conditions. Further, multiple talker conditions can benefit learners as they help learners to better generalize learning to new talkers (Logan et al. 1991). Thus, talker variability during the novel category formation process can operate at an initial cost but end up helping the learner to generalize to new talkers. Talker variability exposes the learner to a greater range of possible acoustic output due to varying vocal tracts, speaking rates, etc. It is suggested that talker variability during novel tone category training is crucial to the ability to normalize differences in F0 across speakers, and this benefits the generalization of categories to new speakers (Wang et al. 1999). In current study we examine the effect of talker variability on the incidental acquisition of novel tone categories. The studies cited suggest that learners trained in a single talker condition will learn faster at first, but will perform worse when generalizing to new talkers.

Similarly, we examine the effect of segmental familiarity during novel tone category formation training. A hypothesis presented by Liu et al. (2011) is that phonotactic and segmental composition, especially involving novel segments, inhibits the learner's ability to attend to tone. Further, when attention is directed to segments, learners do not improve in tone discrimination ability (Chen and Pederson 2017). Therefore, researchers specifically avoid using segments or phonotactic structures that are not native to the participants' first language or use pseudowords to avoid negative effects from non-native phonological patterns (Wong & Perrachione, 2007; Chandrasekaran et al., 2010). As stated above, we specifically tested this hypothesis in a tone discrimination study and found that native English participants were unable to discriminate between novel tone categories when /ŋ/ onsets were present in the tokens (Wright & Baese-Berk, under review). In the current study we examine the impact of segmental familiarity on the formation of novel tone categories during incidental learning.

## 1.2 AUDITORY CATEGORY LEARNING

As discussed above, organisms constantly categorize input in their environment based on their senses. Stated another way, organisms respond differently to objects and events in their environment based on the way that they have categorized that input. Initial work on human speech categorization observed the way humans categorized speech sounds and concluded that the process of speech sound category learning was unique to the human auditory domain (Liberman, 1957; Liberman et al., 1957). The driving factor behind this perspective was research on categorical perception (Liberman et al., 1967; Kuhl, 1994, 2004). Categorical perception is the ability to identify discrete categories along an acoustic continuum of equal steps. In typical categorical perception studies sounds that differ on an acoustic dimension are presented to participants in equal steps along that dimension. The participant's categorization responses to each sound along the continuum do not vary gradually. Rather, there is an abrupt shift at one point on the continuum where the participant will switch from labeling the stimuli as one category to labeling the stimuli as the other category. The concept that categorical perception was specific to human speech resulted in expectations that human speech was driven by specialized processes and mechanisms (see Diehl, Lotto, & Holt, 2004). These concepts impacted research on the acquisition of novel speech sound categories.

For example, the Perceptual Assimilation Model (PAM; Best, 1995; Best and Tyler, 2007) provided predictions regarding how a person might assimilate pairs of sound categories from other languages based on their first language experience. As a child learns their first language (L1), they develop sound categories specific to that language (Eimas et al., 1971; Kuhl, 1987; Werker, 1989; Kuhl et al., 1992). They learn to differentiate between sound categories based on multiple cues, such as voice onset time, phonation, F1, F2, f0 (pitch) height, and f0 contour, which vary as a function of the specific language. As the child learns their first language, they learn to differentiate sound categories using only features present in their L1. Consequentially, as they gain experience with the L1, the ability to discriminate between contrasts not present in their L1 is reduced. A commonly used example is the difficulty Japanese speakers face when discriminating between the English "r" and "l" (Goto, 1971; MacKain et al., 1981, Sheldon and Strange, 1982). For Japanese speakers this difficulty arises as English differentiates between two speech sound categories, but in Japanese a single category utilizes the same acoustic space (Sheldon and Strange, 1982). When a person learns a second language, they often attempt to map the speech sounds of the L2 onto the available acoustic spaces in their L1. Several speech perception models have been presented to account for this mapping. The Perceptual Assimilation Model (PAM) (e.g., Best, 1995; Best and Tyler, 2007), the Speech Learning Model (SLM) (e.g., Flege, 1995), and the Second Language Linguistic Perception Model (L2LP) (e.g., Escudero, 2005) all predict that L2 learners will try to map L2 categories to L1 categories that are closest to them in native acoustic space. However, this mapping can occur in various ways. For example, two target L2 sound categories may be mapped onto a single L1 category, as in the Japanese example above. This can create difficulty in learning the two competing phonemes. It is much easier when two L2 sound categories map onto two different L1 categories. It can be the case, however, that there may not be an L1 category for the L2 phoneme to map onto and the resulting acquisition of the phoneme can vary widely from poor to excellent (Best, 1995).

As discussed, early investigation into novel sound category acquisition was impacted by categorical perception research and the focus of the field of speech perception on speech sound category formation as a speech-specific phenomenon. However, later research demonstrated that categorical perception is not specific to the auditory domain or to humans (Kuhl & Miller, 1978; Kuhl, 1985; Beale & Keil, 1995; Bimler & Kirkland, 2001; Krumhansl, 1991; Livingston et al., 1998; Kluender et al., 2012). The understanding that categorical perception is not unique to human speech corresponded with a greater interest in investigating domain-general processes and mechanisms involved in categorization across modalities.

Psychological research on how humans form novel categories is extensive (Bruner et al., 1956; Smith & Medin, 1981; Nosofsky, 1986; Estes, 1994; Ashby and Maddox, 2005, 2010; Chandrasekaran et al., 2014a, 2014b). The majority of category learning research has focused on visual categorization (see Cohen & Lefebvre, 2005). However, research on auditory categorization has been expanding, resulting in investigations of the applicability of visual categorization research to auditory categorization (Samuel, 1982; Maddox, Molis, & Diehl, 2002; Nearey, 1990; Johnson, 1997). By examining categorization across sensory domains, a more generalized picture has emerged suggesting that the processes involved in category learning may differ depending on the way a person learns the target categories (Ashby & Maddox, 2011; Richler & Palmeri, 2014). Thus, an important factor when learning novel sound categories is the method a person uses to acquire the novel sound categories. Until recently the majority of the research examining how people learn novel sound categories has focused on auditory category learning via learning paradigms that incorporate explicit instruction and feedback.

*Explicit category learning* occurs when learners are made aware of the categories they are learning. With explicit instruction, they learn rules that govern which category a given stimulus belongs to. Learners then apply the rule-based knowledge of the categories as they learn the target categories. Explicit feedback on performance is typically provided throughout training to let learners know if their application of the rules is accurate. By contrast, *implicit category learning* occurs when there are no instructions about the categories. Therefore, the learner does not have a conscious awareness of rules that govern category membership and thus, does not make a conscious effort to apply rules during category learning (see Reber, 1989).

Research focused on a single learning methodology limits our knowledge of the processes involved in the wider range of situations humans experience during auditory category learning. For example, we might "know" that certain factors impact learning, but it may be that they impact acquisition only under the particular methodology used, and that methodology might not be the optimal methodology for the particular learning situation. Limitations on our research could lead to misconceptions that can become rooted in societal knowledge. For example, we may "know" that older adults over a certain age cannot learn novel sound categories as well as younger populations. That is, we can make the mistake of generalizing knowledge that may only be specific to one learning methodology.

In the last two decades there has been a growing interest in novel sound category learning under various methodologies, leading to results that may differ from methodologies that only incorporate explicit instruction and feedback. Research on incidental and passive learning, for example, has resulted in new insight to difficult questions that have arisen in the field of category learning and has resulted in new models incorporating neural mechanisms and processes activated across learning methodologies (see Chandrasekaran et al., 2014). The current research contributes to the expanding knowledge of ways in which alternative learning methodologies result in novel auditory category learning by examining how multiple factors modulate the formation of novel sound categories during incidental auditory category learning.

### 1.2.1 Incidental auditory category learning

Unlike most of the previous novel tone category formation studies that used explicit categorization training (e.g., Wang et al. 1999; Wong and Perrachione 2007), the experiments in the present study utilize an incidental category formation paradigm. Much has been learned from novel tone category studies that use explicit categorization training. However, as discussed, results from explicit categorization training may have limited applicability, especially when learning novel categories in natural environments. For example, in everyday life, humans are rarely directed to look for specific sound categories and apply rules about categories to perceived sounds in an effort to learn to differentiate those sounds. Incidental category learning paradigms are thought to more closely approximate category learning that occurs in a human's natural environment (see Roark et al., 2020 for review).

The incidental learning paradigm used in the current study builds on the Systematic Multimodal Association Reaction Time (SMART) paradigm developed in Wade and Holt (2005) and Gabay et al. (2015). Gabay et al. (2015) successfully used an incidental learning paradigm to test category learning of four synthesized frequency categories. Instead of explicit instruction explaining tone categories combined with feedback during training, participants form tone categories incidentally while focused on a visual detection task. In each trial, listeners heard one of the synthesized frequencies repeated five times and then saw a visual target appear on the screen in one of four rectangles—the rectangles remained in place for the duration of the experiment. When the visual target appeared, the participants pressed a corresponding key. They were instructed to respond as quickly as possible (see Figure 1).

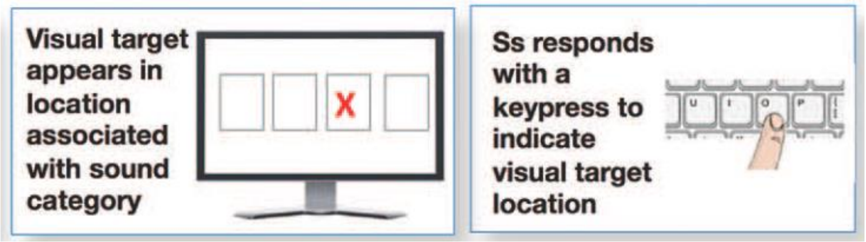
Visual target appears in location associated with sound category
X
Ss responds with a keypress to indicate visual target location
U
I
O
P

Figure 1. Incidental auditory learning paradigm used in Gabay et al. (2015). After hearing the five auditory stimuli, the visual target appears, and learners respond by pressing the key matching the location of the visual target (image from Gabay et al. 2015).

On each trial, the auditory categories in the stimuli were matched to visual locations. As participants discovered that auditory stimuli predicted the location of the visual target, auditory categories were developed and reinforced. Learning occurs even though participants make little effort on each trial. Participants simply see the visual target on the screen and then they respond with the keyboard or mouse. They are not consciously trying to learn. Therefore, it may seem that learning in this paradigm is passive learning. However, learning is not passive. There is a feedback mechanism incorporated in the incidental learning paradigm (Schultz et al., 1993, 1997; Gabay et al. 2015; Ashby & Casale, 2003; Sutton & Barto, 2005; Lim et al., 2014, Reynolds & Wickens, 2002).

In the incidental learning paradigm, learning occurs when the participant begins to use the auditory tokens as clues that reveal the upcoming location of the visual targets. They begin to use the auditory clues to predict where the visual target would appear. Then, participants receive feedback when the visual target appears and their prediction is proven to be correct or incorrect. Participants use this auditory-to-visual correspondence on each trial as reinforcing feedback to refine their categorical judgments of the following auditory stimuli. As they become more confident in their predictions, they move the mouse cursor to the location where they think the visual target will appear. When it appears where they predicted, they are rewarded by being able to click on the visual target faster. If they are wrong in their prediction, they will have to move the cursor to the location of the visual target and their reaction time will be slower.

Evidence of learning in incidental learning paradigms can come from several measures. Response times across training blocks become faster as the auditory-to-visual mapping is discovered. In some studies, learning is also measured by randomizing the auditory-to-visual mapping on a later training block, which has the effect of drastically slowing response times and the response time cost is measured. Further, typically a posttest with novel auditory stimuli is included. On the posttest, participants predict where the visual target would appear after only hearing the auditory stimuli. The experiments in the current study adapt the SMART task for natural tone categories. The adaptation is discussed in more detail in the description of the methodology of the first experiment in Section 3.3.

Incidental auditory category learning studies claim that sound category learning during incidental category learning better approximates sound learning in natural environments and predict that incidental learning is better suited for natural speech sound categories. However, almost all incidental sound category learning studies use synthesized speech sounds rather than natural speech tokens. In the current study we investigate the acquisition of novel speech sound categories through an incidental learning paradigm using natural tokens. The use of natural tokens allows us to investigate multiple factors known to impact novel speech sound category formation in explicit learning paradigms.

## 1.3 Current research

The goal of this dissertation is to examine the perceptual formation of novel tone categories with natural tokens through an incidental learning paradigm. Using natural tokens extends the applicability of research on the incidental formation of novel sound categories and permits the investigation of a number of factors known to impact the perceptual formation of novel sound categories during explicit learning. Specifically, in Experiment 1 we test the impact of within-trial token variability on novel tone category formation. In Experiment 2 we test the impact of talker variability on novel tone category formation. We also test a Control Condition to provide a baseline for the effect of age on the task in order to better compare the impact of age across conditions. In Experiment 3 we test the impact of segmental familiarity on novel tone category formation. In Experiment 4 we test the impact of production during perceptual learning on the perceptual formation of novel tone categories. We also test the modulation of segmental familiarity on the impact of production during perceptual learning.

### 1.3.1 Structure of the dissertation

The studies in this dissertation use natural tokens from multiple talkers. The use of natural tokens results in potentially large amounts of variation between auditory tokens. Acoustic differences between tokens could impact results. Therefore, it is valuable to examine acoustic differences among stimuli in detail. In Chapter 2 we present a characterization of the stimuli, describing differences regarding duration and F0. Chapter 3 through Chapter 6 present four experimental studies performed to analyze different factors that impact the incidental formation of novel tone categories. In the first experiment in Chapter 3 we investigate the role of token variability within trial, comparing trials that contain identical tokens with trials that contain variable tokens. By examining the impact of token variability on the incidental formation of novel tone categories we test the hypothesis that high token variability in close proximity to the audio-to-visual correspondence benefits learners by aiding in categorization and generalization to novel tokens (see Gabay et al., 2015). In the second experiment, in Chapter 4, we examine the impact of talker variability during training on the ability to generalize learning to novel tokens and novel talkers. By examining talker variability across trials, we test the hypothesis that exposure to multiple talkers during training aids in the ability to generalize to novel talkers. Further, in Experiment 2 we also examine a Control Condition where participants have no ability to learn the audio-to-visual correspondence and therefore receive no reinforcing feedback. Therefore, participants should not be able to have faster reaction times across training blocks. By examining a condition that includes no audio-to-visual correspondence, we test the impact of age on the task alone to observe a baseline effect of age on the task. In the third experiment, in Chapter 5, we include conditions containing tokens with different vowels to investigate the impact of segmental familiarity on novel tone category learning. By examining conditions with familiar and unfamiliar segments, we test potential impacts to perceptual learning from increased attentional load stemming from novel segments. In Chapter 6 we investigate the impact of production during perceptual learning, as well as the impact of segmental familiarity during production on perceptual learning. By examining production by participants immediately after auditory perception and the corresponding motor response on each trial, we test the impact that the anticipation of production during the audio-to-visual reinforcement has on perceptual learning. Further, we test the additional impact that the lack of segmental familiarity during motor planning has on perceptual learning. In Chapter 7 we present a summary of the findings and novel contributions to the field as well as future directions of this research.

### 1.3.2 Hypotheses explored in the current research

In the current study we investigate factors that impact the incidental formation of novel natural speech sound categories. Our hypotheses consider predictions from incidental auditory learning and the formation of natural sound categories. Below, we present hypotheses for each experiment.

Experiment 1, Chapter 3

One hypothesis we consider is that the incidental learning of novel tone categories will result in substantially better learning in a shorter amount of time compared to explicit learning methodologies.[5] We also hypothesize that token variability within trial will matter for incidental learning (Gabay et al., 2015). Specifically, variable tokens within trial will result in greater learning than identical tokens within trial.

Experiment 2, Chapter 4

We hypothesize that talker variability will matter for incidental acquisition of novel tone categories. Specifically, training on multiple talkers, compared to training on a single talker, will result in greater similarity in accuracy scores between Posttest 1, where participants generalize to novel tokens from the same talker(s) and Posttest 2, where participants generalize to novel tokens from novel talkers (Lively et al., 1993). However, it may be that overall, learners trained on a single talker could learn more accurately than learners trained on multiple talkers (Perrachione et al., 2011).

Experiment 3, Chapter 5

We hypothesize that segmental familiarity will matter for learning (Liu et al., 2011). Specifically, we expect that results from the two conditions with familiar segments, the /ma/ Condition and the /mi/ Condition will not differ but that a lack of familiarity would negatively impact learning in the /mɯ/ Condition. However, the impact of segmental familiarity on novel tone category formation may differ under the reflexive learning paradigm in the current study compared to the reflective learning paradigms used in previous studies.

Experiment 4, Chapter 6

We hypothesize that production during perceptual learning will matter and that segmental familiarity in the produced token will matter. Specifically, we predict that perceptual learning will be hindered when participants produce the tokens compared to the Perception Only Condition. Further, we expect that the effort to produce unfamiliar segments will increase the inhibitory effect of production on perceptual learning.

[5] In this study we do not directly compare explicit and incidental learning. The hypothesis, based on previous incidental learning studies (Wade & Holt, 2005; Gabay et al., 2015; Roark et al., 2020) is that categories will be formed in a single session during incidental learning rather than over the course of multiple days or weeks, which has been required for the formation of four new tone categories during explicit learning paradigms.

# II. STIMULI

As discussed in Chapter 1, we use natural tokens to test the incidental formation of novel tone categories. Specifically, we use four Thai tone categories produced by six native Thai talkers in /ma/, /mi/, and /mɯ/ syllables. In Section 2.2 we provide a characterization of the stimuli, including details regarding the recording of the stimuli. In Section 2.2.1 we provide an analysis of token duration across tone categories, syllable types, and talkers. In Section 2.2.2 we provide an analysis of the F0 contours that comprise each tone category, provide details for each talker's productions. We also compare F0 range across tone categories, syllable types, and talkers.

## 2.1 CHARACTERIZATION OF THE STIMULI

In the present experiments, stimuli were natural tokens that were recorded from six talkers, who were Thai females in their 20s and 30s and were living in the United States at the time of recording. Tokens from four talkers were used in Experiment 1 and Experiment 2. Tokens from all six talkers were used in Experiment 3 and Experiment 4.

Due to Covid-19 restrictions, recordings of the stimuli were done remotely. A Shure SM35 microphone and a Zoom H4N Pro audio recorder were sent to each talker for recording stimuli. After receiving the recording equipment, video sessions were held to explain recording instructions. Talkers were instructed to record the stimuli in a quiet setting. They were provided spreadsheets with the stimuli they were to record, which contained the syllables /ma/, /mi/, and /mɯ/ with the five Thai tones in the order that they are traditionally practiced in Thai schools: T33 (mid), T21 (low), T241 (falling), T45 (high), T315 (rising). In this way, all Thai talkers were very familiar with the pronunciation and cadence of the token sets. The set of five tokens was then repeated ten times to give ten unique productions of each token. Tokens for all experiments were recorded by each talker in a single recording session.

Tokens were normalized to an average intensity of 70 dB, and noise was reduced in Praat (Boersma & Weenik, 2015). Following Gabay et al. (2015), tokens were from four categories. The tone categories used in all experiments were based on four Thai tones: T45 (high rising), T241 (high falling), T315 (low rising), and T21 (low falling), using tone notation from Chao (1930). I excluded Thai tone T33 as I wanted to train participants on only four categories, and I wanted to maximize differences in each category. The four chosen tones provide a contrast between the categories with one high rising tone category, one high falling tone category, one low rising tone category, and one low falling tone category. Schematics for the five Thai tone categories, as seen in Figure 2, are presented by Reid et al. (2015).

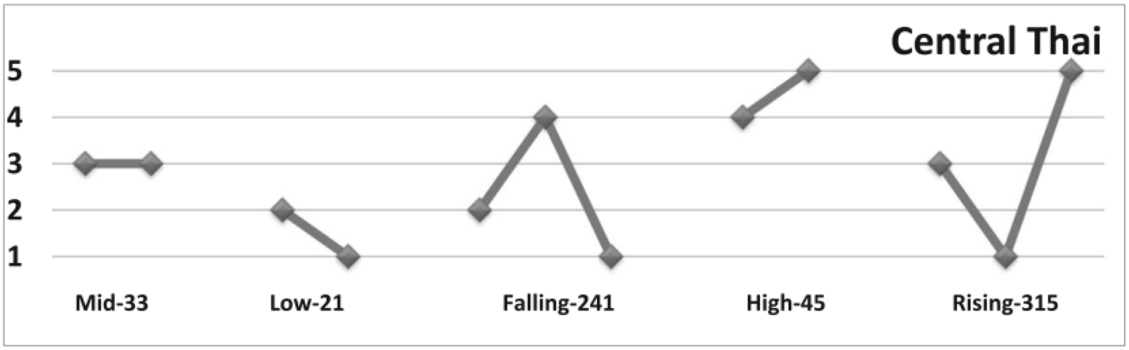

Figure 2. Schematics of Thai tones (Reid et al. 2015)

Tokens from all four tone categories were produced in the syllables /ma/, /mi/, and /mɯ/. Ten exemplars of each tone category were recorded from all four talkers. Typically, half of the exemplars were used for training, and half of the exemplars were used to test generalization of learning to new exemplars on Posttest 1. This is described in more detail in Section 3.2, 4.2, 5.2, and 6.2. Following Gabay et al. (2015), auditory stimuli in each trial consisted of five concatenated tokens, which, in most conditions, were randomly selected without duplication. However, due to the difficult circumstances of the recordings, a few productions by the talkers were not usable, resulting at times in only eight or nine exemplars of a category, rather than the normal ten. In these situations, where only four tokens were available for a trial, one randomly selected token was duplicated. These occurrences are listed in Section 3.2, 4.2, 5.2, and 6.2.

Due to Covid-19 restrictions, all experiments were run online. So, to minimize potential problems during auditory playback across a range of devices, browsers, and internet configurations, each set of five tokens within trial for the Variable Token Condition was randomly selected and concatenated before the experiments and then uploaded as single auditory files.

All tokens were individually inspected for abnormalities. Despite the difficult circumstances requiring that the audio recordings be done in the talkers' homes or offices, there were no noticeable abnormalities, such as clicks or pops, found in any of the stimuli used in the experiments. Further, after tokens were normalized for peak intensity and noise was reduced, there were no instances of obvious background noise (e.g., people talking or doors closing) found in any of the tokens.

### 2.1.1 Duration

Figure 3 illustrates duration for all talkers across tone categories and syllable types. The four boxplots in each of the three charts represent the distribution of durations for tokens from each tone of the chart's syllable type, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean duration for tokens of the specific tone. The dashed lines in each of the three charts represent the aggregated mean duration for all tokens of the chart's syllable type. The letters represent the means of the six individual talkers for the specific tone and syllable type.

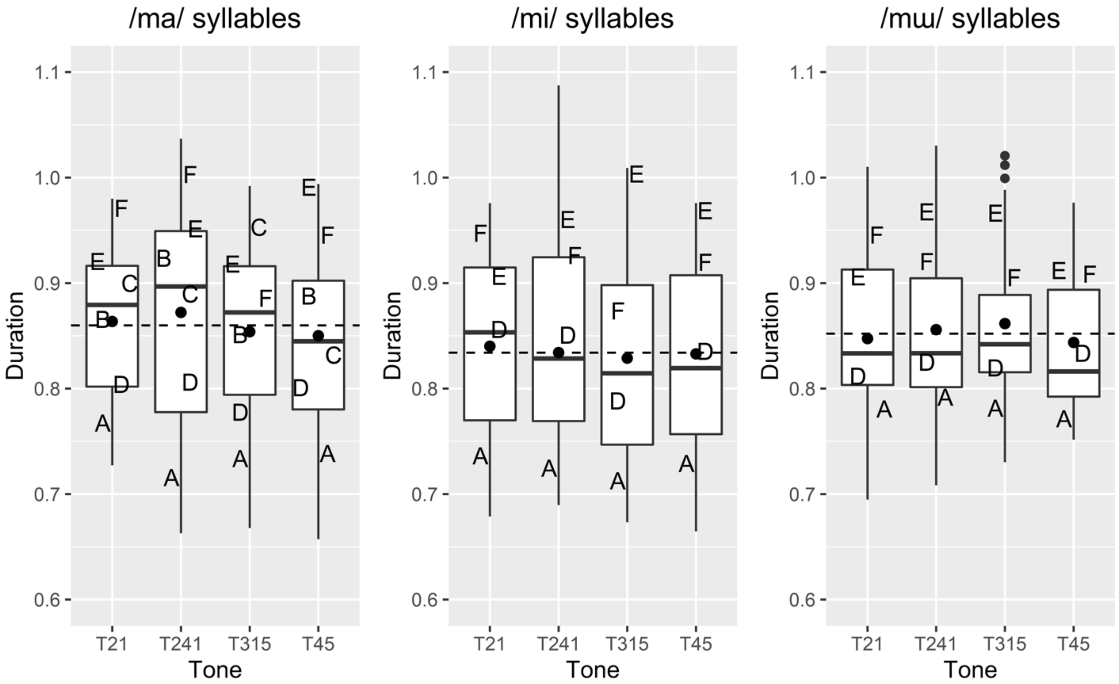

Figure 3. Aggregated duration means for each syllable type (dashed lines), for each tone (dots inside the box plots), and for each talker (letters).

To test differences in duration, I compared several mixed models. To determine whether an interaction between tone category and syllable type made a significant contribution to model fit, I compared models with and without an interaction, and results indicated a nonsignificant interaction ($X^2$ (6) = 4.86, *p* = .56).

duration ~ syllable*tone + (1|talker)
duration ~ syllable + tone + (1|talker)

To test whether duration differed as a function of tone category, I compared models with and without tone category, and results indicated that duration did not significantly differ as a function of tone category ($X^2$ (3) = 4.60, *p* = .20).

duration ~ syllable + tone + (1|talker)
duration ~ syllable + (1|talker)

Also, to test whether duration differed as a function of syllable type (e.g., /ma/, /mi/, /mɯ/), I compared models with and without syllable type, and results indicated that duration significantly differed as a function of syllable type ($X^2$ (2) = 8.20, *p* = .017). Bonferroni corrected post-hoc comparisons revealed that /mi/ syllables were shorter than /mɯ/ syllables (β = -.018, SE = .006, *t* = -2.84, *p* = .015), but /ma/ syllables did not differ from /mi/ syllables (β = .007, SE = .006, *t* = 1.07, *p* = .86) or from /mɯ/ syllables (β = .011, SE = .006, *t* = -1.74, *p* = .25).

duration ~ syllable + tone + (1|talker)
duration ~ tone + (1|talker)

To test whether duration differed as a function of talker, I compared models with and without talker, and as expected from the visualization in Figure 3, duration significantly differed as a function of talker ($X^2$ (1) = 449.68, *p* < 0.001).

duration ~ syllable + tone + (1|talker)
duration ~ syllable + tone

Figure 3 illustrates that Talker A had the shortest durations while Talkers E and F had the longest durations. Table 1 provides the mean, standard deviation, min, max, and range for each talker's durations across syllable types to illustrate differences in duration at the syllable level. Table 1 quantifies the expectation illustrated in Figure 3, that Talker A had the shortest mean durations across syllable types and Talkers E and F had the longest mean durations.

Table 2 provides comprehensive summary statistics for the six talkers, showing the mean, standard deviation, min, max, and range for each talker's durations for each tone category in each syllable type.

Table 1. Summary statistics for duration across syllable types

| Talker | Syllable | n | Mean | SD | Min | Max | Range |
|---|---|---|---|---|---|---|---|
| A | /ma/ | 40 | 0.76 | 0.06 | 0.66 | 0.92 | 0.26 |
| A | /mi/ | 40 | 0.74 | 0.04 | 0.66 | 0.83 | 0.17 |
| A | /mɯ/ | 40 | 0.79 | 0.04 | 0.69 | 0.9 | 0.21 |
| B | /ma/ | 40 | 0.89 | 0.04 | 0.81 | 0.99 | 0.17 |
| C | /ma/ | 32 | 0.90 | 0.05 | 0.81 | 0.99 | 0.18 |
| D | /ma/ | 20 | 0.79 | 0.02 | 0.76 | 0.83 | 0.07 |
| D | /mi/ | 20 | 0.82 | 0.03 | 0.75 | 0.87 | 0.12 |
| D | /mɯ/ | 20 | 0.83 | 0.02 | 0.79 | 0.87 | 0.09 |
| E | /ma/ | 16 | 0.95 | 0.04 | 0.88 | 1.01 | 0.12 |
| E | /mi/ | 20 | 0.95 | 0.05 | 0.84 | 1.09 | 0.25 |
| E | /mɯ/ | 20 | 0.95 | 0.05 | 0.85 | 1.03 | 0.18 |
| F | /ma/ | 20 | 0.95 | 0.04 | 0.88 | 1.04 | 0.16 |
| F | /mi/ | 20 | 0.91 | 0.03 | 0.84 | 0.96 | 0.12 |
| F | /mɯ/ | 20 | 0.91 | 0.06 | 0.82 | 1.01 | 0.19 |

Table 2. Summary statistics for duration across syllable types and tone categories

| Talker | Syllable | Tone | n | Mean | SD | Min | Max | Range |
|---|---|---|---|---|---|---|---|---|
| A | /ma/ | T21 | 10 | 0.79 | 0.07 | 0.73 | 0.92 | 0.19 |
| A | /ma/ | T241 | 10 | 0.73 | 0.05 | 0.66 | 0.81 | 0.14 |
| A | /ma/ | T315 | 10 | 0.75 | 0.06 | 0.67 | 0.87 | 0.20 |
| A | /ma/ | T45 | 10 | 0.76 | 0.04 | 0.66 | 0.84 | 0.18 |
| A | /mi/ | T21 | 10 | 0.75 | 0.04 | 0.68 | 0.83 | 0.15 |
| A | /mi/ | T241 | 10 | 0.74 | 0.04 | 0.69 | 0.82 | 0.13 |
| A | /mi/ | T315 | 10 | 0.73 | 0.03 | 0.67 | 0.77 | 0.10 |
| A | /mi/ | T45 | 10 | 0.75 | 0.05 | 0.66 | 0.82 | 0.15 |
| A | /mɯ/ | T21 | 10 | 0.78 | 0.06 | 0.69 | 0.9 | 0.21 |
| A | /mɯ/ | T241 | 10 | 0.77 | 0.04 | 0.71 | 0.83 | 0.13 |
| A | /mɯ/ | T315 | 10 | 0.80 | 0.05 | 0.73 | 0.89 | 0.16 |
| A | /mɯ/ | T45 | 10 | 0.79 | 0.02 | 0.75 | 0.81 | 0.06 |
| B | /ma/ | T21 | 10 | 0.88 | 0.03 | 0.83 | 0.95 | 0.11 |
| B | /ma/ | T241 | 10 | 0.93 | 0.04 | 0.89 | 0.99 | 0.10 |
| B | /ma/ | T315 | 10 | 0.87 | 0.03 | 0.82 | 0.93 | 0.10 |
| B | /ma/ | T45 | 10 | 0.87 | 0.03 | 0.81 | 0.92 | 0.10 |
| C | /ma/ | T21 | 8 | 0.89 | 0.03 | 0.85 | 0.93 | 0.08 |
| C | /ma/ | T241 | 8 | 0.91 | 0.04 | 0.87 | 0.99 | 0.12 |
| C | /ma/ | T315 | 8 | 0.94 | 0.05 | 0.84 | 0.99 | 0.15 |
| C | /ma/ | T45 | 8 | 0.85 | 0.03 | 0.81 | 0.89 | 0.08 |

Table 2. (continued).

| Talker | Syllable | Tone | n | Mean | SD | Min | Max | Range |
|---|---|---|---|---|---|---|---|---|
| D | /ma/ | T21 | 5 | 0.79 | 0.02 | 0.76 | 0.80 | 0.05 |
| D | /ma/ | T241 | 5 | 0.79 | 0.03 | 0.76 | 0.83 | 0.07 |
| D | /ma/ | T315 | 5 | 0.8 | 0.02 | 0.77 | 0.83 | 0.05 |
| D | /ma/ | T45 | 5 | 0.79 | 0.01 | 0.78 | 0.80 | 0.02 |
| D | /mi/ | T21 | 5 | 0.84 | 0.03 | 0.80 | 0.86 | 0.06 |
| D | /mi/ | T241 | 5 | 0.83 | 0.02 | 0.81 | 0.87 | 0.06 |
| D | /mi/ | T315 | 5 | 0.81 | 0.04 | 0.75 | 0.86 | 0.11 |
| D | /mi/ | T45 | 5 | 0.82 | 0.02 | 0.79 | 0.85 | 0.06 |
| D | /mɯ/ | T21 | 5 | 0.83 | 0.02 | 0.80 | 0.84 | 0.04 |
| D | /mɯ/ | T241 | 5 | 0.84 | 0.03 | 0.82 | 0.87 | 0.05 |
| D | /mɯ/ | T315 | 5 | 0.84 | 0.02 | 0.82 | 0.85 | 0.04 |
| D | /mɯ/ | T45 | 5 | 0.82 | 0.02 | 0.79 | 0.84 | 0.06 |
| E | /ma/ | T21 | 4 | 0.94 | 0.03 | 0.91 | 0.98 | 0.07 |
| E | /ma/ | T241 | 4 | 0.97 | 0.04 | 0.91 | 1.01 | 0.09 |
| E | /ma/ | T315 | 4 | 0.92 | 0.03 | 0.88 | 0.95 | 0.07 |
| E | /ma/ | T45 | 4 | 0.97 | 0.03 | 0.93 | 0.99 | 0.07 |
| E | /mi/ | T21 | 5 | 0.92 | 0.03 | 0.89 | 0.98 | 0.08 |
| E | /mi/ | T241 | 5 | 0.94 | 0.09 | 0.84 | 1.09 | 0.25 |
| E | /mi/ | T315 | 5 | 0.99 | 0.02 | 0.96 | 1.01 | 0.05 |
| E | /mi/ | T45 | 5 | 0.95 | 0.02 | 0.91 | 0.98 | 0.06 |
| E | /mɯ/ | T21 | 5 | 0.92 | 0.05 | 0.85 | 0.99 | 0.14 |
| E | /mɯ/ | T241 | 5 | 0.95 | 0.06 | 0.87 | 1.03 | 0.16 |
| E | /mɯ/ | T315 | 5 | 0.98 | 0.03 | 0.95 | 1.02 | 0.07 |
| E | /mɯ/ | T45 | 5 | 0.93 | 0.04 | 0.88 | 0.98 | 0.10 |
| F | /ma/ | T21 | 5 | 0.95 | 0.02 | 0.92 | 0.97 | 0.05 |
| F | /ma/ | T241 | 5 | 0.99 | 0.04 | 0.95 | 1.04 | 0.09 |
| F | /ma/ | T315 | 5 | 0.90 | 0.02 | 0.88 | 0.93 | 0.05 |
| F | /ma/ | T45 | 5 | 0.96 | 0.03 | 0.91 | 0.99 | 0.07 |
| F | /mi/ | T21 | 5 | 0.93 | 0.03 | 0.88 | 0.95 | 0.07 |
| F | /mi/ | T241 | 5 | 0.91 | 0.05 | 0.84 | 0.96 | 0.12 |
| F | /mi/ | T315 | 5 | 0.89 | 0.01 | 0.87 | 0.90 | 0.03 |
| F | /mi/ | T45 | 5 | 0.90 | 0.02 | 0.88 | 0.92 | 0.04 |
| F | /mɯ/ | T21 | 5 | 0.93 | 0.08 | 0.82 | 1.01 | 0.19 |
| F | /mɯ/ | T241 | 5 | 0.94 | 0.06 | 0.87 | 1.01 | 0.14 |
| F | /mɯ/ | T315 | 5 | 0.89 | 0.06 | 0.84 | 1.00 | 0.16 |
| F | /mɯ/ | T45 | 5 | 0.89 | 0.04 | 0.84 | 0.96 | 0.12 |

In addition to differences across talkers, there were also some differences within talker. For each talker, I performed ANOVAs examining tone category, syllable type, and an interaction between the two. Durations for Talker A are shown in Figure 4. Results from a two-way ANOVA indicated that duration significantly differed as a function of syllable type [*F*(2, 108) = 7.89, *p* < .001, $\eta^2_p$ = .13, $\eta^2_G$ = .12]. However, duration did not differ as a function of tone category [*F*(3, 108) = 1.12, *p* = .34, $\eta^2_p$ = .03, $\eta^2_G$ = .03], and the interaction between syllable type and tone category was nonsignificant [*F*(6, 108) = .94, *p* = .47, $\eta^2_p$ = .05, $\eta^2_G$ = .04]. Bonferroni corrected post-hoc comparisons revealed that /mi/ syllables were shorter than /mɯ/ syllables (β = -.042, SE = .011, t ratio = -3.89, *p* < .001), and /ma/ syllables were shorter than /mɯ/ syllables (β = -.029, SE = .011, t ratio = -2.66, *p* = .027). However, /ma/ syllables did not differ from /mi/ syllables (β = .013, SE = .011, t ratio = 1.23, *p* = 0.66).

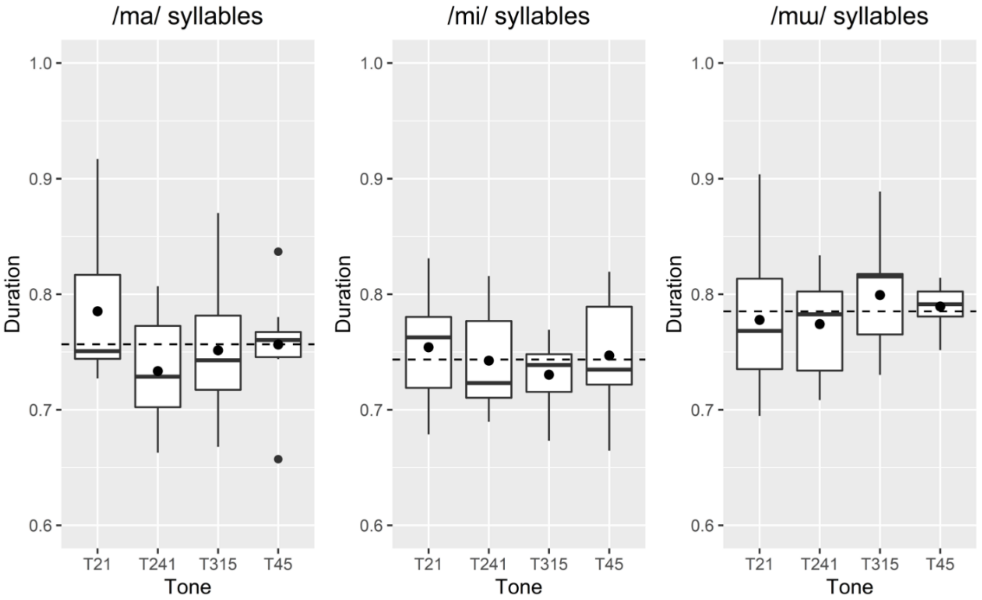

Figure 4. Aggregated duration means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker A.

Talker B was one of three talkers in the multitalker condition in the second experiment, which only used /ma/ syllables. Therefore, /mi/ and /mɯ/ tokens from Talker B were not used or analyzed. Durations for /ma/ syllables for Talker B are shown in Figure 5. Results from a one-way ANOVA indicated that duration significantly differed as a function of tone category [*F*(3, 36) = 7.02, *p* < .001, $\eta^2_p$ = .37, $\eta^2_G$ = .37]. Bonferroni corrected post-hoc comparisons revealed that T241 durations were longer than T21 durations (β = -.046, SE = .015, t ratio = -3.03, *p* = .027), T315 durations (β = .06, SE = .015, t ratio = 3.98, *p* = .002), and T45 durations (β = .059, SE = .015, t ratio = 3.92, *p* = 0.002).

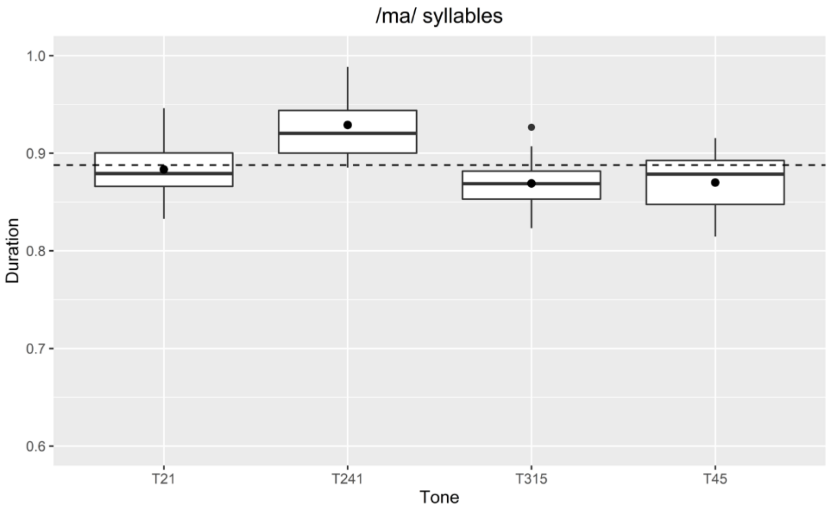

Figure 5. Aggregated duration means for /ma/ syllables (dashed lines) and tone category (dots inside the box plots) for Talker B.

Talker C was also one of three talkers in the multitalker condition in the second experiment, which only used /ma/ syllables. Therefore, /mi/ and /mɯ/ tokens from Talker C were not used or analyzed. Durations for /ma/ syllables for Talker C are shown in Figure 6. Results from a one-way ANOVA indicated that duration significantly differed as a function of tone category [$F(3, 28) = 8.36$, $p < .001$, $\eta^2_p = .47$, $\eta^2_G = .47$]. Bonferroni corrected post-hoc comparisons revealed that T45 durations were shorter than T241 durations (β = .058, SE = .018, t ratio = 3.31, $p = .015$) and T315 durations (β = .086, SE = .018, t ratio = 4.89, $p < .001$).

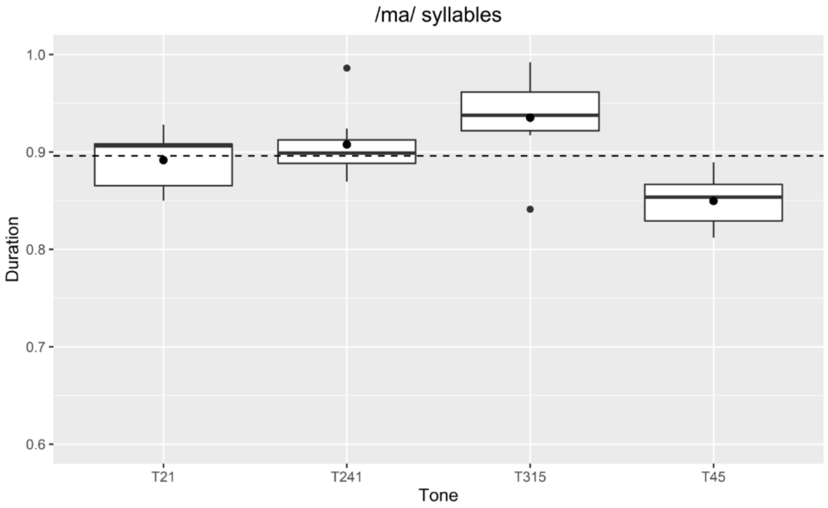

Figure 6. Aggregated duration means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker C.

Durations for Talker D are shown in Figure 7. Results from a two-way ANOVA indicated that duration significantly differed as a function of syllable type [$F(2, 48) = 17.24$, $p < .001$, $\eta^2_p = .42$, $\eta^2_G = .37$]. However, duration did not significantly differ as a function of tone category [$F(3, 48) = 1.06$, $p = .37$, $\eta^2_p = .06$, $\eta^2_G = .03$], and the interaction between syllable type and tone category was nonsignificant [$F(6, 48) = 1.05$, $p = .41$, $\eta^2_p = .12$, $\eta^2_G = .07$]. Bonferroni corrected post-hoc comparisons revealed that /ma/ syllables were shorter than /mi/ syllables (β = -.034, SE = .008, t ratio = -4.51, $p < .001$) and /mɯ/ syllables (β = -.042, SE = .008, t ratio = -5.51, $p < .001$). However, /mi/ syllables did not differ from /mɯ/ syllables (β = -.008, SE = .008, t ratio = -1.00, $p = 0.96$).

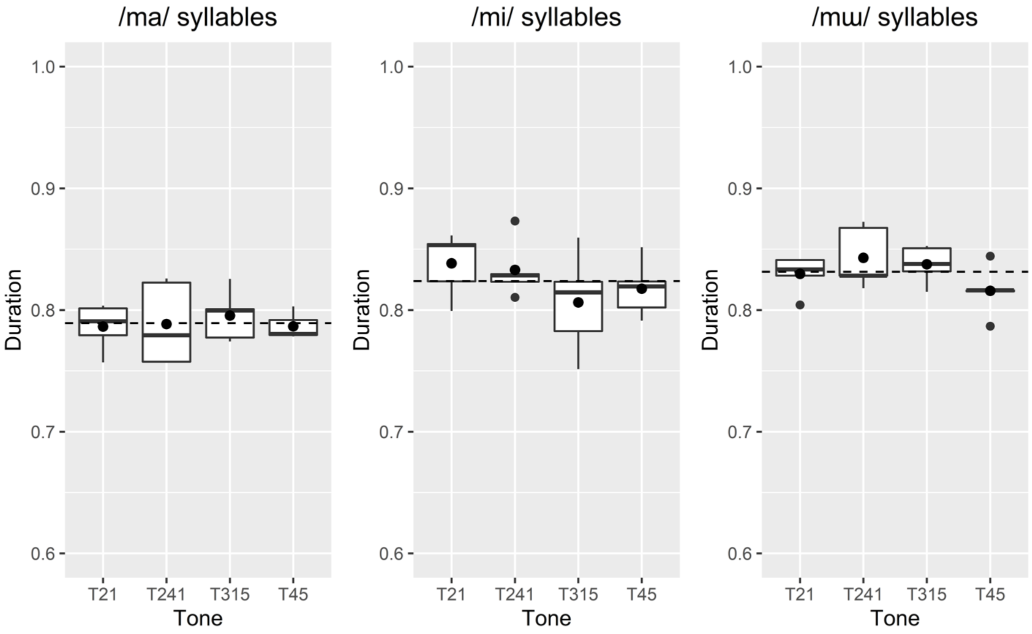

Figure 7. Aggregated duration means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker D.

Durations for Talker E are shown in Figure 8. Results from a two-way ANOVA indicated that duration did not significantly differ as a function of syllable type [$F(2, 44) = .05$, $p = .95$, $\eta^2_p = .002$, $\eta^2_G = .002$], nor as a function of tone category [$F(3, 44) = 1.75$, $p = .17$, $\eta^2_p = .11$, $\eta^2_G = .09$], and the interaction between syllable type and tone category was nonsignificant [$F(6, 44) = 1.40$, $p = .24$, $\eta^2_p = .16$, $\eta^2_G = .15$].

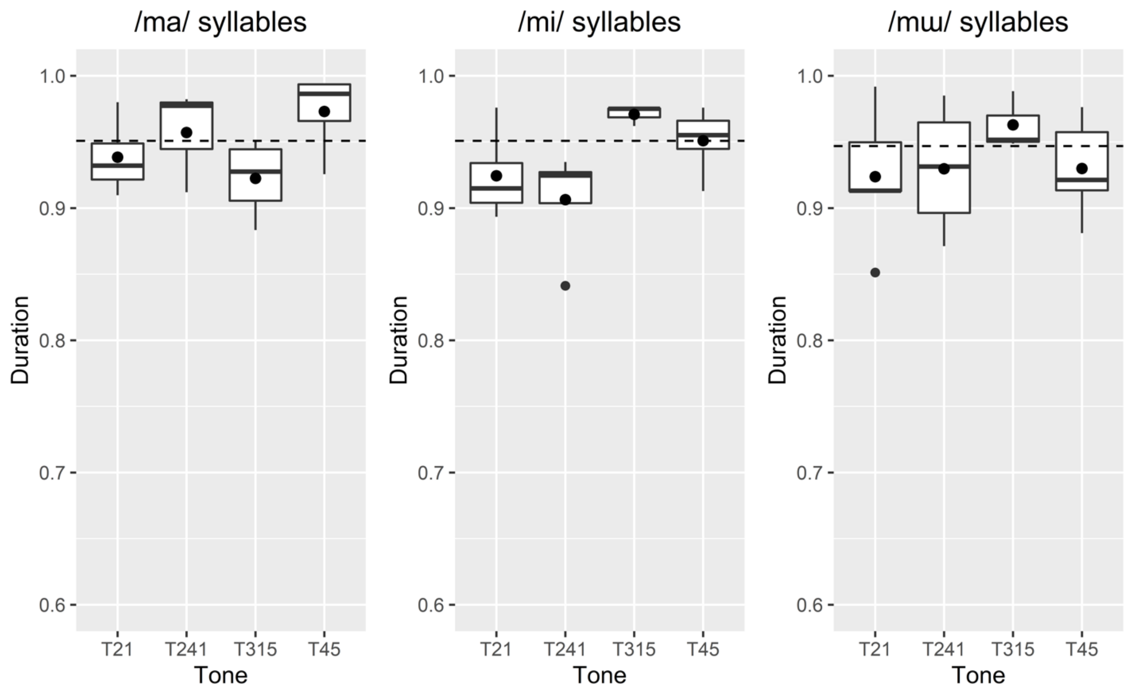

Figure 8. Aggregated duration means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker E.

Durations for Talker F are shown in Figure 9. Results from a two-way ANOVA indicated that duration significantly differed as a function of syllable type [$F(2, 48) = 5.77$, $p = .006$, $\eta^2_p = .19$, $\eta^2_G = .15$], and as a function of tone category [$F(3, 48) = 3.94$, $p = .01$, $\eta^2_p = .20$, $\eta^2_G = .16$]. However, the interaction between syllable and tone was nonsignificant [$F(6, 48) = .77$, $p = .60$, $\eta^2_p = .09$, $\eta^2_G = .06$]. Bonferroni corrected post-hoc comparisons revealed that /ma/ syllables were longer than /mi/ syllables (β = -.042, SE = .014, t ratio = 3.06, *p* = .010), and /ma/ syllables were longer than /mɯ/ syllables (β = -.039, SE = .014, t ratio = 2.81, *p* = .022). However, /mi/ syllables did not differ from /mɯ/ syllables (β = -.003, SE = .014, t ratio = -.25, *p* = 1.00). Also, T241 was longer than T315 (β = .05, SE = .016, t ratio = 3.17, *p* = .016).

The tokens used in the current studies were natural tokens. Duration was not controlled. Therefore, there were differences in duration. Controlling for talker, duration differed as a function of syllable type but not as a function of tone category. Specifically, /mi/ syllables were shorter than /mɯ/ syllables. Within talker there were individual differences in duration. For talker A duration differed as a function of syllable type, with /mɯ/ syllables being longer than /ma/ or /mi/ syllables. For talker B duration differed as a function of tone category, with tone T241 being longer than the other tones. For talker C duration differed as a function of tone category, with tone T315 being longer than the other tones. For talker D duration differed as a function of syllable type, with /ma/ syllables being shorter than /mi/ or /mɯ/ syllables. For talker E duration did not differ as a function of syllable type or tone category. For talker F duration differed as a function of syllable type and tone category, with /ma/ syllables being longer than /mi/ or /mɯ/ syllables and tone T241 being longer than tone T315. Overall, there were no consistent differences in durations across tone categories that might aid in an interpretation of the results from the current study. The duration differences across syllable type could affect results. Each condition in the current studies uses a single syllable type. So, it may be that participants exposed to /mɯ/ syllables have an advantage or disadvantage due to the longer duration of the syllable type. This will be considered in the analysis of the results.

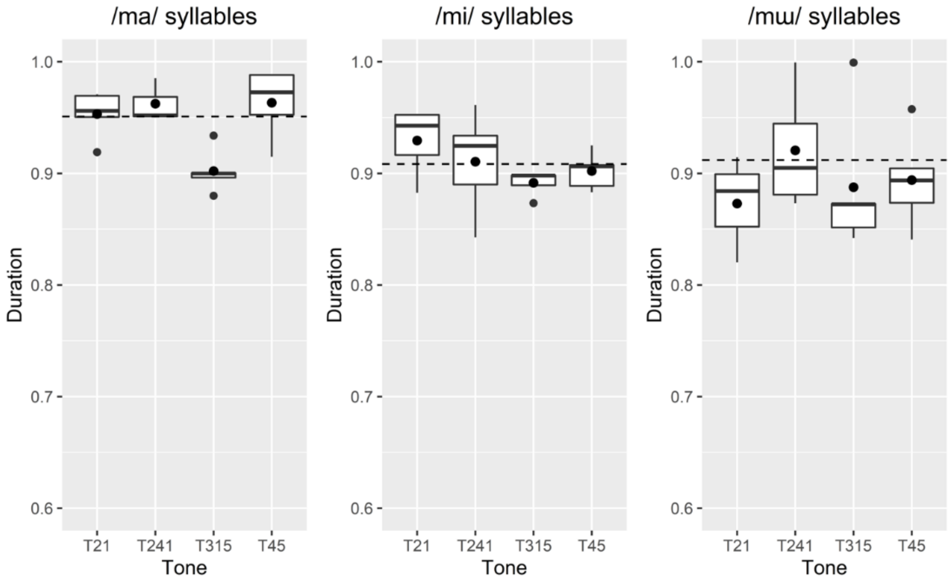

Figure 9. Aggregated duration means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker F.

### 2.1.2 F0

The following section presents an analysis of F0 contours and F0 range across talkers and within talker. I investigate F0 contours of the tone categories across talkers and systematic differences in F0 range of the tone categories across talkers and syllable types to provide analyses of differences between talkers or aberrations within talker that may impact learning in the current experiments.

Figure 10 illustrates the tone contours from the stimuli taken from each talker with normalized time. The contours represent means extracted from every five percent of the duration of the tone bearing unit across all tokens. The initial and final portions – about ten percent – of the durations were not used as they included large numbers of missing and randomly jittered values due to transitions to silence. F0 values were extracted using Praat (Boersma & Weenink, 2015), and the values were carefully inspected. Values were sometimes missing or randomly jittered due creaky voice, which occurred predominantly on lower F0 contours. These values were then measured manually. The light grey areas bordering the F0 contours represent ± 1 standard error of the mean.

A visual inspection of the F0 contours reveals individual differences in the realization of the tone categories. From the differences in ranges of Hertz on the y axis, it is easy to see that talkers differed somewhat in the F0 range that they used across all tones and for each tone category. F0 range specifically will be discussed in detail below. There are also individual differences in the shape of the F0 contours. For example, the crest of the T241 contour and trough of the T315 contour occur at different durations across the contours, the shape of the T45 rising contour differs, and the T21 ending may fall, rise slightly, or level off.

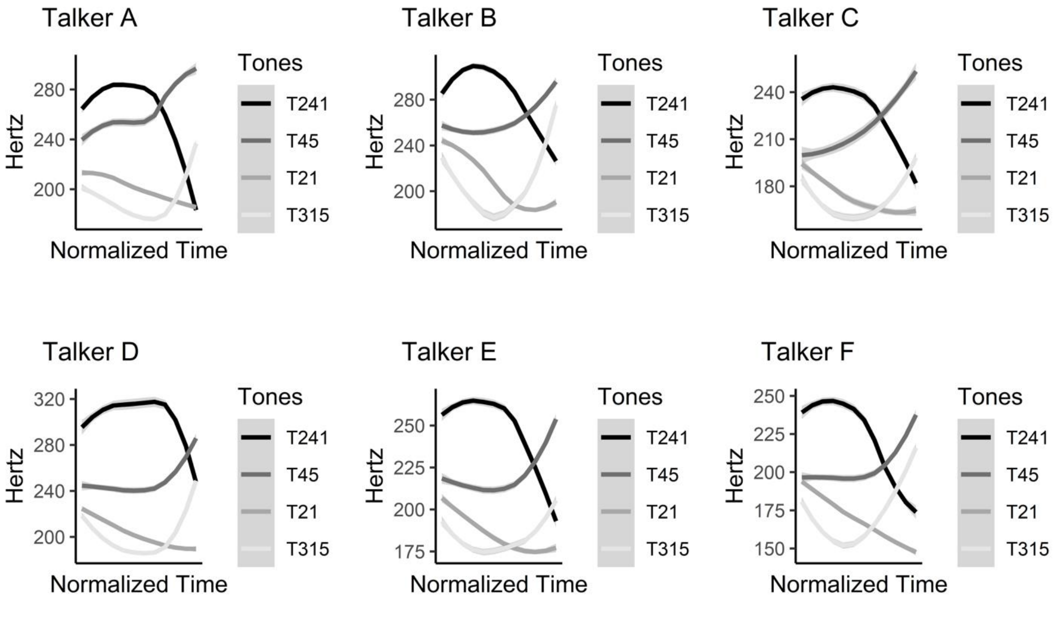

Figure 10. Mean F0 contours and ± 1 standard error of the mean for each tone category for each talker across normalized time.

Figure 11 illustrates F0 range values for all talkers across tone categories and syllable types. The four boxplots in each of the three charts represent the distribution of F0 ranges for tokens from each tone of the chart’s syllable type, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean F0 range for tokens of the specific tone. The dashed lines in each of the three charts represent the aggregated mean F0 range for all tokens of the chart’s syllable type. The letters represent the means of the six individual talkers for the specific tone and syllable type.

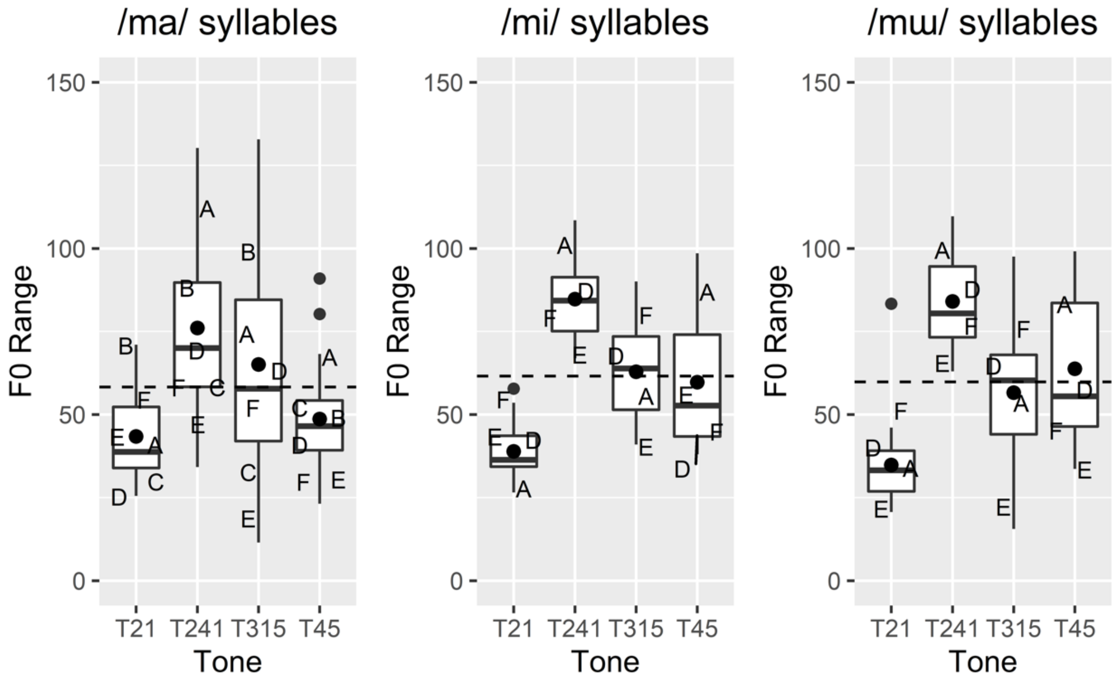

Figure 11. Aggregated F0 range means for each syllable type (dashed lines), for each tone (dots inside the box plots), and for each talker (letters).

The overall F0 range across all tones and the F0 range for each tone is a primary feature of F0 that differs across the speakers of a language and thus may impact learners’ ability to perceive and learn tone categories. Therefore, it is expected that F0 range will differ across talkers and across tones. A visual inspection of the differences between talkers and tones in Figure 11 seems to confirm this hypothesis. However, it is unclear whether F0 range will differ across syllable types for all talkers or within talker. To test differences in F0 range across syllables, talkers, and tones, I compared several mixed models.

I compared models with and without talker, and as expected, F0 range significantly differed as a function of talker ($X^2$ (1) = 106.96, *p* < 0.001).

F0 range ~ syllable + tone + (1|talker)
F0 range ~ syllable + tone

I also compared models with and without tone category to test whether F0 range differed as a function of tone category, and as expected, results indicated that F0 range significantly differed as a function of tone category ($X^2$ (3) = 232.17, *p* = .20). Bonferroni corrected post-hoc comparisons revealed that the F0 range of each tone was significantly different from each other tone. In a visual inspection of Figure 11 it appears that T315 and T45 have a very similar F0 range, but the difference between the two is still significant (β = 6.4, SE = 2.3, *t* = 2.78, *p* = .034). Pairwise comparisons are presented in Table 3.

F0 range ~ syllable + tone + (1|talker)
F0 range ~ syllable + (1|talker)

Table 3. Bonferroni corrected pairwise comparisons for F0 range across tone categories

| Tone | β | *SE* | *t* | *p* |
|---|---|---|---|---|
| T21 – T241 | -40.74 | 2.30 | -17.68 | < .001 |
| T21 – T315 | -22.34 | 2.30 | -9.70 | < .001 |
| T21 – T45 | -15.94 | 2.30 | -6.92 | < .001 |
| T241 – T315 | 18.40 | 2.30 | 7.99 | < .001 |
| T241 – T45 | 24.80 | 2.30 | 10.77 | < .001 |
| T315 – T45 | 6.40 | 2.30 | 2.78 | .034 |

Also, to test whether F0 range differed as a function of syllable type (e.g., /ma/, /mi/, /mɯ/), I compared models with and without syllable type, and results indicated that F0 range significantly differed as a function of syllable type ($X^2$ (2) = 8.62, *p* = .013). Bonferroni corrected post-hoc comparisons revealed that the F0 range of /mi/ syllables was wider than /ma/ syllables (β = -6.35, SE = 2.23, *t* = -2.85, *p* = .014), but /ma/ syllables did not differ from /mɯ/ syllables (β = -4.61, SE = 2.23, *t* = -2.07, *p* = .12) and /mi/ syllables did not differ from /mɯ/ syllables (β = 1.75, SE = 2.21, *t* = .79, *p* = 1).

F0 range ~ syllable + tone + (1|talker)
F0 range ~ tone + (1|talker)

A main interest was to examine F0 range for each tone category across syllable types, and so I compared models with and without an interaction between tone category and syllable type to determine if an interaction made a significant contribution to model fit. Results indicated a significant interaction between tone category and syllable type ($X^2$ (6) = 32.04, *p* < .001).

F0 range ~ syllable * tone + (1|talker)
F0 range ~ syllable + tone + (1|talker)

To further investigate the interaction between tone category and syllable type, I used subsets of the data to measure each tone category across syllable types. For each tone category I compared models with and without syllable type to determine if syllable type made a significant contribution to model fit. If F0 range differed as a function of syllable type for a tone category, I performed Bonferroni corrected post-hoc comparisons to further investigate differences across syllable types.

F0 range ~ syllable + (1|talker)
F0 range ~ (1|talker)

Syllable type did not make a significant contribution to model fit as a predictor for T241 F0 range ($X^2$ (2) = 3.76, *p* = .15), T21 F0 range ($X^2$ (2) = 3.92, *p* = .14), or T315 F0 range ($X^2$ (2) = 4.24, *p* = .12). However, syllable type did make a significant contribution to model fit for T45 ($X^2$ (2) = 28.10, *p* < .001). Bonferroni corrected post-hoc comparisons revealed that for T45 the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -11.93, SE = 2.93, *t* = -4.08, *p* < .001) and /mɯ/ syllables (β = -16.00, SE = 2.93, *t* = -5.47, *p* < .001). However, for T45 the F0 range of /mi/ syllables did not differ from /mɯ/ syllables (β = -4.06, SE = 2.90, *t* = -1.40, *p* = .50). Figure 11 illustrates these differences in F0 range across syllables for T45, with /ma/ syllables illustrating lower F0 range than /mi/ or /mɯ/ syllables.

Figure 11 illustrates F0 s for each talker for each tone category across the three syllable types. A visual inspection of Figure 11 indicates that patterns emerge for each talker across syllable types and that some talkers have consistently wider or narrower F0 ranges than other talkers. For example, Talker B consistently has wider F0 ranges across tone categories than Talker E. Table 4 provides the F0 mean, standard deviation, min, max, and range for each talker's productions across tone categories and syllable types. Table 4 is ordered by talker and by tone category to facilitate the comparison of F0 range values of the tone categories within talker.

Table 4. Summary statistics for F0 range across syllable types and tone categories, ordered by talker and tone category to facilitate comparison of F0 range for each tone category across syllable types

| Talker | Tone | Syllable | n | Mean | SD | Min | Max | Range |
|---|---|---|---|---|---|---|---|---|
| A | T21 | /ma/ | 120 | 196.85 | 12.7 | 171.87 | 218.28 | 46.41 |
| A | T21 | /mi/ | 120 | 203.64 | 11.5 | 173.55 | 227.76 | 54.21 |
| A | T21 | /mɯ/ | 120 | 200.98 | 11.97 | 179.79 | 268.83 | 89.04 |
| A | T241 | /ma/ | 120 | 259.74 | 35.66 | 165.12 | 307.18 | 142.06 |
| A | T241 | /mi/ | 120 | 259.4 | 30.9 | 179.77 | 291.34 | 111.57 |
| A | T241 | /mɯ/ | 120 | 261.15 | 31.43 | 178.86 | 291.97 | 113.11 |
| A | T315 | /ma/ | 120 | 188.05 | 20.23 | 162.48 | 274.56 | 112.08 |
| A | T315 | /mi/ | 120 | 196.49 | 17.96 | 177.3 | 269.31 | 92.01 |
| A | T315 | /mɯ/ | 120 | 193.21 | 18.05 | 169.88 | 278.17 | 108.29 |
| A | T45 | /ma/ | 120 | 257.26 | 20.7 | 229.21 | 329.43 | 100.22 |
| A | T45 | /mi/ | 120 | 268.89 | 23.95 | 223.73 | 347.11 | 123.38 |
| A | T45 | /mɯ/ | 120 | 265.91 | 25.89 | 214.54 | 351.67 | 137.13 |
| B | T21 | /ma/ | 120 | 208.25 | 23.35 | 169.41 | 251.16 | 81.75 |
| B | T241 | /ma/ | 120 | 282.22 | 28.49 | 218.1 | 321.99 | 103.89 |
| B | T315 | /ma/ | 120 | 207.68 | 29.82 | 163.06 | 295.87 | 132.82 |
| B | T45 | /ma/ | 120 | 262.78 | 15.83 | 245.08 | 315.73 | 70.66 |
| C | T21 | /ma/ | 96 | 173.28 | 11.7 | 146 | 198.58 | 52.58 |
| C | T241 | /ma/ | 96 | 226.27 | 21.37 | 176.12 | 252.8 | 76.68 |
| C | T315 | /ma/ | 96 | 171.79 | 13.15 | 157 | 213.24 | 56.24 |
| C | T45 | /ma/ | 96 | 218.36 | 19.42 | 191.14 | 270.85 | 79.71 |
| D | T21 | /ma/ | 60 | 197.08 | 10.62 | 181.48 | 224.54 | 43.06 |
| D | T21 | /mi/ | 60 | 205.64 | 14.25 | 184.2 | 231.91 | 47.7 |
| D | T21 | /mɯ/ | 60 | 205.31 | 12.23 | 188.05 | 232.37 | 44.32 |
| D | T241 | /ma/ | 60 | 287.12 | 18.87 | 190.48 | 301.34 | 110.86 |
| D | T241 | /mi/ | 60 | 314.78 | 24.01 | 246.63 | 337.94 | 91.31 |
| D | T241 | /mɯ/ | 60 | 308.26 | 23.59 | 223.84 | 330.75 | 106.92 |
| D | T315 | /ma/ | 60 | 198.96 | 18.92 | 179.03 | 258.52 | 79.48 |
| D | T315 | /mi/ | 60 | 204.23 | 21.28 | 184.45 | 273.2 | 88.75 |
| D | T315 | /mɯ/ | 60 | 203.69 | 18.82 | 184.67 | 264.83 | 80.16 |
| D | T45 | /ma/ | 60 | 238.27 | 13.1 | 227.17 | 280.38 | 53.21 |
| D | T45 | /mi/ | 60 | 257.44 | 14.03 | 244.79 | 296.09 | 51.29 |
| D | T45 | /mɯ/ | 60 | 253.18 | 15.72 | 238.84 | 295.7 | 56.86 |
| E | T21 | /ma/ | 48 | 184.04 | 12.32 | 162 | 210.08 | 48.08 |
| E | T21 | /mi/ | 60 | 183.89 | 13.08 | 166.36 | 210.9 | 44.54 |
| E | T21 | /mɯ/ | 60 | 188.21 | 10.61 | 168.52 | 218.07 | 49.56 |

Table 4. (continued).

| Talker | Tone | Syllable | n | Mean | SD | Min | Max | Range |
|---|---|---|---|---|---|---|---|---|
| E | T241 | /ma/ | 48 | 239.63 | 19.92 | 192.77 | 263.13 | 70.37 |
| E | T241 | /mi/ | 60 | 248.64 | 26.9 | 190.4 | 277.5 | 87.1 |
| E | T241 | /mɯ/ | 60 | 247.83 | 25.78 | 191.05 | 272.71 | 81.66 |
| E | T315 | /ma/ | 48 | 183.41 | 8.47 | 164.37 | 206.79 | 42.42 |
| E | T315 | /mi/ | 60 | 185.58 | 14.18 | 167.89 | 231.91 | 64.01 |
| E | T315 | /mɯ/ | 60 | 182.47 | 9.18 | 171.95 | 213.66 | 41.71 |
| E | T45 | /ma/ | 48 | 213.18 | 11.71 | 200.66 | 248.44 | 47.78 |
| E | T45 | /mi/ | 60 | 225.25 | 15.54 | 209.89 | 273.68 | 63.79 |
| E | T45 | /mɯ/ | 60 | 223.65 | 12.92 | 211.55 | 263.36 | 51.81 |
| F | T21 | /ma/ | 60 | 166.07 | 14.45 | 142 | 197.6 | 55.6 |
| F | T21 | /mi/ | 60 | 170.13 | 16.33 | 138 | 196.23 | 58.23 |
| F | T21 | /mɯ/ | 60 | 170.87 | 15.16 | 145 | 195.47 | 50.47 |
| F | T241 | /ma/ | 60 | 213.7 | 24.15 | 170.21 | 237.51 | 67.3 |
| F | T241 | /mi/ | 60 | 229.88 | 31.23 | 173.57 | 271.37 | 97.8 |
| F | T241 | /mɯ/ | 60 | 221.17 | 29.13 | 149.1 | 255.21 | 106.11 |
| F | T315 | /ma/ | 60 | 172.24 | 17.68 | 142 | 214.91 | 72.91 |
| F | T315 | /mi/ | 60 | 170.39 | 22.33 | 141.41 | 228.58 | 87.17 |
| F | T315 | /mɯ/ | 60 | 176.12 | 21.27 | 146.88 | 226.11 | 79.23 |
| F | T45 | /ma/ | 60 | 197.25 | 11.52 | 187.27 | 229.37 | 42.1 |
| F | T45 | /mi/ | 60 | 205 | 14.31 | 191.62 | 250.28 | 58.65 |
| F | T45 | /mɯ/ | 60 | 210.96 | 15.97 | 189.94 | 255.25 | 65.31 |

A comparison of F0 range for each tone category across syllable types reveals some differences within talker. Figure 12 illustrates mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker A across normalized time. A visual inspection of the contours in Figure 12 reveals differences in the shape of T45, with /ma/ syllables differing from /mi/ and /mɯ/ syllables. There were systematic differences in Talker A's productions of T45 across syllable types. Figure 13 illustrates Talker A's productions of T45 in /ma/, /mi/, and /mɯ/ syllables, with F0 illustrated by dotted lines and intensity illustrated by solid lines. The F0 contours of /mi/ and /mɯ/ syllables differ from /ma/ syllables, along with intensity. Talker A had two methods for producing T45. These methods were slightly interchangeable, but the majority of the productions followed the patterns shown in Figure 13. It is possible that these differences impacted perceptual learning if learners found one method to be more salient than the other method. This difference will be considered in the discussion of the results of the corresponding experiments.

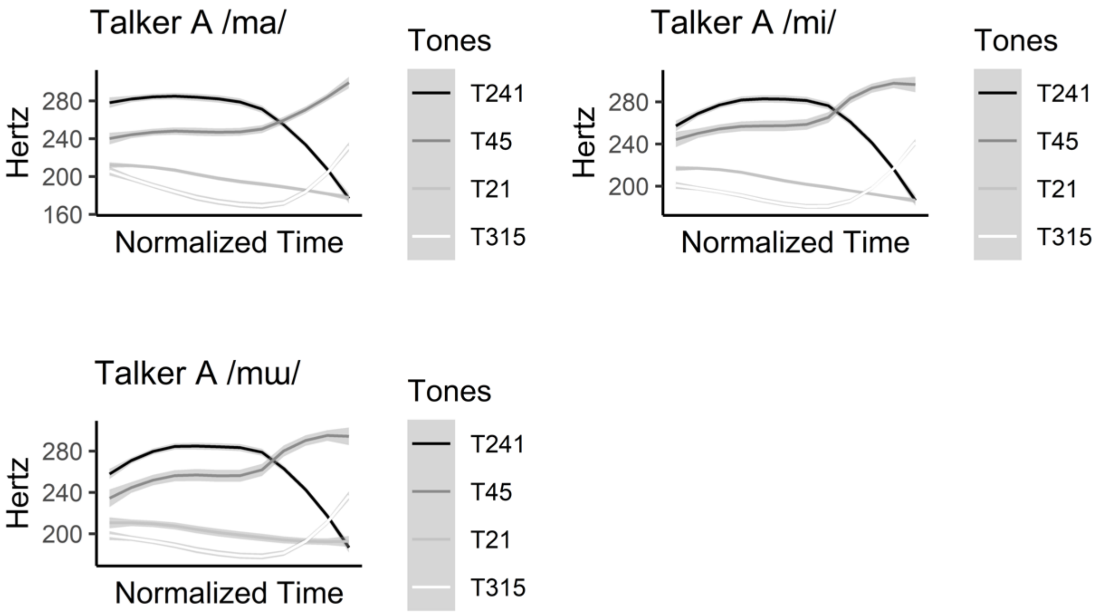

Figure 12. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker A across normalized time.

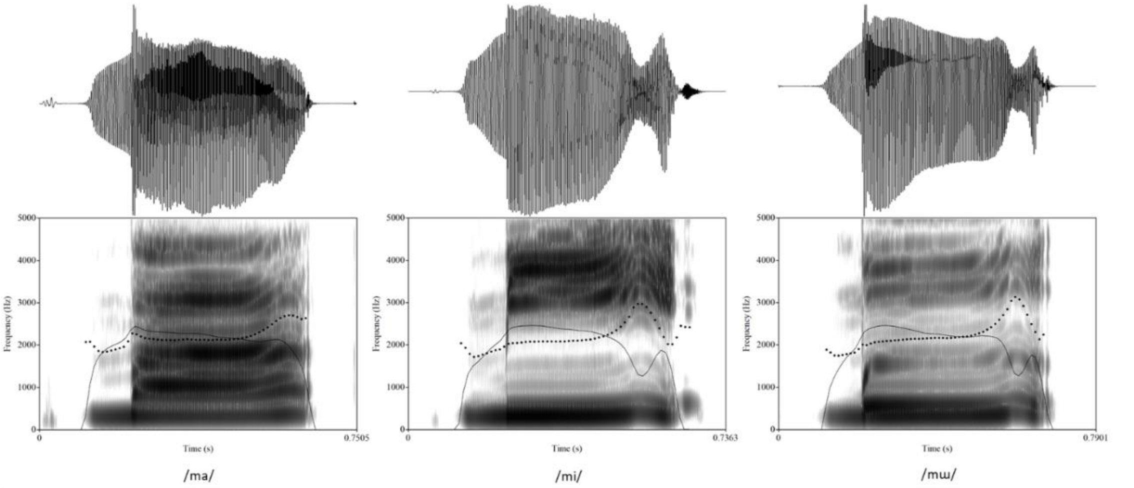

Figure 13. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker A across normalized time.

F0 ranges for each syllable type for Talker A are shown in Figure 14. The four boxplots in each of the three charts represent the distribution of F0 ranges for tokens from each tone of the chart's syllable type, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean F0 range for tokens of the specific tone. The dashed lines in each of the three charts represent the aggregated mean F0 range for all tokens of the chart's syllable type. An obvious difference observed in Figure 14 regards the F0 range for T45 in /ma/ syllables compared with the F0 range of T45 in /mi/ and /mɯ/ syllables, with T45 in /ma/ syllables having a narrower F0 range. This difference is also observable in Table 4. The F0 range difference between syllables is likely due to the difference between production methods shown in Figure 12.

Results from a two-way ANOVA examining F0 range across syllable types for Talker A indicated that, as expected, F0 range differs as a function of tone category [$F(3, 108) = 122.08$, $p < .001$, $\eta^2_p = .77$, $\eta^2_G = .72$]. Overall F0 range did not significantly differ as a function of syllable type [$F(2, 108) = .28$, $p = .76$, $\eta^2_p = .005$, $\eta^2_G = .001$]. However, the interaction between syllable type and tone category was significant [$F(6, 108) = 5.28$, $p < .001$, $\eta^2_p = .22$, $\eta^2_G = .06$].

To further investigate the interaction between tone category and syllable type for Talker A, I used subsets of the data to measure each tone category across syllable types. For each tone category I compared models with and without syllable type to determine if syllable type made a significant contribution to model fit. If F0 range differed as a function of syllable type for a tone category, I performed Bonferroni corrected post-hoc comparisons to further investigate differences across syllable types.

F0 range ~ syllable
F0 range ~ 1

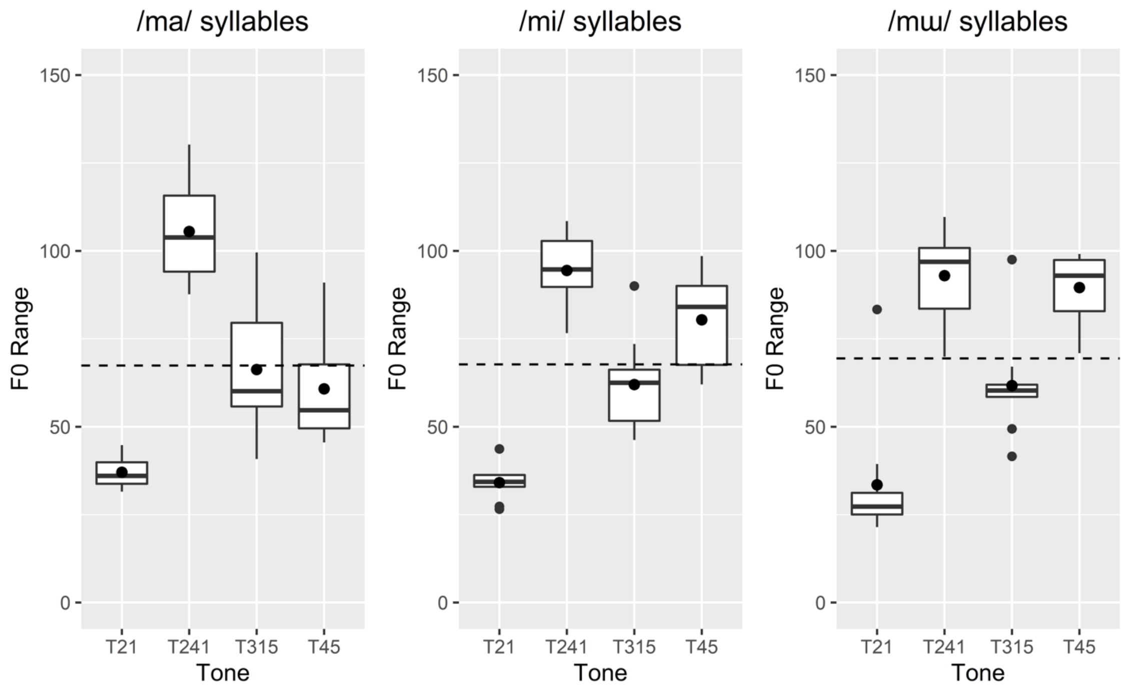

Figure 14. Aggregated F0 range means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker A.

Syllable type did not make a significant contribution to model fit as a predictor for T241 F0 range (*F* (2) = 2.96, *p* = .07), T21 F0 range (*F* (2) = .29, *p* = .75), or T315 F0 range (*F* (2) = .28, *p* = .76). However, syllable type did make a significant contribution to model fit for T45 (*F* (2) = 13.42, *p* < .001). Bonferroni corrected post-hoc comparisons revealed that for T45 for Talker A the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -19.64, SE = 5.68, *t* = -3.46, *p* = .006) and /mɯ/ syllables (β = -28.83, SE = 5.68, *t* = -5.07, *p* < .001). However, for T45 the F0 range of /mi/ syllables did not differ from /mɯ/ syllables (β = -9.19, SE = 5.68, *t* = -1.62, *p* = .35). Figure 14 illustrates these differences in F0 range across syllables for T45, with /ma/ syllables illustrating lower F0 range than /mi/ or /mɯ/ syllables.

Talker B was one of three talkers in the multitalker condition in the second experiment, which only used /ma/ syllables. Therefore, /mi/ and /mɯ/ tokens from Talker B were not used or analyzed and comparisons were not investigated. Figure 15 illustrates Talker B's productions of the four tone categories with the solid lines representing the mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/ syllables across normalized time. F0 range for each tone category for /ma/ syllables for Talker B are shown in Figure 16.

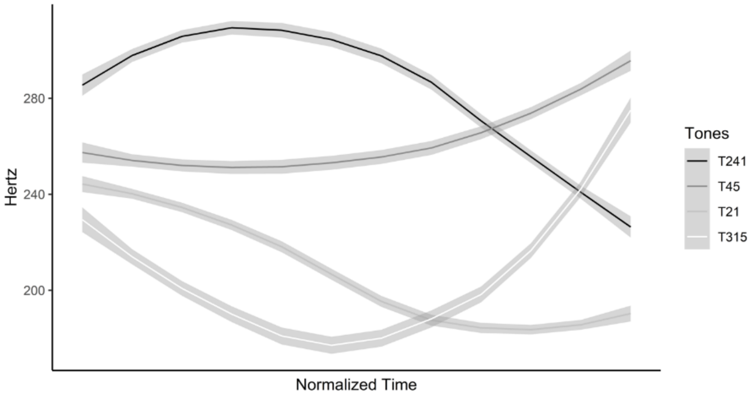

Figure 15. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/ syllables for talker B across normalized time.

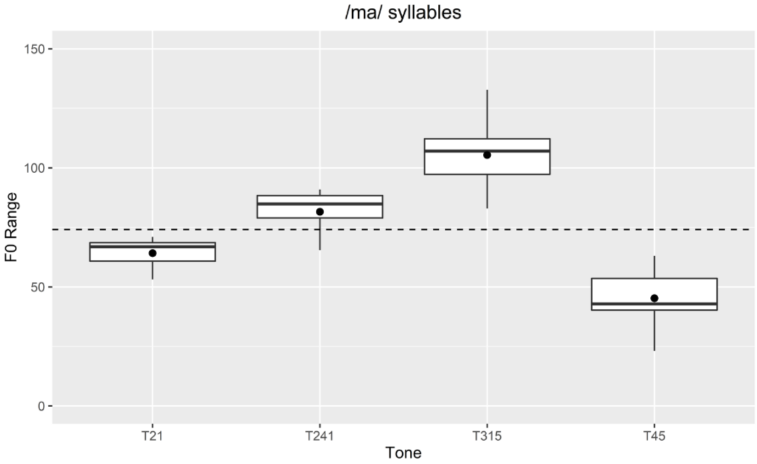

Figure 16. Aggregated F0 range means for /ma/ syllables (dashed lines) and tone category (dots inside the box plots) for Talker B.

Talker C was also one of three talkers in the multitalker condition in the second experiment, which only used /ma/ syllables. So, like Talker B, /mi/ and /mɯ/ tokens from Talker C were not used or analyzed and comparisons were not investigated. Figure 17 illustrates Talker C's productions of the four tone categories with the solid lines representing the mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/ syllables across normalized time. F0 range for each tone category for /ma/ syllables for Talker C are shown in Figure 18.

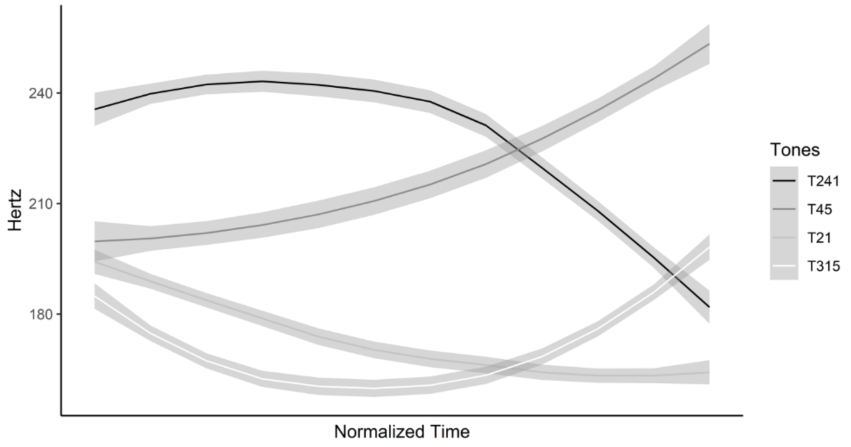

Figure 17. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/ syllables for talker C across normalized time.

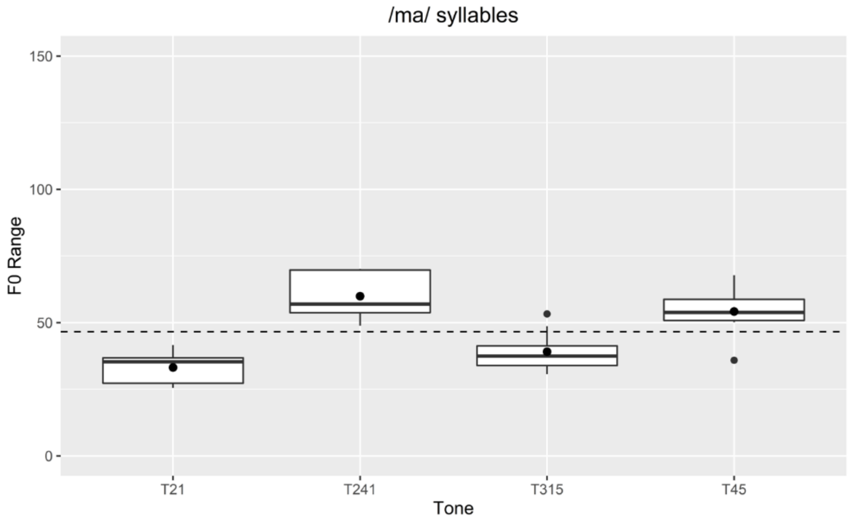

Figure 18. Aggregated F0 range means for /ma/ syllables (dashed lines) and tone category (dots inside the box plots) for Talker C.

Figure 19 illustrates mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker D across normalized time. A visual inspection of the tone contours suggests that Talker D’s productions of the F0 contours across syllable types was consistent.

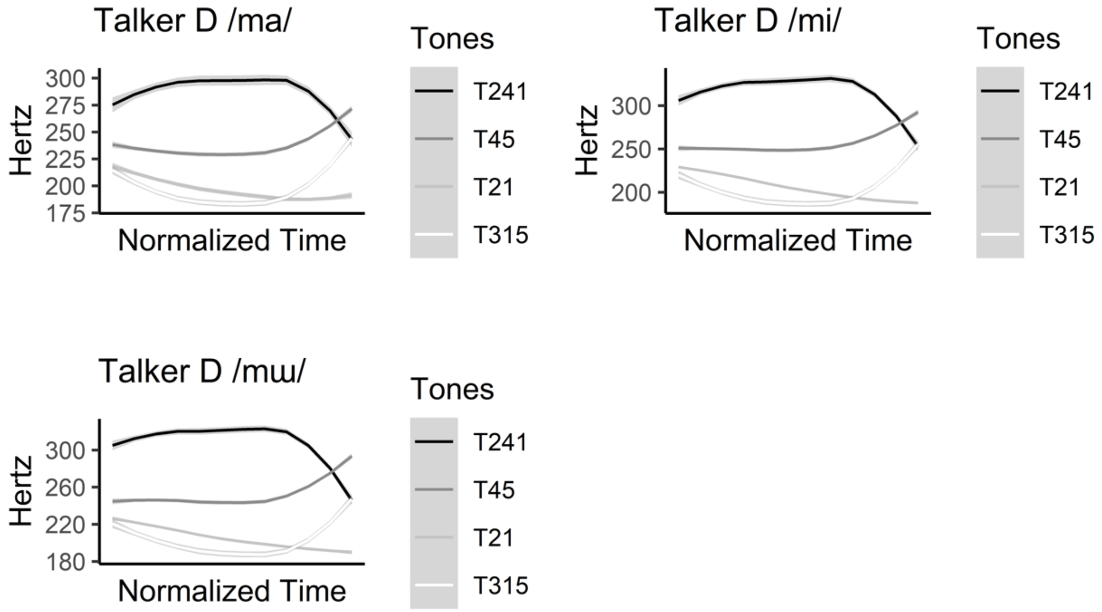

Figure 19. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker D across normalized time.

F0 ranges for each syllable type for Talker D are shown in Figure 20. The four boxplots in each of the three charts represent the distribution of F0 ranges for tokens from each tone of the chart's syllable type, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean F0 range for tokens of the specific tone. The dashed lines in each of the three charts represent the aggregated mean F0 range for all tokens of the chart's syllable type.

Results from a two-way ANOVA examining F0 range across tone category and syllable type for Talker D indicated that, as expected, F0 range differs as a function of tone category [$F(3, 48) = 37.35$, $p < .001$, $\eta^2_p = .70$, $\eta^2_G = .64$]. Also, F0 range significantly differed as a function of syllable type [$F(2, 48) = 3.98$, $p < .001$, $\eta^2_p = .14$, $\eta^2_G = .05$], but the interaction between syllable type and tone category was not significant [$F(6, 48) = 1.09$, $p = .38$, $\eta^2_p = .12$, $\eta^2_G = .04$]. Bonferroni corrected post-hoc comparisons revealed that the F0 range of /ma/ syllables was narrower than /mi/ syllables ($\beta = -9.32$, SE = 3.53, $t = -2.64$, $p = .03$), but /ma/ syllables did not differ from /mɯ/ syllables (β = -7.71, SE = 3.53, *t* = -2.18, *p* = .10) and /mi/ syllables did not differ from /mɯ/ syllables (β = 1.61, SE = 3.53, *t* = .46, *p* = 1).

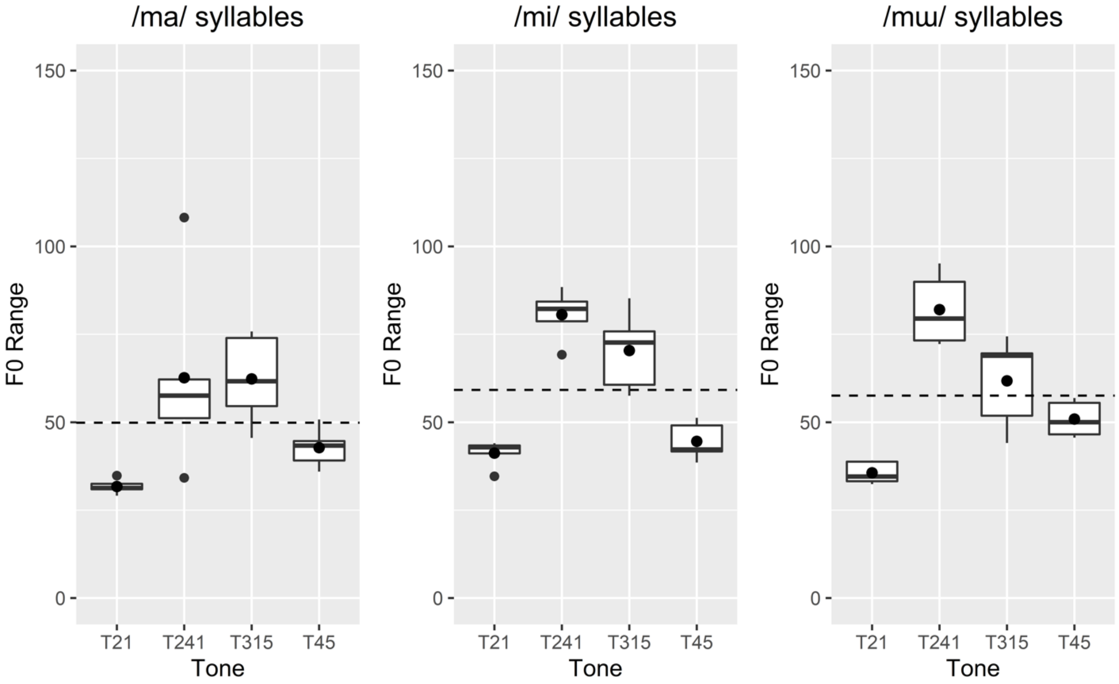

Figure 20. Aggregated F0 range means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker D.

Figure 21 illustrates mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker E across normalized time. A visual inspection of the tone contours suggests that Talker E's productions of the F0 contours across syllable types was consistent.

F0 ranges for each syllable type for Talker E are shown in Figure 22. The four boxplots in each of the three charts represent the distribution of F0 ranges for tokens from each tone of the chart's syllable type, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean F0 range for tokens of the specific tone. The dashed lines in each of the three charts represent the aggregated mean F0 range for all tokens of the chart's syllable type.

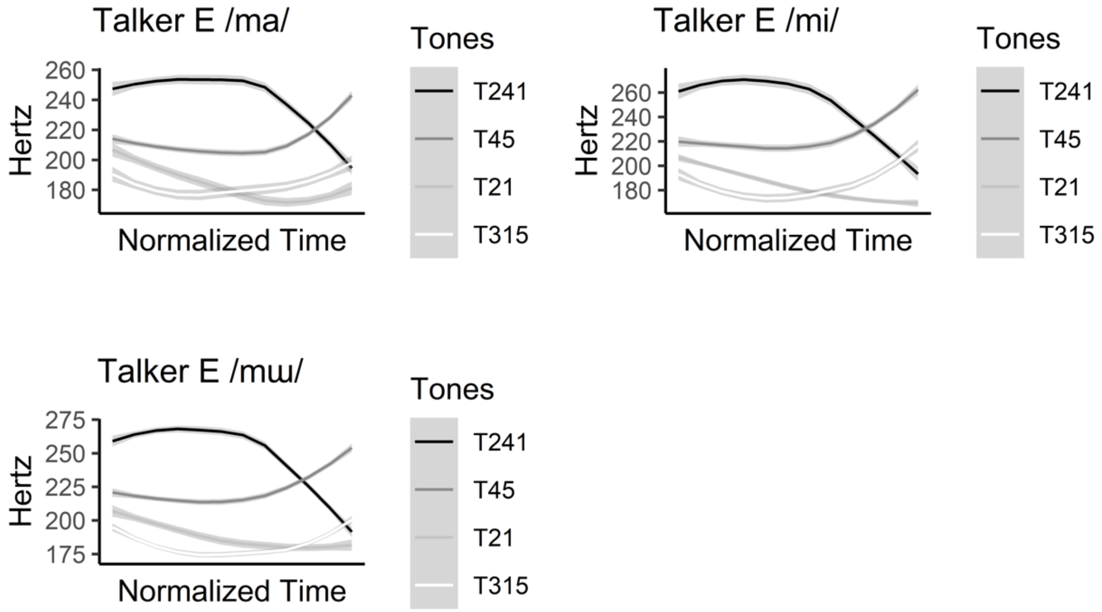

Figure 21. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker E across normalized time.

Results from a two-way ANOVA examining F0 range across tone category and syllable type for Talker E indicated that, as expected, F0 range differs as a function of tone category [$F(3, 44) = 75.98$, $p < .001$, $\eta^2_p = .84$, $\eta^2_G = .69$]. Also, F0 range significantly differed as a function of syllable type [$F(2, 44) = 16.33$, $p < .001$, $\eta^2_p = .43$, $\eta^2_G = .10$], and the interaction between syllable type and tone category was significant [$F(6, 44) = 4.32$, $p = .002$, $\eta^2_p = .37$, $\eta^2_G = .08$]. Bonferroni corrected post-hoc comparisons revealed that the F0 range of /ma/ syllables was narrower than /mi/ syllables ($\beta = -12.31$, SE = 2.31, $t = -5.34$, $p < .001$), but /ma/ syllables did not differ from /mɯ/ syllables ($\beta = -2.93$, SE = 2.31, $t = -1.27$, $p = .63$). Also, the F0 range of /mi/ syllables was wider than /mɯ/ syllables ($\beta = 9.38$, SE = 2.17, $t = 4.32$, $p < .001$).

To further investigate the interaction between tone category and syllable type for Talker F, I used subsets of the data to measure each tone category across syllable types. For each tone category I compared models with and without syllable type to determine if syllable type made a significant contribution to model fit. If F0 range differed as a function of syllable type for a tone category, I performed Bonferroni corrected post-hoc comparisons to further investigate differences across syllable types.

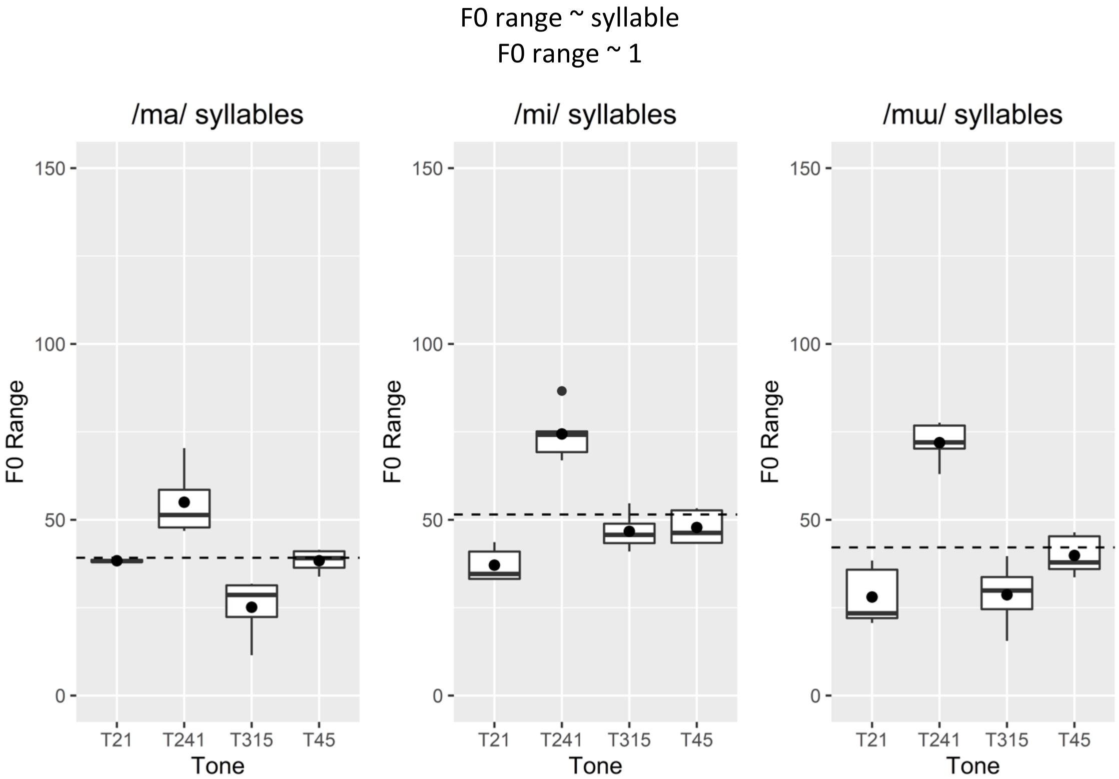

Figure 22. Aggregated F0 range means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker E.

Syllable type made a significant contribution to model fit as a predictor for F0 range for all tones: T241 (*F* (2) = 7.33, *p* = .009), T21 (*F* (2) = 4.38, *p* = .04), T315 (*F* (2) = 9.66, *p* = .004), T45 (*F* (2) = 5.11, *p* = .03). Bonferroni corrected post-hoc comparisons revealed that for T241 for Talker A the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -19.44, SE = 5.43, *t* = -3.58, *p* = .01) and /mɯ/ syllables (β = -16.96, SE = 5.43, *t* = -3.12, *p* = .03). However, the F0 range of /mi/ syllables did not differ from /mɯ/ syllables (β = 2.48, SE = 5.12, *t* = .48, *p* = 1). Although syllable type made a significant contribution to model fit for T21, Bonferroni corrected post-hoc comparisons revealed no differences. The F0 range of /ma/ syllables did not differ from /mi/ syllables (β = 1.24, SE = 3.92, *t* = .32, *p* = 1) or /mɯ/ syllables (β = 10.29, SE = 3.92, *t* = 2.62, *p* = .07), and /mi/ syllables did not differ from /mɯ/ syllables (β = 9.05, SE = 3.70, *t* = 2.45, *p* = .10). For T315 the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -21.59, SE = 5.43, *t* = -3.98, *p* = .007) but did not differ from /mɯ/ syllables (β = -3.55, SE = 5.43, *t* = -.65, *p* = 1). Further, the F0 range of /mi/ syllables was wider than /mɯ/ syllables (β = 18.04, SE = 5.12, *t* = 3.53, *p* = .01). For T45 the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -

9.46, SE = 3.28, *t* = -2.89, *p* = .04) but did not differ from /mɯ/ syllables (β = -1.50, SE = 3.28, *t* = -.46, *p* = 1). Further, /mi/ syllables did not differ from /mɯ/ syllables (β = 7.96, SE = 3.09, *t* = 2.58, *p* = .08). Figure 22 illustrates these differences in F0 range across syllable types.

Figure 23 illustrates mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker F across normalized time. A visual inspection of the tone contours suggests that Talker F’s productions of the F0 contours across syllable types was mostly consistent. The trough of T315 occurs at different durations across syllable types. An inspection of the individual tokens reveals that differences in voice quality are likely the cause of the differences in T315. All /ma/ syllables, as illustrated in Figure 24, were produced with long durations of creaky voice. In Figure 24 F0 is illustrated by a dotted line and intensity is illustrated by a solid line. Creaky voice occurs in the middle of the syllable and disrupts the F0 contour. Creaky voice did occur on most of Talker F’s productions of /mi/ and /mɯ/, but some did not have creaky voice, and when creaky voice occurred the duration was not as long as on /ma/ syllables. Voice quality is a feature that can be used to distinguish tone categories. For example, in the four tone categories used in the current studies, creaky voice occurs on low tones. Participants could use creaky voice, especially creaky voice as prominent as that used by Talker F, as a cue to identify and learn the tone categories.

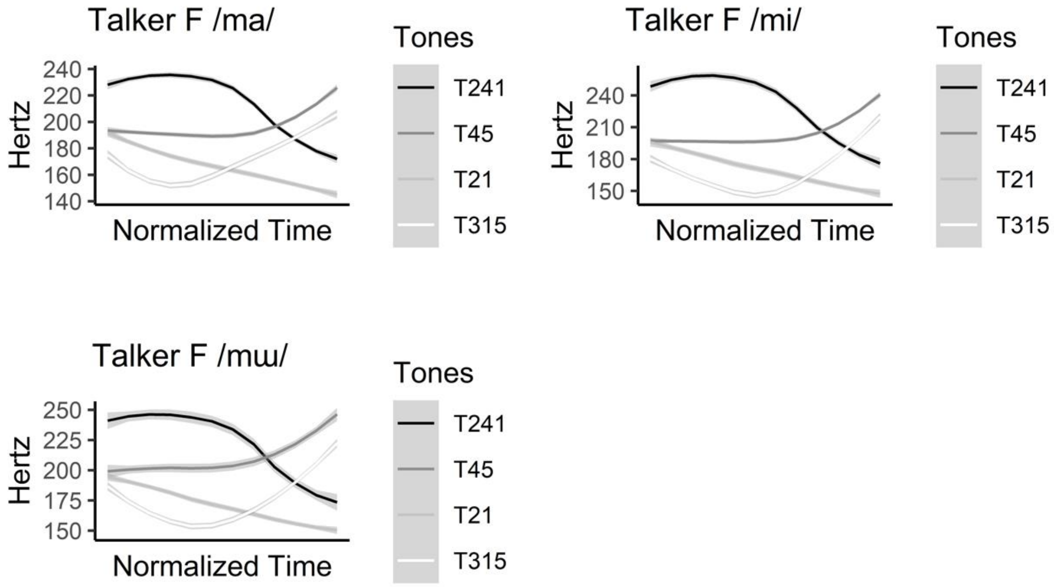

Figure 23. Mean F0 contours and ± 1 standard error of the mean for each tone category for /ma/, /mi/, and /mɯ/ syllables for Talker F across normalized time.

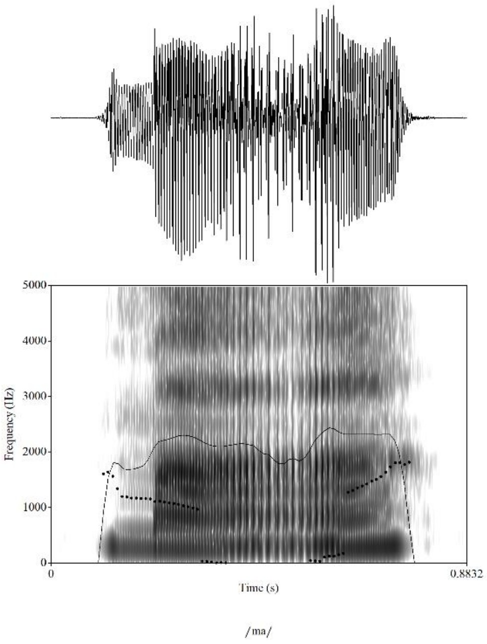

Figure 24. T315 produced by Talker F in a /ma/ syllable illustrating creaky voice occurring on lower F0 ranges.

F0 ranges for each syllable type for Talker F are shown in Figure 25. The four boxplots in each of the three charts represent the distribution of F0 ranges for tokens from each tone of the chart's syllable type, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean F0 range for tokens of the specific tone. The dashed lines in each of the three charts represent the aggregated mean F0 range for all tokens of the chart's syllable type.

Results from a two-way ANOVA examining F0 range across tone category and syllable type for Talker F indicated that, as expected, F0 range differs as a function of tone category [*F*(3, 48) = 73.73, $p < .001$, $\eta^2_p = .82$, $\eta^2_G = .70$]. Also, F0 range significantly differed as a function of syllable type [*F*(2, 48) = 14.74, $p < .001$, $\eta^2_p = .38$, $\eta^2_G = .09$], and the interaction between syllable type and tone category was significant [*F*(6, 48) = 2.66, $p = .026$, $\eta^2_p = .25$, $\eta^2_G = .05$]. Bonferroni corrected post-hoc comparisons revealed that the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -10.69, SE = 2.16, $t = -4.94$, $p < .001$) and /mɯ/ syllables (β = -9.56, SE = 2.16, *t*

= -4.42, *p* < .001). The F0 range of /mɯ/ syllables did not differ from /mɯ/ syllables (β = 1.13, SE = 2.16, *t* = .52, *p* = 1). The narrower F0 range of ma syllables may impact perceptual learning.

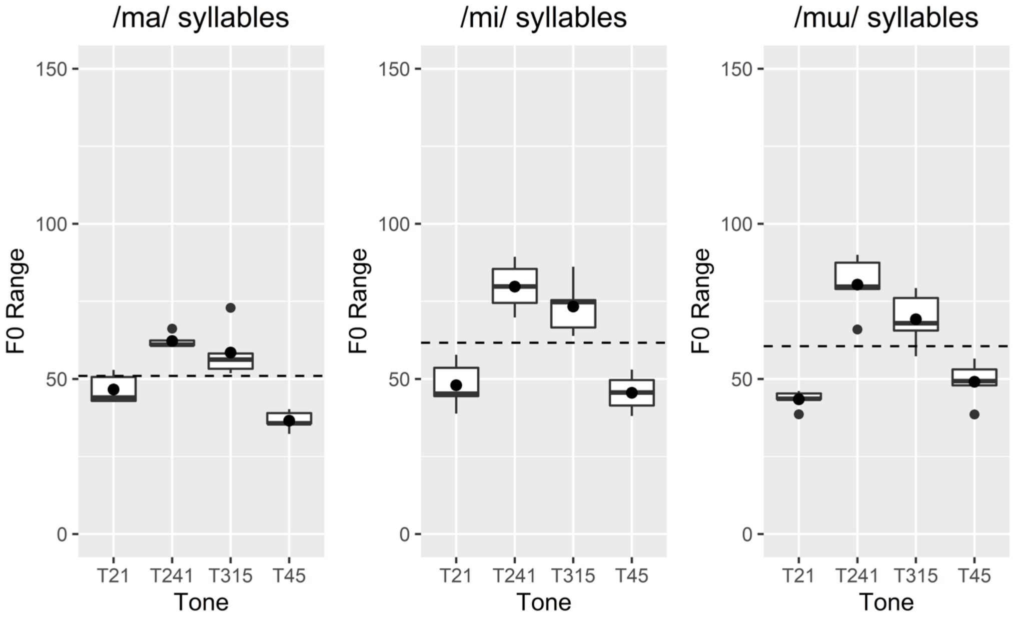

Figure 25. Aggregated F0 range means for each syllable type (dashed lines) and tone category (dots inside the box plots) for Talker F.

To further investigate the interaction between tone category and syllable type for Talker F, I used subsets of the data to measure each tone category across syllable types. For each tone category I compared models with and without syllable type to determine if syllable type made a significant contribution to model fit. If F0 range differed as a function of syllable type for a tone category, I performed Bonferroni corrected post-hoc comparisons to further investigate differences across syllable types.

F0 range ~ syllable
F0 range ~ 1

Syllable type did not make a significant contribution to model fit as a predictor for T21 F0 range (*F* (2) = .92, *p* = .43). However, syllable type did make a significant contribution to model fit for T241 F0 range (*F* (2) = 10.22, *p* = .003), T315 F0 range (*F* (2) = 3.95, *p* = .048), and T45 (*F* (2) = 6.78, *p* = .01). Bonferroni corrected post-hoc comparisons revealed that for T241 for Talker F the F0 range of /ma/ syllables was narrower than /mi/ syllables (β = -17.55, SE = 4.57, *t* = -3.84, *p* = .007) and /mɯ/ syllables (β = -18.18, SE = 4.57, *t* = -3.98, *p* = .006). However, the F0 range of

/mi/ syllables did not differ from /mɯ/ syllables (β = -.64, SE = 4.57, *t* = -.14, *p* = 1). Although syllable type made a significant contribution to model fit for T315, Bonferroni corrected post-hoc comparisons revealed no differences. The F0 range of /ma/ syllables did not differ from /mi/ syllables (β = -14.86, SE = 5.46, *t* = -2.72, *p* = .056) or /mɯ/ syllables (β = -10.75, SE = 5.46, *t* = -1.97, *p* = .22), and /mi/ syllables did not differ from /mɯ/ syllables (β = 4.12, SE = 5.46, *t* = .75, *p* = 1). For T45 the F0 range of /ma/ syllables did not differ from /mi/ syllables (β = -9.00, SE = 3.51, *t* = -2.56, *p* = .07), but /ma/ syllables were narrower than /mɯ/ syllables (β = -12.52, SE = 3.51, *t* = -3.57, *p* = .01). Further, the F0 range of /mi/ syllables did not differ from /mɯ/ syllables (β = -3.53, SE = 3.51, *t* = -1.01, *p* = 1). Figure 25 illustrates these differences in F0 range across syllable types.

Overall, as expected, the F0 ranges of the stimuli used in the current experiments differed across talkers and across tone categories. However, there were overall patterns in F0 ranges across tone categories. Typically, T21 had the narrowest F0 range and T241 had the widest F0 range. T315 had the greatest variation in F0 ranges across talkers. F0 ranges also differed across syllable types, often with /ma/ syllables having a narrower F0 range than /mi/ or /mɯ/ syllables. When comparing individual tone categories across syllable types, only T45 differed, with /ma/ syllables having narrower F0 range than /mi/ or /mɯ/ syllables. There were also some differences in the shapes of the F0 contours. Talker A employed two methods for the production of T45, consistently using one method for /ma/ syllables and another method for /mi/ and /mɯ/ syllables. Talker F displayed differences in the trough of T315 as a result of differences in the amount of creaky voice present across syllables, with /ma/ syllables having longer durations of creaky voice than /mi/ or /mɯ/ syllables. These differences will be considered in the analysis of the results of the corresponding experiments.

# III. TOKEN VARIABILITY

## 3.1 INTRODUCTION

One challenge when listening to an interlocutor is that the interlocutor's productions of a particular sound category can have features that widely vary from production to production. Variability in productions could be affected by the phonotactic environment of the sound category, but even if the phonotactic environment is the same, features are likely to vary, as pointed out in the productions used in the current study in Chapter 2. Therefore, the task of the listener is the task of *categorization*, which refers to the process of identifying which features of the target sound category are salient to the category and which are unimportant so that the sound can be perceived as the intended category.[6] Thus, the learner must develop the ability to separate salient acoustic features from unimportant features so that when they hear novel productions of the target category, they will be able to recognize the sound as the intended category, a process called *generalization*. Generalization has become an important test of categorization. If learners are able to generalize to novel tokens, then they display higher levels of category learning. Therefore, previous research has concluded that exposure during training to a wide range of variability is vital for category development (e.g., Bradlow et al., 1997; Iverson et al., 2005; Jamieson & Morosan, 1989; Lively et al., 1993; Wang et al., 1999). However, the manner in which variability is encountered during training significantly impacts the ability to acquire novel sound categories. In Experiment 1 we investigate within-trial variability and across-trial variability to examine the impact of the temporal distribution of acoustic variability on incidental auditory learning.

### 3.1.1 Incidental learning

The incidental acquisition of novel sound categories is a relatively new area of investigation in the field of speech perception and production. Initial investigations sought to understand factors driving incidental learning using nonspeech auditory categories (Wade & Holt, 2005; Seitz et al., 2010; Lim & Holt, 2011; Vlahou et al., 2012; Emberson et al., 2013; Gabay et al.,

[6] See Section 1.2.2 for a discussion on categorization.

2015; Lim et al., 2019; Roark et al., 2020). Experiment 1 extends the investigation of factors driving incidental learning into natural speech sound categories.

As discussed in Section 1.2, traditionally, studies investigating novel tone category formation require learners to return to the lab over several days or weeks for training sessions to develop behavioral mastery of four novel tone categories (Francis et al., 2008; Chandrasekaran, 2010; Wong Puisan & Lam Ka Yu, 2021). These training sessions typically include explicit instruction regarding the target categories and feedback on performance. The difference between the time course of learning for incidental and explicit learning paradigms is notable and may be attributed to differences in the learning systems engaged by the paradigms (Tricomi et al., 2006; Lim et al., 2013). These differences may also result in more robust learning in incidental paradigms (Wiener et al., 2019). Traditional paradigms and incidental paradigms are somewhat similar but also have several differences. As discussed in section 1.2, incidental learning is not passive learning. There is a feedback mechanic incorporated in the incidental learning paradigm. In the incidental paradigm learning occurs when the participant realizes that the auditory tokens provide clues regarding the location of the visual targets. Then they begin to use those clues to predict where the visual target would appear. On a trial they hear the sounds and are predicting where the target will be when it appears. This provides implicit feedback telling them if they were right or wrong in their prediction. They use that feedback to refine their categorical judgments of the following auditory stimuli. Therefore, one difference is that feedback is delayed in traditional paradigms compared to the feedback received in an incidental paradigm. Thus, Gabay et al. (2015) hypothesized that token variability within trial, due to the close temporal proximity to the feedback mechanism in the incidental paradigm, would result in better categorization and generalization than variability spread across trials. Their results indicated that within trial variability was substantially better for category learning than identical tokens within trial.

### 3.1.2 The impact of within-trial token variability on sound category learning

When a participant hears multiple tokens close together on the same trial, they are able to practice token normalization. That is, they are able to compare tokens to each other and determine what features are similar across tokens. This aids in the extraction of the salient features of the category. If the variability from the unimportant acoustic features is spread out temporally, token normalization is much more difficult. Explicit sound category learning studies often contain a limited number of auditory tokens on each trial. By contrast, incidental learning paradigms typically include multiple auditory tokens on each trial (Wade & Holt, 2005). For example, Gabay et al. (2015) included five auditory tokens on each trial and they hypothesized that the composition of the auditory tokens on each trial might matter for learning. Therefore, they tested one condition that contained identical tokens on each trial and one condition that contained variable tokens on each trial. As mentioned, their findings indicated that participants learn much better from variable tokens on each trial. They concluded that variable tokens within trial temporally places auditory exemplar variability in closer proximity to the mechanic in the paradigm that drives learning. Specifically, variable auditory tokens within each trial allows participants to better refine their categorization of stimuli by aiding in the extraction of salient acoustic features from the various exemplars as they identify the acoustic characteristics that are essential to the specific category. When this process occurs in close proximity to the learning reinforcement mechanic, learning is enhanced.

### 3.1.3 Current experiment

In the current experiment we investigate whether an incidental learning paradigm using natural auditory tokens will result in the formation of novel tone categories and the ability to generalize learning to novel tokens and novel talkers. Further, we determine the impact of acoustic variability within trial on incidental perceptual learning. By investigating within-trial variability and across-trial variability, we examine whether the proximity of the acoustic variability to the visuomotor associations impacts the incidental learning of novel tone categories.

Based on previous research, we expect that participants will be able to acquire four natural novel tone categories in a single session via incidental learning. We also expect that reaction times across blocks during training will get faster if they are learning the categories. Further, we expect that accuracy scores at test will correlate with reaction times during training. We also expect that within-trial variability in the Variable Token Condition will result in greater learning than across-trial variability in the Identical Token Condition.

## 3.2 METHODS

### 3.2.1 Participants

Participants were recruited online on the Prolific online research platform. All participants self-identified as being monolingual English speakers and identified as being native English speakers from America, Canada, the United Kingdom, South Africa, Australia, or New Zealand. Participants that reported significant language learning experience, that reported hearing impairments, or that did not use the right equipment (headphones and an external mouse) were excluded from the study.

In Experiment 1, participants were recruited for two separate conditions. In the Identical Token Condition, 25 participants were recruited (11 female, 14 male). No participants were excluded for not meeting the inclusion criteria in this condition. Participants in this condition spoke a variety of English dialects (14 American, 2 Australian, 5 British, 1 Canadian, 1 Irish, 3 New Zealand).[7] Ages ranged from 18 to 63 with a mean of 34.52 and standard deviation of 13.87.[8]

In the Variable Token Condition, 29 participants were recruited. Four participants were excluded for using the wrong equipment or for hearing impairments, leaving 25 participants (13 female, 11 male, 1 non-binary). Participants in this condition spoke a variety of English dialects (6 American, 2 Australian, 14 British, 1 Canadian, 1 Irish, and 1 NA). Ages ranged from 19 to 56 with a mean of 29.08 and standard deviation of 9.45. All participants were paid for their participation through Prolific.

### 3.2.2 Stimuli

Stimuli used in Experiment 1 were natural tokens recorded from four female native Thai speakers[9]. Figure 26 illustrates the four tone categories as produced by the four talkers in the current study. Talker A stimuli were used during training and on Posttest 1. Talker D, Talker E,

[7] It is not expected that experience with specific English dialects would aid in novel tone category acquisition over other dialects. English dialects do not use F0 information contrastively at the lexical level. Further, experience with other regional languages used in proximity to the specific dialect should not be a factor as participation was limited to those that identified as being monolingual English speakers.

[8] Age is considered as a covariate during analysis and is reported in the results.

[9] The four tone categories, as produced by each talker, are illustrated and characterized in more detail in Chapter 2.

and Talker F stimuli were used for Posttest 2. The contours in Figure 26 represent means extracted from each token produced by the individual talkers who recorded the stimuli for the current experiment. The light grey areas bordering the F0 contours represent ± 1 standard error of the mean.

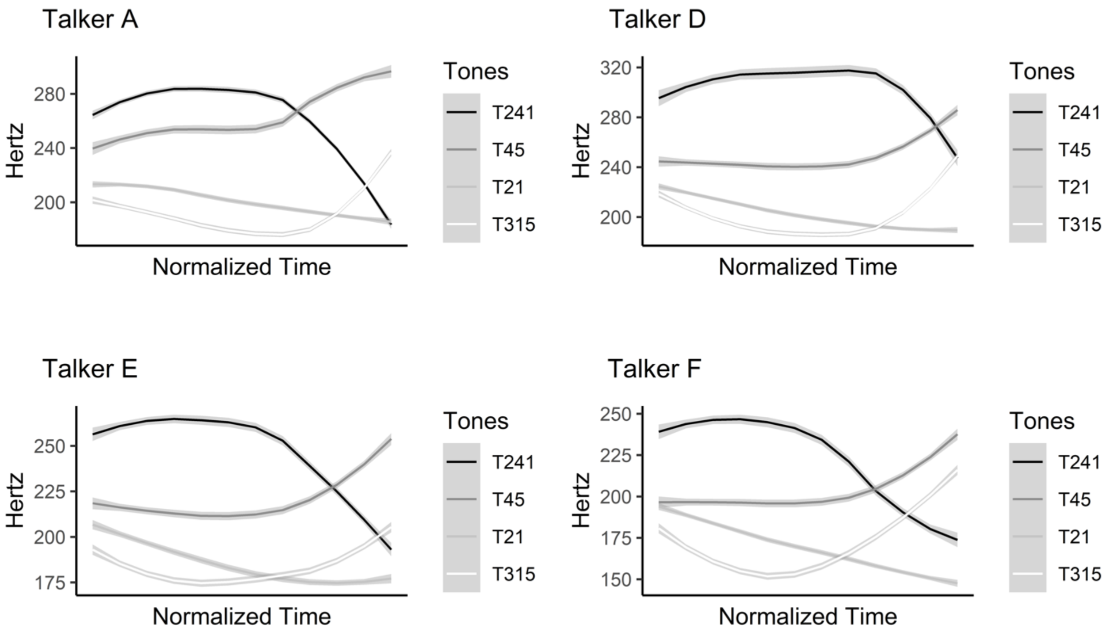

Figure 26. Mean F0 contours and ± 1 standard error of the mean for each tone category for each talker across normalized time.

Tokens from all four categories were produced in the syllable /ma/. Ten exemplars of each category were recorded from all four talkers. Half of the exemplars of each category from Talker A were used for training, and half of the exemplars were used to test generalization of learning to new exemplars on Posttest 1. Five tokens from Talker D and Talker E and four tokens from Talker F[10] were used to test generalization of learning to new speakers on Posttest 2. Following Gabay et al. (2015), auditory stimuli in each trial consisted of five concatenated tokens. In the Identical Token Condition, the five concatenated tokens within trial were identical. In the Variable Token Condition, the five concatenated tokens were randomly selected.

[10] Due to Covid restrictions, which led to talkers recording themselves, some tokens were not usable. In these situations, trials were still comprised of five randomly selected tokens with one token being duplicated.

## 3.3 PROCEDURE

In Experiment 1, two groups of participants were exposed to four novel Thai tone categories through an incidental learning paradigm, which was developed based on previous incidental learning paradigms (Gabay et al., 2015; Lim et al., 2013; Lim & Holt, 2011; Wade & Holt, 2005). As in Gabay et al. (2015), participants received a brief introduction that made sure they were using the right equipment (i.e., an external mouse and headphones) and then introduced the task, but did not include information regarding the target auditory categories. Participants were then trained via the incidental learning paradigm and went through four training blocks. Then the first two posttests were introduced to prepare the participants for the task on the posttests, which differed slightly from the training task. After completing posttests 1 and 2, participants were given instructions for posttest 3, which tested production of the tone categories[11], and participants completed posttest 3. Finally, participants completed a language background questionnaire.

### 3.3.1 Training

Before training began participants received a short introduction to the task. They were told that they would hear a sound repeated several times. After that, they saw four boxes appear and one box would have an X inside it. They were then instructed to use their mouse to click on the X as fast as they could. Then they were to move their mouse to the center target on the screen to start the next trial. Before beginning the training blocks, the participants performed eight practice trials.

The training section of the experiment included four blocks with 48 trials in each block and a thirty second break between each block. Each training block contained the same trials, but the order of the trials across blocks was randomly selected by the experiment. On each trial participants first heard the five auditory stimuli from Talker A. In the Identical Token Condition, these tokens were the same auditory stimulus repeated five times. In the Variable Token Condition, the five stimuli were composed of five different auditory stimuli played in a random order that was compiled before the experiment. Across trials, in the Incidental Token Condition, participants heard six different productions of each tone category randomly selected by the

[11] Posttest 3 elicited productions of the tone categories from the participants for analyses of correlations between production and perceptual learning. However, analyses of the production data will not be included in the present work.

experiment, for a total of twenty-four trials. This random selection was then repeated for a total of forty-eight trials. Across trials, in the Variable Token Condition, participants heard six different concatenations of five randomly selected productions of each tone category, which were selected prior to the experiment. The order of the auditory stimuli across trials was randomly selected by the experiment. This resulted in twenty-four trials, which were repeated once for a total of forty-eight trials. Immediately after the auditory stimuli played in each trial, four boxes appeared on the screen, and one of the boxes had an X in the box, as illustrated in Figure 27.

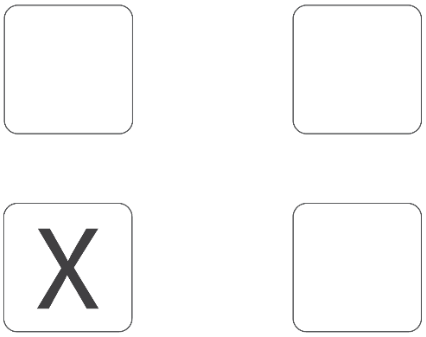

Figure 27. Example of a visual target displayed on a training trial.

Participants had been instructed to click on the X as fast as they could. After clicking on the visual target, the visual stimulus disappeared. Clicking on an empty box did not progress the trial. In this way, the participant was forced to respond correctly. After the visual stimulus disappeared, a visual prompt was displayed in the middle of the screen, as shown in Figure 28.

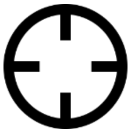

Figure 28. Example of a circle displayed on a training trial prompting the participant to move their cursor back to the middle of the screen.

Participants had been instructed to move their cursor back to the visual prompt in the middle of the screen to advance to the next trial. By arranging the visual target in a 2 x 2 grid, as shown in Figure 27, and having the participants bring the cursor back to the middle of the screen, I was able to track mouse movement, which permits the measurement of the participant's decision space as well as a confusion matrix that investigates which categories sound more similar to the participant[12]. Besides mouse tracking, I also measured reaction time from the initial appearance of the boxes to the time the participant clicked on the X. Initially participants would not be able to use the auditory stimuli to predict the appearance of the X, but as they learned the mapping, they would come to predict where the X would appear, and their reaction times would become faster.

### 3.3.2 Testing

After the training trials, participants received a brief introduction to the test trials. Participants were told that the task would change some. They would hear the sound and boxes would appear, as shown in Figure 29, but an X would not appear. They had to click on the box where they thought the X should appear. After clicking on a box, the trial ended and the next trial began.

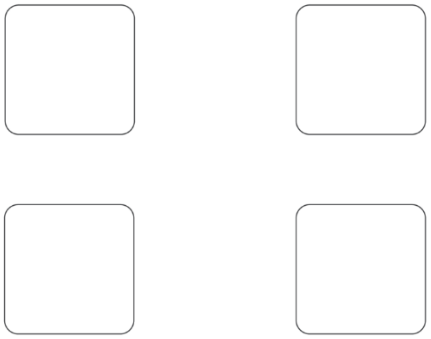

Figure 29. Example of a visual target displayed on a test trial in Posttest 1 and Posttest 2.

#### *3.3.2.1 Posttest 1: Generalization to new tokens*

Posttest 1 measured generalization to new tokens from Talker A. It was composed of thirty-six trials. Like the training blocks, on each trial participants first heard the five auditory stimuli. As in training, in the Identical Token Condition, the five stimuli were composed of the same auditory stimulus repeated five times, and in the Variable Token Condition the five stimuli were composed of five different auditory stimuli played in a random order that was compiled before the experiment. However, the tokens were new tokens that were not used during training.

[12] An analysis of mouse tracking data is not included in the dissertation. Future analyses and description of the current work will analyze and consider mouse tracking data and report results.

Across trials, in the Identical Token Condition, participants heard three different productions of each tone category randomly selected by the experiment, making twelve trials. This random selection was then repeated three times for a total of thirty-six trials. Across trials, in the Variable Token Condition, participants heard three different concatenations of five randomly selected productions of each tone category, which were selected prior to the experiment. The order of the auditory stimuli across trials was randomly selected by the experiment. This resulted in twelve trials, which were repeated three times for a total of thirty-six trials. Accuracy on each trial was measured.

#### *3.3.2.2 Posttest 2: Generalization to new talkers*

Posttest 2 measured generalization to new talkers. Before it began participants were told that they would do the same task but that now they would hear different voices saying the sounds. Everything was the same as Posttest 1 except for the stimuli, which came from Talker D, Talker E, and Talker F. In the Identical Token Condition, where the five sounds within trial were identical repetitions of a single sound, three different productions of each tone category from each talker were used, for a total of thirty-six trials (3 productions X 4 tones X 3 talkers). In the Variable Token Condition, where the five sounds within trial were randomly selected productions, three different concatenations of each tone category from each talker were used for a total of thirty-six trials (3 concatenations X 4 tones X 3 talkers). The order of presentation across trials was randomly selected by the experiment. Accuracy on each trial was measured.

#### *3.3.2.3 Posttest 3: Production of the tone categories*

After the two posttests that tested perceptual learning, posttest 3 tested production of the four tone categories. To accomplish this more explicit instruction was required. Participants were told that each box during the training and Posttest 1 and 2 had a unique pitch pattern associated with it and that now they would be recorded producing the pitch patterns that went with each box. They were told that in posttest 3 the four boxes would appear and one box would have the X in it, as shown in Figure 30. They were to say the box's pitch pattern with 'ma' a single time. Together with the visual target a button with a microphone and a button with a stop signal on it appeared. The participant clicked on the microphone button to begin the recording and then clicked on the stop button to end the recording. The trial automatically ended after the stop button was pressed. Thirty-six trials were conducted[13].

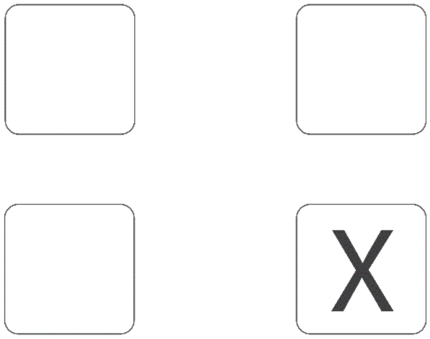

Figure 30. Example of a visual target displayed on a trial from posttest 3.

## 3.4 Results

Category learning was assessed with four measures. During training participants' reaction times were measured to investigate learning across training blocks. Mouse tracking was also used during training to permit the investigation of changes in the participant's decision space over the course of learning[14]. Also, by investigating the deviations towards other choices I measure the perceptual similarity of the categories and determine the time course of the perceptual separation of the analogous categories. During Posttest 1 participants' accuracy scores were measured to test generalization to novel tokens from the same talker. During Posttest 2 participants' accuracy scores were measured to test generalization to novel tokens from novel talkers. During posttest 3 participants' productions were recorded to test correlations between perceptual learning and production accuracy across the experiment's conditions[15].

### 3.4.1 Training reaction times

As in Gabay et al. (2015), the first measure of category learning uses changes in visual target detection time as a metric. Across the four training blocks, the auditory stimuli on each trial correlates with one of the four visual targets that follow the stimuli. For example, T241 always

[13] An analysis of production data is not included in the dissertation. Future analyses and description of the current work will analyze and consider production data and report results.

[14] An analysis of mouse tracking data is not included in the dissertation. Future analyses and description of the current work will analyze and consider mouse tracking data and report results.

[15] An analysis of production data is not included in the dissertation. Future analyses and description of the current work will analyze and consider production data and report results.

occurs with a visual target in the top right quadrant. As participants learn the auditory-to-visual mapping, they begin to use the auditory stimuli to predict the location of the visual target. In this way they become faster at clicking on the visual target. As in Gabay et al. (2015), I expect that if participants are able to use an incidental auditory-to-visual mapping task to learn natural sound categories, then visual target detection times will become faster across training blocks. As discussed, the two conditions in this experiment are designed to test the impact of category exemplar variability on incidental auditory-to-visual category learning. Although both conditions contain the same tokens and therefore the same overall variability, the auditory stimuli in Identical Token Condition only contains identical repetitions of one token within a trial. Therefore, the variability of the exemplars is spread out across trials. In the Variable Token Condition, the auditory stimuli contain five different tokens within trial. Therefore, in the Variable Token Condition the exposure to exemplar variability occurs in close proximity to the visual detection task. By comparing these two conditions, I test the impact of proximity of exemplar variability to the visuomotor associations on natural sound category learning. Proximity of exemplar variability to the visuomotor associations is operationalized in the current study as the temporal distance that variable productions are from the visual detection task, and that temporal distance is either within a trial, as in the Variable Token Condition, or across trials, as in the Identical Token Condition. It is expected, following results from Gabay et al. (2015), that high variability in closer proximity to the visuomotor associations will result in more robust learning. So, although I expect that both conditions will result in faster visual target detection times across training blocks, I predict that reaction times will be faster in the Variable Token Condition.

It is important to note that for Experiment 1, and for all experiments in the current work, the study was conducted online rather than in a lab. In a lab there is control over the environment, which results in control over the computer interface as well as peripherals. Conducting the experiment online permits participants to have multiple screens or devices available for working on other tasks while doing the experiment. There may also be other distractors such as other people present or food and drink available. These external factors may result in differences in reaction times across training blocks that might not be experienced in a controlled lab setting, thereby potentially adding noise to the present data. Therefore, it is expected that results, especially from measures that correlate with each other, would potentially be even stronger in a controlled environment.

### *3.4.1.1 Analysis*

Visual target detection times were measured from the end of the auditory stimuli to the time the participant clicked on the visual target. Reaction times greater than 1,500 ms were excluded from analyses. For each condition, I compare reaction times across training blocks by comparing a full model and a reduced model without training block. I then conduct contrast coded linear mixed-effects regressions to compare each training block to the subsequent training block to examine changes in reaction times from block to block. Also, as differences in age can affect learning and hearing ability (Kiessling et al., 2003; Clinard et al., 2010), I conduct model comparisons to examine age as a fixed effect. Finally, I compare reaction times across training blocks across the two conditions by comparing a full model with an interaction between condition and training block and a reduced model without an interaction.

### *3.4.1.2 Reaction Times*

Results indicated that participants in both conditions became faster across training blocks. Figure 31 illustrates log-transformed reaction times across training blocks for the Identical Token Condition, where participants heard identical tokens within trial. The four boxplots in each of the three charts represent the distribution of reaction times for each block, with the solid line in the middle of each box representing the median, the bottom and top of the box representing the first and third quartiles, and the whiskers representing the furthest value at no more than 1.5 times the interquartile range. The dots in the boxes represent the mean reaction time for the specific block, illustrating that reaction times in the Identical Token Condition become faster across blocks.

To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in The Identical Token Condition ($X^2$ (3) = 22.43, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

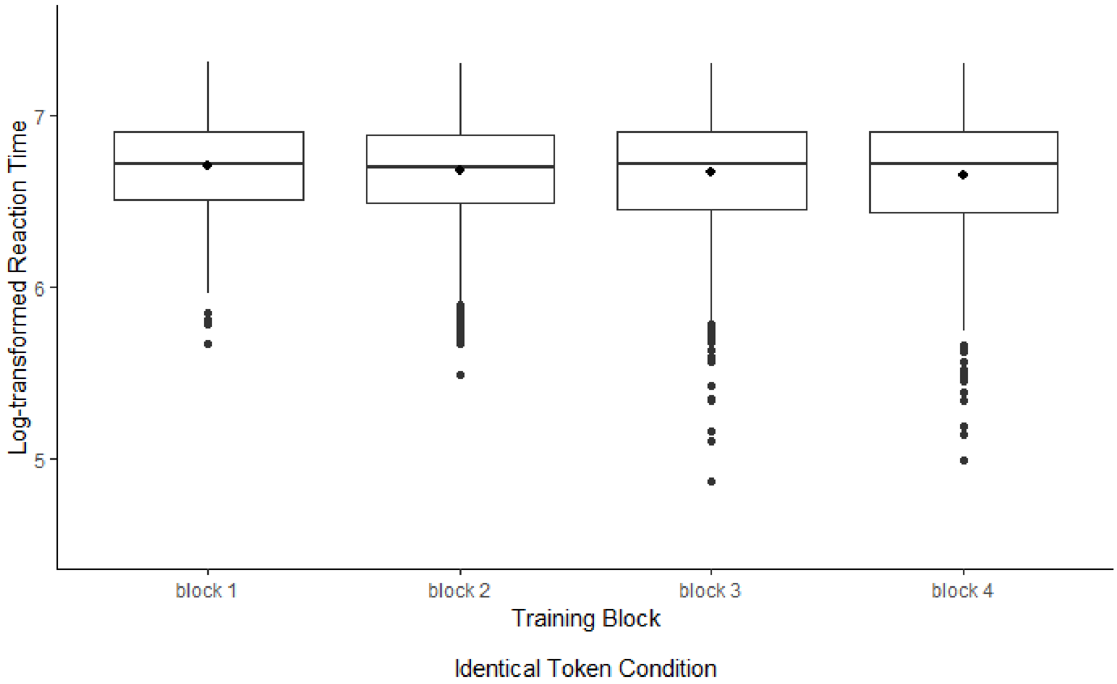

Figure 31. Log-transformed reaction times across training blocks in the Identical Token Condition.

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.71, *SD* = .27) were significantly slower than block 2 (*M* = 6.68, *SD* = .30; β = -.025, *t* = -2.58, p < .01), reaction times in block 2 did not differ from block 3 (*M* = 6.67, *SD* = .34; β = -.003, *t* = -.27, p = .79), and reaction times in block 3 did not differ from block 4 (*M* = 6.66, *SD* = .35; β = -.018, *t* = 2.37, p = .07).

To test whether reaction times differed as a function of age, I compared models with and without age, controlling for training block, and results indicated that reaction time significantly differed as a function of age in the Identical Token Condition ($X^2$ (1) = 5.47, *p* = .019).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ training_block + (1|participant)

Figure 32 illustrates log-transformed reaction times as a function of age. Mean reaction times across blocks for each participant are illustrated as dots with error bars illustrating 95% confidence intervals. If participants are learning the categories, quantified as faster reaction times across training blocks, then darker blocks will be lower on the y axis in Figure 32 and lighter blocks will be higher. This is evident in the youngest participant, who displayed learning and had the fastest reaction times, which occurred in block 3 and block 4. However, the oldest participant also displayed learning, but their reaction times were not as fast as the younger participants. So, although faster reaction times display learning within participant, overall, reaction times differ as a function of age in the Identical Token Condition. Also, the linear regression lines illustrate the point at which participants are learning the categories. For younger participants, block 1 is slower but block 2, block 3, and block 4 do not differ, indicating that category acquisition is occurring around block 2. However, for older participants, this is not the case. Block 1 does not differ from block 2. Block 3 begins to differ, and block 4 is faster, indicating that category learning is occurring later for older participants, around block 3 or block 4.

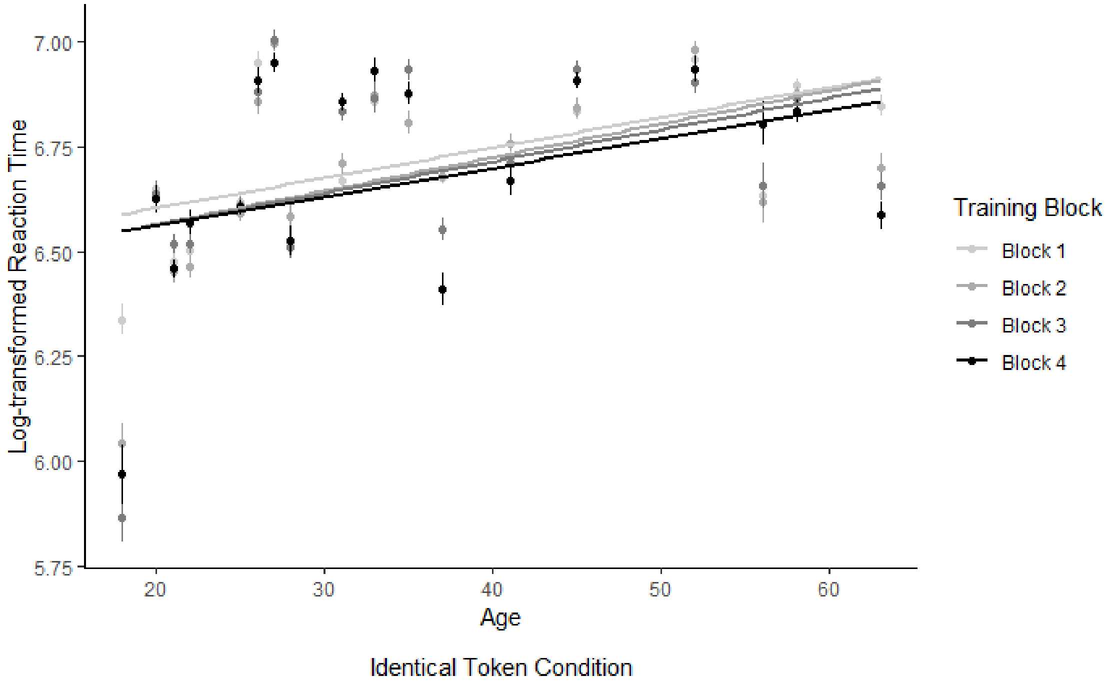

Figure 32. Log-transformed reaction times across age in the Identical Token Condition.

Figure 33 illustrates log-transformed reaction times across training blocks for the Variable Token Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block. As in the Identical Token Condition, reaction times in the Variable Token Condition become faster across blocks.

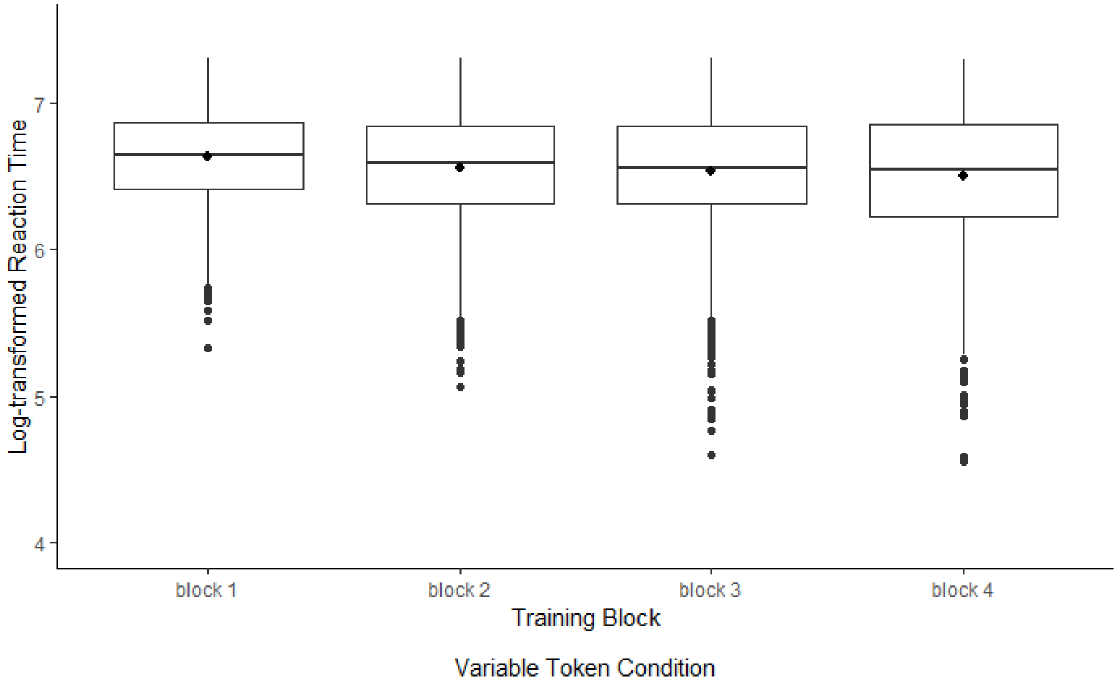

Figure 33. Log-transformed reaction times across training blocks in the Variable Token Condition.

To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the Variable Token Condition ($X^2$ (3) = 114.05, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.63, *SD* = .31) were significantly slower than block 2 (*M* = 6.56, *SD* = .39; β = -.065, *t* = -5.55, p < .001), reaction times in block 2 did not differ from block 3 (*M* = 6.54, *SD* = .42; β = -.022, *t* = -1.89, p = .06), and reaction times in block 3 were significantly slower than block 4 (*M* = 6.51, *SD* = .47; β = -.012, *t* = -2.98, p = .003).

To test whether reaction times differed as a function of age, I compared models with and without age, controlling for training block, and results indicated that reaction time did not significantly differ as a function of age in the Variable Token Condition ($X^2$ (1) = 1.55, *p* = .21).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ training_block + (1|participant)

Figure 34 illustrates log-transformed reaction times as a function of age in the Variable Token Condition. In the Variable Token Condition, few participants over forty and having none of those participants exhibit learning led to results being uninformative regarding the time course of learning across age groups.

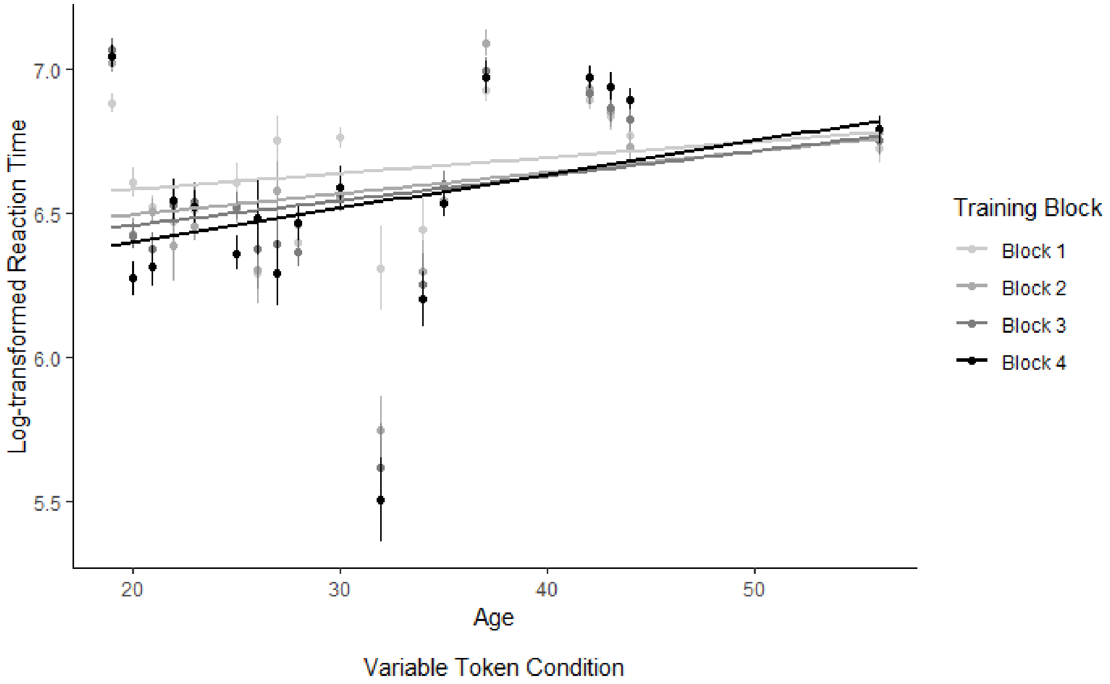

Figure 34. Log-transformed reaction times across training blocks in the Variable Token Condition.

Finally, I compared reaction times across the two conditions. As mentioned, it was expected that learning would be more robust in the Variable Token Condition, where tokens within trial were variable. Figure 35 illustrates mean reaction times across training blocks for the Identical Token Condition and the Variable Token Condition with whiskers illustrating 95% confidence intervals. Table 5 provides the mean and standard deviation of response times for both conditions.

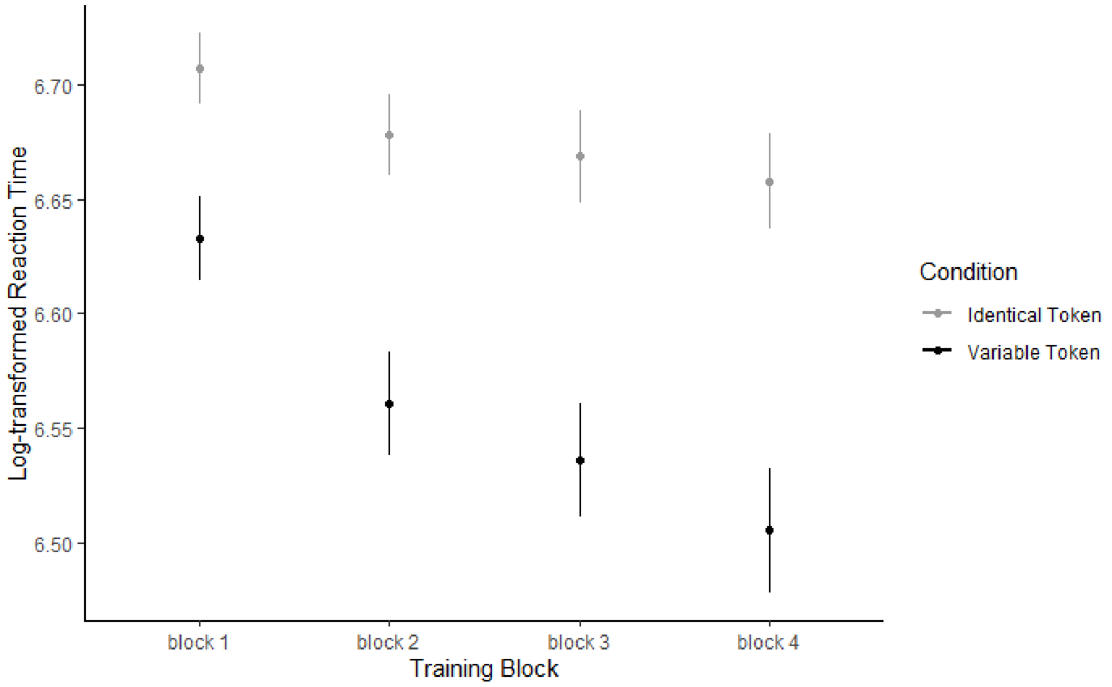

Figure 35. Log-transformed mean reaction times across training blocks for the Identical Token Condition and the Variable Token Condition. Error bars represent 95% confidence intervals.

Table 5. Summary statistics for reaction times for the Identical Token Condition with identical within trial tokens and the Variable Token Condition with variable within trial tokens

| Condition | Block 1 (mean, SD) | Block 2 (mean, SD) | Block 3 (mean, SD) | Block 4 (mean, SD) |
|---|---|---|---|---|
| Identical Token | 6.71, .27 | 6.68, .30 | 6.67, .34 | 6.66, .35 |
| Variable Token | 6.63, .31 | 6.56, .39 | 6.54, .42 | 6.51, .47 |

To test whether reaction times differed across conditions, I compared models with and without an interaction between condition and training block. Results indicated that reaction time differs across training blocks as a function of condition ($X^2$ (3) = 28.63, *p* < .001).

reaction_time ~ condition * training_block + age + (1|participant)
reaction_time ~ condition + training_block + age + (1|participant)

By comparing reaction times across training blocks as a function of condition, I tested the impact of proximity of exemplar variability to the visuomotor associations on natural sound category learning. In the Identical Token Condition, stimuli within trial contained identical tokens and therefore less variability immediately before the visuomotor associations than the Variable Token Condition, which contained variable tokens within trial. Thus, the greater variability of tokens immediately before the visual detection task in the incidental learning paradigm resulted in greater learning across training blocks. So, as expected, both conditions resulted in faster visual target detection times across training blocks, but reaction times were faster in the Variable Token Condition.

### 3.4.2 Generalization to new tokens and talkers

Posttest 1 tested participants' ability to generalize to new tokens from the same talker, and Posttest 2 tested generalization to new talkers. Generalization is the ability to use past learning in present situations that are similar (e.g., Kruschke, 2005). If participants learned the four tone categories during training, then it is expected that they will be able to accurately identify the categories in novel tokens from the same talker. It is also expected that they will be able to identify the categories in novel tokens from novel talkers but with less accuracy due to greater variance in the signal as a result of multiple talkers. The structure of both posttests is identical and both measure identification accuracy of the target tone category. If participants have learned the categories they should be able to accurately identify in which box the visual target should have appeared based solely on hearing the auditory stimuli, and therefore, their accuracy scores will be higher. As in the reaction time metric from the training blocks, it is expected that the closer proximity of exemplar variability to the visuomotor associations in the Variable Token Condition will result in more robust category learning, which will be evident from accuracy scores on Posttest 1 and Posttest 2.

#### *3.4.2.1 Analysis*

Accuracy scores for both conditions were measured on Posttest 1 and Posttest 2. For each condition, I compare accuracy scores on both posttests to chance using one sample t-tests. To test whether accuracy scores differ as a function of condition, I conduct model comparisons with and without condition for each posttest. To test whether there is a correlation between the learning measures, I conduct correlation tests between reaction times during training and accuracy scores at test for each condition. Finally, I conduct model comparisons to examine age as a fixed effect for both conditions on Posttest 1 and Posttest 2.

#### *3.4.2.2 Accuracy*

Figure 36 illustrates mean proportion correct scores with 95% confidence intervals for the Identical Token Condition and the Variable Token Condition on Posttest 1 and Posttest 2. The figure suggests that participants in both conditions accurately identified the target categories above chance on Posttest 1 and on Posttest 2, and that participants in the Variable Token Condition may have performed better on Posttest 1 than participants in the Identical Token Condition.

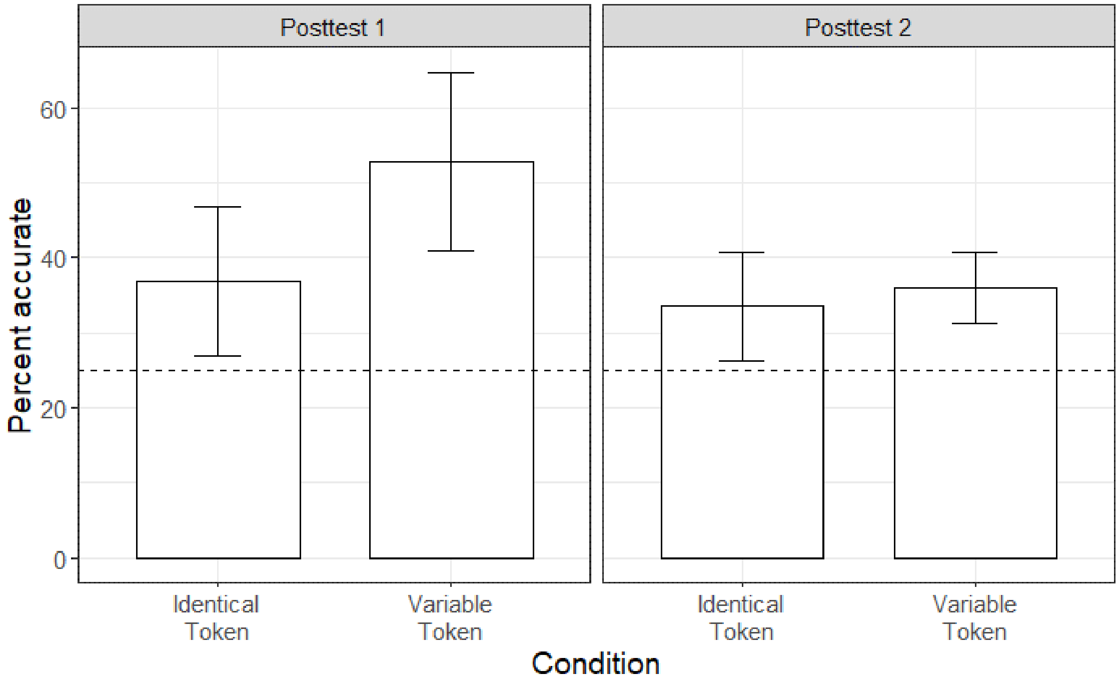

Figure 36. Mean proportion correct for the Identical Token Condition and the Variable Token Condition on Posttest 1 and Posttest 2. Error bars represent 95% confidence intervals. The dashed line represents chance at 25%.

To test whether accuracy scores differed from chance, I examined accuracy scores within condition on Posttest 1 and Posttest 2. In the Identical Token Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24)$ = 2.46, $p$ = .01, ($M$ = 36.83, $SE$ = 4.80) and on Posttest 2, $t(24)$ = 2.46, $p$ = .01, ($M$ = 33.56, $SE$ = 3.48). Also, in the Variable Token Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24)$ = 4.80, $p$ < .001, ($M$ = 52.83, $SE$ = 5.80) and on Posttest 2 , $V$ = 247, $p$ < .001, ($Mdn$ = 34.72)[16].

[16] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 2 ($W$ = .88, $p$ = .008).

To test whether accuracy scores differed across conditions on Posttest 1 and Posttest 2, I compared models with and without condition for each posttest. Results indicated that accuracy scores on Posttest 1 differ as a function of condition ($X^2$ (1) = 4.42, *p* = .035). However, accuracy scores on Posttest 2 did not differ as a function of condition ($X^2$ (1) = .47, *p* = .49).

accuracy ~ condition + age + (1|participant)
accuracy ~ age + (1|participant)

Overall, participants in both conditions accurately identified the target categories above chance on Posttest 1 and on Posttest 2, indicating that both conditions resulted in learning and that learning generalized to novel tokens on Posttest 1 and novel talkers on Posttest 2. A comparison of conditions on Posttest 1 indicated that participants in the Variable Token Condition, more accurately identified the target categories than participants in the Identical Token Condition, indicating that high variability within trial led to more robust generalization to novel tokens than identical tokens within trial. However, the benefit from high variability within trial did not result in more robust generalization to novel talkers over and above exposure to identical tokens within trial.

During training, greater learning was measured through reaction times becoming faster across training blocks. At test, greater learning was measured through higher accuracy scores. It was expected that faster reaction times at the end of training would correlate with higher accuracy scores at test. Figure 37 illustrates the correlation between reaction times on block 4 and accuracy scores on Posttest 1, suggesting a relationship between the two measures.

Spearman's rho correlation coefficient[17] was used to assess the relationship between reaction times on training block 4 and accuracy scores on Posttest 1. The relationship between the two measures was significant in the Identical Token Condition (*r* = -.49, *p* = .01), and the Variable Token Condition (*r* = -.69, *p* < .001). The correlation between the two measures across conditions suggests that faster reaction times in training relates to better accuracy on the generalization test and that both measures reliably assess category learning.

[17] A Shapiro-Wilk normality test indicated the data for Condition 1 were not normally distributed (*W* = .86, *p* = .002; *W* = .90, *p* = .02). Therefore, we conducted the non-parametric Spearman's test for Condition 1. Although the data for Condition 2 were normally distributed (*W* = .93, *p* = .08; *W* = .96, *p* = .37), Spearman's test was used for consistency. Pearson's correlation coefficient was also significant for Condition 2 (*r*(23) = -.67, *p* < .001).

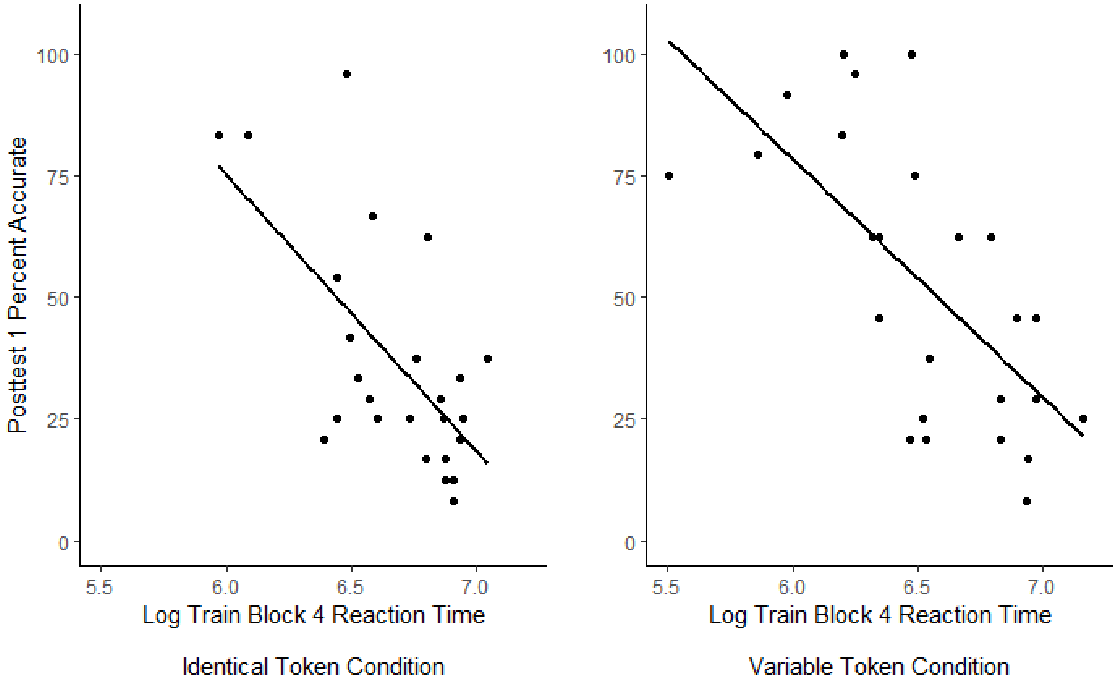

Figure 37. Relationship between two measures assessing category learning across conditions with log transformed reaction times on training block 4 on the x axis and accuracy scores on Posttest 1 on the y axis.

To test whether accuracy scores at test differed as a function of age, I compared models with and without age, and results indicated that accuracy scores did not significantly differ as a function of age in the Identical Token Condition on Posttest 1 ($X^2$ (1) = .61, *p* = .44) or on Posttest 2 ($X^2$ (1) = 1.80, *p* = .18). Further, accuracy scores did not significantly differ as a function of age in the Variable Token Condition on Posttest 1 ($X^2$ (1) = .44, *p* = .51) or on Posttest 2 ($X^2$ (1) = 3.62, *p* = .057).

accuracy ~ age + (1|participant)
accuracy ~ (1|participant)

Figure 38 illustrates accuracy scores on Posttest 1 and Posttest 2 as a function of age across conditions. The model comparison demonstrated that accuracy scores did not differ as a function of age. However, Figure 38 suggests the possibility of different trends in accuracy scores across age groups, with accuracy scores disproportionately impacted by variability within trial. It may be that younger learners benefit more from higher variability within trial than older learners.

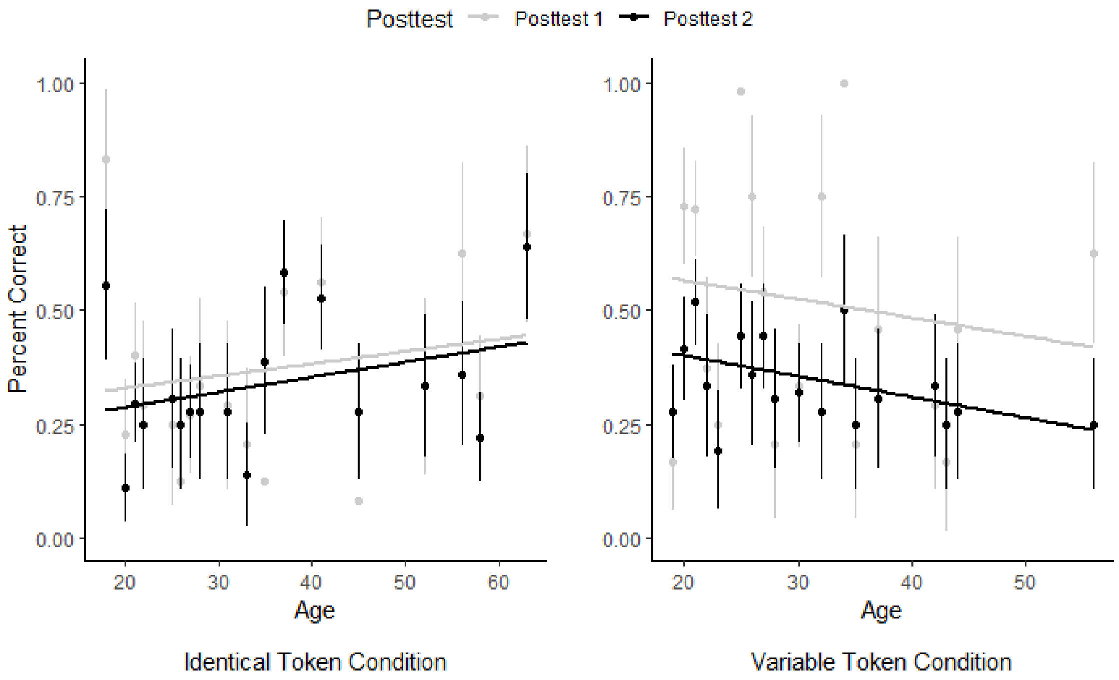

Figure 38. Accuracy scores on Posttest 1 and Posttest 2 across age in the Identical Token Condition and the Variable Token Condition.

## 3.5 Discussion

In Experiment 1 we investigated whether an incidental learning paradigm using naturally produced auditory tokens would result in the formation of novel tone categories and the ability to generalize learning to novel tokens and to novel talkers. We also examined the impact of acoustic variability within trial on incidental perceptual learning. By investigating within trial variability, we examine the impact of the temporal distribution of acoustic variability on incidental auditory learning. Results indicated that participants were successful in using the incidental paradigm to develop four novel tone categories in a single session. Results also indicated that high variability of tokens within trial resulted in more robust learning than identical tokens within trial. Below, we describe the implications of these results in more detail.

### 3.5.1 Incidental learning with natural tokens

The present study extends the investigation of factors driving incidental learning into natural speech sound categories and finds that adults with no prior experience with the target tone categories can use an incidental learning paradigm with natural tokens to form four novel tone categories after 30 minutes of training with up to 100% accuracy. Participants did not achieve this success due to experience. All participants were monolingual English speakers with little or no experience learning another language and had no experience learning the tone categories. Further, participants did not succeed because of age. Both younger and older participants demonstrated substantial learning. Also, participants did not succeed because the task of category learning was easy. Learning to accurately perceive tone categories in other languages is known to be very difficult for native English speakers (Ke & Reed, 1995; Sun, 1998; Hao, 2012; Hao, 2018). Further, participants did not have any particular motivation to acquire the target categories. Participants were unaware of the categories during training, had not made efforts in their lives to acquire other languages, and several participants expressed that the learning paradigm was not particularly engaging. It is most likely that participants learned the four tone categories despite a lack of motivation, instead of because of a surplus of motivation. Indeed, it could be that increasing motivation and engagement in the task could further improve accuracy. In this task participants were only clicking on an X in one of four boxes on the screen, which is not particularly interesting. Some incidental category learning studies increase engagement in the task by embedding the incidental learning paradigm in a video game (Wade & Holt, 2005; Wiener et al., 2019). In these experiments, participants respond during the task by shooting aliens that appear on the screen. Auditory tokens predict the location and type of alien that appears. If participants are able to learn the audio-to-visual mapping they are rewarded by being able to better keep up with the pace of the game as the speed of the ships increases and by being able to maximize the limited range of their weapon (Wiener et al., 2019). It may be that novel tone learning can be increased beyond the results of the current study by increasing engagement through similar mechanics.

The results from Experiment 1 suggest that adults attempting to learn the tone categories in tonal languages do not fail because they are too old or are unmotivated or lack experience. Rather, it may be that the learning methodologies typically employed by adults do not facilitate the formation of novel tone categories. Adults may have greater success learning novel tone categories through an incidental paradigm such as the one used in the current study. Behavioral studies examining novel tone category formation across learning paradigms (e.g., explicit, incidental, and passive) could address this question by training participants in each paradigm to the point of behavioral mastery and measuring the time course of learning, as well as retention after a set period of time past behavioral mastery. Alternatively, learning could be measured across paradigms by examining the development of sensory plasticity, which is measured through the frequency-following response, a neurophonic potential encoding acoustic details along the early auditory pathway (see Reetzke et al., 2018). A study could measure the time course of the development of sensory plasticity across learning paradigms and investigate differences in retention after a set period of time.

### 3.5.2 Within trial variability

Experiment 1 measured learning through two measures: change in reaction times across training blocks and posttest accuracy. These measures were correlated, with the training reaction times predicting accuracy on posttest 1, which measured generalization to novel tokens. Training reaction times and posttest accuracy indicated that participants in both the Identical Token Condition and in the Variable Token Condition learned the four target tone categories. However, as expected from results from Gabay et al. (2015), results in the two conditions were not the same. Reaction times across training blocks in the Variable Token Condition were faster than reaction times in the Identical Token Condition. Further, accuracy scores on Posttest 1 were higher in the Variable Token Condition than in the Identical Token Condition. These results replicate results from Gabay et al. (2015), which found that variable tokens within trial resulted in greater learning than identical tokens within trial and extend these results to naturally produced tone categories.

Gabay et al. (2015) hypothesized that the difference in learning occurs due to the proximity of the token variability to the visuomotor association. In the Variable Token Condition, token variability occurs within trial. In the Identical Token Condition, token variability occurs across trials. The visuomotor association occurs at the end of each trial. Therefore, token variability in the Variable Token Condition occurs in closer proximity to the visuomotor association in the visual detection task. The visual detection task is the binding signal that drives learning in the paradigm. Under this account, in the incidental paradigm, learning occurs when the participant begins to use the auditory clues to predict where the visual target will appear.

After this process begins, when the participant hears the sounds on each trial, they make predictions regarding where the target will appear. Then, the appearance of the target provides implicit feedback telling them if they were right or wrong in their prediction. They use that feedback to refine their categorization of the auditory stimuli. Gabay et al. (2015) argue that variable tokens within trial temporally places auditory exemplar variability in closer proximity to the mechanic in the paradigm that drives learning. Specifically, variable auditory tokens within each trial allows participants to better refine their categorization of stimuli by aiding in the extraction of salient acoustic features from the various exemplars as they identify the acoustic characteristics that are essential to the specific category. When this process occurs in close proximity to the learning reinforcement mechanic, learning is enhanced. In the Identical Token Condition, the acoustic variability is spread out across trials, making it more difficult to extract the salient acoustic features of each tone category and is not tightly coupled to the learning reinforcement mechanic. However, the benefit of high variability in close temporal proximity to the learning reinforcement mechanic may be an untested hypothesis. Results in Gabay et al. (2015) and in Experiment 1 in the present study only indicate that token variability matters for category learning and that variable tokens in close temporal proximity result in greater category learning. It may or may not be that proximity of the variability to the learning mechanic matters for category learning. Experiments that do not use learning reinforcement mechanics (e.g., passive learning paradigms) may be needed to investigate the impact of the temporal proximity of token variability on novel sound category acquisition.

### 3.5.3 Generalization to novel talkers

Posttest 2 tested generalization to novel talkers, and as expected from results from Gabay et al. (2015), participants in both conditions were able to generalize to novel talkers. In Gabay et al. (2015) participants that heard variable tokens within trial and identical tokens within trial were able to learn the sound categories and generalize learning to novel exemplars. Generalization to novel tokens indicates categorization (Palmeri & Gauthier, 2004; Holt & Lotto, 2010). In speech categorization a listener must generalize across acoustically variant sounds to determine which features are salient to a specific type of sound and use those salient features to classify novel sounds. In Gabay et al. (2015) and in Experiment 1, if the learners had not learned the sound categories during training, they would not have been accurate on novel tokens at test. Even though talkers that share the same language background differ widely in the acoustic realizations of their productions, previous tone acquisition studies indicate that learners can generalize learning of novel tone categories to novel talkers (Wang et al., 1999; Qin & Zhang, 2020). Therefore, it was expected that learners in Experiment 1 would be able to generalize learning to novel talkers as well as novel tokens from the talker they were trained on. However, due to the differences in productions between talkers (see Chapter 2), it was expected that participants would perform worse on Posttest 2 than on Posttest 1.

Participants in both conditions were able to successfully generalize learning to novel talkers, but they were less accurate than when generalizing to novel tokens from the same talker. This is likely due to variations in productions across talkers. As discussed in Holt and Lotto (2010), acoustic variations between talkers arise from differences in anatomy and physiology (Fant, 1966), speaking rate (Gay, 1978; Miller & Baer, 1983), and the environments that the stimuli were recorded in, which were not controlled in this study (Houtgast & Steeneken, 1973; Kuttruff, 2016). Differences in the stimuli from the talkers in Experiment 1 are outlined in detail in Chapter 2 and include variations in F0 range, F0 contour shape, and syllable duration. When moving from tokens used over the training blocks and on Posttest 1, which all came from the same talker, to tokens on Posttest 2, which came from three new talkers, the potential variation of acoustic features increases exponentially and categorization of the stimuli became more difficult. It may be that if participants experience greater acoustic variability by hearing multiple talkers during training, then their results when generalizing to novel talkers will be more similar to their results when generalizing to novel tokens from the talkers they were trained on. This topic is addressed in Experiment 2.

### 3.5.4 Stimuli effects

Chapter 2 characterizes the stimuli used in the experiments in this study and discusses differences in the stimuli that may affect results in the experiments. The stimuli from Talker A differed from the talkers used on Posttest 2 in a few ways. Talker A had shorter syllable durations than the other talkers. There were also individual differences in durations across tone categories, with Talker A differing from the other talkers. Talker A's productions of T21 were longer than the other tone categories. Talker D's productions did not differ across tone categories. Talker E's productions of T45 were relatively longer and productions of T315 were relatively shorter than the other tone categories. Talker F's productions of T315 were shorter than the other tone categories. Further, Talker A's F0 contour for T45 also differed some from the F0 contours of the other talkers. There were also differences in the amount of creaky voice that each talker used on T21 and T214. It is possible that these differences made it more difficult to generalize to novel talkers on Posttest 2 as participants might have expected Talker A's idiosyncrasies to be features of the tone categories across talkers. It may be that accuracy scores on Posttest 2 could be higher if the stimuli from the talker that participants heard in training was more similar to the stimuli from the talkers they heard on Posttest 2.

### 3.5.5 Learning differences as a function of age

Participation in the study was not limited based on age, and therefore ages ranged from 18 to 63. Results in the Identical Token Condition suggested that reaction times differed as a function of age, with reaction times being slower for older participants. Reaction times in the Variable Token Condition did not differ as a function of age. It was expected that older participants would have slower reaction times across training blocks than younger participants. The task used in the study includes auditory perception when listening to stimuli, working memory when using processing auditory stimuli and using the stimuli to predict the location of the visual target, visuomotor control when identifying the visual target, and hand motor control when directing the mouse cursor to the visual target. Cognitive function and motor control processes generally slow across the lifespan (e.g., Salthouse, 1985). This tendency is especially evident in temporal tasks that involve reaction time as a measure (e.g., Lima et al., 1991). It is likely that reaction times did not differ as a function of age in the Variable Token Condition due to the small number of older participants in that condition.

Expectations regarding age as a predictor of accuracy on the posttests, however, were less clear. It was possible that a general cognitive slowing across the lifespan might result in information decay during the processing of the auditory signal, resulting in lower accuracy scores (Salthouse, 1996). However, not all cognitive functions are adversely impacted by age. Language comprehension ability remains stable across the lifespan for healthy individuals (Madden, 1988; Burke et al., 2012), but this is particular to lexical items. The processing of nonlexical items is negatively impacted by age (Lima et al., 1991). Further, reduced frequency following response (FFR) amplitude and increased non-stimulus neural activity among adults over 40 (Skoe et al., 2015), suggest that novel tone learning may be more challenging for older participants. Therefore, it was expected that older participants' accuracy scores would be lower than younger participants' scores, but accuracy results on the posttests did not indicate a relationship between age and learning. However, there were few older participants compared to younger participants, reducing the ability to statistically compare differences in learning across different age groups. However, there are a few observations that may be made from the results. In Experiment 1 we can see that individuals across all ages were able to learn from the incidental paradigm and form the novel tone categories. Therefore, if challenges resulted from declined cognitive processing ability for older participants, it did not hinder them from using the incidental learning paradigm to form novel tone categories. However, in both conditions, no individuals over 40 achieved accuracy scores as high as participants under 40. This was particularly true for the Variable Token Condition, which resulted in six of the participants under 40 achieving scores near 75% accuracy or higher. Further, high variability within trial was especially beneficial for younger participants compared to older participants.

## 3.6 Conclusion

In Experiment 1 we investigated the role of token variability within trial, comparing trials that contained identical tokens with trials that contain variable tokens from the same talker. By examining the impact of token variability on the incidental formation of novel tone categories we tested the hypothesis that high token variability in close proximity to the reinforcement learning mechanism benefits learners by aiding in categorization and generalization to novel tokens. Results indicated that native English participants with no prior experience with the target tone categories can use an incidental learning paradigm with natural tokens to form four novel tone categories after 30 minutes of training with very high, even perfect, accuracy. These findings extend the investigation of factors impacting incidental learning into natural speech sound categories, confirming hypotheses suggesting that incidental learning is an effective means of learning natural speech sound categories. Further, the examination of token variability within trial replicated the results of previous studies, indicating that presenting five different tokens on each trial resulted in greater learning than presenting five identical tokens on each trial. As predicted by previous categorization research, high variability in close temporal proximity to a response resulted in greater learning. Similarly, as predicted by incidental category formation research, high variability of tokens in close proximity to the mechanism in the incidental learning paradigm that drives learning resulted in greater learning than when the variability was spread out across trials. Further, our results, replicating previous studies, demonstrated that the two measures of reaction time during training and accuracy at test are correlated and provide consistent measures of learning. However, we also demonstrated that additions to the paradigm, such as age as a factor, can disrupt the correlation between measures.

In Experiment 1 we also tested the ability to generalize to novel talkers. Participants were able to generalize learning to novel talkers but as expected, they were less accurate when categorizing stimuli from novel talkers. The difficulty generalizing to novel talkers was expected because, as illustrated in Chapter 2, stimuli from multiple talkers presents a much wider range of acoustic variability across multiple dimensions. We concluded that to prepare for generalization to novel talkers, exposure to a wider range of acoustic features during training may be required.

# IV. TALKER VARIABILITY

## 4.1 Introduction

In Experiment 1 we found that higher token variability within trial resulted in greater acquisition of novel tone categories. However, results indicated a sharp decline when generalizing learning to novel talkers. These results demonstrated that training had not prepared learners for categorization under conditions with greater variability (e.g., multiple talkers). Therefore, in Experiment 2 we examine the impact of training with multiple talkers. Will participants better generalize to novel talkers if they are trained on multiple talkers compared to a single talker? Further, if we increase variability during training, will learners still be able to acquire the categories as effectively as they did in the single talker condition?

In Experiment 2 we also include a Control Condition where the audio-to-visual correspondence of tone categories to a visual target on the screen is removed during training by randomizing the correspondence of tone categories and visual categories from trial to trial. This will effectively remove reinforcement learning from the paradigm as the auditory tokens will not map to the visual targets. By including the Control Condition, we are able to investigate the effect of age on the task. That is, if participants are not able to learn a mapping, they will not be able to respond faster across training blocks. Therefore, the Control Condition will allow us to investigate a baseline effect for age. We expect that there will be a linear relationship between age and reaction times during training. This will provide a baseline that we can use to analyze the effect of age on training reaction times in other conditions. Further, participants in the Control Condition will not have learned the audio-to-visual mapping during training. That is, they will not have learned where the visual target should appear after hearing the auditory stimuli. For example, they will not know that the low tone occurs with the visual target in the bottom right box. Therefore, at test they will not be able to accurately associate the tone categories with the visual targets as participants in the other conditions did.

### 4.1.1 The impact of talker variability on sound category learning

A challenging yet common task humans face in auditory perception is the need to identify speech sound categories across interlocutors. As discussed in Chapter 3, multiple productions of the same sound category by a single talker can contain a range of acoustic variability.

Productions from multiple talkers introduces an even wider range of acoustic features for listeners to generalize across. Therefore, due to the lack of invariance in the acoustic signal between talkers, generalization of sound categories across multiple talkers is very challenging, especially when learning novel sound categories. That is, there are numerous cues that might distinguish an individual category and each speaker of a language varies in their production of those cues. For example, as indicated in Chapter 2, a common production effect as F0 drops lower during productions of tone categories is the occurrence of creaky voice. There is a wide range in the amount of creaky voice that may occur. Some talkers may produce creaky voice on every production of a low tone category while others may produce none, and there can be a wide range in between. Therefore, to better generalize to novel talkers, it is typically beneficial to be exposed to a range of productions from multiple talkers during training (Jamieson & Morosan, 1989; Lively et al., 1993; Bradlow et al., 1997; Wang et al., 1999; Barcroft & Sommers, 2005; Iverson et al., 2005; Brooks et al., 2006). Previous research suggests that training people on multiple talkers helps them to generalize better to novel talkers. For example, Lively et al. (1993) found that participants trained with stimuli from multiple talkers resulted in greater categorization of sound categories after training.

As discussed, previous research indicates that training on multiple talkers helps learners generalize learning to novel talkers. Further, greater token variability in Experiment 1 improved category learning. Considering these results, should it be expected that further increasing variability through the addition of multiple talkers during training will improve learning? What if we also increased segmental and phonotactic variability? Would learning continue to improve? The underlying question is, is there a limit to the benefit of variability during novel category learning? Is there a point where learning is hindered by an amount of variability that reduces the learner’s ability to attend to the salient features of the category? According to Reverse-Hierarchy Theory (RHT; Ahissar & Hochstein, 2004; Ahissar et al., 2008) perceptual learning occurs when listeners identify the correct perceptual level (e.g., pitch contour) and attend to meaningful input. One hypothesis is that large amounts of variability during initial category learning will inhibit learners from attending to the correct perceptual level. This may be the case for studies with results suggesting that exposure to multiple talkers reduces perceptual learning (Mullenix & Pisoni, 1990; Magnuson & Nusbaum, 2007; Perrachione et al., 2011; Bradley, 2017). During the formation of novel tone categories with explicit learning paradigms, high variability through multiple talkers can reduce learning compared to low variability conditions (Perrachione et al., 2011). The impact of talker variability on novel tone formation during incidental learning has not yet been studied. In the current experiment we ask whether high variability from multiple talkers will also impact incidental learning.

In Experiment 2 we examine these questions by testing the impact of talker variability across trials during training on the ability to generalize learning to novel tokens from the same talkers and to novel tokens from novel talkers. However, without previous research investigating talker variability during incidental category learning, it is difficult to know if high talker variability in the present experiment will hinder the initial formation of novel tone categories.

### 4.1.2 Unsupervised learning

Experiment 2 also contains a Control Condition, which is a passive listening condition with no ability to learn the audio-to-visual correspondence and therefore no reinforcement learning. By examining a condition that includes no audio-to-visual correspondence and no reinforcement, participants should not be able to respond faster across blocks. This will allow us to test the impact of age on the task alone to observe a baseline effect of age on the task. However, this also means that at test we will not be able to measure learning in the same way that other conditions are measured. At test participants will not have learned which tone category is assigned to which visual target. Therefore, it is most likely that they will attempt determine their own auditory to visual mapping and this will likely differ for each participant. Using the same measure as the other conditions would only measure those participants that happen to choose a mapping that aligns with our predetermined mapping.

Since there will be no reinforcement during training, any tone category formation that occurs in this condition will be due to passive exposure to the stimuli. Here, *passive* indicates that their participants did not receive instruction regarding the tone categories and they did not receive feedback, whether explicit or via an implicit reinforcement learning mechanic. However, participants are aware of the number of categories due to the presence of the four boxes when the visual target appears. Therefore, this fits the definition of unsupervised category learning, where the learner is told the number of categories to be learned but does not receive feedback (Ashby et al., 1999). Therefore, this use of passive exposure differs from some other studies in small but potentially important ways. For example, the passive condition in Roark et al. (2020) did not include a motor response, but the audio-to-visual correspondence was left intact. Therefore, the reinforcement learning mechanism was present in the paradigm. Participants were still able to make predictions and see if the predictions were accurate. Therefore, their passive condition would not meet the definition of unsupervised category learning. Roark et al. (2020) concluded that a motor response was not necessary for learning, but the audio-to-visual correspondence was necessary. They also stated that passive accumulation of acoustic input regularities was insufficient for learning. Results and expectations from Roark et al. (2020) follow expectations found in the COVIS model (Ashby et al., 1998) [18]. In the research regarding the COVIS model, especially regarding incidental learning with information-integration categories, a key factor that drives learning is the nature and timing of the feedback on each trial (Ashby & Casale, 2003; Ashby et al., 1999). If there is no reinforcement from the audio-to-visual mapping, learning will not occur. Further, Ashby and Casale (2003) state that there is no evidence that people can learn information-integration categories without feedback. Therefore, it is not expected that participants' responses will indicate any sign of learning in the Control Condition. To be clear here, signs of learning in this condition would occur if participants show consistency in their audio-to-visual mapping at test, whatever mapping they decide to use.

### 4.1.3 Current experiment

Experiment 2 contains three conditions: Single Talker Condition, Multi-talker Condition, Control Condition. By examining talker variability during training, we investigate the impact of talker variability on the ability to form novel tone categories and generalize learning to new talkers. We also compare the Single Talker Condition to a Control Condition where there is no auditory to visual mapping during training, meaning that the auditory stimuli and visual targets are randomly selected each trial. This condition will provide a baseline for the effect of age on the incidental learning task.

It is expected that participants in the Single Talker Condition and the Multi-talker Condition will learn, but that generalization to novel tokens from the same talker(s) on Posttest 1 might be

[18] The COmpetition between Verbal and Implicit Systems model (COVIS; Ashby et al., 1998, 2011; Chandrasekaran et al., 2014a) is a dual-learning systems model of speech category learning. COVIS posits there is a reflective learning system that is activated during explicit, rule-based learning, and a reflexive learning system that is activated during implicit learning.

more robust in the Single Talker Condition than in the Multi-talker Condition. Further, we expect that there will be a greater difference between scores on Posttest 1 and Posttest 2 for the Single Talker Condition than the Multi-talker Condition. That is, accuracy scores will likely decrease more when generalizing to novel talkers on Posttest 2 for participants in the Single Talker Condition than for participants in the Multi-talker Condition.

We expect that participants in the Control Condition will not learn the tone categories. Further, during training their reaction times will not get faster and they will not display accuracy at test.

## 4.2 Methods

### 4.2.1 Participants

As in Experiment 1, participants were recruited online via Prolific. All participants self-identified as being monolingual English speakers and identified as being native English speakers from America, Canada, the United Kingdom, South Africa, Australia, or New Zealand. Participants that reported significant language learning experience, that reported hearing impairments, or that did not use the right equipment (headphones and an external mouse) were excluded from the study.

In the Single Talker Condition, where participants heard a single talker across trials during training, 29 participants were recruited[19]. Four participants were excluded for using the wrong equipment or for hearing impairments, leaving 25 participants (13 female, 11 male, 1 non-binary). Participants in this condition spoke a variety of English dialects (6 American, 2

[19] The Single Talker Condition in the present experiment is the Variable Token Condition from Experiment 1. Therefore, descriptions of the Single Talker Condition here in Experiment 2 are a restatement of details from the Variable Token Condition in Experiment 1. Primary differences in the description of the Single Talker Condition in the present chapter arise from the differences in the comparisons made across the experiments. Experiment 1 compared token variability within and across trials. Experiment 2 examines talker variability across trials, comparing training on a single talker in the Single Talker Condition with multiple talkers in the Multiple Talker Condition.

Australian, 14 British, 1 Canadian, 1 Irish, and 1 NA) [20]. Ages ranged from 19 to 56 with a mean of 29.08 and standard deviation of 9.45[21].

In the Multi-talker Condition, where participants heard multiple talkers across trials during training, 27 participants were recruited. Two participants were excluded for using the wrong equipment, leaving 25 participants (14 female, 11 male). Participants spoke a variety of English dialects (4 American, 14 British, 2 Canadian, and 5 NA). Ages ranged from 19 to 59 with a mean of 34.80 and standard deviation of 13.92.

Experiment 2 also contains a Control Condition, where audio-to-visual mapping was randomized, making it impossible to learn an audio-to-visual mapping scheme during training. In the Control Condition, 25 participants were recruited (13 female, 11 male, 1 NA). No participants were excluded. Participants spoke a variety of English dialects (5 American, 3 Australian, 13 British, 2 Canadian, 1 New Zealand, 1 South African). Ages ranged from 19 to 62 with a mean of 30.87 and a standard deviation of 11.78. All participants were paid for their participation through Prolific.

### 4.2.2 Stimuli

Stimuli used in the Single Talker Condition and the Control Condition were the same stimuli used in the Variable Token Condition in Experiment 1[22]. In each condition in experiment 2, the set of five tokens within trial contained random tokens, constructed as described in experiment 1. However, in the multitalker condition, during training, participants heard stimuli from three talkers, randomly presented across trials.

Tokens from all four tone categories were produced in the syllable /ma/. Ten exemplars of each category were recorded from all six talkers. In the Single Talker Condition and the Control Condition, half of the exemplars of each category from Talker A were used for training, and half of the exemplars were used to test generalization of learning to new exemplars on Posttest 1.

[20] It is not expected that experience with specific English dialects would aid in novel tone category acquisition over other dialects. English dialects do not use F0 information contrastively at the lexical level. Further, experience with other regional languages used in proximity to the specific dialect should not be a factor as participation was limited to those that identified as being monolingual English speakers.

[21] Age is considered as a covariate during analysis and is reported in the results.

[22] Properties of the stimuli are discussed in detail in Chapter 2.

Five exemplars from Talker D and Talker E and four exemplars from Talker F[23] were used to test generalization of learning to new speakers on Posttest 2 in the Single Talker Condition and the Control Condition. In the Multi-talker Condition, five exemplars of each tone category from Talker A and Talker B and four exemplars from Talker C were used for training, and five different exemplars from Talker A and Talker B and four different exemplars from Talker C were used to test generalization of learning to new exemplars on Posttest 1. Posttest 2 stimuli were identical to the other conditions.

## 4.3 PROCEDURE

The procedure for Experiment 2 was the same as the procedure for Experiment 1. The primary difference regards the stimuli and the Control Condition. In Experiment 2, three groups of participants were exposed to four novel Thai tone categories through an incidental learning paradigm. Participants went through four training blocks with forty-eight trials in each block. Then, Posttest 1 tested generalization to novel tokens from the same talker(s) over thirty-six trials, and Posttest 2 tested generalization to novel talkers over thirty-six trials. Posttest 3 tested production of the tone categories over thirty-six trials. Finally, participants completed a language background questionnaire.

### 4.3.1 Training

Participants in each condition were trained with the incidental paradigm described in Experiment 1. On each trial participants heard five sounds and then clicked on a visual target, an 'X', that appeared in one of four boxes. Participants were trained across four training blocks with forty-eight trials in each block. For all conditions, auditory stimuli in each trial consisted of five concatenated exemplars. The concatenations were randomly selected prior to subject running. However, the presentation of trials was randomly selected by the experiment. In the Single Talker Condition and the Control Condition, training was composed of six different concatenations of each tone category from Talker A for a total of twenty-four trials (6 concatenations X 4 tones X 1 talker). These twenty-four trials were duplicated on each training block for a total of forty-eight trials per block. In the Multi-talker Condition, training was

[23] Due to Covid restrictions, which led to talkers recording themselves, some tokens were not usable. In these situations, trials were still comprised of five randomly selected tokens with one token being duplicated.

composed of four different concatenations of each tone category from each talker for a total of forty-eight trials (4 concatenations X 4 tones X 3 talkers), which were repeated on each training block.

The Control Condition differed from the other conditions in that the audio-to-visual mapping was randomized. Therefore, the auditory stimuli did not provide consistent clues regarding the location of the visual target, making it impossible to develop an audio-to-visual mapping scheme over the course of training. Therefore, they completed training having heard the same stimuli as the Single Talker Condition, but did not experience the reinforcement from the incidental learning that participants in the Single Talker Condition experienced.

For all conditions, reaction times were measured to examine learning across the four training blocks. It is expected that faster reaction times across training blocks will occur for those that learn the target tone categories and that faster reaction times will correlate with performance at test. Further, mouse tracking was conducted to examine changes in decision space over time as participants acquire the tone categories[24].

### 4.3.2 Testing

The testing procedure for Experiment 2 was the same as Experiment 1. Participants heard five sounds and then saw four boxes appear without a visual target. They then chose which box the target should appear in. Posttest 1 tested generalization to novel tokens and Posttest 2 tested generalization to novel talkers. Posttest 1 and Posttest 2 were the same for all conditions. Unlike the training blocks, in the Control Condition, the mapping was not randomized. The mapping scheme was present and consistent, like the Single Talker Condition and the Multi-talker Condition. The difference for participants in the Control Condition is that they had no clues during training to help them learn the mapping scheme. Further, since there was no explicit or implicit feedback regarding the mapping scheme during the posttests, they also had no clues during the tests to help them learn the auditory-to-visual mapping scheme that was being used to test them. However, as mentioned, it is possible that they will develop their own mapping scheme and that their mapping scheme may overlap or be the same as the mapping scheme predetermined by the test.

[24] An analysis of mouse tracking data is not included in the dissertation. Future analyses and description of the current work will analyze and consider mouse tracking data and report results.

#### *4.3.2.1 Posttest 1: Generalization to new tokens*

Posttest 1 trials for the Single Talker Condition and the Control Condition were composed of three different concatenations of each tone category from Talker A for a total of twelve trials (3 concatenations X 4 tones X 1 talker). These twelve trials were repeated three times on Posttest 1 for a total of thirty-six trials. Posttest 1 trials for the Multi-talker Condition were composed of three different concatenations of each tone category from Talker A, Talker B, and Talker C for a total of thirty-six trials (3 concatenations X 4 tones X 3 talker).

#### *4.3.2.2 Posttest 2: Generalization to new talkers*

Posttest 2 trials for all conditions were composed of three different concatenations of each tone category from each Talker D, Talker E, and Talker F, for a total of thirty-six trials (3 concatenations X 4 tones X 3 talkers).

#### *4.3.2.3 Posttest 3: Production of the tone categories*

Experiment 2 also contained a third posttest, which was conducted in the same way as Experiment 1. Participants saw the visual target appear in one of the four boxes and recorded themselves saying the target tone with the syllable /ma/. Thirty-six trials were conducted.

## 4.4 RESULTS

### 4.4.1 Training reaction times

Experiment 1 tested the impact of token variability on novel tone category learning, finding that token variability within trial resulted in more robust learning than token variability across trials. In Experiment 2 all conditions utilize token variability within trial. The main variable measured in Experiment 2 is talker variability across trials. In the Single Talker Condition, all training trials contained auditory stimuli from a single talker. In the Multi-talker Condition, training trials contained auditory stimuli from one of three talkers. Experiment 1 found that participants that learn the target tone categories have reaction times that get faster across training blocks. It is expected that participants in the Multi-talker Condition, will learn the tone categories and will have faster reaction times across training blocks. However, due to greater variability in the auditory signal from acoustic variations across talkers[25], it is expected that reaction times will be slower across training blocks for the Multi-talker Condition than for the Single Talker Condition.

Experiment 2 also contains a Control Condition. The primary purpose of the Control Condition is to measure reaction times across training blocks for comparison with the Single Talker Condition and the Multi-talker Condition. Participants in the Control Condition receive no clues regarding the audio-to-visual mapping scheme. Therefore, there is nothing that will enable them to predict where the visual target will appear. This should make it impossible for any participant in the Control Condition to have reaction times that become faster across blocks. Since participants in the Single Talker Condition and the Multi-talker Condition will receive auditory clues regarding the appearance of the visual targets, those that learn the audio-to-visual mapping should have reaction times that get faster across training blocks.

#### *4.4.1.1 Analysis*

As in Experiment 1, visual target detection times were measured from the end of the auditory stimuli to the time the participant clicked on the visual target. Reaction times greater than 1,500 ms were excluded from analyses. For each condition, I compare reaction times across training blocks by comparing a full model and a reduced model without training block. I then conduct contrast coded linear mixed-effects regressions to compare each training block to the subsequent training block to examine changes in reaction times from block to block. Further, I compare reaction times across training blocks across the three conditions by comparing a full model with an interaction between condition and training block and a reduced model without an interaction, followed up by post-hoc comparisons of each condition with each other condition. Finally, as differences in age can affect learning and hearing ability (Kiessling et al., 2003; Clinard et al., 2010), I conduct model comparisons to examine age as a fixed effect.

#### *4.4.1.2 Reaction Times*

Results indicated that reaction times from participants in the Single Talker Condition, became faster across training blocks. However, reaction times from participants in the Multi-talker Condition and the Control Condition did not become faster. Figure 39 illustrates log-transformed reaction times across training blocks for the Single Talker Condition. The four boxplots in each of

[25] See Chapter 2 for a detailed characterization of the auditory stimuli.

the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block.

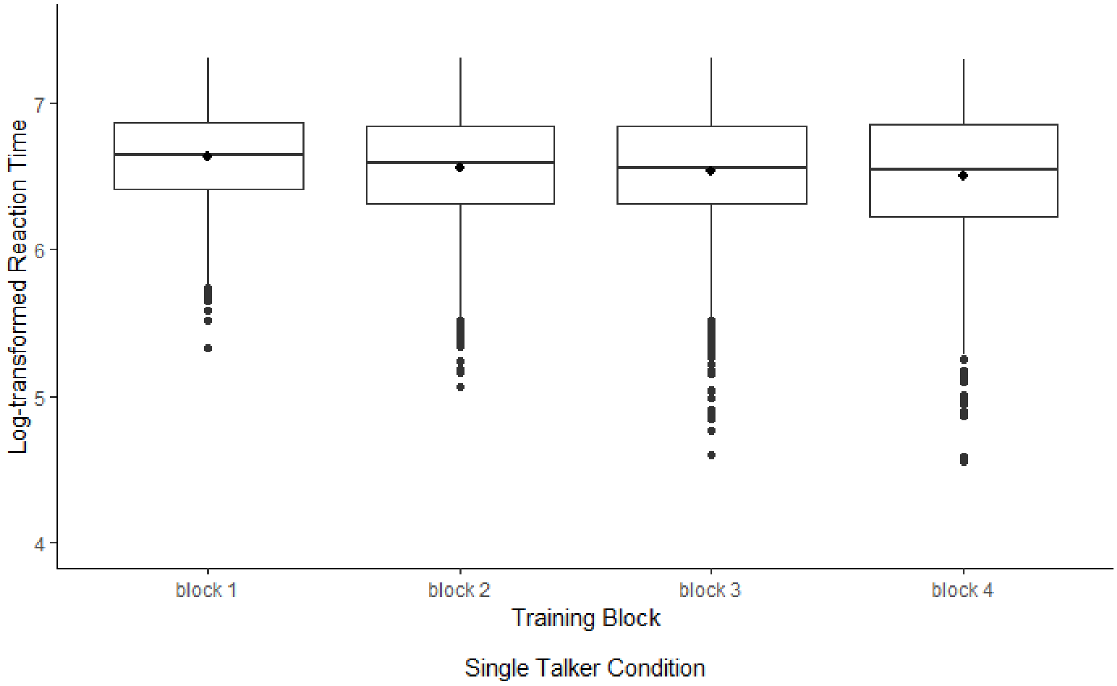

Figure 39. Log-transformed reaction times across training blocks in the Single Talker Condition.

To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the Single Talker Condition (X$^2$ (3) = 114.05, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.63, *SD* = .31) were significantly slower than block 2 (*M* = 6.56, *SD* = .39; β = -.065, *t* = -5.55, p < .001), reaction times in block 2 did not differ from block 3 (*M* = 6.54, *SD* = .42; β = -.022, *t* = -1.89, p = .06), and reaction times in block 3 were significantly slower than block 4 (*M* = 6.51, *SD* = .47; β = -.012, *t* = -2.98, p = .003).

Figure 40 illustrates log-transformed reaction times across training blocks for the Multi-talker Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block.

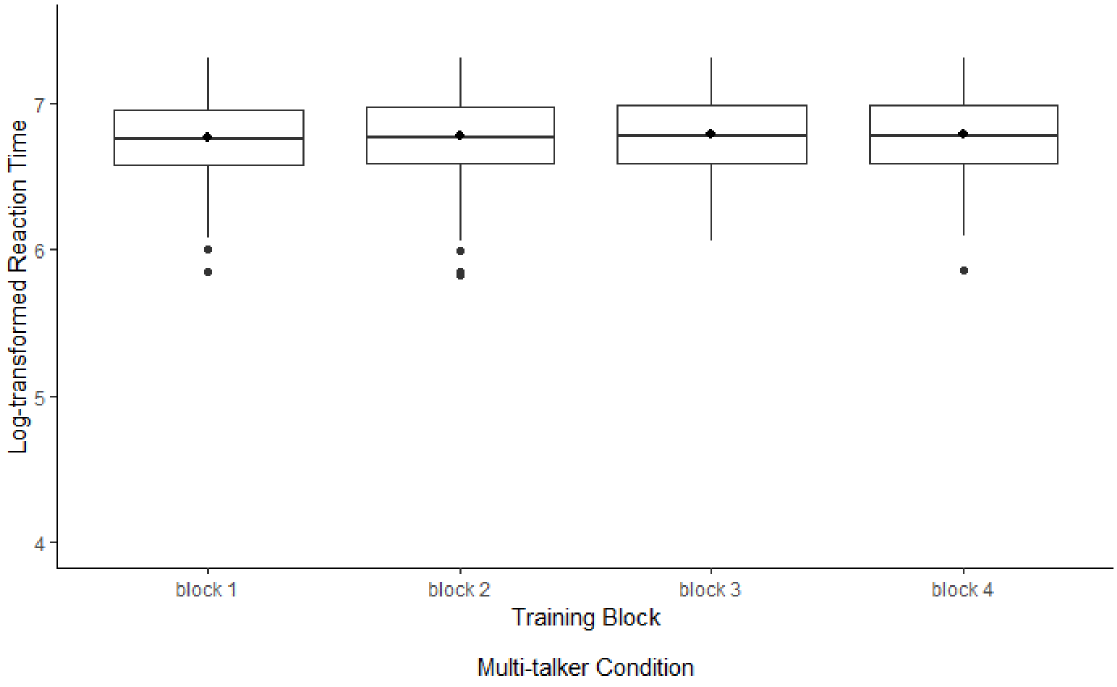

Figure 40. Log-transformed reaction times across training blocks in the Multi-talker Condition.

As a whole, participants' reaction times in the Multi-talker Condition did not get faster across training blocks. Instead, they became slightly slower across training blocks. To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the Multi-talker Condition ($X^2$ (3) = 40.60, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.77, *SD* = .26) did not differ from block 2 (*M* = 6.78, *SD* = .26; β = .012, *t* = 1.56, p = .12), reaction times in block 2 differed significantly from block 3 (*M* = 6.79, *SD* = .26; β = .024, *t* = 2.97, p = .003), and reaction times in block 3 did not differ from block 4 (*M* = 6.79, *SD* = .26; β = .01, *t* = 1.17, p = .24).

Figure 41 illustrates log-transformed reaction times across training blocks for the Control Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block. Figure 41 suggests that, as expected, participants' reaction times did not get faster across training blocks.

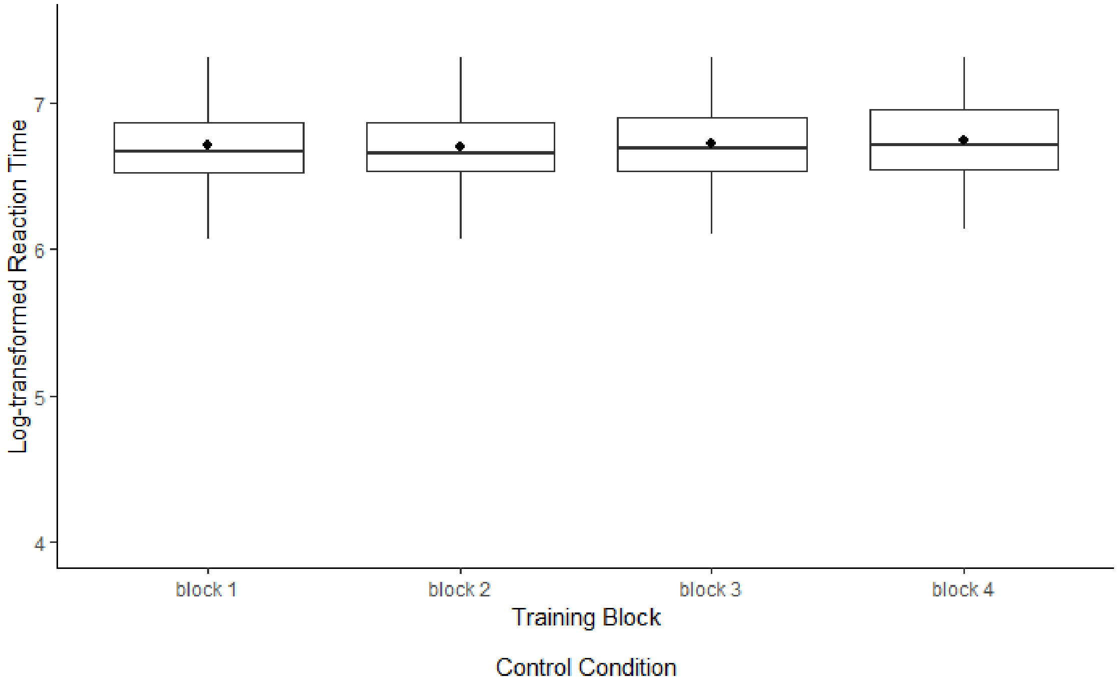

Figure 41. Log-transformed reaction times across training blocks in the Control Condition.

As a whole, participants' reaction times in the Control Condition became slower across training blocks. To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the Multi-talker Condition ($X^2$ (3) = 52.80, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.70, *SD* = .26) did not differ from block 2 (*M* = 6.70, *SD* = .25; β = -.006, *t* = -.71, p = .48), reaction times in block 2 differed significantly from block 3 (*M* = 6.72, *SD* = .26; β = .029, *t* = 3.44, p < .001), and reaction times in block 3 differed significantly from block 4 (*M* = 6.75, *SD* = .27; β = .025, *t* = 3.06, p = .002).

I compared reaction times across the three conditions. Figure 42 illustrates mean reaction times across training blocks for each condition with whiskers illustrating 95% confidence intervals. Table 6 provides the means and standard deviations of response times for the three conditions. It was expected that learning would be more robust in the Single Talker Condition but that participants in the Multi-talker Condition would still learn. However, as illustrated in Figure 42 and described in Table 6, reaction times in the Multi-talker Condition did not become faster across training blocks. In the Control Condition, as illustrated in Figure 42 and Table 6, reaction times condition slowed substantially across training blocks.

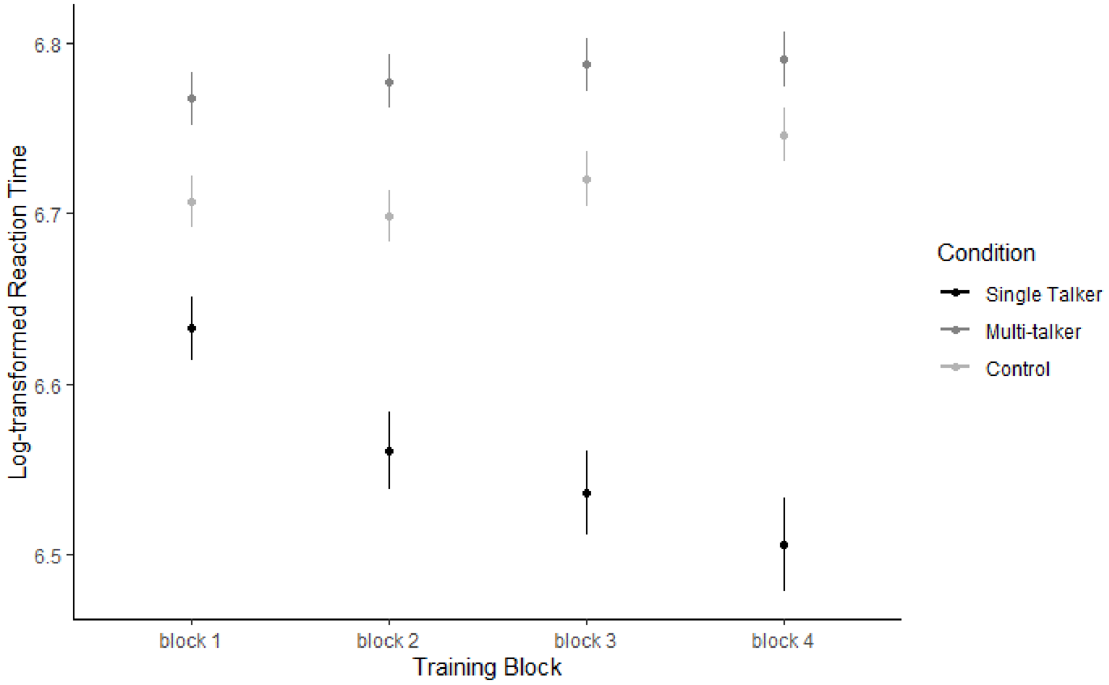

Figure 42. Log-transformed mean reaction times across training blocks for the Single Talker Condition, the Multi-talker Condition, and the Control Condition. Error bars represent 95% confidence intervals.

Table 6. Summary statistics for reaction times for the Single Talker Condition, the Multi-talker Condition, and the Control Condition

| Condition | Block 1 (mean, SD) | Block 2 (mean, SD) | Block 3 (mean, SD) | Block 4 (mean, SD) |
|---|---|---|---|---|
| Single Talker | 6.63, .31 | 6.56, .39 | 6.54, .42 | 6.51, .47 |
| Multi-talker | 6.77, .26 | 6.78, .26 | 6.79, .26 | 6.79, .26 |
| Control | 6.70, .26 | 6.70, .25 | 6.72, .26 | 6.75, .27 |

To test whether reaction times differed across conditions, I compared models with and without an interaction between condition and training block. Results indicated that reaction time differs across training blocks as a function of condition ($X^2$ (6) = 229, *p* < .001).

reaction_time ~ condition * training_block + age + (1|participant)
reaction_time ~ condition + training_block + age + (1|participant)

Bonferroni corrected post-hoc comparisons revealed that reaction times in the Single Talker Condition differed from the Multi-talker Condition (β = -.185, SE = .066, *z* = -2.81, *p* = .015) and the Control Condition (β = -.171, SE = .066, *z* = -2.59, *p* = .029), but the Multi-talker Condition did not differ from the Control Condition (β = .014, SE = .066, *z* = .22, *p* = 1).

By comparing reaction times across training blocks as a function of condition, I tested the impact of talker variability on natural sound category learning. In the Single Talker Condition stimuli across trials contained tokens from a single talker and therefore less overall variability in the acoustic signal than the Multi-talker Condition. The greater variability in the acoustic signal from multiple talkers in the Multi-talker Condition resulted in slower reaction times across training blocks. Slower reaction times were expected in the Multi-talker Condition. However, it was expected that participants in the Multi-talker Condition would still exhibit learning, but their reaction times across training blocks was not indicative of learning. Further, it was expected that reaction times in the Control Condition would not get faster, and as expected, they did not get faster. Instead, they got slower across training blocks.

I also tested whether reaction times differed as a function of age in each condition by comparing models with and without age, controlling for training block. Results from the Single Talker Condition indicated that reaction times did not significantly differ as a function of age ($X^2$ (1) = 1.55, *p* = .21).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ training_block + (1|participant)

Figure 43 illustrates log-transformed reaction times as a function of age in the Single Talker Condition. Mean reaction times across blocks for each participant are illustrated as dots with error bars illustrating 95% confidence intervals. If participants are learning the categories, quantified as faster reaction times across training blocks, then darker blocks will be lower on the y axis in Figure 43 and lighter blocks will be higher. In the Single Talker Condition, none of the participants over forty exhibited faster reaction times across blocks, which led to results being uninformative regarding the time course of learning across age groups in this condition.

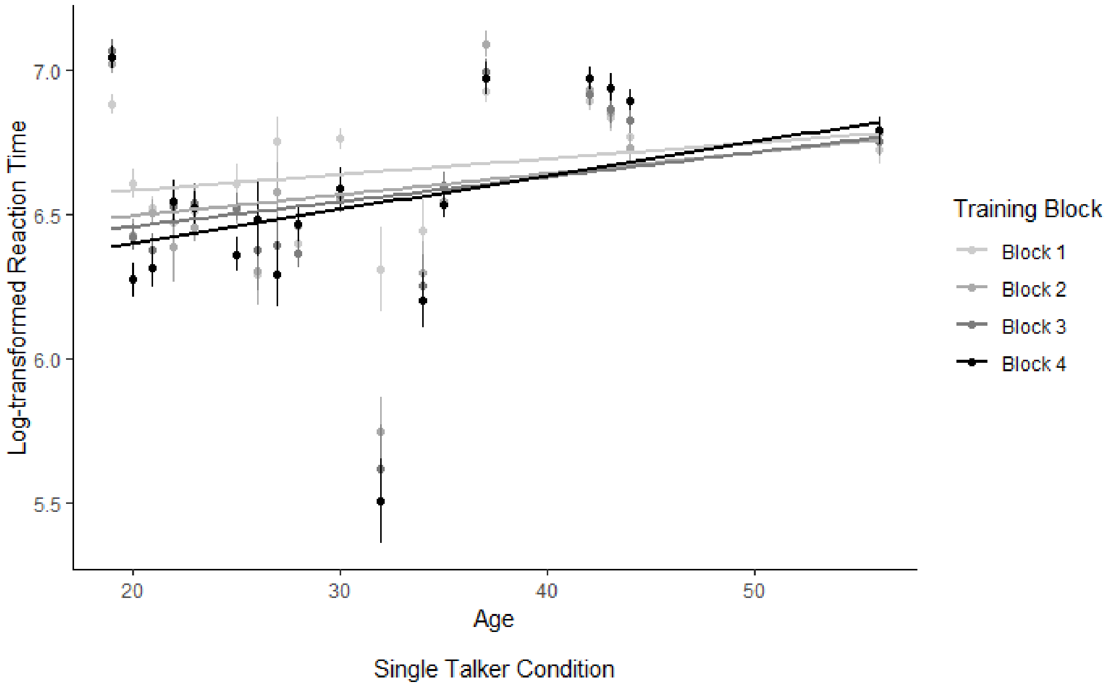

Figure 43. Log-transformed reaction times across training blocks in the Single Talker Condition.

Results from the Multi-talker Condition indicated that reaction time significantly differed as a function of age ($X^2$ (1) = 7.02, *p* = .008). Figure 44 illustrates log-transformed reaction times as a function of age in the Multi-talker Condition. Few participants in the Multi-talker Condition exhibited signs of learning. However, some of the oldest participants had reaction times that became faster across training blocks. Although their reaction times became faster across blocks, overall, their reaction times were not as fast as the younger participants, even those that did not display learning. So, although faster reaction times display learning within participant, overall, reaction times differ as a function of age in the Multi-talker Condition.

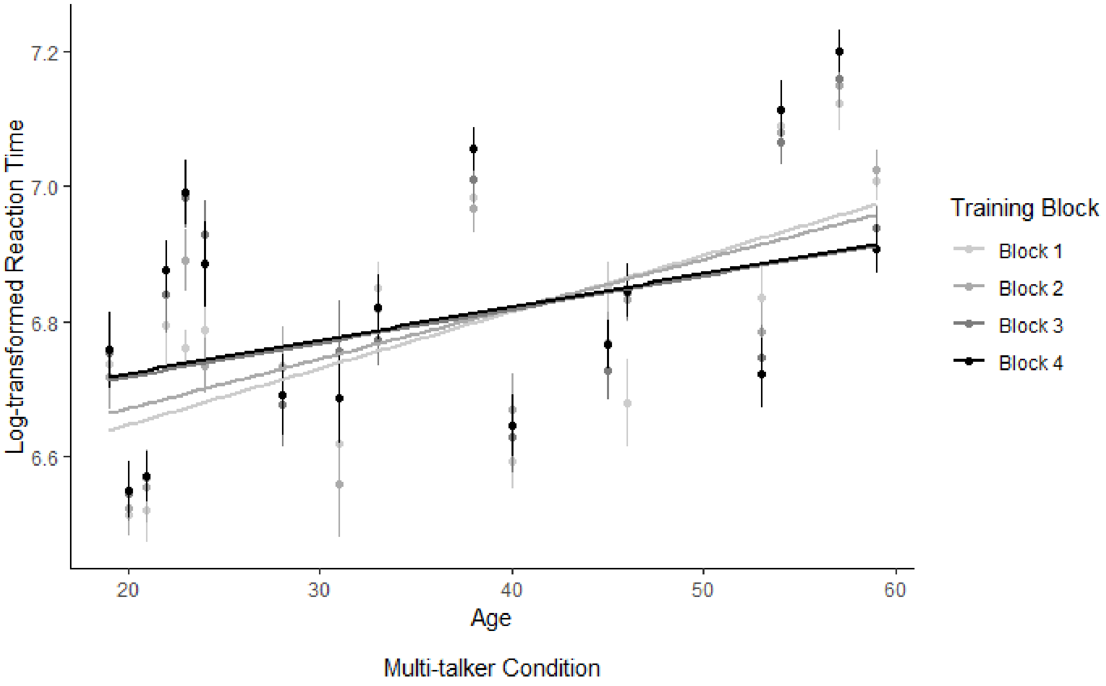

Figure 44. Log-transformed reaction times across age in the Multi-talker Condition.

Results from the Control Condition indicated that reaction time significantly differed as a function of age ($X^2$ (1) = 8.21, *p* = .004). Figure 45 illustrates log-transformed reaction times as a function of age in the Control Condition. As expected, participants in the Control Condition did not show signs of learning. Therefore, Figure 45 provides a clearer understanding of the baseline effect of age on reaction times during the task and the effect of training block on reaction times during the task. Overall, younger participants perform the task faster than older participants, and reaction times from participants tend to get slower across training blocks.

In Experiment 2 I measured the reaction times of participants across training blocks in three conditions. In the Single Talker Condition, reaction times became faster across training blocks, indicating that participants learned the novel tone categories and were able to use that learning to predict the locations of the visual targets. By contrast, in the Multi-talker Condition reaction times did not get faster across training blocks. Rather, they became slightly slower, indicating that relatively few participants learned the novel tone categories. In the Control

Condition reaction times also became slower across training blocks. Results also indicated that age has an effect on reaction times during the experiment. Reaction times from older participants tend to be slower than younger participants.

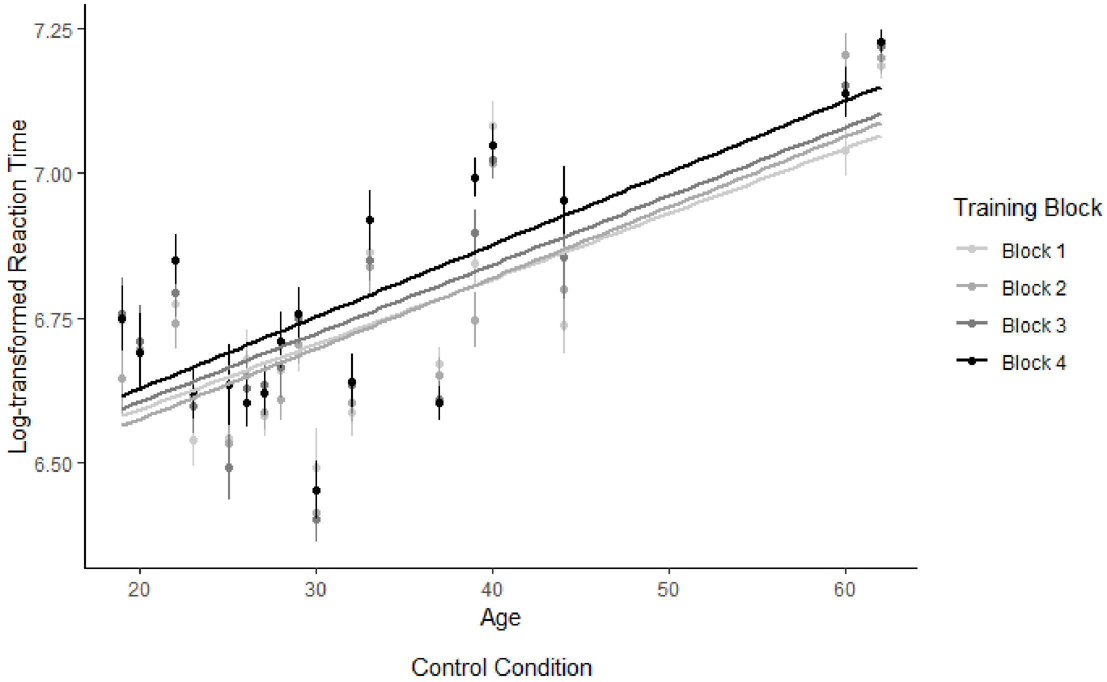

Figure 45. Log-transformed reaction times across age in the Control Condition.

### 4.4.2 Generalization to new tokens and new talkers

As in Experiment 1, Posttest 1 tested participants' ability to generalize to new tokens from the same talker(s), and Posttest 2 tested generalization to new talkers. The structure of both posttests is identical and both measure identification accuracy of the target tone category. If participants have learned the categories they should be able to accurately identify in which box the visual target should have appeared based solely on hearing the auditory stimuli, and therefore, their accuracy scores will be higher. Experiment 1 confirmed that participants that hear a single talker during training are able to accurately identify the four novel tone categories on Posttest 1. However, when they hear novel talkers on Posttest 2, they are less accurate. The Multi-talker Condition in the present experiment trained participants on multiple talkers with the expectation that greater variability in the acoustic signal from multiple talkers during training may result in lower accuracy on Posttest 1 compared with the Single Talker Condition, but should also result in more equivalent generalization to new talkers on Posttest 2. In the

Control Condition participants did not learn the audio-to-visual mapping during training. They also received no clues regarding the audio-to-visual mapping at test.

#### *4.4.2.1 Analysis*

Accuracy scores for all conditions were measured on Posttest 1 and Posttest 2. For each condition, I compare accuracy scores on both posttests to chance using one sample t-tests. To test whether accuracy scores differ as a function of condition, I conduct model comparisons with and without condition for each posttest. To test whether there is a correlation between the learning measures, I conduct correlation tests between reaction times during training and accuracy scores at test for each condition. Finally, I conduct model comparisons to examine age as a fixed effect for all conditions on Posttest 1 and Posttest 2.

#### *4.4.2.2 Accuracy*

Figure 46 illustrates mean proportion correct scores with 95% confidence intervals for the Single Talker Condition, the Multi-talker Condition, and the Control Condition on Posttest 1 and Posttest 2. As discussed, results from participants in the Control Condition may show accuracy if they chose an auditory-to-visual mapping scheme that matched the predetermined scheme used in the experiment. We included the Control Condition in Figure 46 to see if this might be the case. The figure suggests that participants in all conditions, including the Control Condition, accurately identified the target categories above chance on Posttest 1 and on Posttest 2, and that participants in the Single Talker Condition performed better on Posttest 1 than participants in the Multi-talker Condition.

To test whether accuracy scores differed from chance, I examined accuracy scores within condition on Posttest 1 and Posttest 2. In the Single Talker Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24) = 4.80$, $p < .001$, ($M = 52.83$, $SE = 5.80$) and on Posttest 2, $V = 247$, $p < .001$, ($Mdn = 34.72$)[26]. In the Multi-talker Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24) = 3.20$, $p = .002$, ($M = 30.89$, $SE = 1.84$) and on Posttest 2 , $t(24) = 2.88$, $p = .004$, ($M = 30.89$, $SE = 2.04$). In the Control Condition participants were also able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24) =$

[26] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 2 ($W = .88$, $p = .008$).

2.68, *p* = .007, (*M* = 30.78, *SE* = 2.16) and on Posttest 2 , *t*(24) = 2.18, *p* = .019, (*M* = 28.89, *SE* = 1.79).

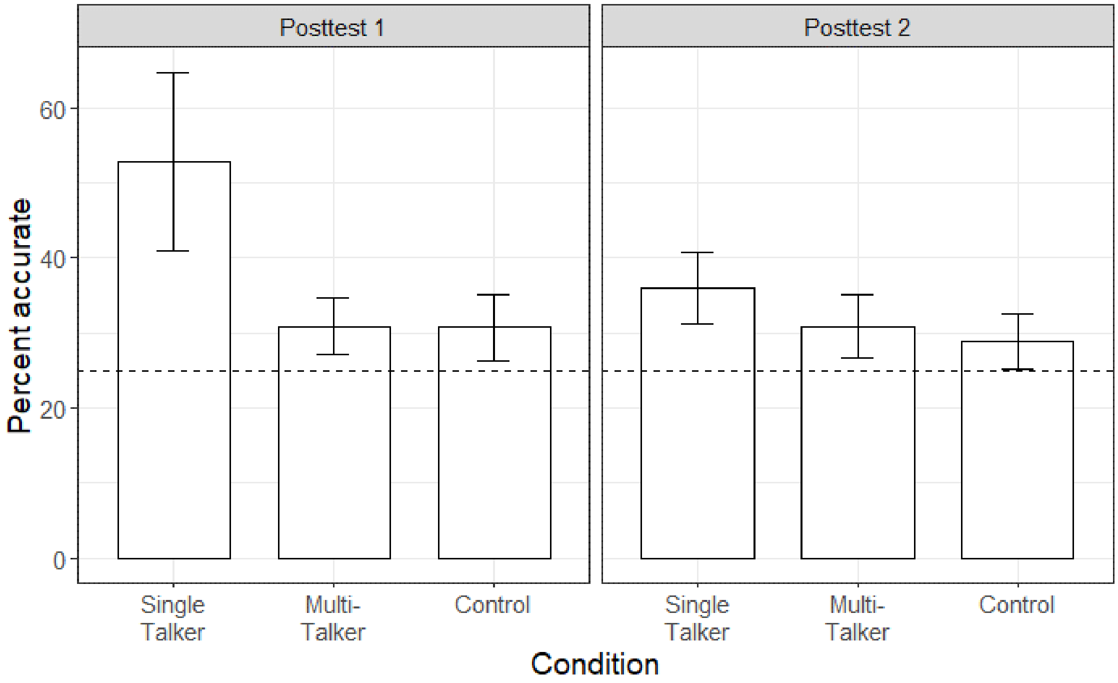

Figure 46. Mean proportion correct for all conditions on Posttest 1 and Posttest 2. Error bars represent 95% confidence intervals. The dashed line represents chance at 25%.

To test whether accuracy scores differed across conditions on Posttest 1 and Posttest 2, I compared models with and without condition for each posttest. Results indicated that accuracy scores differ as a function of condition on Posttest 1 ($X^2$ (2) = 20.26, *p* < .001) and on Posttest 2 ($X^2$ (2) = 6.19, *p* = .045).

accuracy ~ condition + age + (1|participant)
accuracy ~ age + (1|participant)

However, Bonferroni corrected post-hoc comparisons did not reveal a difference between individual conditions. The Single Talker Condition did not differ from the Multi-talker Condition (β = -.051, SE = .03, *t* = -1.70, *p* = .21) or from the Control Condition (β = -.072, SE = .03, *t* = -2.40, *p* = .05). Further, the Multi-talker Condition did not differ from the Control Condition (β = -.021, SE = .03, *t* = -.69, *p* = .77).

Overall, participants in all three conditions accurately identified the target categories above chance on Posttest 1 and on Posttest 2. In the Single Talker Condition and the Multi-talker

Condition, this indicates that participants learned to identify the tone categories and that learning generalized to novel tokens on Posttest 1 and novel talkers on Posttest 2. A comparison of conditions on Posttest 1 indicated that participants in the Single Talker Condition, more accurately identified the target categories than participants in the Multi-talker Condition, indicating that less variability in the acoustic signal during initial exposure to novel tone categories led to more robust generalization to novel tokens from the same talker(s). However, the benefit from exposure to only one talker during training did not result in more robust generalization to novel talkers over and above exposure to multiple talkers during training. Rather, there was a substantial decrease in performance across the posttests for the Single Talker Condition. Performance in the Multi-talker Condition, on the other hand, remained the same across posttests.

As mentioned, results from participants in the Control Condition were included in these measurements to examine whether participants may have chosen an auditory-to-visual mapping scheme that matched the predetermined scheme used in the experiment. This possibility was unlikely. However, results indicated that some of the participants did consistently map the tone categories to the same visual targets predetermined by the experiment. When we examine individual scores, it becomes clearer that some participants in the Control Condition were consistent in this mapping. Figure 47 illustrates mean proportion correct scores with 95% confidence intervals for the Control Condition on Posttest 1 and Posttest 2. The dots represent individual scores, illustrating that some participants were able to accurately identify the tone categories. Consistently mapping the tone categories to a visual target suggests that participants in the Control Condition were able to reliably categorize the auditory tokens. This occurred despite not having learned the audio-to-visual mapping during training. These results suggest that passive auditory exposure to the novel tone categories during an unrelated task may be sufficient exposure for the perceptual formation of novel tone categories.

During training, greater learning was measured through reaction times becoming faster across training blocks. At test, greater learning was measured through higher accuracy scores. It was expected that faster reaction times at the end of training would correlate with higher accuracy scores at test for the Single Talker Condition and the Multi-talker Condition. Figure 48 illustrates the correlation between reaction times on block 4 and accuracy scores on Posttest 1, suggesting a relationship between the two measures in the Single Talker Condition and possibly in the Multi-talker Condition, but not in the Control Condition.

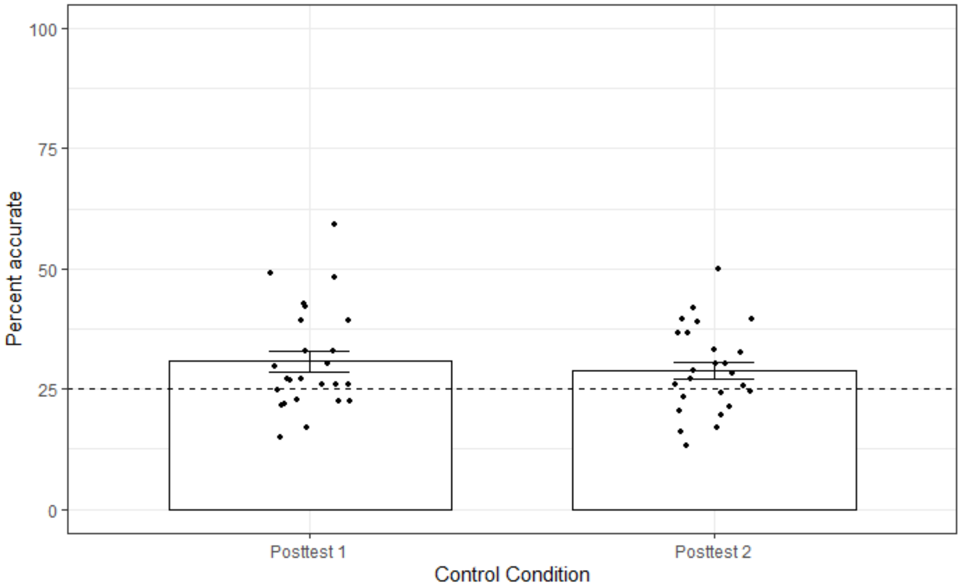

Figure 47. Mean proportion correct for the Control Condition on Posttest 1 and Posttest 2. Error bars represent 95% confidence intervals. The dashed line represents chance at 25%. The dots represent jittered accuracy scores from individual participants.

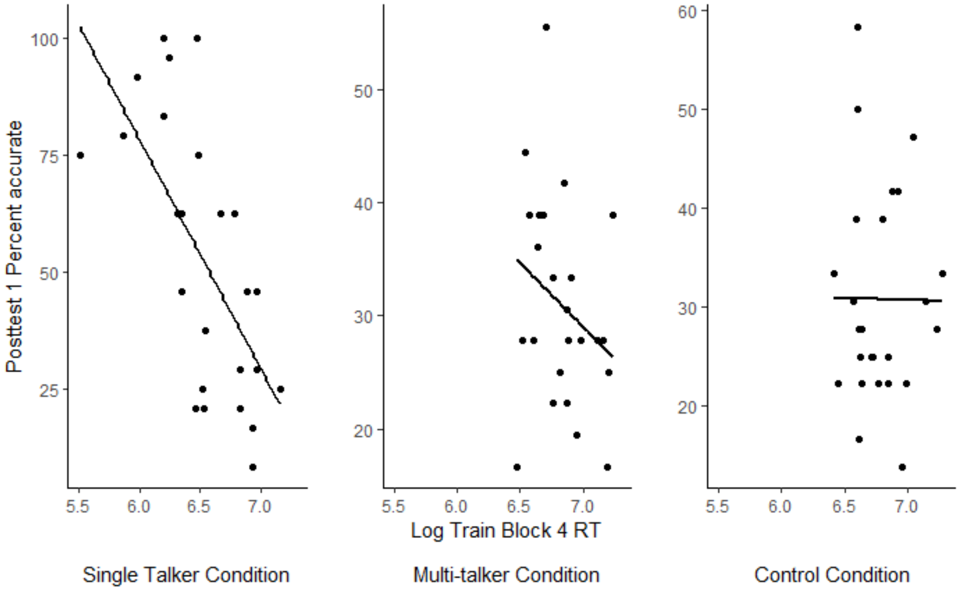

Figure 48. Relationship between two measures assessing category learning across conditions with log transformed reaction times on training block 4 on the x axis and accuracy scores on Posttest 1 on the y axis.

Pearson's correlation coefficient was used to assess the relationship between reaction times on training block 4 and accuracy scores on Posttest 1. The relationship between the two measures was significant in the Single Talker Condition (*r*(23) = -.67, *p* < .001), but not significant in the Multi-talker Condition (*r*(23) = -.27, *p* = .18) or the Control Condition (*r*(23)= -.008, *p* = .97). The significant correlation between the two measures on the Single Talker Condition suggests that faster reaction times in training relates to better accuracy on the generalization test. The correlation in the Multi-talker Condition is confounded by age. Those that displayed learning on Posttest 1 were primarily older participants, whose reaction times, even if they have learned and are getting faster across blocks, are still likely to be slower than younger participants.

To test whether accuracy scores at test differed as a function of age for each condition, I compared models with and without age, and results indicated that accuracy scores did not significantly differ as a function of age in the Single Talker Condition on Posttest 1 ($X^2$ (1) = .44, *p* = .51) or on Posttest 2 ($X^2$ (1) = 3.62, *p* = .057). Accuracy scores did not significantly differ as a function of age in the Multi-talker Condition on Posttest 1 ($X^2$ (1) = 1.22, *p* = .27) or on Posttest 2 ($X^2$ (1) = .16, *p* = .68). Further, accuracy scores did not significantly differ as a function of age in the Control Condition on Posttest 1 ($X^2$ (1) = .002, *p* = .97) or on Posttest 2 ($X^2$ (1) = 2.69, *p* = .10).

accuracy ~ age + (1|participant)
accuracy ~ (1|participant)

Figure 49 illustrates accuracy scores on Posttest 1 and Posttest 2 as a function of age across conditions. The model comparison demonstrated that accuracy scores did not differ as a function of age.

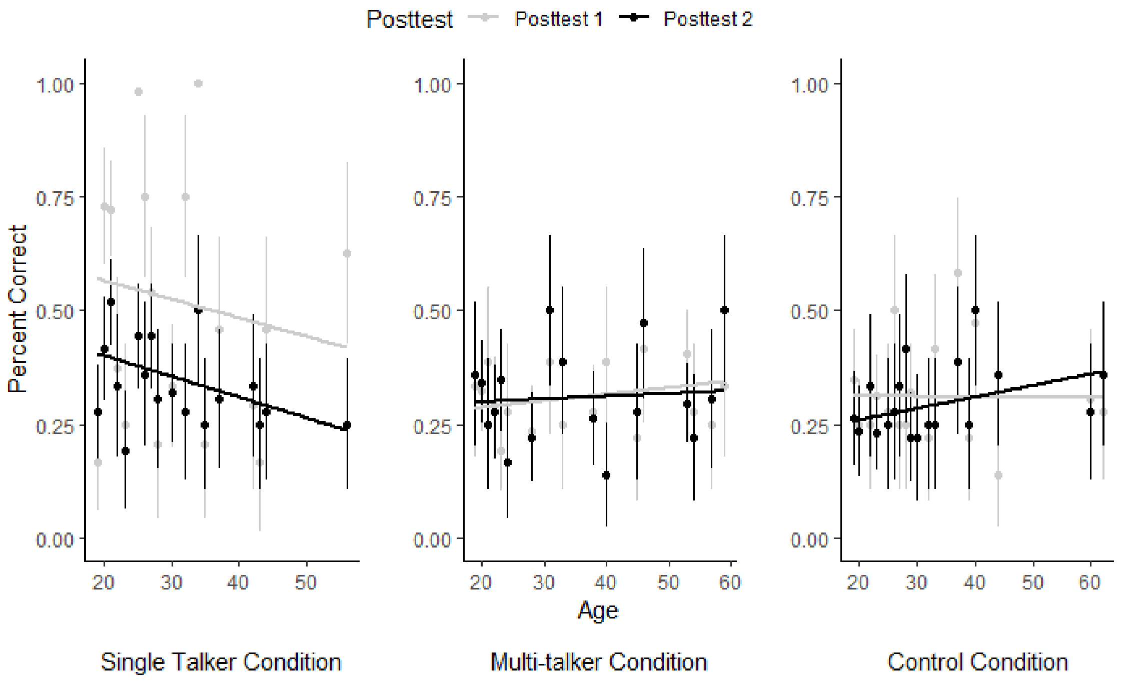

Figure 49. Accuracy scores on Posttest 1 and Posttest 2 across age in the Single Talker Condition and the Multi-talker Condition.

## 4.5 Discussion

In Experiment 2, I examined the impact of talker variability during training on the incidental perceptual learning of novel tone categories. I compared a Single Talker Condition with a Multi-talker Condition that contained stimuli from three different talkers. I also compared the results from these conditions to a Control Condition where the incidental auditory-to-visuomotor correspondence that reinforces learning was not available during training. Results indicated that reaction times from participants in the Single Talker Condition, became faster across training blocks. However, reaction times from participants in the Multi-talker Condition and the Control Condition did not become faster. However, participants in all conditions were able to accurately identify the target categories above chance on Posttest 1 and on Posttest 2. Further, participants in the Single Talker Condition were more accurate on Posttest 1 than participants in the Multi-talker Condition. Below I discuss the implications of these results for categorization and perceptual learning.

### 4.5.1 Incidental learning and passive learning

It was expected that participants in the Control Condition would not learn to consistently distinguish the tone categories. However, results indicated that some participants in the Control Condition were able to categorize the tone categories above chance. These results were somewhat surprising considering that the Control Condition did not contain the learning reinforcement available in the two incidental learning conditions; the Single Talker Condition and the Multi-talker Condition. It is likely that the ability to categorize the novel tone categories arose from passive exposure to the stimuli, rather than the intended incidental learning for the other two conditions.

There is a difference between incidental learning and passive exposure that led to the expectation that participants would not be able to form novel tone categories in the Control Condition. Incidental learning is not passive, nor is it without feedback. The auditory-to-visuomotor correspondence reinforces learning by providing feedback on each trial (see Gabay et al., 2015). The auditory tokens provide cues that participants use to predict the location of the visual target. Participants receive feedback when the visual target appears and their prediction is proven to be correct or incorrect. As they become more confident in their predictions, they move the mouse cursor to the location where they think the visual target will appear. When it appears where they predicted, they are rewarded by being able to click on the visual target faster. If they are wrong in their prediction, they will have to move the cursor to the location of the visual target and their reaction time will be slower. This learning reinforcement works in part because participants are motivated to get through the experiment as fast as they can – they are paid the same amount whether they finish in fifty minutes or an hour. Further, they are instructed at the beginning to click on the visual target as fast as they can. That is, their learning is reinforced.

The learning reinforcement described here is a form of reinforcement learning, which is goal-directed, meaning that learning is driven by the participant's desire to achieve a goal (Sutton & Barto, 2005). In this case the goal is to minimize prediction errors (i.e., reward prediction error). Behavioral actions leading to rewards are reinforced, while behaviors leading to punishment become modified. Lim et al. (2014) argue that the learning reinforcement utilized in goal-directed learning has a neural basis that may not occur during passive exposure to stimuli or in explicit training paradigms. They argue that dopamine neurons in the basal ganglia can serve as a teaching signal to drive reinforcement learning. Dopamine neurons have been shown to be sensitive to reward prediction, firing when predictions are rewarded and depressed when predictions fail (Schultz et al., 1993, 1997). This process can lead to modulations in synaptic plasticity of cortico-striatal pathways (Reynolds & Wickens, 2002). None of this reinforcement was available to participants in the Control Condition.

Another aspect of the incidental paradigm used in this study that reinforces learning is the motor movement that occurs when the participant moves the mouse to the visual target and clicks on the visual target. The motor movement provides a motor response that links together with the auditory token and the visual target, providing an auditory-to-visuomotor correspondence. The motor movement included in the incidental paradigm in the present study may increase learning. However, the extent to which the motor response reinforces learning is unclear. Results from Roark et al. (2020) indicate that incidental learning is not dependent on motor movement. In their study, Roark and colleagues tested the effect of the motor response on incidental category learning by having one group respond to every trial by pushing the space bar rather than a key that corresponded to the audio-to-visual mapping. This kept the audio-to-visual correspondence but removed the reinforcement from the motor response. Results did not differ from the baseline group. Therefore, it was clear that participants could learn from the auditory-to-visual mapping alone.

Participants in the Control Condition were not able to benefit from the reinforcement that occurs during our incidental learning paradigm because they were unable to make accurate predictions regarding the location of the visual target. By randomizing the audio-to-visual mapping on each trial, the auditory-to-visuomotor correspondence was unavailable to these participants to make predictions, thus their predictions could not be rewarded as in the other conditions. Slower reaction times across training blocks support the proposition that there was nothing that the participants could use to accurately predict the location of the visual target on a given trial. That is, they were not able to react faster across training blocks. The lack of any auditory-to-visuomotor correspondence resulted in participants only experiencing the auditory stimuli passively without learning reinforcement, which differs from the passive condition in Roark et al. (2020), which contained an audio-to-visual correspondence with no motor response. Roark et al. (2020) concluded that a motor response was not necessary for learning, but the audio-to-visual correspondence was necessary. They also stated that passive accumulation of acoustic input regularities was insufficient for learning. Their conclusion primarily stems from the Misalignment Condition in their study, where there was an audio-to-visual correspondence, but the visual targets were different colors and did not match the visual target's location. The participants responded by pushing a button corresponding to the color rather than to the location. At test, participants had to guess the location of the visual target based on the auditory stimuli. They were not successful. Roark et al. (2020) state this as evidence that the audio-to-visual correspondence is necessary for learning. However, this conclusion may not be supported by the design and results of their study. Their results indicate that it is possible to distract participants from attending to the audio-to-visual correspondence, which can result in participants being incapable of learning the mapping. To determine whether the audio-to-visual correspondence is necessary or the extent to which it benefits novel sound category formation, the audio-to-visual correspondence needs to be removed from the paradigm, as it was in the Control Condition in the present study. In the present study, passive exposure without audio-to-visual correspondence or a reinforcing motor response was sufficient for categorization to occur in the Control Condition. Again, categorization refers to the ability to consistently make decisions about an object's type (see Palmieri & Gauthier, 2004; Holt & Lotto, 2010). Participants in the Control Condition did this by being consistent in their assignment of the auditory tokens to the visual targets. It is important to be clear here. All that can be concluded from the present study's results is that it was possible for participants to develop the ability to consistently categorize the stimuli from passive exposure alone. We cannot make a conclusion regarding the extent of their learning or compare that learning to the other conditions. We do not know the full extent of learning in the Control Condition because the posttests only granted correct accuracy scores to those that chose an audio-to-visual mapping that matched the predetermined mapping of the experiment. Participants that chose other mappings were counted as incorrect even though they may have also learned the categories as well as those that chose the predetermined mapping. The Control Condition was primarily designed to create a baseline effect of age for reaction times during training. Future research should investigate the extent to which participants can learn from passive exposure, directly comparing passive learning to learning that includes reinforcement from an audio-to-visual correspondence.

There are factors in the current study's design that likely benefitted novel tone category formation in the Control Condition. The ability to form novel sound categories from passive exposure may vary depending on the complexity and similarity of the sound categories (Wade & Holt, 2005; Emberson et al., 2013; LeBovidge, 2018). To learn from passive exposure, sound categories need to be perceptually distinct (Emberson et al., 2013). It is likely that the distinctiveness of each category in the current study aided in the formation of the novel tone categories in the Control Condition. As discussed in Chapter 2, the four Thai tones used in the current study were selected to maximize differences in each category. T45 is high and rises. T315 is low and rises. T241 is high and falls. T21 is low and falls. These four tones provide a contrast between categories, with one high rising tone category, one low rising tone category, one high falling tone category, and one low falling tone category. To summarize, success in forming novel sound categories from passive exposure alone may be moderated by the distinctiveness of the categories, and the distinct nature of the categories used in Experiment 2 may have aided in the passive acquisition of the novel tone categories.

Another factor that may have led to successful novel tone category formation in the Control Condition is the use of high-variability stimuli in close temporal proximity. As discussed in Experiment 1, high-variability stimuli aids in generalization, particularly in the ability to identify the salient acoustic features of the category while learning to ignore the features that are not important for the category. In Experiment 1, I also discussed the conclusion from Gabay et al. (2015), that stated that high-variability stimuli in close temporal proximity to the audio-to-visual correspondence was the “representational glue” that binds the category exemplars together during incidental training, leading to greater category development. However, results from the Control Condition suggest that high-variability stimuli in close temporal proximity may also benefit category learning in learning paradigms that do not contain audio-to-visual learning reinforcement. Thus, it may be that the benefit of high-variability stimuli in close temporal proximity is not paradigm specific. Rather, this type of high-variability training may benefit category learning across paradigms. Thus, the temporal proximity of acoustic variability may be a key factor in the ability to form novel sound categories from passive exposure alone.

### 4.5.2 Correlation between measures

In both conditions in Experiment 1 and in the Single Talker Condition in Experiment 2, reaction times during the final block of training were significantly correlated with accuracy scores at test, indicating that the two measures serve as predictors of learning. However, when participants do not learn, as in the Control Condition in Experiment 2, the measures are not correlated. Further, in conditions where few learn, such as the Multi-talker Condition in Experiment 2, the correlation between the measures becomes unclear. The correlation in the Multi-talker Condition was further confounded by age. Those that learned in the Multi-talker Condition were older participants, and even though older participants learn and therefore become faster across training blocks, their overall reaction times still tend to remain slower than younger participants.

If the incidental learning paradigm is working properly without interference, we see a strong correlation between reaction times during training and posttest scores. However, as indicated by the results in Experiment 2 and in Roark et al. (2020), it is possible for the correlation between measures to become less clear. This occurred in the Multi-talker Condition due to relatively few participants learning and those that learned were older, which resulted in reaction times of older participants becoming faster than baseline and being more similar to younger participants' reaction times. In this way the correlation between measures was disrupted by the impact of age on reaction times. Roark et al. (2020) did not report correlations between measures. However, a post-hoc interpretation of their reaction times during training and accuracy scores at test suggests that the correlation between reaction time and posttest accuracy was likely disrupted in the Irrelevant Feature Condition and in the Misalignment Condition in Roark et al. (2020). In the Irrelevant Feature Condition there was an additional distractor feature that was irrelevant to the task. This addition resulted in an impact to reaction times over and above their baseline condition, but it did not impact accuracy, suggesting that the correlation between the two measures was skewed. In the Misalignment Condition in Roark et al. (2020) the auditory categories were not linked to the task-relevant feature, creating an audio-to-visual misalignment rather than an audio-to-visual correspondence. This disruption resulted in reaction times that remained similar to other conditions but accuracy at posttest suffered completely, again suggested that the relationship between the two measures was skewed. Therefore, if there is not a strong correlation between the two measures, then there may be a factor in the experiment design or differences among participants that is resulting in slower reaction times or reduced accuracy. Consequently, it is likely that the degree of the correlation between the two measures is informative. If there is a clear linear relationship between the two measures, then the paradigm is functioning without additional distractors that are skewing the results. If the correlation is slightly skewed, then the added element in the paradigm is a minor complicating factor for the experiment, but if the correlation is skewed by a large amount, then the added element is a larger complicating factor for the experiment. Therefore, it may be that the degree of correlation between measures could serve as a proxy for the degree of complication caused by the additional element.

### 4.5.3 Talker variability

Results from Experiment 2 indicated that participants in the Single Talker Condition and in the Multi-talker Condition were able to categorize the novel tone categories above chance. However, learning in the Single Talker Condition was more robust than in the Multi-talker Condition on Posttest 1, where participants generalized learning to novel tokens from the same talker(s). As discussed in Section 4.1, there was some expectation that the Single Talker Condition would indicate more robust generalization to novel tokens from the same talker(s) than the Multi-talker Condition. Further, it was expected that there would be a greater difference between Posttest 1 and Posttest 2, where participants generalized to novel talkers, in the Single Talker Condition than in the Multi-talker Condition. As expected, participants in the Single Talker Condition exhibited a sharp decline in categorization accuracy when exposed to multiple new talkers in Posttest 2. By contrast, accuracy scores on Posttest 2 did not differ from accuracy scores on Posttest 1 for participants in the Multi-talker Condition.

The substantial drop in accuracy on Posttest 2 for participants in the Single Talker Condition is likely due to the increase in variability in the acoustic signal that occurs with stimuli from multiple talkers. In Section 3.5.3 and Section 1.2, I discuss the task of perceptual categorization. To form novel sound categories with natural speech stimuli a listener must generalize across acoustically variant sounds to determine which features are salient for the category. If a listener only hears a very limited set of exemplars of a category, when faced with greater variation in the acoustic signal, they will not be as successful at generalizing their learning to the novel exemplars. In one study, for example, French participants that were trained on stimuli with low variability were not very successful at identifying the English /θ/ and /ð/ (Jamieson & Morosan, 1989). By contrast, participants that were trained on high variability stimuli were more successful at identifying /θ/ and /ð/ (Jamieson & Morosan, 1986). The conclusion was that higher stimulus variability aids in the formation of novel sound categories by helping learners attend to the salient differences between the categories and ignore the unimportant differences between stimuli of the same category. Results from Lively et al., (1993) support these conclusions. Training on multiple talkers resulted in improved accuracy and generalization to a new talker, but training on a single talker did not result in generalization to a new talker.

Results from Experiment 2 support previous findings that novel tone category formation training with limited variability, as in the Single Talker Condition can result in reduced categorization ability when generalizing to novel talkers. However, overall learning in the Single Talker Condition was more robust compared to the Multi-talker Condition, as illustrated by reaction times and accuracy scores on Posttest 1. Reaction times became faster across blocks in the Single Talker Condition, but remained steady in the Multi-talker Condition. Further, accuracy scores on Posttest 1 were significantly higher in the Single Talker Condition. Overall, fewer participants learned in the Multi-talker Condition than in the Single Talker Condition and those that learned did not achieve accuracy scores as high as the learners in the Single Talker Condition. There are several aspects of the Multi-talker Condition that may have resulted in reduced tone category formation: 1) the inherent acoustic variability stemming from the use of natural tokens from multiple talkers, 2) perceptual difficulties specific to the population, or 3) difficulties arising from the design of the experiment. It also may be that a combination of these factors resulted in a reduced capacity to attend to the salient features of each tone category.

In Experiment 1, greater within-trial variability enhanced learning, but in Experiment 2, greater talker variability across trials hindered learning. High-variability training is known to enhance perceptual learning as it aids in the generalization necessary for categorization (Jamieson & Morosan, 1989; Lively et al., 1993; Bradlow et al., 1997; Wang et al., 1999; Barcroft & Sommers, 2005; Iverson et al., 2005; Brooks et al., 2006). However, not all variability is the same. In some situations, variability can hinder speech perception (Mullenix & Pisoni, 1990; Magnuson & Nusbaum, 2007; Perrachione et al., 2011, Bradley, 2017). While additional within-trial stimuli variability benefitted learning in Experiment 1, across-trial multi-talker stimuli variability hindered learning in Experiment 2. It is possible that the amount of acoustic variability in the productions across the three talkers in the Multi-talker condition was sufficient to reduce participants' ability to attend to the salient acoustic features of each tone category and reduce learning. Results from Wong et al. (2004) indicate a processing cost when listening to auditory tokens from multiple talkers (also see Kaganovich et al., 2006; Creel et al., 2008). Specifically, processing speech from multiple talkers can result in greater activation of brain regions associated with speech perception and slower reaction times. Results from Perrachione et al.

(2011) suggest that exposure to multiple talkers can reduce perceptual learning. In presenting their conclusions, Perrachione et al. (2011) present an interpretation rooted in Reverse-Hierarchy Theory (RHT; Ahissar & Hochstein, 2004; Ahissar et al., 2009), which states that perceptual learning occurs when listeners identify the correct perceptual level (e.g., pitch contour) and attend to meaningful input. The difficulty that multiple talkers present is that the correct perceptual level for the target categories is obscured by the greater number of uninformative cues (see Chapter 2 for acoustic differences across talkers). Perrachione, et al. (2011) also demonstrate that the effect of talker variability on speech perception is modulated by the individual perceptual abilities of the listener. For example, individuals with higher pretraining pitch contour perception abilities were capable of benefitting from higher variability across stimuli, whereas those with lower initial perceptual abilities were hindered by high-variability training. However, results from Experiment 2 do not support this conclusion. If they did, we would have expected at least some of the participants in the Multi-talker condition to have accuracy scores comparable to participants in the Single Talker Condition, but all scores from those that learned in the Multi-talker Condition were low. In this respect, incidental learning may differ from explicit training with feedback. Incidental learning may be modulated less by the individual perceptual abilities of the listener.

Another factor potentially impacting the incidental acquisition of novel tone categories is the language background of the participants. Participants in the current studies were native English speakers with little or no experience learning other languages. Magnuson and Nusbaum (2007) specifically found that differences in pitch in a mixed talker condition resulted in slower processing of the stimuli than a blocked talker condition with native English participants. Further, native English listeners tend to attend to the level of the pitch over the slope of the pitch. Guion and Pederson (2007) found that cue weighting is different for naïve native English and native Japanese listeners compared to native Mandarin listeners when listening to Mandarin tones. Mandarin participants utilize the level of the pitch and the slope of the pitch, while native English and native Japanese listeners primarily attend to the level of the pitch. The three talkers used in the Multi-talker Condition varied in pitch range. Therefore, participants heard the tone category on one trial produced by a talker with a particular pitch range, and then on a subsequent trial heard the tone category produced in a different pitch range. It is possible that the differences in pitch range across trials led the participants, being native English speakers, to over-attend to differences between pitch ranges on trials, distracting them from the pitch contours, which were necessary to attend to for tone category development. It is possible that this perceptual difficulty would not be problematic for listeners that have tonal L1s. Further, if this hypothesis is correct, native English participants would likely benefit from within-trial talker variability rather than across-trial talker variability. Within-trial talker variability should instruct native English listeners to ignore differences in pitch range of the talkers and attend to differences in pitch contour.

Thus, the difficulty of the Multi-talker Condition may be due to the experiment design. If a population tends to attend to features that could potentially distract them from the salient features of the target categories, then accommodations may need to be made in the experiment design (see Perrachione et al., 2011). Experiment 1 found that participants improved substantially in novel tone acquisition when exposed to trials with variable tokens from the same talker compared to trials with identical tokens from the same talker. Having variable tokens aids in perceptual categorization by helping the learner to determine which features are salient to the sound category and which are unimportant to the category. In the Multi-talker Condition in Experiment 2, each trial consisted of variable stimuli, but they were from a single talker. Therefore, talker variability in the Multi-talker Condition occurred across trials. Barcroft and Sommers (2005) found that participants learned novel lexical targets better when hearing repetitions of stimuli that did not come from the same talker but from multiple talkers. Similarly, participants may learn the target tone categories better by hearing multiple talkers' productions within trial. If native English listeners are over-attending to pitch level or pitch range between talkers and need to learn that the variations between talkers' pitch levels is unimportant, then it is likely that talker variability within trial would train native English learners to ignore differences in pitch level across talkers and attend to pitch contours instead. Participants would hear the consistencies in the pitch contours, identify them as being salient to the category, and be trained to ignore differences in pitch height across talkers. When talker variability is only found across trials, this process is more difficult due to the temporal distance between the most variable exemplars of the category. If within-trial talker variability results in greater learning and greater ability to generalize to novel talkers, then it may indicate that the perceptual categorization mechanism that generalizes salient features of the category across the range of exemplars is most efficient when the full range of features found in the category exemplars occurs in close temporal proximity during training.

Overall results from Experiment 2 suggest that high-variability training with multiple talkers can both help and hinder the incidental learning of novel tone categories. Multi-talker training has the potential to help learners better generalize to tokens from novel talkers. However, initial exposure to multiple talkers may hinder the learners' ability to attend to the salient features of the novel sound category (see Barcroft & Sommers, 2005). The conclusion that there is a greater initial cost to perceptual training on multiple talkers but also a greater potential payoff when generalizing to novel talkers may relate to speech perception more generally. This finding is consistent with other work (Lee and Baese-Berk, under review) which found that exposure to multiple talkers initially slows non-native English speakers' perception of native English speakers, but results in greater adaptation to novel talkers.

The initial learning deficit found in the Multi-talker Condition 1 in Experiment 2 may have occurred due to several factors, such as the inherent acoustic variability stemming from the use of natural tokens from multiple talkers, perceptual difficulties specific to the population in the study, or difficulties arising from the design of the experiment. To further investigate differences between incidental learning involving single and multiple talkers, it would be of interest to compare the results of Experiment 2 with an experiment examining talker variability within trial, rather than across trials.

Another possibility is that training with multiple talkers requires more time to result in more robust learning. When hearing multiple talkers, participants indicate greater speech processing challenges. Goldinger (1990) found that participants selected slower word presentation rates when listening to multiple talkers. Further, faster presentation rates resulted in better lexical processing in a single talker condition compared to a multiple talker condition (Goldinger et al., 1991). Hearing multiple talkers may simply require more time in the incidental learning paradigm to achieve similar results as the Single Talker Condition. This is likely due to the wider range of acoustic features found across productions from multiple talkers. For example, when productions from multiple talkers are randomized across trials, substantial differences between talkers in vowel space and pitch have been noted to slow processing (Magnuson & Nusbaum, 2007). In the present study, participants only heard about one thousand tokens over the course of thirty minutes. If participants in a single talker condition and a multiple talker condition were trained longer, it may be that accuracy on novel tokens from the same talker(s) would become move equivalent, and in that case, it would be expected that participants in the multiple talker condition would better generalize to novel talkers. If additional training led to improvements in the Multi-talker Condition over the Single Talker Condition, then it may suggest a more general rule that greater variability in the stimuli requires more time for category development to occur, which would suggest that the task of categorization becomes more difficult with the amount of variation. This hypothesis could be tested further by increasing variability through the inclusion of variable syllable types produced by multiple talkers. Such comparison may provide details regarding the time course of category learning across a wider range of variability encountered by those seeking to acquire the target categories during language acquisition. The time course of learning may have a linear relationship with the number of talkers, but there is evidence that the difficulty of the task plateaus. For example, Mullennix and Pisoni (1999) found that stimuli from four talkers and sixteen talkers resulted in the same amount of perceptual interference.

Researchers involved in language acquisition may be interested in training programs that result in the greatest accuracy across potential variability in the shortest amount of time. Recent work investigating the application of incidental learning to real world language acquisition provides insight into this concern. Wiener et al. (2019) found that scaffolding learning by beginning with acoustically simpler categories in an implicit learning paradigm resulted in improved categorization and more native-like Mandarin tone productions than explicit speech training. We expect that further investigation of the effect of scaffolding from lower to higher levels of acoustic variability during incidental learning would prove beneficial for language acquisition pedagogy. For example, a study over several training periods that contains the Single Talker Condition and the Multi-talker Condition from Experiment 2, as well as a condition that begins with a single talker and increases the number of talkers over several days of training may show that increasing talker variability over time could result in a better ability to generalize to novel tokens from the same talkers and novel tokens from new talkers in the shortest amount of time.

### 4.5.4 Learning differences as a function of age

In Experiment 2 participants' ages ranged from 18 to 62, permitting some observation of the effect of age on the incidental learning of novel tone categories. In the Control Condition participants were not able to learn and become faster across training blocks. Thus, the Control Condition provided a clear effect of age on training reaction times for the task used in the paradigm, with older participants having slower reaction times than younger participants. As discussed in Section 3.5.5, this result was expected since the task involves a reaction time measure to a multimodal response known to be slower across the lifespan (Salthouse, 1985; Lima et al., 1991). The linear relationship between age and reaction times tends to hold even when more participants above the age of 40 learn than participants below the age of 40, as in the Multi-talker Condition.

In all three conditions, accuracy scores at test indicated that participants of all ages were able to learn. As discussed in Section 3.5.5, there were not enough participants across age groups for statistical comparison, but some general observations can be made from the data.[27] As in Experiment 1, stimuli variability across conditions disproportionately impacted different age groups. In Experiment 1 younger participants were substantially more accurate, compared to older participants, when tokens within trial were variable. That is, high-variability tokens within trial from the same talker helped younger participants to learn the novel tone categories. However, in Experiment 2, high-variability tokens across trials in the Multi-talker Condition seemed to disproportionately hinder learning for younger participants. Further, older participants seemed to benefit from the greater variability in the Multi-talker Condition. These results may have important implications for understanding the underlying processes of perceptual categorization during incidental learning and how those processes change across the lifespan.

As discussed, results from the Single Talker Condition in Experiment 2 found that younger and older participants were able to learn the target tone categories. Maddox et al. (2013) found a similar result. Older participants were able to learn just as well as younger participants through incidental training. However, they also found that explicit training resulted in reduced learning for older participants. The Competition between Verbal and Implicit Systems model (COVIS; Ashby et al., 1998, 2011; Chandrasekaran et al., 2014) posits that there are different neural structures engaged during explicit and implicit category learning paradigms. It may be that the neural mechanisms and processes used during explicit learning may differ from those

[27] When discussing age groups for general observations, I consider those under 40 to be younger and those over 40 to be older.

used during incidental learning. Thus, neural mechanisms and processes used during incidental learning may be impacted less by age related cognitive decline. However, the processes and mechanisms may still be subject to age related effects. The results from the Multi-talker Condition suggest that there may be differences in the incidental learning of novel sound categories across the lifespan. Specifically, older participants seemed to be less impacted by multi-talker variability across trials than younger participants. Older participants may have been impacted less by multiple talkers due to reduced sensitivity to pitch compared to younger participants. Clinard et al. (2010) found that the ability to discriminate frequencies becomes poorer as age increases. Further, they found that the neural representation of frequency, as measured by the frequency-following response (FFR), shows a decline for higher pitch ranges but are intact for lower pitch ranges (also see Skoe et al., 2015; Anderson et al. 2012). One of the main differences between talkers, as illustrated in Chapter 2, is pitch range. If older participants' perception of the talkers' pitch is compressed compared to younger participants, then differences in pitch range between talkers may not be as well perceived by older participants. Therefore, they may not have been as sensitive to differences in talkers' pitch ranges as younger participants.

## 4.6 Conclusion

Experiment 2 directly tested the impact of talker variability on novel tone category formation and the ability to generalize learning to novel tokens from the same talker(s) and novel tokens from new talkers. By examining talker variability across trials, we tested the hypothesis that exposure to multiple talkers during training aids in the ability to generalize to novel talkers. Results from Experiment 2 demonstrated that participants exposed to stimuli from a single talker during training and participants exposed to multiple talkers across trials during training are able to learn novel tone categories above chance. However, talker variability during training impacts the ability to generalize learning to novel tokens and novel talkers in an incidental learning paradigm. Specifically, hearing stimuli from a single talker during training results in substantially more robust generalization to novel tokens from the same talker(s) than hearing stimuli from multiple talkers. Further, if participants hear a single talker during training, there is a sharp decline in accuracy when generalizing to novel talkers. By contrast, if participants are trained on multiple talkers during training, there is little or no difference when generalizing learning to novel talkers. That is, accuracy scores were the same when generalizing to novel tokens from the same talkers and when generalizing to novel tokens from novel talkers for participants trained on multiple talkers.

In the Control Condition in Experiment 2, we examined a condition where there was no audio-to-visual correspondence. Therefore, participants had no ability to predict where the visual target would appear during training. Thus, they did not experience reinforcement learning and did not respond faster across training blocks. By examining a condition that includes no audio-to-visual correspondence and no reinforcement learning, we were able to test the impact of age on the task alone and observe a baseline effect of age on the task. There was a linear effect of age on reaction times, with older participant having slower reaction times than younger participants.

Surprisingly, results from the Control Condition demonstrated that participants in a passive listening condition, where there was no ability to learn the audio-to-visual correspondence and therefore no reinforcement learning, were also able to form novel tone categories. This result was surprising because research on incidental learning suggests that reinforcement is necessary for learning because it is the “glue” that binds the signals together during reinforcement learning. However, we demonstrated that participants can consistently categorize novel tone categories after passive exposure alone. We hypothesized that a key factor in the ability to form novel tone categories from passive exposure is the use of stimuli that contains multiple variable tokens in close temporal proximity.

# V. SEGMENTAL FAMILIARITY

## 5.1 Introduction

An important factor that impacts novel tone category acquisition is the ability to attend to the salient features of the novel tone categories. As discussed, in natural speech there are numerous features that may distract participants from attending to the important features. Some of those features may interact in different ways depending on the learner's language background. For example, if a learner is already familiar with the use of a particular feature for determining sound category membership, they may process novel stimuli differently than learners that do not have the same familiarity. One example comes from novel tone category learning. Learners that use F0 for category membership in their first language display differences in processing novel tone categories (Guion & Pederson, 2007). Further, familiarity with the segmental or phonotactic composition of the tokens can impact perception. Thus, the current study examines the impact of segmental familiarity on the incidental formation of novel tone categories under the expectation that the presence of unfamiliar segments in the stimuli may negatively impact perceptual learning.

### 5.1.1 The impact of segmental familiarity on sound category learning

As discussed, an important aspect of learning novel sound categories is the ability to attend to the salient acoustic features between the categories. Further, it is possible to be distracted from attending to the salient features of the category by a number of factors. The phonotactic environment of the target sound, may inhibit attention to the target acoustic features (Guion & Pederson, 2007; Liu et al., 2011; Wright & Baese-Berk, under review). It is hypothesized that inhibition from complex stimuli involving unfamiliar segmental and suprasegmental features may occur due to increased attentional loads during perceptual learning, thereby increasing the difficulty of learning the novel sound categories. Liu et al. (2011) addresses the question of whether segments and suprasegmentals should be learned together or separately, suggesting that learning both components of a syllable at the same time presents an increased level of difficulty. They tested learners on the acquisition of novel tone categories across different levels of segmental difficulties and found that tone learning suffered under higher levels of segmental difficulty. These results suggest that learners can be distracted from learning tone categories by difficulties presented by the segments. Liu et al. (2011) concluded that discrimination among temporally integrated features, such as segments and tones, is challenging.

This line of thought is further supported by studies indicating differences in the neural processing of native and non-native segments. For example, Peltola et al. (2003) investigated neural response patterns, measured through mismatch negativity (MMN)[28], to English vowels by naïve Finns, Finnish students of English, and native English speakers. They found that the processing of segments familiar to the native language differed from the processing of segments unfamiliar to the native language. Further, they found that even though Finnish students of English had extensive classroom exposure to English, they still did not process the segments like native English speakers. Therefore, Experiment 3 investigates whether processing differences of familiar and unfamiliar segments impacts novel tone category formation during reflexive learning.

Another factor that may be impacted by segmental familiarity is attention. Due to experience with the L1, listeners may also be endogenously oriented to focus on different features in novel stimuli. For example, tone perception studies show that native English listeners weigh pitch cues differently than Mandarin listeners (Guion & Pederson, 2007). Also, Chen and Pederson (2017) found that when attention is directed to segments, learners do not improve in tone discrimination. It may be that native English speakers, due to lack of experience with lexical pitch, are endogenously oriented to direct their attention to segmental composition during auditory perception. If this is the case, then the unfamiliar segment in the /mɯ/ Condition in Experiment 3 may result in a greater attentional load, thereby distracting participants from attending to the tone categories. Thus, we ask whether conditions composed of tokens with familiar segments would permit greater endogenous orientation to f0 information, resulting in greater tone acquisition than conditions with unfamiliar segments.

[28] Mismatch negativity (MMN) is an auditory event-related potential that occurs when a standard sound is presented repeatedly and then interrupted by a deviant sound, permitting the investigation of the extent to which a person hears two sounds as the same or different (see Näätänen, 1992). This is often taken as evidence for listeners perceiving distinctions between similar sounds, even if behaviorally they classify them identically.

### 5.1.2 Current experiment

In the current experiment, we examine the impact of segmental familiarity during training on the incidental perceptual learning of novel tone categories by comparing three conditions that contain tokens produced in different syllables. Two conditions, the /ma/ Condition and the /mi/ Condition, are comprised of segments more familiar to the participants' language background experience. One condition, the /mɯ/ Condition, contains a segment unfamiliar to the participants' language background experience. We expect that results in the two familiar conditions will not differ from each other. However, we expect that a lack of familiarity in the /mɯ/ Condition will negatively impact learning.

## 5.2 Methods

### 5.2.1 Participants

As in the other experiments, participants were recruited online via Prolific. All participants self-identified as being monolingual English speakers and identified as being native English speakers from America, Canada, the United Kingdom, South Africa, Australia, or New Zealand. Participants that reported significant language learning experience, that reported hearing impairments, or that did not use the right equipment (headphones and an external mouse) were excluded from the study.

In /ma/ Condition, where participants heard a single talker produce tokens using the syllable /ma/ during training, 29 participants were recruited[29]. Four participants were excluded for using the wrong equipment or for hearing impairments, leaving 25 participants (13 female, 11 male, 1 non-binary). Participants spoke a variety of English dialects (6 American, 2 Australian,

[29] The /ma/ Condition in the present experiment is the Variable Token Condition from Experiment 1 and the Single Talker Condition from Experiment 2. Therefore, descriptions of the /ma/ Condition in the present experiment are a restatement of details from Experiment 1 and Experiment 2. Primary differences in the description of the /ma/ Condition in the present chapter arise from the differences in the comparisons made across the experiments. Experiment 1 compared token variability within and across trials. Experiment 2 examined talker variability across trials. Experiment 3 compares segmental variability across conditions.

14 British, 1 Canadian, 1 Irish, and 1 NA)[30]. Ages ranged from 19 to 56 with a mean of 29.08 and standard deviation of 9.45[31].

In the /mi/ Condition, where participants heard a single talker produce tokens using the syllable /mi/ during training, 25 participants were recruited (16 female, 9 male). No participants were excluded. Participants spoke a variety of English dialects (3 American, 1 Australian, 18 British, 1 Canadian, 2 Irish). Ages ranged from 20 to 54 with a mean of 33.30 and standard deviation of 9.72.

In the third condition, where participants heard a single talker produce tokens using the syllable /mɯ/ during training, 25 participants were recruited (13 female, 10 male, 1 transman, 1 transwoman). No participants were excluded. Participants spoke a variety of English dialects (4 American, 4 Australian, 11 British, 3 Canadian, 1 Irish, 2 NA). Ages ranged from 19 to 57 with a mean of 30.00 and standard deviation of 12.78. All participants were paid for their participation.

### 5.2.2 Stimuli

Stimuli in experiment 3 had the same composition as the Variable Token Condition in Experiment 1. The five auditory tokens in each trial were randomly selected before the experiment. In Experiment 3, for the three conditions, all auditory tokens for training and for Posttest 1 came from Talker A. In the /ma/ Condition, participants heard tokens produced in the syllable /ma/. In the /mi/ Condition, participants heard tokens produced in the syllable /mi/. In the /mɯ/ Condition, participants heard tokens produced in the syllable /mɯ/. In each condition, half of the exemplars of each category from Talker A were used for training, and half of the exemplars were used to test generalization of learning to new exemplars on Posttest 1.

## 5.3 Procedure

The procedure for Experiment 3 was the same as the procedure for the other experiments, and each condition in Experiment 3 was conducted identically. In the present experiment the primary difference regards the stimuli. In Experiment 3, three groups of participants were

[30] It is not expected that experience with specific English dialects would aid in novel tone category acquisition over other dialects. English dialects do not use F0 information contrastively at the lexical level. Further, experience with other regional languages used in proximity to the specific dialect should not be a factor as participation was limited to those that identified as being monolingual English speakers.

[31] Age is considered as a covariate during analysis and is reported in the results.

exposed to four novel Thai tone categories through an incidental learning paradigm. Participants went through four training blocks with forty-eight trials in each block. Then, Posttest 1 tested generalization to novel tokens from the same talker(s) over thirty-six trials, and Posttest 2 tested generalization to novel talkers over thirty-six trials. Posttest 3 tested production of the tone categories over thirty-six trials. Finally, participants completed a language background questionnaire.

### 5.3.1 Training

Participants in each condition were trained with the incidental paradigm described in Experiment 1. On each trial participants heard five sounds and then clicked on a visual target, an 'X', that appeared in one of four boxes. Participants were trained across four training blocks with forty-eight trials in each block. For all conditions, auditory stimuli in each trial consisted of five concatenated exemplars. The concatenations were randomly selected prior to subject running. However, the presentation of trials was randomly selected by the experiment. In each condition, training was composed of six different concatenations of each tone category from Talker A for a total of twenty-four trials (6 concatenations X 4 tones X 1 talker). These twenty-four trials were duplicated on each training block for a total of forty-eight trials per block.

For all conditions, reaction times were measured to examine learning across the four training blocks. It is expected that faster reaction times across training blocks will occur for those that learn the target tone categories and that faster reaction times will correlate with performance at test. Further, mouse tracking was conducted to examine changes in decision space over time as participants acquire the tone categories[32].

### 5.3.2 Testing

The testing procedure for Experiment 3 was the same as the procedure for the other experiments. Participants heard five sounds and then saw four boxes appear without a visual target. They then chose which box the target should appear in. Posttest 1 tested generalization to novel tokens and Posttest 2 tested generalization to novel talkers. Posttest 1 and Posttest 2 were the same for all conditions.

[32] An analysis of mouse tracking data is not included in the dissertation. Future analyses and description of the current work will analyze and consider mouse tracking data and report results.

#### *5.3.2.1 Posttest 1: Generalization to new tokens*

Posttest 1 trials for each condition were composed of three different concatenations of each tone category from Talker A for a total of twelve trials (3 concatenations X 4 tones X 1 talker). These twelve trials were repeated three times on Posttest 1 for a total of thirty-six trials.

#### *5.3.2.2 Posttest 2: Generalization to new talkers*

Posttest 2 trials for all conditions were composed of three different concatenations of each tone category from each Talker D, Talker E, and Talker F, for a total of thirty-six trials (3 concatenations X 4 tones X 3 talkers).

#### *5.3.2.3 Posttest 3: Production of the tone categories*

Experiment 3 also contained a third posttest, which was conducted in the same way as Experiment 1. Participants saw the visual target appear in one of the four boxes and recorded themselves saying the target tone with the syllable /ma/. Thirty-six trials were conducted[33].

## 5.4 RESULTS

### 5.4.1 Training reaction times

Experiment 1 tested the impact of token variability on novel tone category learning, finding that token variability within trial resulted in more robust learning than token variability across trials. Experiment 2 tested the impact of talker variability on novel tone category learning, finding that a single talker during training resulted in more robust learning than multiple talkers during training. In each condition in Experiment 3 all trials contain variable auditory tokens, and auditory tokens during training are from a single talker. Experiment 3 tests the impact of segmental familiarity on novel tone perception. If familiarity with the segments used in the stimuli impact novel tone category learning, it is expected that segments familiar to the participants (i.e., /ma/ and /mi/) will result in learning outcomes that differ from unfamiliar segments (i.e., /mɯ/).

Experiment 1 and Experiment 2 found that participants that learn the target tone categories have reaction times that get faster across training blocks. It is expected that

[33] An analysis of production data is not included in the dissertation. Future analyses and description of the current work will analyze and consider production data and report results.

participants in the /mi/ Condition, will learn the tone categories and will have faster reaction times across training blocks. It is also expected that participants in the /mɯ/ Condition will learn the tone categories and will have faster reaction times across training blocks. However, it is expected that reaction times will be slower across training blocks for the /mɯ/ Condition, as a lack of familiarity with /ɯ/ adds to the attentional load during perceptual learning.

#### *5.4.1.1 Analysis*

As in the previous experiments, visual target detection times were measured from the end of the auditory stimuli to the time the participant clicked on the visual target. Reaction times greater than 1,500 ms were excluded from analyses. For each condition, I compare reaction times across training blocks by comparing a full model and a reduced model without training block. I then conduct contrast coded linear mixed-effects regressions to compare each training block to the subsequent training block to examine changes in reaction times from block to block. Further, I compare reaction times across training blocks across the three conditions by comparing a full model with an interaction between condition and training block and a reduced model without an interaction, followed up by post-hoc comparisons of each condition with each other condition. Finally, as differences in age can affect learning and hearing ability (Kiessling et al., 2003; Clinard et al., 2010), I conduct model comparisons to examine age as a fixed effect.

#### *5.4.1.2 Reaction Times*

Results indicated that reaction times from participants in all conditions became faster across training blocks. Figure 50 illustrates log-transformed reaction times across training blocks for the /ma/ Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block.

To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the /ma/ Condition ($X^2$ (3) = 114.05, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

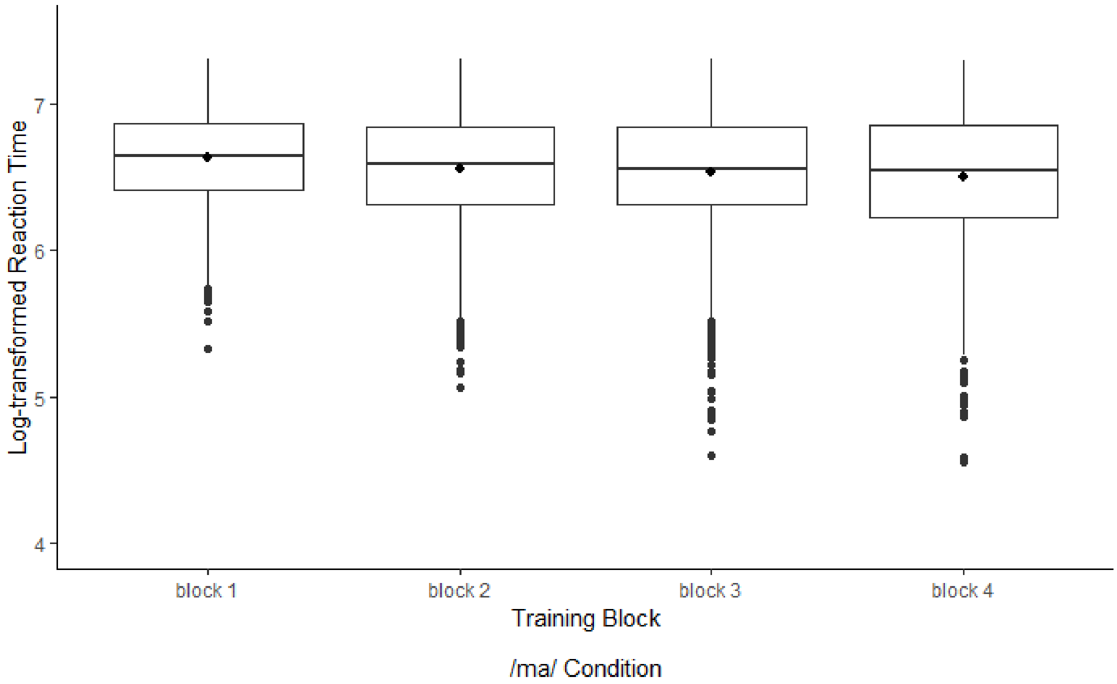

Figure 50. Log-transformed reaction times across training blocks in the /ma/ Condition.

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.63, *SD* = .31) were significantly slower than block 2 (*M* = 6.56, *SD* = .39; β = -.065, *t* = -5.55, p < .001), reaction times in block 2 did not differ from block 3 (*M* = 6.54, *SD* = .42; β = -.022, *t* = -1.89, p = .06), and reaction times in block 3 were significantly slower than block 4 (*M* = 6.51, *SD* = .47; β = -.012, *t* = -2.98, p = .003).

Figure 51 illustrates log-transformed reaction times across training blocks for the /mi/ Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block.

As a whole, participants' reaction times in the /mi/ Condition became faster across training blocks. To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the /mi/ Condition ($X^2$ (3) = 72.20, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

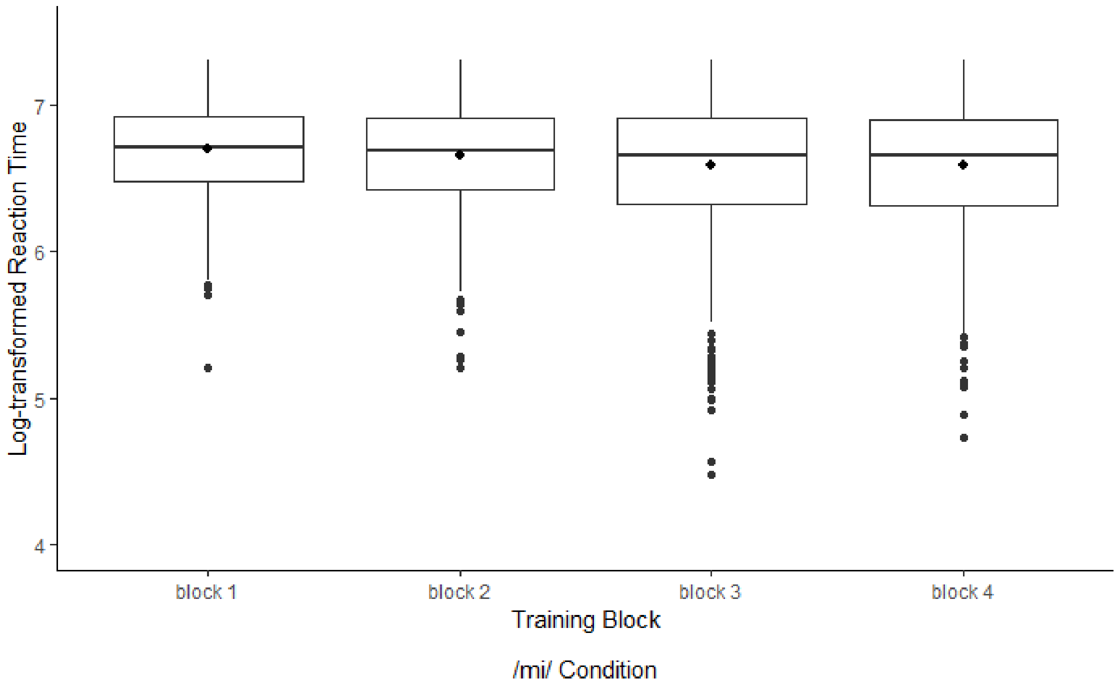

Figure 51. Log-transformed reaction times across training blocks in the /mi/ Condition.

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.70, *SD* = .31) differed significantly from block 2 (*M* = 6.65, *SD* = .36; β = -.027, *t* = -2.15, p = .03), reaction times in block 2 differed significantly from block 3 (*M* = 6.59, *SD* = .44; β = .068, *t* = -5.44, p < .001), and reaction times in block 3 did not differ from block 4 (*M* = 6.60, *SD* = .43; β = .019, *t* = 1.50, p = .13).

Figure 52 illustrates log-transformed reaction times across training blocks for the /mɯ/ Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block.

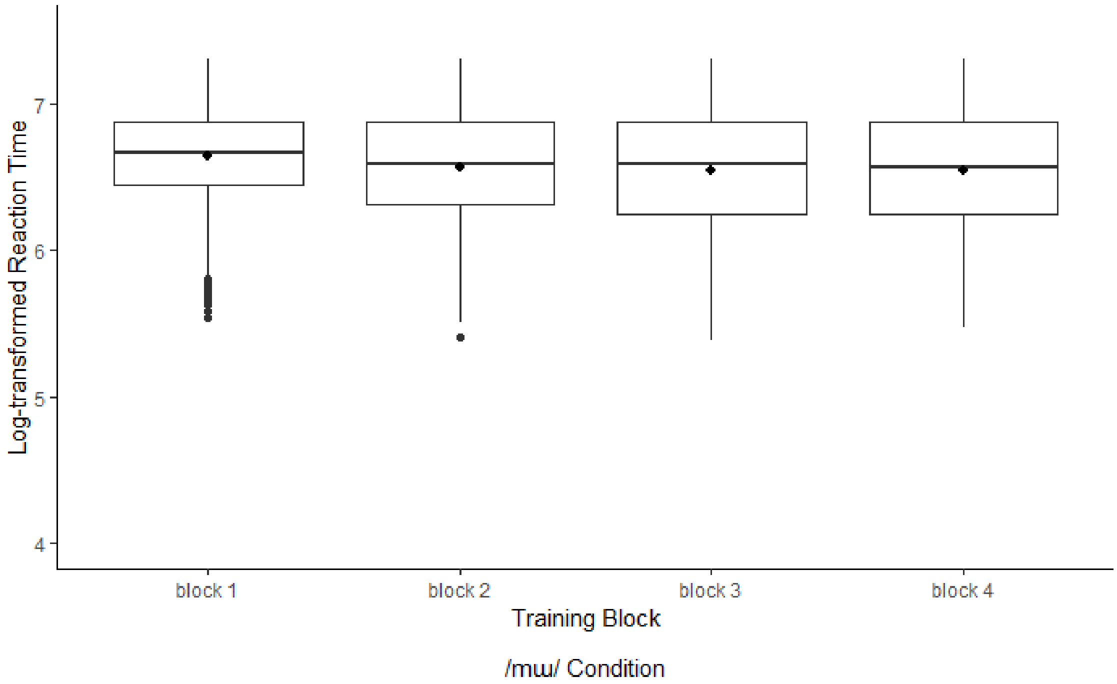

Figure 52. Log-transformed reaction times across training blocks in the /mɯ/ Condition.

As a whole, participants' reaction times in the /mɯ/ Condition also became faster across training blocks. To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the Multi-talker Condition (X$^2$ (3) = 92.40, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.65, *SD* = .32) differed significantly from block 2 (*M* = 6.56, *SD* = .40; β = -.087, *t* = -7.46, p < .001), reaction times in block 2 did not differ from block 3 (*M* = 6.55, *SD* = .41; β = -.005, *t* = -.45, p = .65), and reaction times in block 3 did not differ from block 4 (*M* = 6.55, *SD* = .42; β = -.005, *t* = -.41, p = .69).

I compared reaction times across the three conditions. Figure 53 illustrates mean reaction times across training blocks for each condition with whiskers illustrating 95% confidence intervals. Table 7 provides the means and standard deviations of response times for the three conditions. It was expected that reaction times would be more similar in the /ma/ and /mi/ conditions. It was expected that reaction times from participants in the /mɯ/ Condition would become faster across training blocks, but not be as fast as the other conditions. However, as illustrated in Figure 53 and described in Table 7, the change in reaction times across training blocks was very similar for all conditions.

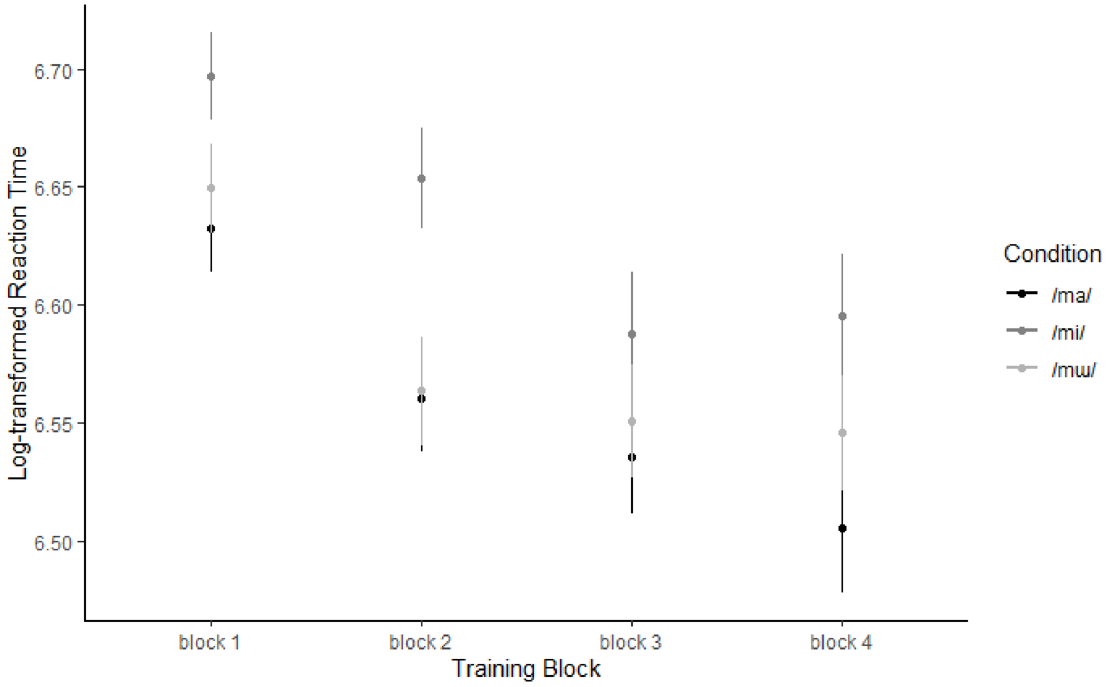

Figure 53. Log-transformed mean reaction times across training blocks for the /ma/ Condition, the /mi/ Condition, and the /mɯ/ Condition. Error bars represent 95% confidence intervals.

Table 7. Summary statistics for reaction times for the /ma/ Condition, the /mi/ Condition, and the /mɯ/ Condition

| Condition | Block 1 (mean, SD) | Block 2 (mean, SD) | Block 3 (mean, SD) | Block 4 (mean, SD) |
|---|---|---|---|---|
| /ma/ | 6.63, .31 | 6.56, .39 | 6.54, .42 | 6.51, .47 |
| /mi/ | 6.70, .31 | 6.65, .36 | 6.59, .44 | 6.60, .43 |
| /mɯ/ | 6.65, .32 | 6.56, .40 | 6.55, .41 | 6.55, .42 |

To test whether reaction times differed across conditions, I compared models with and without an interaction between condition and training block. Results indicated that reaction time differs across training blocks as a function of condition ($X^2$ (6) = 26.71, *p* < .001).

reaction_time ~ condition * training_block + age + (1|participant)
reaction_time ~ condition + training_block + age + (1|participant)

However, Bonferroni corrected post-hoc comparisons revealed that reaction times in the /ma/ Condition did not differ from the /mi/ Condition (β = -.063, SE = .084, *z* = -.74, *p* = 1) or the /mɯ/ Condition (β = -.016, SE = .082, *z* = -.20, *p* = 1), and the /mi/ Condition did not differ from the /mɯ/ Condition (β = .047, SE = .084, *z* = .55, *p* = 1).

By comparing reaction times across training blocks as a function of condition, I tested the impact of segmental familiarity on natural sound category learning. In the /ma/ Condition and the /mi/ Condition, stimuli across trials contained tokens composed of segments common to the participants' L1. It was expected that in the /mɯ/ Condition, which contained tokens composed of segments unfamiliar to the participants, reactions times would differ from the other conditions. However, the change in reaction times across training blocks was similar in all three conditions.

I also tested whether reaction times differed as a function of age in each condition by comparing models with and without age, controlling for training block. Results from the /ma/ Condition indicated that reaction times did not significantly differ as a function of age ($X^2$ (1) = 1.55, *p* = .21).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ training_block + (1|participant)

Figure 54 illustrates log-transformed reaction times as a function of age in the /ma/ Condition. Mean reaction times across blocks for each participant are illustrated as dots with error bars illustrating 95% confidence intervals. If participants are learning the categories, quantified as faster reaction times across training blocks, then darker blocks will be lower on the y axis in Figure 54 and lighter blocks will be higher. In the /ma/ Condition, none of the participants over forty exhibited faster reaction times across blocks, which led to results being uninformative regarding the time course of learning across age groups in this condition.

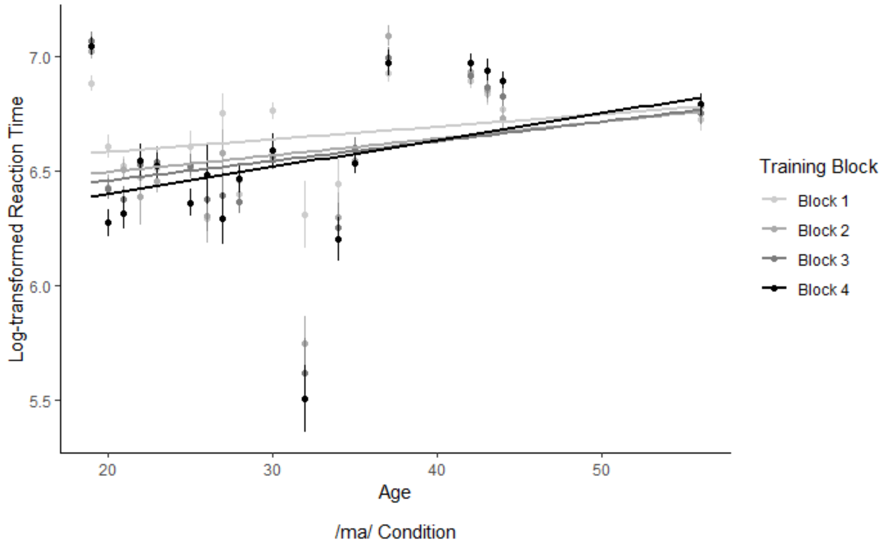

Figure 54. Log-transformed reaction times across training blocks in the /ma/ Condition.

Results from the /mi/ Condition indicated that reaction times did not significantly differ as a function of age ($X^2$ (1) = .69, *p* = .41). Figure 55 illustrates log-transformed reaction times as a function of age in the /mi/ Condition. Results from the /mi/ Condition continue to support the trend that, out of those that learn, reaction times from older participants tend to be slower overall than younger participants. Further, the oldest and the third oldest participants' results indicated learning and that category acquisition most likely occurred around the beginning of the third block.

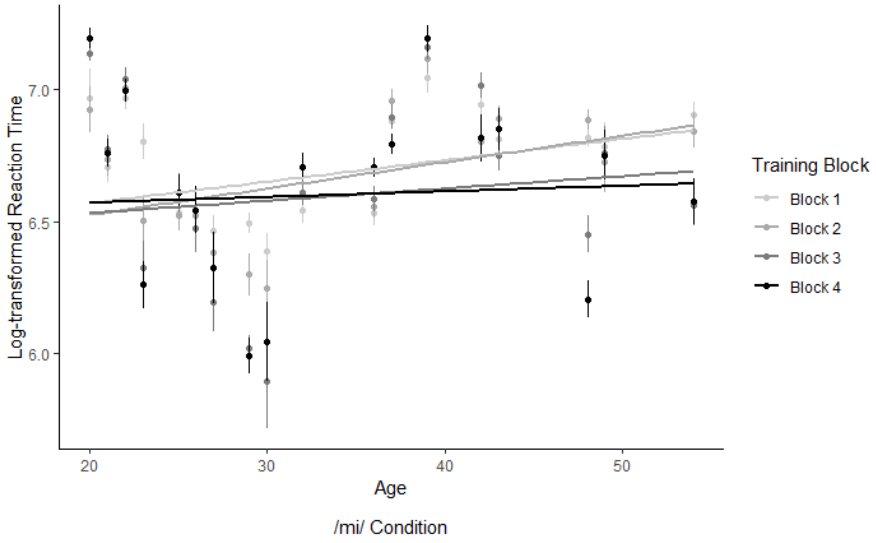

Figure 55. Log-transformed reaction times across age in the /mi/ Condition.

Results from the /mɯ/ Condition indicated that reaction times did not significantly differ as a function of age ($X^2$ (1) = .74, *p* = .39). Figure 56 illustrates log-transformed reaction times as a function of age in the /mɯ/ Condition. As in the /mi/ Condition, results from the /mɯ/ Condition continue to support the trend that, out of those that learn, reaction times from older participants tend to be slower overall than younger participants.

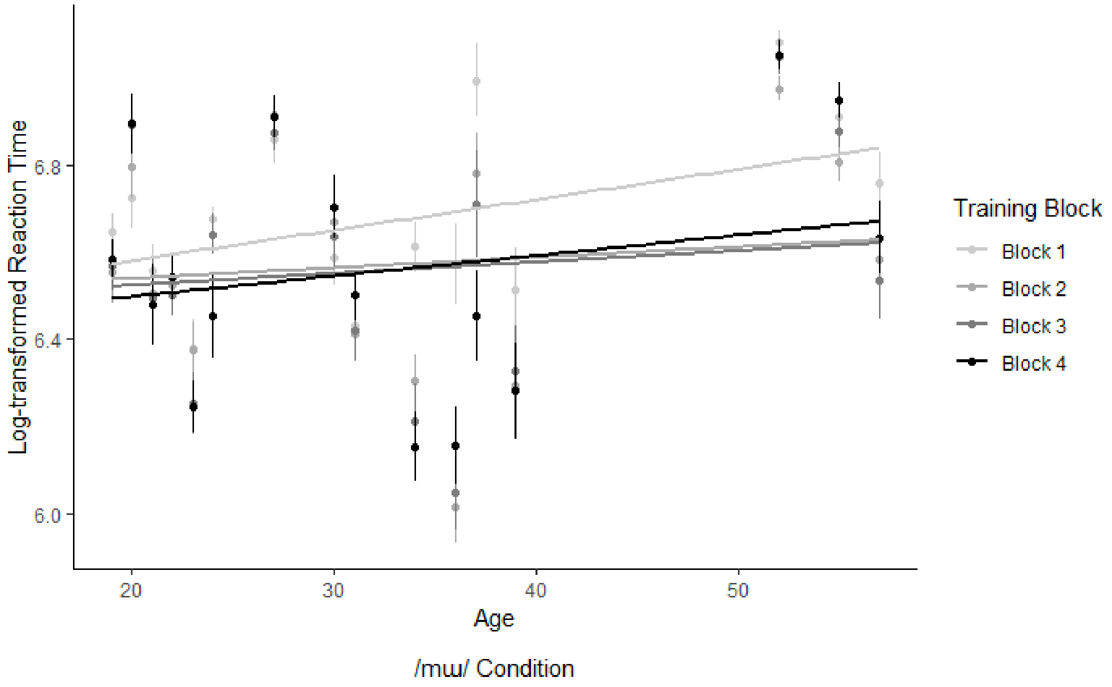

Figure 56. Log-transformed reaction times across age in the /mɯ/ Condition.

In Experiment 3 I measured the reaction times of participants across training blocks in three conditions. In all three conditions, reaction times became faster across training blocks, indicating that participants learned the novel tone categories and were able to use that learning to predict the locations of the visual targets. Results did not indicate that the /mɯ/ Condition differed from the other conditions. Statistically, when accounting for all participants, including those that learned and those that didn't learn, results indicated that age has no effect on reaction times during the experiment. Learning, then, can alter the baseline effect of age on the visual detection task, which was illustrated in the Control Condition in Experiment 2, and showed that reaction times in the task become slower among older participants. If we look only at those that learn in Figures (54-56), we see that the effect of age does tend to result in slower reaction times for older participants.

### 5.4.2 Generalization to new tokens and new talkers

As in the other experiments, Posttest 1 tested participants' ability to generalize to new tokens from the same talker, and Posttest 2 tested generalization to new talkers. The structure of both posttests is identical and both measure identification accuracy of the target tone category. If participants have learned the categories they should be able to accurately identify in which box the visual target should have appeared based solely on hearing the auditory stimuli, and therefore, their accuracy scores will be higher. Experiment 1 confirmed that participants that hear a single talker during training are able to accurately identify the four novel tone categories on Posttest 1. However, when they hear novel talkers on Posttest 2, they are less accurate. The Multi-talker Condition in Experiment 2 trained participants on multiple talkers, resulting in lower accuracy on Posttest 1, but accuracy on Posttest 1 did not differ from Posttest 2. In the present experiment it is expected that all conditions will result in the ability to generalize to novel tokens, measured in accuracy scores above chance on Posttest 1 and the ability to generalize to novel talkers, measured in accuracy scores above chance on Posttest 2. However, it is expected that, due to segmental familiarity, accuracy scores on both measures will differ for participants in the /mɯ/ Condition compared to the /ma/ Condition and /mi/ Condition.

#### *5.4.2.1 Analysis*

Accuracy scores for all conditions were measured on Posttest 1 and Posttest 2. For each condition, I compare accuracy scores on both posttests to chance using one sample t-tests. To test whether accuracy scores differ as a function of condition, I conduct model comparisons with and without condition for each posttest. To test whether there is a correlation between the learning measures, I conduct correlation tests between reaction times during training and accuracy scores at test for each condition. Finally, I conduct model comparisons to examine age as a fixed effect for all conditions on Posttest 1 and Posttest 2.

#### *5.4.2.2 Accuracy*

Figure 57 illustrates mean proportion correct scores with 95% confidence intervals for the /ma/ Condition, the /mi/ Condition, and the /mɯ/ Condition on Posttest 1 and Posttest 2. The figure suggests that participants in all conditions accurately identified the target categories above chance on Posttest 1 and on Posttest 2 and that accuracy across conditions likely did not differ.

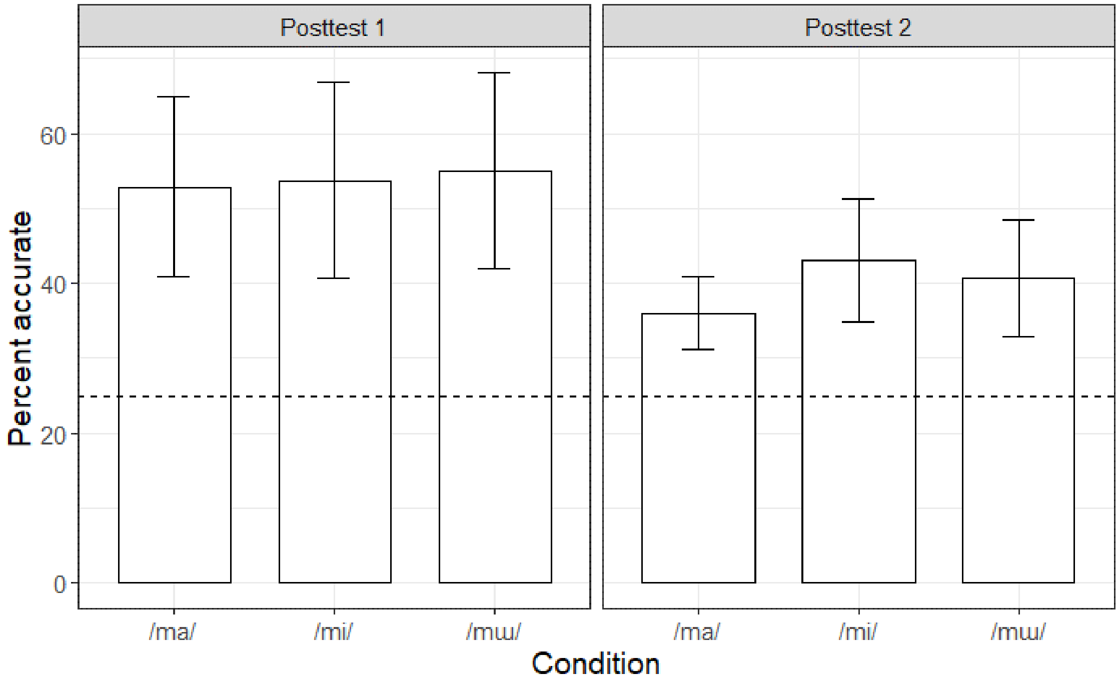

Figure 57. Mean proportion correct for all conditions on Posttest 1 and Posttest 2. Error bars represent 95% confidence intervals. The dashed line represents chance at 25%.

To test whether accuracy scores differed from chance, I examined accuracy scores within condition on Posttest 1 and Posttest 2. In the /ma/ Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24) = 4.80$, $p < .001$, ($M = 52.83$, $SE = 5.80$) and on Posttest 2, $V = 247$, $p < .001$, ($Mdn = 34.72$)[34]. In the /mi/ Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $V = 255$, $p < .001$, ($Mdn = 57.67$) and on Posttest 2, $V = 229$, $p < .001$, ($Mdn = 44.44$)[35]. In the /mɯ/ Condition participants were also able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $V = 247$, $p < .001$, ($Mdn = 34.72$)[36] and on Posttest 2, $t(24) = 4.13$, $p < .001$, ($M = 40.67$, $SE = 3.79$).

[34] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 2 ($W = .88$, $p = .008$).

[35] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 1 ($W = .87$, $p = .004$) and on Posttest 2 ($W = .84$, $p = .001$).

[36] On Posttest 1 a Shapiro-Wilk normality test indicated the data were not normally distributed ($W = .88$, $p = .007$), and therefore a Wilcoxon signed rank test was used.

To test whether accuracy scores differed across conditions on Posttest 1 and Posttest 2, I compared models with and without condition for each posttest. Results indicated that accuracy scores did not differ as a function of condition on Posttest 1 ($X^2$ (2) = .11, *p* = .95) or on Posttest 2 ($X^2$ (2) = 3.06, *p* = .22).

accuracy ~ condition + age + (1|participant)
accuracy ~ age + (1|participant)

Overall, participants in all three conditions accurately identified the target categories above chance on Posttest 1 and on Posttest 2, indicating that all conditions resulted in learning and that learning generalized to novel tokens on Posttest 1 and novel talkers on Posttest 2. A comparison of conditions on Posttest 1 indicated that participants in the /ma/ and /mi/ conditions did not more accurately identify the target categories than participants in the /mɯ/ Condition, indicating that less segmental familiarity in the /mɯ/ Condition did not result in less robust learning or generalization. As all three conditions produced equivalent results, the additions of the /mi/ Condition and the /mɯ/ Condition provide two internal replications of the results from the /ma/ Condition, adding confidence that the results in the /ma/ condition are not spurious.

During training, greater learning was measured through reaction times becoming faster across training blocks. At test, greater learning was measured through higher accuracy scores. It was expected that faster reaction times at the end of training would correlate with higher accuracy scores at test for all conditions. Figure 58 illustrates the correlation between reaction times on block 4 and accuracy scores on Posttest 1, suggesting a relationship between the two measures across all conditions.

Spearman's rho correlation coefficient[37] was used to assess the relationship between reaction times on training block 4 and accuracy scores on Posttest 1. The relationship between the two measures was significant in the /ma/ Condition (*r* = -.69, *p* < .001), the /mi/ Condition (*r* = -.67, *p* < .001), and in the /mɯ/ Condition (*r* = -.72, *p* < .001). The correlation between the two

[37] A Shapiro-Wilk normality test indicated the some of the data in the /mi/ Condition (*W* = .86, *p* = .004; *W* = .93, *p* = .10) and in the /mɯ/ Condition (*W* = .88, *p* = .007; *W* = .95, *p* = .22) were not normally distributed. Therefore, we conducted the non-parametric Spearman's test for all conditions. Although the data for the /ma/ Condition were normally distributed (*W* = .93, *p* = .08; *W* = .96, *p* = .37), Spearman's test was used for consistency. Pearson's correlation coefficient was also significant for the /ma/ Condition (*r*(23) = -.67, *p* < .001).

measures across conditions suggests that faster reaction times in training relates to better accuracy on the generalization test and that both measures reliably assess category learning.

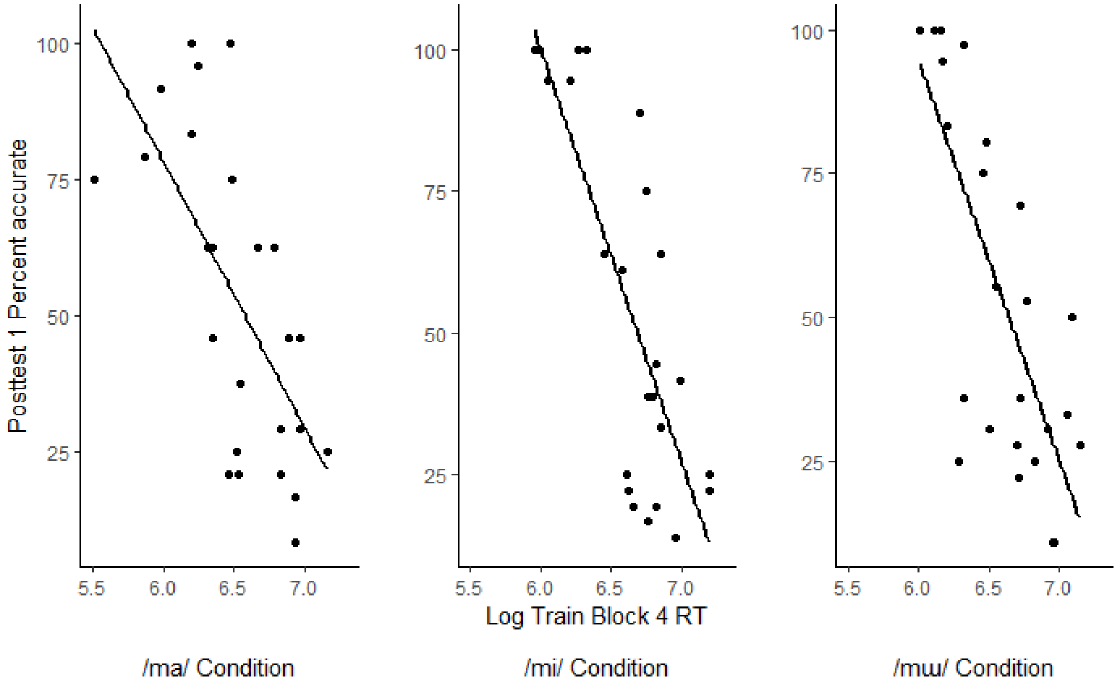

Figure 58. Relationship between two measures assessing category learning across conditions with log transformed reaction times on training block 4 on the x axis and accuracy scores on Posttest 1 on the y axis.

To test whether accuracy scores at test differed as a function of age for each condition, I compared models with and without age, and results indicated that accuracy scores did not significantly differ as a function of age in the /ma/ Condition on Posttest 1 ($X^2$ (1) = .44, *p* = .51) or on Posttest 2 ($X^2$ (1) = 3.62, *p* = .057). Accuracy scores did not significantly differ as a function of age in the /mi/ Condition on Posttest 1 ($X^2$ (1) = .23, *p* = .64) or on Posttest 2 ($X^2$ (1) = .66, *p* = .42). Further, accuracy scores did not significantly differ as a function of age in the /mɯ/ Condition on Posttest 1 ($X^2$ (1) = .84, *p* = .36) or on Posttest 2 ($X^2$ (1) = .51, *p* = .48).

accuracy ~ age + (1|participant)
accuracy ~ (1|participant)

Figure 59 illustrates accuracy scores on Posttest 1 and Posttest 2 as a function of age across conditions. The model comparison demonstrated that accuracy scores did not differ as a function of age. There does appear to be a trend across conditions for younger participants to have higher accuracy scores than older participants. The three conditions did not differ from each other. Therefore, to further investigate the relationship between age and accuracy scores, I aggregated scores across conditions.

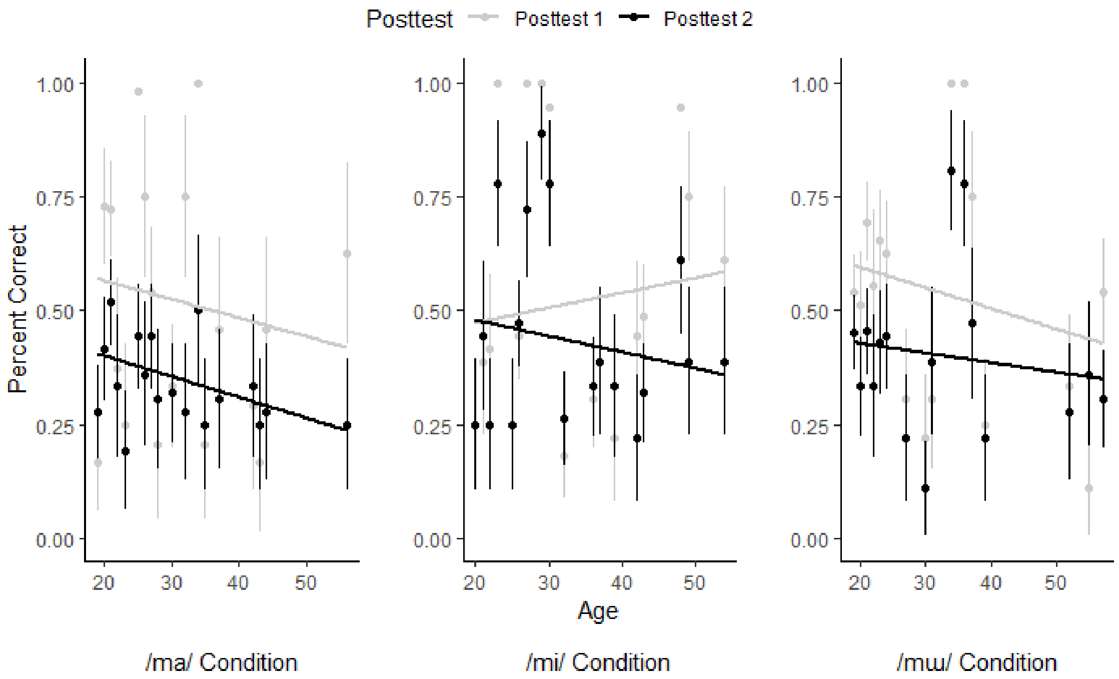

Figure 59. Accuracy scores on Posttest 1 and Posttest 2 across age in the /ma/ Condition, the /mi/ Condition, and the /mɯ/ Condition.

Figure 60 illustrates scores from all three conditions aggregated and suggests that there is a trend for accuracy scores to be higher for younger participants. However, a model comparison with and without age indicated that accuracy scores did not significantly differ as a function of age in the aggregated data on Posttest 1 ($X^2$ (1) = .55, *p* = .46) or on Posttest 2 ($X^2$ (1) = 3.06, *p* = .22).

accuracy ~ age + (1|participant)
accuracy ~ (1|participant)

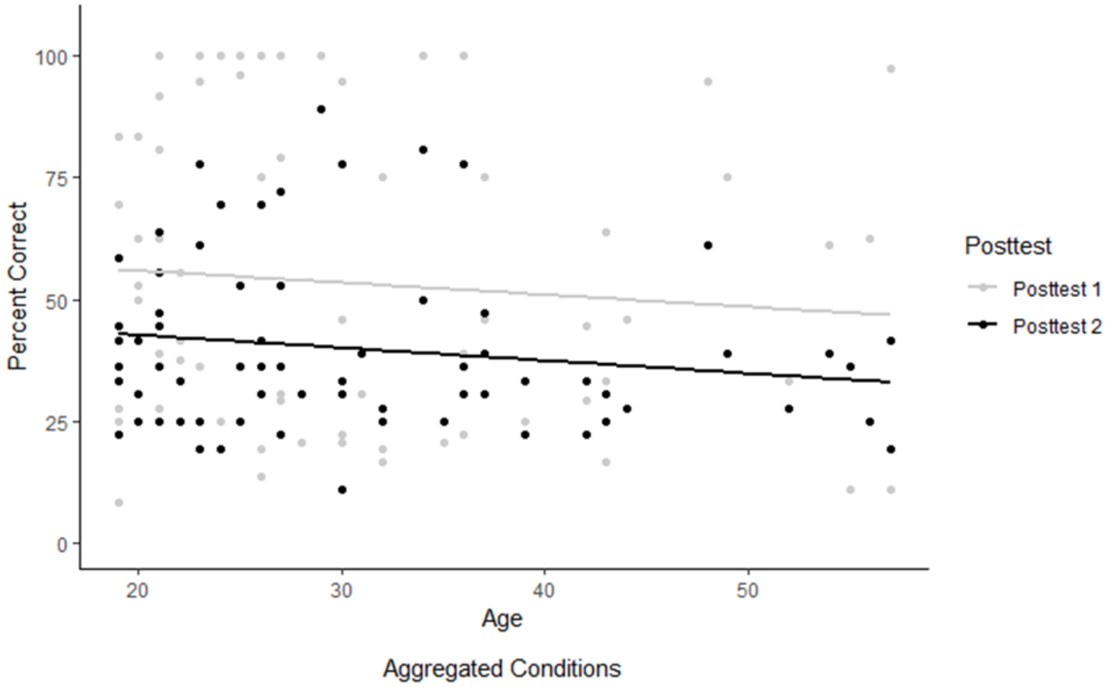

Figure 60. Accuracy scores on Posttest 1 and Posttest 2 across age with conditions aggregated.

## 5.5 Discussion

In Experiment 3, I examined the impact of segmental familiarity during training on incidental perceptual learning of novel tone categories. I compared three conditions that contained tokens produced in different syllables. Two conditions, the /ma/ Condition and the /mi/ Condition, were comprised of segments more familiar to the participants' language background experience. One condition, the /mɯ/ Condition, contained a segment unfamiliar to the participants' language background experience. It was expected that the two familiar conditions' results would not differ, but that a lack of familiarity would negatively impact learning in the /mɯ/ Condition. Results indicated that reaction times from participants in all conditions became faster across training blocks. Further, participants in all three conditions accurately identified the target categories above chance on Posttest 1 and on Posttest 2, indicating that all conditions resulted in learning and that learning generalized to novel tokens on Posttest 1 and novel talkers on Posttest 2. Overall, the three conditions did not differ from each other. Thus, segmental familiarity in the /mɯ/ Condition did not result in less robust learning or generalization than the familiar categories. Below I discuss the implications of these results for categorization and perceptual learning.

### 5.5.1 The effect of segmental familiarity on novel tone category formation

As discussed in Section 1.2 and Section 5.1, an important aspect of learning novel sound categories is the ability to attend to the salient acoustic features between the categories. Multiple factors, such as the phonotactic environment of the target sound, may inhibit attention to the target acoustic features (Guion & Pederson, 2007; Liu et al., 2011). Further, complex stimuli involving unfamiliar segmental and suprasegmental features may increase the attentional load during perceptual learning, thereby increasing the difficulty of learning the novel sound categories (Liu et al., 2011). As discussed in section 5.1, the neural processing of native and non-native segments differs, and this difference occurs with naïve listeners and experienced learners (Peltola et al., 2003).

Therefore, it was hypothesized that differences in the processing of familiar and unfamiliar segments might differentially impact the acquisition of novel tone categories carried by those segments. Specifically, it was expected that conditions with familiar segments, /ma/ and /mi/, may result in learning that differs from a condition with an unfamiliar segment in the primary tone bearing unit (e.g., /mɯ/). However, there were no significant differences between conditions containing familiar and unfamiliar segments. Below I discuss potential explanations for these results.

It is possible that the unfamiliar and familiar conditions resulted in the same amount of learning because listeners perceived and processed the unfamiliar segment, /ɯ/, as a familiar segment, such as /ə/. When a person learns sounds from a second language, they often attempt to map the sounds of the L2 onto the available acoustic spaces in their L1. Several speech perception models have been presented to account for this mapping. For example, the Perceptual Assimilation Model (PAM; e.g., Best, 1995; Best and Tyler, 2007), the Speech Learning Model (SLM; e.g., Flege, 1995), and the Second Language Linguistic Perception Model (L2LP; e.g., Escudero, 2005) all predict that L2 learners will try to map L2 sound categories to L1 categories that are closest to them in native acoustic space. Therefore, one possibility is that participants heard /ɯ/ and mapped it onto the acoustic space of an English vowel and thereby avoided processing difficulties that may arise from the unfamiliar segment. However, this seems unlikely. In the following experiment, Experiment 4, participants were required to produce the tokens they hear on each trial and the experiment recorded their productions. If participants are mapping /ɯ/ onto the acoustic space of an English vowel, then they would likely produce the

English vowel that they map /ɯ/ onto. However, participants primarily produced vowels unlike English vowels in an attempt to approximate /ɯ/.[38]

It is also possible that the unfamiliar segment in the /mɯ/ Condition did not disrupt novel tone category formation as expected due to differences in tone perception task difficulty. As discussed in Chapter 2 and in Section 5.4.1, the tone categories were selected to maximize differences between categories, and it is likely that the distinctiveness of each category made the novel tone category formation task in the current study easier than tasks in other studies that found effects of segmental familiarity on novel tone perception tasks. During novel tone category acquisition with synthesized tone categories, participants perform better at learning tone categories that are more distinct and struggle to learn categories that have greater featural overlap (Liu & Holt, 2011). Further, in novel tone discrimination tasks, the difficulty of the task can be impacted by phonotactic structure of the tokens and by the similarity of the target tones. For example, when participants attempt to discriminate novel tones in tokens that contain different segments, they perform worse than when segments are the same across tokens (Liu et al., 2011). Further, when the target tones are very similar, as in the Thai mid and low tones, native English participants have difficulty discriminating tones (Wayland & Guion, 2004), and discrimination ability across tones that are already difficult to discriminate can be further hindered by the presence of unfamiliar onsets (Wright & Baese-Berk, under review). Therefore, it is possible that the impact of unfamiliar segments on novel tone perception may be modulated by the difficulty of the task, which could be impacted by the distinctiveness of the tone categories and the composition of the carrier token. Therefore, it may be that the incidental tone learning task used in the current study is not particularly difficult compared to tasks such as explicit categorization and the discrimination of novel tones.

An important consideration in understanding the results from the current study in light of previous research is that novel tone category learning may differ from novel tone discrimination, and factors that impact one task may not impact the other task in the same way (Logan et al., 1991; Wayland & Li, 2008; Liu et al., 2011). Further, the impact of factors on novel tone learning during explicit category learning tasks may differ from learning during incidental category learning tasks (Lim et al., 2014). Currently, relatively little is known about the impact of

[38] This suggestion stems from a preliminary investigation of the acoustic data collected from participants in Experiment 4. Due to time constraints a deeper analysis of the data is not presented here.

segmental features on novel tone perception and whether results from explicit tone discrimination and category learning studies might be replicated during incidental learning. The results from Experiment 3 indicate that familiarity with the vowel in the carrier token did not impact incidental learning. It may be that unfamiliar onsets could impact incidental learning (Wright and Baese-Berk, under review) or that variable segments across tokens could impact incidental learning (Liu et al., 2011). Results from Experiment 2 suggest that it is likely that segmental or phonotactic variability across tokens would impact learning. In Experiment 2 variability from multiple talkers reduced learning reaction times and accuracy. It may be that the primary factor to consider during incidental auditory learning is variability, particularly with natural speech tokens. Speech categories are not unidimensional (Lisker, 1986). Therefore, during novel sound category learning, learners must generalize across multiple acoustic cues stemming from talkers, segments, and suprasegmental features (Liberman et al., 1967; Lim et al., 2014). Factors that add to this variability may be more likely to impact learning.

It was expected that the inclusion of unfamiliar segments would increase cognitive load, which would impact perceptual learning. During auditory perception many factors can result in an increase in the amount of effort participants must make to attend to the salient features of the stimuli, which is often referred to as “effortful listening”. For example, the presence of noise in the signal can result in challenges for the hearing impaired (Rabbitt, 1991) or for non-native listeners (Miller et al., 2009). Baese-Berk and Samuel (2016) posit that impairments associated with effortful listening may also arise during perceptual learning. Similarly, it was hypothesized that the presence of the unfamiliar segment may result in increased cognitive load and impairment in perceptual learning. There are theoretical reasons that may explain the lack of impairment in the present study. The COmpetition between Verbal and Implicit Systems model (COVIS; Ashby et al., 1998, 2011) was extended from visual to auditory perceptual domains (Chandrasekaran et al., 2014). COVIS postulates two learning systems for category learning, a reflective learning system and a reflexive learning system. The reflective learning system is explicit in formulating and testing rules during the categorization process using executive attention and working memory and is engaged during explicit learning paradigms. The reflexive learning system, which is engaged during incidental learning paradigms, is implicit in associating stimuli with distinct regions in perceptual space using reinforcing feedback, such as the feedback found in current study.  The two types of learning engage different neural structures, resulting in differences in cognitive load. A key difference is that reflective learning requires the use of working memory and executive attention. Reflexive learning does not. As mentioned above, the impact of a factor on novel sound category acquisition during explicit perceptual tasks may differ from incidental tasks. Therefore, it is possible that the inclusion of unfamiliar segments during novel tone category learning could result in different impacts on the two types of learning. Thus, one hypothesis is that the lack of familiarity may increase the processing challenge for the neural systems that are engaged by working memory and executive attention during reflective learning but not increase the challenge for the systems engaged by reflexive learning tasks.

The use of the unfamiliar /ɯ/ in the /mɯ/ Condition in Experiment 3 did not increase variability across tokens or across trials. Therefore, due to the consistency of the acoustic features of the vowel, even though they were unfamiliar, it was likely that participants were still able to attend to the salient features of the tone categories. Overall, there are few studies that research the interaction between segmental and suprasegmental features during novel tone perception. Results from Experiment 3 add to this growing body of literature, indicating that during incidental learning, learners are equally capable of attending to salient tone category features in tokens that contain familiar and unfamiliar segments. Further, future work will investigate the hypothesis presented here, that phonotactic or segmental variability across tokens or across trials may impact novel tone category learning.

### 5.5.2 Learning differences as a function of age

In Experiment 3 participants' ages ranged from 19 to 57, permitting further observation of the effect of age on the incidental learning of novel tone categories. The Control Condition in Experiment 2, where participants were not able to learn across training blocks, provided a clear linear effect of age on the incidental learning task, with older participants' responses being significantly slower than younger participants' responses. In Experiment 3, results indicated that age had no effect on reaction times during the experiment, which suggests that the incidental learning task can alter the baseline effect of age on reaction times during the incidental learning task. However, this result is confounded by the inclusion of those that learned and became faster across blocks and those that didn't learn and remained at baseline across blocks. If we control for those that learned or didn't learn, there is still a trend for older participants to have slower reaction times.

As mentioned, the /ma/ Condition in Experiment 3 was the Variable Token Condition in Experiment 1 and therefore had the same results regarding age. Further, results from the /mi/ Condition and /mɯ/ Condition in Experiment 3 did not differ from the /ma/ Condition. Therefore, to enable a test of a larger sample size (n=75) for an effect of age on accuracy scores in Experiment 3, I aggregated scores across conditions. Although there was a trend for younger participants to be more accurate, statistical analyses did not indicate a difference across the lifespan. The result that older participants can learn novel tone categories as well as younger participants may be surprising. It may be expected, due to age related cognitive decline, that older participants would perform worse than younger participants (Clinard et al., 2010; Anderson et al. 2012; Skoe et al., 2015). Research regarding neural plasticity across the lifespan suggests that we might see age related deficits in learning. Plasticity, defined as the brain's ability to alter its functional and behavioral capacities by implementing lasting structural changes, decreases across the lifespan. The transition from childhood to adulthood is commonly thought to result in a suppression of plasticity in the human brain (Lindenberger & Lovden, 2019). However, a significant decline in novel tone category learning across the lifespan was not supported by the results of Experiment 3. It may be that the neural processes and mechanisms that are engaged through incidental learning are not impacted by cognitive decline as extensively as processes and mechanisms engaged through other learning paradigms.

These results build on and support findings from Maddox et al. (2013), which provided a first look at the effect of age on the ability to form novel speech sound categories through incidental learning. When learning novel sound categories, older participants' learning ability suffered under reflective, rule-based learning conditions. However, under reflexive, implicit learning conditions, such as the incidental learning paradigm in the current study, older participants performed as well as younger participants. As previously mentioned, reflective learning requires an allocation of working memory and utilizes different neural structures from reflexive learning. Therefore, age-related neural decline does not seem to impact reflexive learning in the same way that it impacts reflective learning. This may be in part due to the enhancement of corticostriatal synaptic plasticity by the reinforcement learning that is a part of the reflexive learning used in the current study's paradigm (see Reynolds & Wickens, 2002).

However, I do want to point out potential differences between younger and older participants that may be beneficial to pursue in future research. Although statistical analyses indicated that accuracy did not differ as a function of age, there was a slight trend for younger participants to do better on generalizing to novel tokens and novel talkers. Further, it seems clear that older participants struggled to do well when generalizing to novel talkers on Posttest 2. Across the three conditions, only one participant over forty scored above fifty percent accuracy on Posttest 2. This may be due to the smaller numbers of participants over forty. However, it also might be that there are age related differences that impact the ability to generalize more broadly across categories.

## 5.6 Conclusion

In Experiment 3 we examined the impact of segmental familiarity during training on the incidental perceptual learning of novel tone categories by comparing three conditions that contained tokens produced in different syllables. Two conditions, the /ma/ Condition and the /mi/ Condition, were comprised of segments more familiar to the participants' language background experience. One condition, the /mɯ/ Condition, contained a segment unfamiliar to the participants' language background experience. By examining conditions with familiar and unfamiliar segments, we tested potential impacts to perceptual learning from increased attentional load stemming from the processing of novel segments. We demonstrated from identical results across three conditions that the presence of an unfamiliar vowel in the auditory stimuli did not impact the incidental formation of novel tone categories. That is, the additional complexity from processing unfamiliar segmental features did not result in reduced learning of the target tone categories. As the results from the three conditions in Experiment 3 did not differ from each other, they provide two internal replications of the study. We also combined results to investigate a potential linear effect of age on novel tone category learning. That is, did accuracy results differ as a function of age. Statistically, there was no difference across the lifespan, meaning that older adults learned as well as younger adults. However, we did note a potential trend for younger adults to be more accurate after training.

# VI. PRODUCTION DURING PERCEPTUAL LEARNING

## 6.1 Introduction

In Experiment 4, we examine the impact of production during training on incidental perceptual learning of novel tone categories. We also examine the impact of segmental familiarity in the learners' productions during training on novel tone category learning. We compare three conditions, a Perception Only Condition that does not contain a production component and two production conditions where participants produce the token on each trial. The /ma/ Production Condition is comprised of segments more familiar to the participants' language background experience. The /mɯ/ Production Condition contains a segment unfamiliar to the participants' language background experience.

We expect that results from the two production conditions will differ from the Perception Only Condition. That is, we predict that the additional production by learners during perceptual learning will result in reduced learning compared to the Perception Only Condition. Further, we expect that the lack of segmental familiarity in the /mɯ/ Production Condition will negatively impact perceptual learning compared to the /ma/ Production Condition.

### 6.1.1 The effect of production on perceptual learning

In speech perception and production, varying views have arisen regarding the effect of production on perceptual learning. Some studies suggest that production during perceptual learning improves learning. Other studies find that production during perceptual learning hinders learning. This area is especially important to classroom language learning situations where teachers must decide whether they will have students repeat words as they hear them. Teachers have been encouraged to have their students repeat utterances as a means of moving through the zone of proximal development (Vygotsky, 1978) towards an ultimately correct production (Duff, 2000).

One logical assumption about the impact of production on perceptual learning is that production during learning would result in improved perceptual ability (Leach & Samuel, 2007; Baese-Berk, 2010). This expectation arose from theories of perception, including direct realism and motor theory, which posit that the underlying basis of perception are the gestures used to produce the sounds (Fowler, 1986; Best, 1995). Thus, perception and production were thought to be very closely connected and the practice of producing sounds should reinforce perceptual learning. In support of the idea that production enhances the learning of words, Gathercole and Conway (1988) found that reading and producing words improved retention beyond reading and hearing the words. MacLeod et al. (2010) also studied and confirmed the "production effect", where producing a word aloud during study improves retention of the word. Forrin et al. (2012) found that the production effect was stronger for full productions. Whispered and mouthed productions were not as beneficial. Zamuner et al. (2016) found that production during the learning of non-words enhances recognition at test and conclude that production is needed during perceptual learning to establish a bidirectional link between the perception and production systems.

Taken together, studies on the benefit of production during the learning of words seem to provide robust evidence for the production effect. However, when learners begin to learn words with phonological systems that differ from their L1, results are not the same. Dahlen and Caldwell-Harris (2013) tested English speaking adults' learning of Turkish words. They found that listeners who rehearsed the words sub-vocally did as well or better than listeners who vocalized the words during learning. They concluded that overt vocalization may actually detract from learning as attention becomes divided between processing the sounds and performing the vocalizations. Results from other studies support this conclusion. Kaushanskaya and Yoo (2011) directly tested the production effect on learning novel words with familiar phonological structures (i.e. structures found in the L1) and on unfamiliar words. They found the production effect for phonologically familiar words, but for words with phonological features that differed from the L1, subvocal rehearsal lead to better recall and recognition than vocal rehearsal. They concluded that there appears to be distinct cognitive processes for each rehearsal type. These results coincide with conclusions from Feldman and Healy (1998) — novel words with L1 phonological structures are easier to learn than novel words with novel structures. Leach and Samuel (2007) also found that production during perception training hindered the learning of words with novel segments. Baese-Berk and Samuel (2016) tested the effect of production hindering the perceptual learning of phonologically novel words by seeing if it was simply production that hindered learning or if it was a production of the target token. They concluded that it was the production of the target token itself, not production in general that created the greatest hindrance to learning. However, producing unrelated items still creates some hindrance to learning. They also found that the disruption can be lessened with experience to the target phonological structure. Taken together these studies suggest that the effect of production during perceptual training differs based on the familiarity or lack of familiarity of the target word's phonological structure to the L1.

### 6.1.2 Current experiment

Experiment 4 contains three conditions: Perception Only Condition, /ma/ Production Condition, /mɯ/ Production Condition. By comparing a Perception Only Condition with production conditions, we investigate the impact of production during incidental perceptual learning on the formation of novel tone categories. We also examine the impact of segmental familiarity on perceptual learning by comparing a production condition that contains familiar segments with a production condition that contains an unfamiliar segment.

It is expected that production of the tokens on each trial in the production conditions will hinder learning. Participants in the production conditions may still be able to learn, but it is expected that learning will be reduced compared to the Perception Only Condition. Further, it is expected that the addition of an unfamiliar segment in the /mɯ/ Production Condition will result in a greater inhibition to learning than the /ma/ Production Condition.

## 6.2 Methods

### 6.2.1 Participants

As in the other experiments, participants were recruited online via Prolific. All participants self-identified as being monolingual English speakers and identified as being native English speakers from America, Canada, the United Kingdom, South Africa, Australia, or New Zealand. Participants that reported significant language learning experience, that reported hearing impairments, or that did not use the right equipment (headphones and an external mouse) were excluded from the study.

In Perception Only Condition, where participants heard a single talker produce tokens using the syllable /ma/ during training and did not produce the tokens during training, 29 participants were recruited[39]. Four participants were excluded for using the wrong equipment or for hearing impairments, leaving 25 participants (13 female, 11 male, 1 non-binary). Participants spoke a variety of English dialects (6 American, 2 Australian, 14 British, 1 Canadian, 1 Irish, and 1 NA)[40]. Ages ranged from 19 to 56 with a mean of 29.08 and standard deviation of 9.45[41].

In the /ma/ Production Condition, where participants heard a single talker produce tokens using the syllable /ma/ and produced the tokens on each trial during training, 26 participants were recruited. One participant was excluded for language experience, leaving 25 participants (14 female, 11 male). Participants spoke a variety of English dialects (2 American, 20 British, 1 Canadian, 2 South African). Ages ranged from 19 to 66 with a mean of 30.88 and standard deviation of 11.87.

In the third condition, where participants heard a single talker produce tokens using the syllable /mɯ/ and produced the tokens on each trial during training, 25 participants were recruited (15 female, 10 male). No participants were excluded. Participants spoke a variety of English dialects (1 American, 20 British, 1 Canadian, 1 Irish, 1 New Zealand, 1 Scottish). Ages ranged from 20 to 55 with a mean of 34.64 and standard deviation of 11.13. All participants were paid for their participation.

### 6.2.2 Stimuli

Stimuli in experiment 4 had the same composition as the stimuli used in Experiment 3. In each condition in experiment 4, the set of five tokens within trial contained random tokens, constructed as described in experiment 1. In Experiment 4, for the three conditions, all auditory tokens for training and for Posttest 1 came from Talker A. In the Perception Only Condition,

[39] The /ma/ Condition in the present experiment is the Variable Token Condition from Experiment 1, the Single Talker Condition from Experiment 2, and the /ma/ Condition from Experiment 3. Therefore, descriptions of the /ma/ Condition in the present experiment are a restatement of details from Experiment 1, Experiment 2, and Experiment 3. Primary differences in the description of the /ma/ Condition in the present chapter arise from the differences in the comparisons made across the experiments. Experiment 1 compared token variability within and across trials. Experiment 2 examined talker variability across trials. Experiment 3 compared segmental variability across conditions. Experiment 4 examines the impact of production during training on the perceptual formation of novel tone categories.

[40] It is not expected that experience with specific English dialects would aid in novel tone category acquisition over other dialects. English dialects do not use F0 information contrastively at the lexical level. Further, experience with other regional languages used in proximity to the specific dialect should not be a factor as participation was limited to those that identified as being monolingual English speakers.

[41] Age is considered as a covariate during analysis and is reported in the results.

participants heard tokens produced in the syllable /ma/. In the /ma/ Production Condition, participants heard tokens produced in the syllable /ma/. In the /mɯ/ Production Condition, participants heard tokens produced in the syllable /mɯ/. In each condition, half of the exemplars of each category from Talker A were used for training, and half of the exemplars were used to test generalization of learning to new exemplars on Posttest 1. As in Experiment 3, stimuli used in Posttest 2 came from Talker D, Talker E, and Talker F.

## 6.3 PROCEDURE

The procedure for the Perception Only Condition in Experiment 4 is the same as the procedure for the other experiments. In the two production conditions participants also recorded themselves producing the tokens that they heard on each trial during training. Further, the two production conditions differ in the stimuli used, with /ma/ tokens used in the /ma/ Production Condition and /mɯ/ tokens used in the /mɯ/ Production Condition.

In Experiment 4, three groups of participants were exposed to four novel Thai tone categories through an incidental learning paradigm. Participants went through four training blocks with forty-eight trials in each block. Then, Posttest 1 tested generalization to novel tokens from the same talker(s) over thirty-six trials, and Posttest 2 tested generalization to novel talkers over thirty-six trials. Posttest 3 tested production of the tone categories over thirty-six trials. Finally, participants completed a language background questionnaire.

### 6.3.1 Training

Participants in each condition were trained with the incidental paradigm described in Experiment 1. On each trial participants heard five sounds and then clicked on a visual target, an ‘X’, that appeared in one of four boxes. Participants were trained across four training blocks with forty-eight trials in each block. For all conditions, auditory stimuli in each trial consisted of five concatenated exemplars. The concatenations were randomly selected prior to subject running. However, the presentation of trials was randomly selected by the experiment. In each condition, training was composed of six different concatenations of each tone category from Talker A for a total of twenty-four trials (6 concatenations X 4 tones X 1 talker). These twenty-four trials were duplicated on each training block for a total of forty-eight trials per block.

The /ma/ Production Condition and /mɯ/ Production Condition differed from the Perception Only Condition in Experiment 4 and from all other conditions in the previous experiments. On every trial during training, participants in the two production conditions went through the trial exactly as all the other conditions. However, after clicking on the visual target, two buttons appeared. One button had a microphone icon on it and one button had a stop icon on it. They clicked on the microphone button to start the recording. They then produced the token that they had heard on the trial a single time. Then they clicked on the stop button to stop the recording. After clicking on the stop button the target in the middle of the screen appeared to prompt them to move their mouse cursor back to the center of the screen.

For all conditions, reaction times from the end of the auditory stimuli to the participant's selection of the visual target were measured to examine learning across the four training blocks. It is expected that faster reaction times across training blocks will occur for those that learn the target tone categories and that faster reaction times will correlate with performance at test. Further, mouse tracking was conducted to examine changes in decision space over time as participants acquire the tone categories[42].

### 6.3.2 Testing

The testing procedure for Experiment 4 was the same as the procedure for the other experiments. Participants heard five sounds and then saw four boxes appear without a visual target. They then chose which box the target should appear in. Posttest 1 tested generalization to novel tokens and Posttest 2 tested generalization to novel talkers. Posttest 1 and Posttest 2 were the same for all conditions.

#### *6.3.2.1 Posttest 1: Generalization to new tokens*

Posttest 1 trials for each condition were composed of three different concatenations of each tone category from Talker A for a total of twelve trials (3 concatenations X 4 tones X 1 talker). These twelve trials were repeated three times on Posttest 1 for a total of thirty-six trials.

[42] An analysis of mouse tracking data is not included in the dissertation. Future analyses and description of the current work will analyze and consider mouse tracking data and report results.

#### *6.3.2.2 Posttest 2: Generalization to new talkers*

Posttest 2 trials for all conditions were composed of three different concatenations of each tone category from each Talker D, Talker E, and Talker F, for a total of thirty-six trials (3 concatenations X 4 tones X 3 talkers).

#### *6.3.2.3 Posttest 3: Production of the tone categories*

Experiment 4 also contained a third posttest, which was conducted in the same way as Experiment 1. Participants saw the visual target appear in one of the four boxes and recorded themselves saying the target tone with the syllable /ma/. Thirty-six trials were conducted[43].

## 6.4 RESULTS

### 6.4.1 Training reaction times

Experiment 1 tested the impact of token variability on novel tone category learning, finding that token variability within trial resulted in more robust learning than token variability across trials. Experiment 2 tested the impact of talker variability on novel tone category learning, finding that a single talker during training resulted in more robust learning than multiple talkers during training. Experiment 3 tested the impact of segmental familiarity on novel tone perception, finding that the lack of familiarity with the tone bearing segment did not impact novel tone category learning. In each condition in Experiment 4 all trials contain variable auditory tokens, and auditory tokens during training are from a single talker. In the present experiment I examine the impact of production during perceptual learning on novel tone category learning and test whether that impact is modulated by familiarity with the tone bearing segment.

Experiment 1, Experiment 2, and Experiment 3 found that participants that learn the target tone categories have reaction times that get faster across training blocks. It is expected that participants in the Perception Only Condition, will learn the tone categories and will have faster reaction times across training blocks. It is also expected that production during perceptual learning in the two production conditions will negatively impact perceptual learning, resulting in reaction times that are slower across training blocks than the reaction times in the Perception

[43] An analysis of production data is not included in the dissertation. Future analyses and description of the current work will analyze and consider production data and report results.

Only Condition. Further, it is expected that reaction times will be slower across training blocks for the /mɯ/ Production Condition than the /ma/ Production Condition.

#### *6.4.1.1 Analysis*

As in the previous experiments, visual target detection times were measured from the end of the auditory stimuli to the time the participant clicked on the visual target. Reaction times greater than 1,500 ms were excluded from analyses. For each condition, I compare reaction times across training blocks by comparing a full model and a reduced model without training block. I then conduct contrast coded linear mixed-effects regressions to compare each training block to the subsequent training block to examine changes in reaction times from block to block. Further, I compare reaction times across training blocks across the three conditions by comparing a full model with an interaction between condition and training block and a reduced model without an interaction, followed up by post-hoc comparisons of each condition with each other condition. Finally, as differences in age can affect learning and hearing ability (Kiessling et al., 2003; Clinard et al., 2010), I conduct model comparisons to examine age as a fixed effect.

#### *6.4.1.2 Reaction Times*

Results indicated that reaction times from participants in the Perception Only Condition and the /ma/ Production Condition became faster across training blocks, but reaction times from participants in the /mɯ/ Production Condition did not become faster across training blocks. Figure 61 illustrates log-transformed reaction times across training blocks for the Perception Only Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block.

To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the Perception Only Condition ($X^2$ (3) = 114.05, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

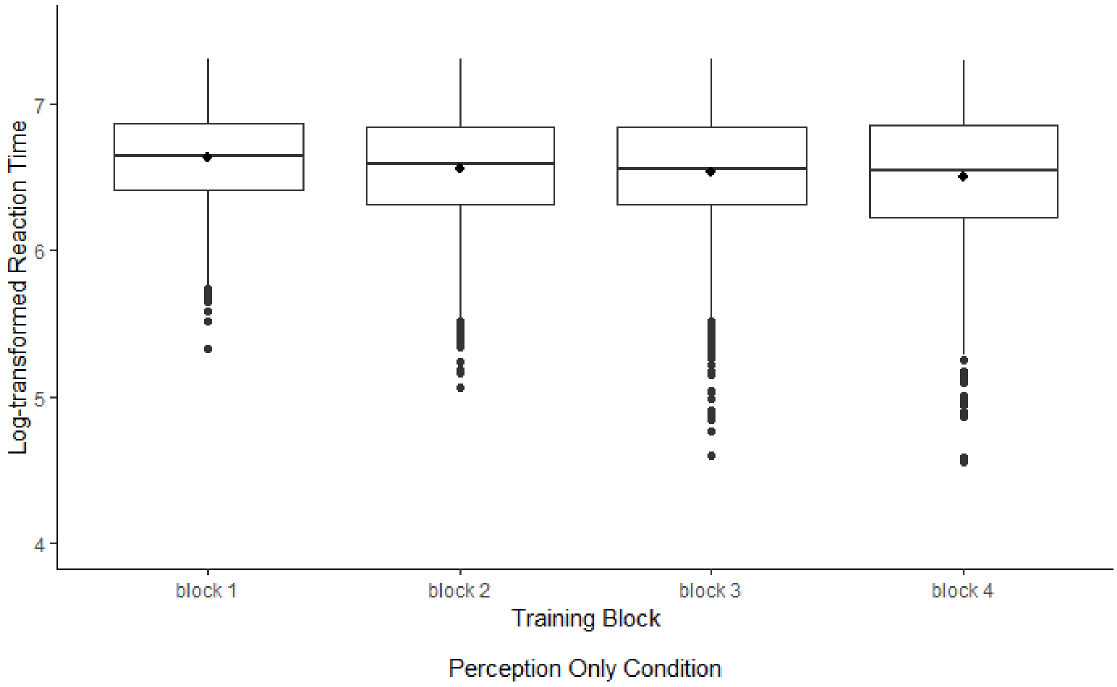

Figure 61. Log-transformed reaction times across training blocks in the Perception Only Condition.

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.63, *SD* = .31) were significantly slower than block 2 (*M* = 6.56, *SD* = .39; β = -.065, *t* = -5.55, p < .001), reaction times in block 2 did not differ from block 3 (*M* = 6.54, *SD* = .42; β = -.022, *t* = -1.89, p = .06), and reaction times in block 3 were significantly slower than block 4 (*M* = 6.51, *SD* = .47; β = -.012, *t* = -2.98, p = .003).

Figure 62 illustrates log-transformed reaction times across training blocks for the /ma/ Production Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block. Figure 62 suggests that, as a whole, participants' reaction times in the /ma/ Production Condition may have become faster across training blocks. To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time significantly differed as a function of training block in the /ma/ Production Condition ($X^2$ (3) = 120.94, *p* < .001).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ age + (1|participant)

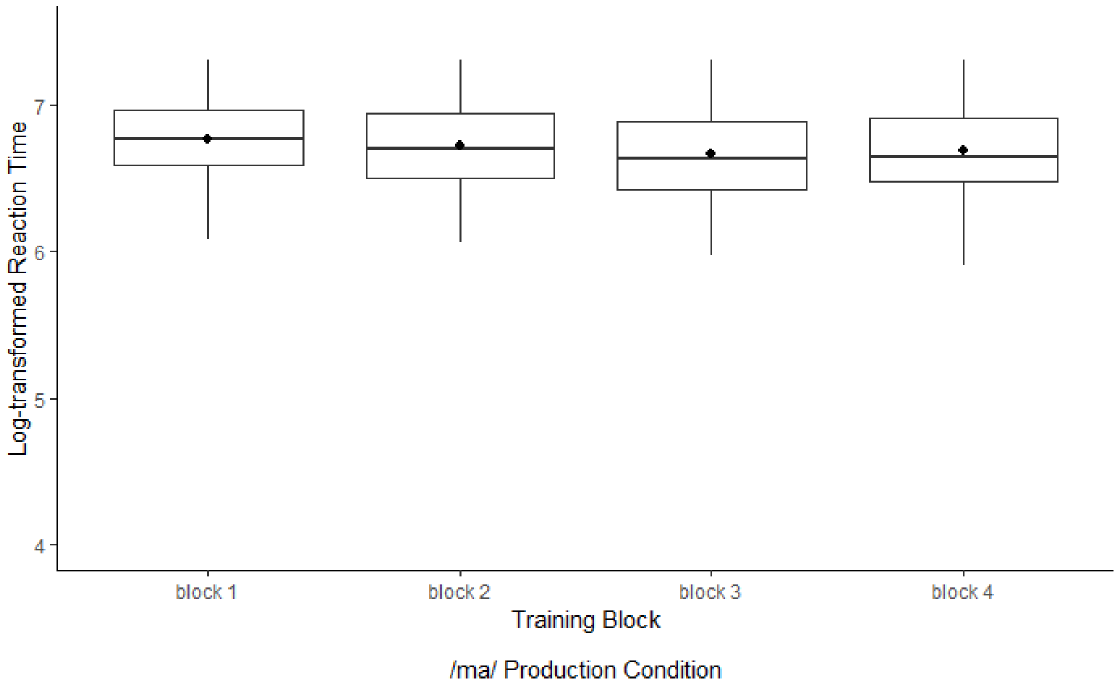

Figure 62. Log-transformed reaction times across training blocks in the /ma/ Production Condition.

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.77, *SD* = .26) differed significantly from block 2 (*M* = 6.72, *SD* = .28; β = -.042, *t* = -4.70, p < .001), reaction times in block 2 differed significantly from block 3 (*M* = 6.66, *SD* = .30; β = -.051, *t* = -5.80, p < .001), and reaction times in block 3 differed significantly from block 4 (*M* = 6.69, *SD* = .29; β = .023, *t* = 2.60, p = .009).

Figure 63 illustrates log-transformed reaction times across training blocks for the /mɯ/ Production Condition. The four boxplots in each of the three charts represent the distribution of reaction times for each block, and the dots in the boxes represent the mean reaction time for the specific block. Figure 63 suggests that as a whole, participants' reaction times in the /mɯ/ Production Condition may not have become faster across training blocks. To test whether reaction times differed as a function of training block, I compared models with and without training block, controlling for participant age, and results indicated that reaction time did not significantly differ as a function of training block in the Multi-talker Condition ($X^2$ (3) = 2.82, *p* = .42).

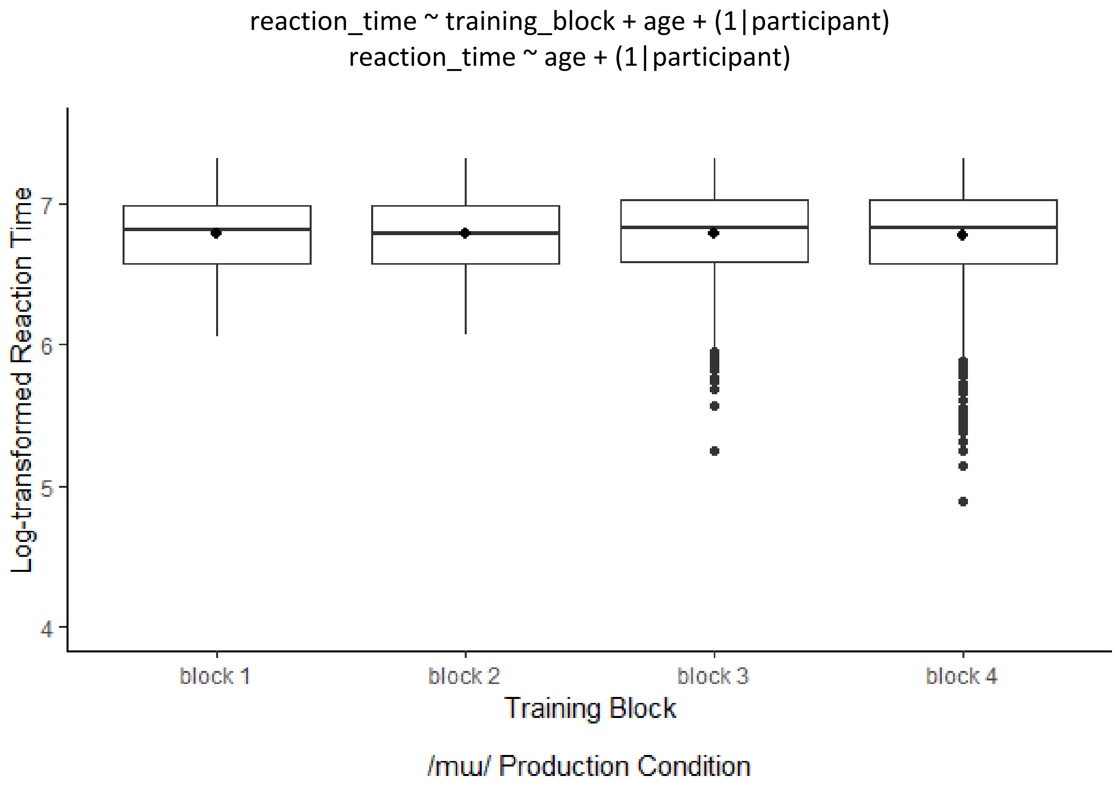

Figure 63. Log-transformed reaction times across training blocks in the /mɯ/ Production Condition.

A contrast coded linear mixed-effects regression comparing block 1 with block 2, block 2 with block 3, and block 3 with block 4 indicated that reaction times in block 1 (*M* = 6.79, *SD* = .27) did not differ from block 2 (*M* = 6.78, *SD* = .27; β = -.004, *t* = -.48, p = .63), reaction times in block 2 did not differ from block 3 (*M* = 6.79, *SD* = .31; β = .014, *t* = 1.54, p = .12), and reaction times in block 3 did not differ from block 4 (*M* = 6.77, *SD* = .35; β = -.012, *t* = -1.34, p = .18).

I compared reaction times across the three conditions. Figure 64 illustrates mean reaction times across training blocks for each condition with whiskers illustrating 95% confidence intervals. Table 8 provides the means and standard deviations of response times for the three conditions. It was expected that reaction times would be the fastest in the Perception Only Condition and that reaction times in the /ma/ Production Condition and the /mɯ/ production condition would be slower. Further, it was expected that reaction times from participants in the /mɯ/ Production Condition would be slower than reaction times from participants in the /ma/ Production Condition. As illustrated in Figure 64 and described in Table 8, the change in reaction times across training blocks differed according to our expectations. Reaction times in the Perception Only Condition were the fastest, followed by the /ma/ Production Condition, and the /mɯ/ Production Condition was the slowest.

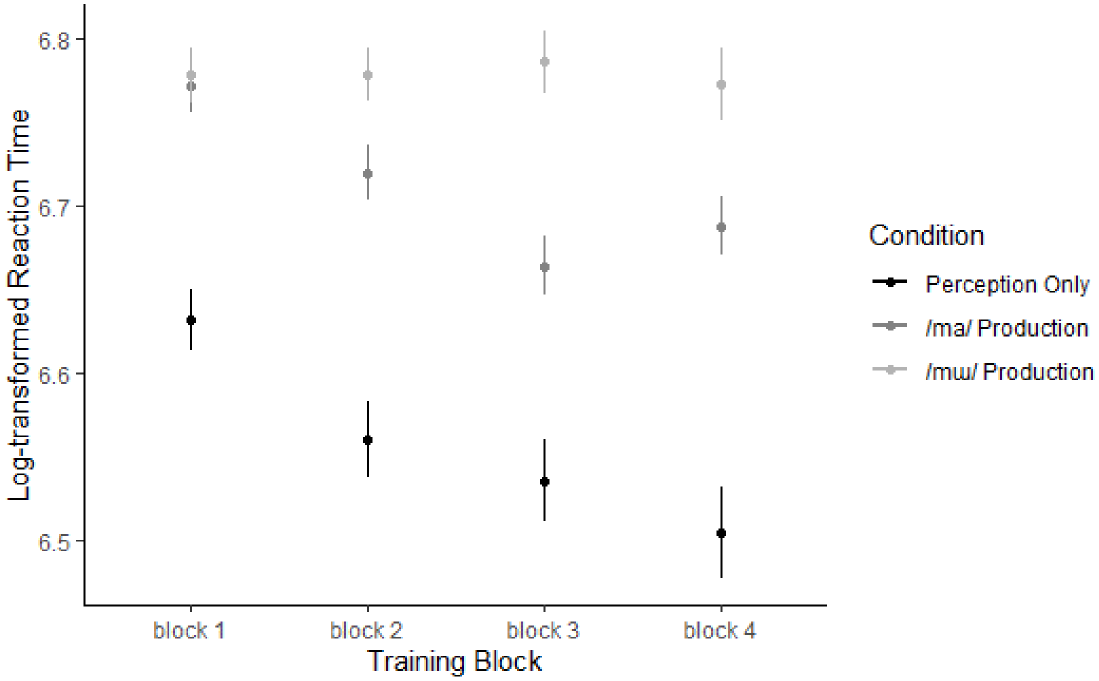

Figure 64. Log-transformed mean reaction times across training blocks for the Perception Only Condition, the /ma/ Production Condition, and the /mɯ/ Production Condition. Error bars represent 95% confidence intervals.

Table 8. Summary statistics for reaction times for the Perception Only Condition, the /ma/ Production Condition, and the /mɯ/ Production Condition

| Condition | Block 1 (mean, SD) | Block 2 (mean, SD) | Block 3 (mean, SD) | Block 4 (mean, SD) |
|---|---|---|---|---|
| Perception Only | 6.63, .31 | 6.56, .39 | 6.54, .42 | 6.51, .47 |
| /ma/ Production | 6.77, .26 | 6.72, .28 | 6.66, .30 | 6.69, .29 |
| /mɯ/ Production | 6.79, .27 | 6.78, .27 | 6.79, .31 | 6.77, .35 |

To test whether reaction times differed across conditions, I compared models with and without an interaction between condition and training block. Results indicated that reaction time differs across training blocks as a function of condition ($X^2$ (6) = 106.17, *p* < .001).

reaction_time ~ condition * training_block + age + (1|participant)
reaction_time ~ condition + training_block + age + (1|participant)

Bonferroni corrected post-hoc comparisons revealed that reaction times in the Perception Only Condition did not differ from the /ma/ Production Condition (β = -.148, SE = .067, *z* = -2.22, *p* = .08), but they did differ from the /mɯ/ Production Condition (β = -.173, SE = .068, *z* = -2.55, *p* = .03). The /ma/ Production Condition did not differ from the /mɯ/ Production Condition (β = -.026, SE = .067, *z* = -.38, *p* = 1).

By comparing reaction times across training blocks as a function of condition, I tested the impact of production during perceptual learning on novel tone category formation. In the Perception Only Condition, reaction times from participants became faster across training blocks, indicating tone category formation occurred. Reaction times in the /ma/ Production condition were not as fast as those in the Perception Only Condition, but they did get faster across training blocks, indicating some learning of the tone categories likely occurred. By contrast, reaction times in the /mɯ/ Production Condition did not get faster across training blocks, suggesting a potential impact of segmental familiarity, with the production of unfamiliar segments negatively impacting the perceptual formation of novel suprasegmental categories.

I also tested whether reaction times differed as a function of age in each condition by comparing models with and without age, controlling for training block. Results from the Perception Only Condition indicated that reaction times did not significantly differ as a function of age ($X^2$ (1) = 1.55, *p* = .21).

reaction_time ~ training_block + age + (1|participant)
reaction_time ~ training_block + (1|participant)

Figure 65 illustrates log-transformed reaction times as a function of age in the Perception Only Condition. Mean reaction times across blocks for each participant are illustrated as dots with error bars illustrating 95% confidence intervals. If participants are learning the categories, quantified as faster reaction times across training blocks, then darker blocks will be lower on the y axis in Figure 65 and lighter blocks will be higher. In the /ma/ Condition, none of the participants over forty exhibited faster reaction times across blocks, which led to results being uninformative regarding the time course of learning across age groups in this condition.

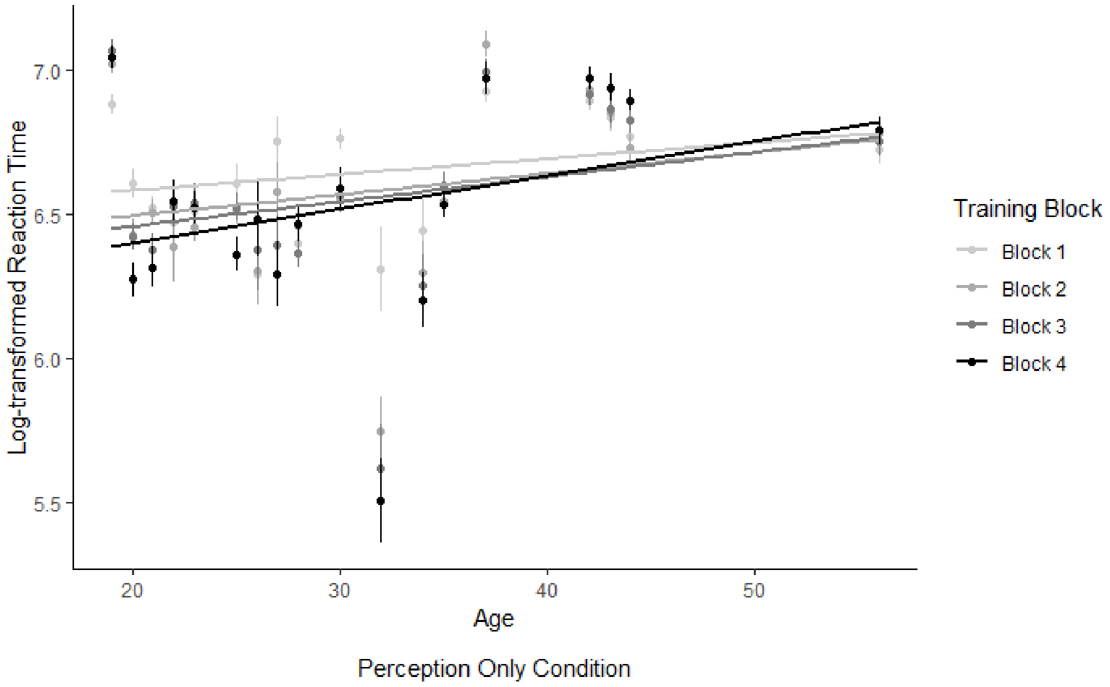

Figure 65. Log-transformed reaction times across training blocks in the Perception Only Condition.

Results from the /ma/ Production Condition indicated that reaction times did not significantly differ as a function of age ($X^2$ (1) = 2.92, *p* = .09). Figure 66 illustrates log-transformed reaction times as a function of age in the /ma/ Production Condition. Results from the /ma/ Production Condition continue to support the trend that, out of those that learn, reaction times from older participants tend to be slower overall than younger participants. Further, Figure 66 suggests that the two oldest participants and some of the younger participants' results indicated learning and that category acquisition most likely occurred around the beginning of the third block. Figure 66 also suggests that, for those that learned, the third and fourth blocks tend to differ, with the fourth block having slower reaction times than the third block. Having to produce the tokens on each trial resulted in a study that was longer, overall, than the previous experiments. It is possible that slower reaction times on block four are indicative of fatigue. It is also possible that once participants are able to reliably predict the location of the visual target on each trial, they relax and slow down some.

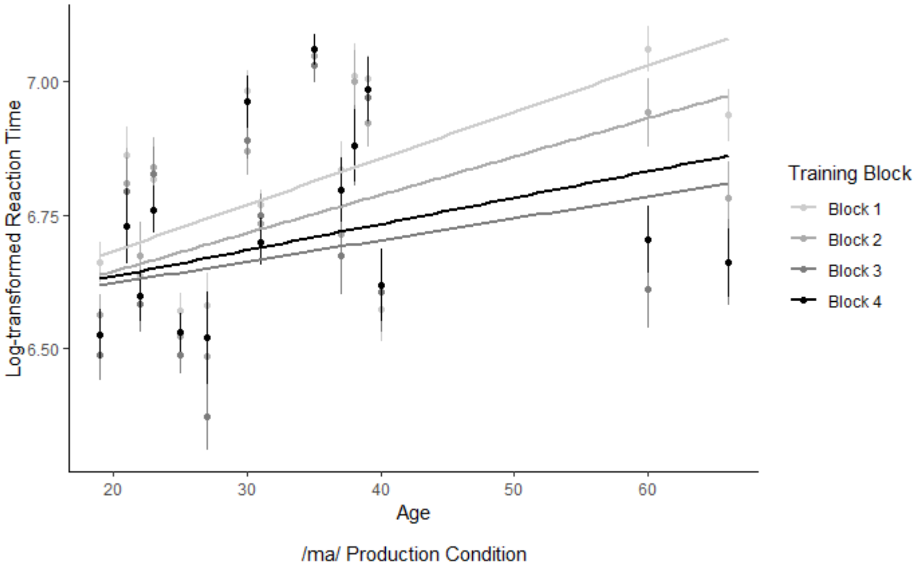

Figure 66. Log-transformed reaction times across age in the /ma/ Production Condition.

Figure 67 illustrates log-transformed reaction times as a function of age in the /mɯ/ Production Condition. Results from the /mɯ/ Production Condition indicated that reaction times differed as a function of age ($X^2$ (1) = 9.86, *p* = .002). Since few participants showed signs of learning in the /mɯ/ Production Condition, results more closely resembled the Control Condition from Experiment 2, indicating a correlation between age and reaction times across training blocks.

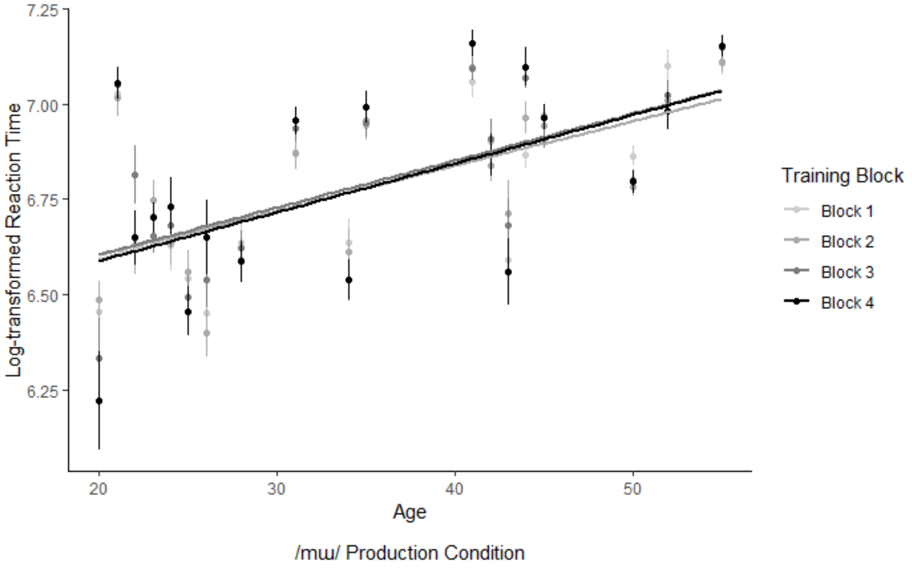

Figure 67. Log-transformed reaction times across age in the /mɯ/ Production Condition.

In Experiment 4 I measured the reaction times of participants across training blocks in three conditions. In two of the conditions, reaction times became faster across training blocks, indicating that participants learned the novel tone categories and were able to use that learning to predict the locations of the visual targets. Reaction times in the Perception Only Condition were the fastest across blocks. Reaction times in the /ma/ Production Condition also got faster across block two and block three but then got slower in block four. Reaction times in the /mɯ/ Production Condition did not get faster across blocks. The effect of age on reaction times differed across conditions in a similar way. In the Perception Only Condition and the /ma/ Production Condition, reaction times did not differ as a function of age. In the /mɯ/ Production Condition, however, age reaction times did differ as a function of age, with older participants being slower than younger participants. The results in the /mɯ/ Production Condition more closely resembled the Control Condition from Experiment 2. The similarity is most likely due to very few participants in this condition showing signs of learning the tone categories.

### 6.4.2 Generalization to new tokens and new talkers

As in the other experiments, Posttest 1 tested participants' ability to generalize to new tokens from the same talker, and Posttest 2 tested generalization to new talkers. The structure of both posttests is identical and both measure identification accuracy of the target tone category. If participants have learned the categories they should be able to accurately identify in which box the visual target should have appeared based solely on hearing the auditory stimuli, and therefore, their accuracy scores will be higher. Experiment 1 confirmed that participants that hear a single talker during training are able to accurately identify the four novel tone categories on Posttest 1. However, when they hear novel talkers on Posttest 2, they are less accurate. The Multi-talker Condition in Experiment 2 trained participants on multiple talkers, resulting in lower accuracy on Posttest 1, but accuracy on Posttest 1 did not differ from Posttest 2. Experiment 3 indicated that familiarity with the tone bearing segment did not impact perceptual learning. An impact was expected due to increased attentional load from the unfamiliar segment. In Experiment 4, in two conditions, participants produced the tokens that they heard during training. Posttest 1 and Posttest 2, however, were identical to previous experiments. It is expected that production during perceptual learning will result in less learning and lower accuracy scores at test. Of the two production conditions, it is expected that the effort to produce unfamiliar segments during training will result in an increased planning load and will have a negative effect on learning, compared with the production of familiar segments.

### *6.4.2.1 Analysis*

Accuracy scores for all conditions were measured on Posttest 1 and Posttest 2. For each condition, I compare accuracy scores on both posttests to chance using one sample t-tests. To test whether accuracy scores differ as a function of condition, I conduct model comparisons with and without condition for each posttest. To test whether there is a correlation between the learning measures, I conduct correlation tests between reaction times during training and accuracy scores at test for each condition. Finally, I conduct model comparisons to examine age as a fixed effect for all conditions on Posttest 1 and Posttest 2.

### *6.4.2.2 Accuracy*

Figure 68 illustrates mean proportion correct scores with 95% confidence intervals for the Perception Only Condition, the /ma/ Production Condition, and the /mɯ/ Production Condition on Posttest 1 and Posttest 2. The figure suggests that participants in all conditions accurately identified the target categories above chance on Posttest 1 and perhaps on Posttest 2 as well and that participants in the Perception Only Condition were more accurate on Posttest 1 than participants in the other conditions.

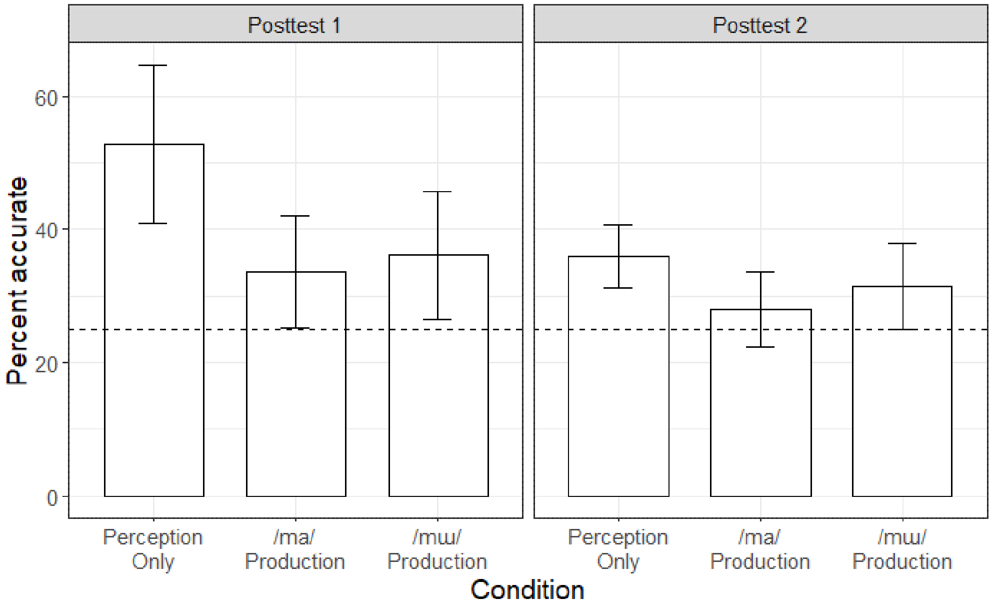

Figure 68. Mean proportion correct for all conditions on Posttest 1 and Posttest 2. Error bars represent 95% confidence intervals. The dashed line represents chance at 25%.

To test whether accuracy scores differed from chance, I examined accuracy scores across participants within condition on Posttest 1 and Posttest 2. In the Perception Only Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $t(24) = 4.80$, $p < .001$, ($M = 52.83$, $SE = 5.80$) and on Posttest 2, $V = 247$, $p < .001$, ($Mdn = 34.72$)[44]. In the /ma/ Production Condition participants were not able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $V = 158$, $p = .07$, ($Mdn = 31.94$) or on Posttest 2, $V = 132$, $p = .29$, ($Mdn = 26.39$)[45]. In the /mɯ/ Production Condition participants were able to match novel sounds to the visual locations at above-chance levels on Posttest 1, $V = 200$, $p = .009$, ($Mdn = 33.33$) and on Posttest 2, $V = 143$, $p = .03$, ($Mdn = 31.94$) [46].

To test whether accuracy scores differed across conditions on Posttest 1 and Posttest 2, I compared models with and without condition for each posttest. Results indicated that accuracy scores differed as a function of condition on Posttest 1 ($X^2$ (2) = 7.30, $p = .03$) but did not differ on Posttest 2 ($X^2$ (2) = 4.76, $p = .09$).

$$\text{accuracy} \sim \text{condition} + \text{age} + (1|\text{participant})$$
$$\text{accuracy} \sim \text{age} + (1|\text{participant})$$

Bonferroni corrected post-hoc comparisons revealed that on Posttest 1, the Perception Only Condition was more accurate than the /ma/ Production Condition (β = .193, SE = .07, $t = 2.72$, $p = .03$) and was more accurate than the /mɯ/ Production Condition (β = .193, SE = .07, $t = 2.71$, $p = .03$). Further, the /ma/ Production Condition did not differ from the /mɯ/ Production Condition (β < .001, SE = .02, $t = .02$, $p = 1$).

The accuracy results were difficult to interpret considering that the two production conditions did not differ from each other on Posttest 1, but the /ma/ Production Condition was not above chance, while the /mɯ/ Production Condition was above chance. A visual inspection of the individual accuracy scores aids in understanding the results. Figure 69 is a replication of Figure 68 illustrated with individual participants' data points. Figure 69 suggests that the two production conditions had similar numbers of participants that displayed learning on Posttest 1.

[44] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 2 ($W = .88$, $p = .008$).

[45] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 1 ($W = .87$, $p = .003$) and on Posttest 2 ($W = .87$, $p = .004$).

[46] A Wilcoxon signed rank test was used as a Shapiro-Wilk normality test indicated the data were not normally distributed on Posttest 1 ($W = .79$, $p < .001$) and on Posttest 2 ($W = .91$, $p = .02$).

The /mɯ/ Production Condition ended up having a higher median than the /ma/ Production Condition due to the difference between the highest achieving participants in each condition.

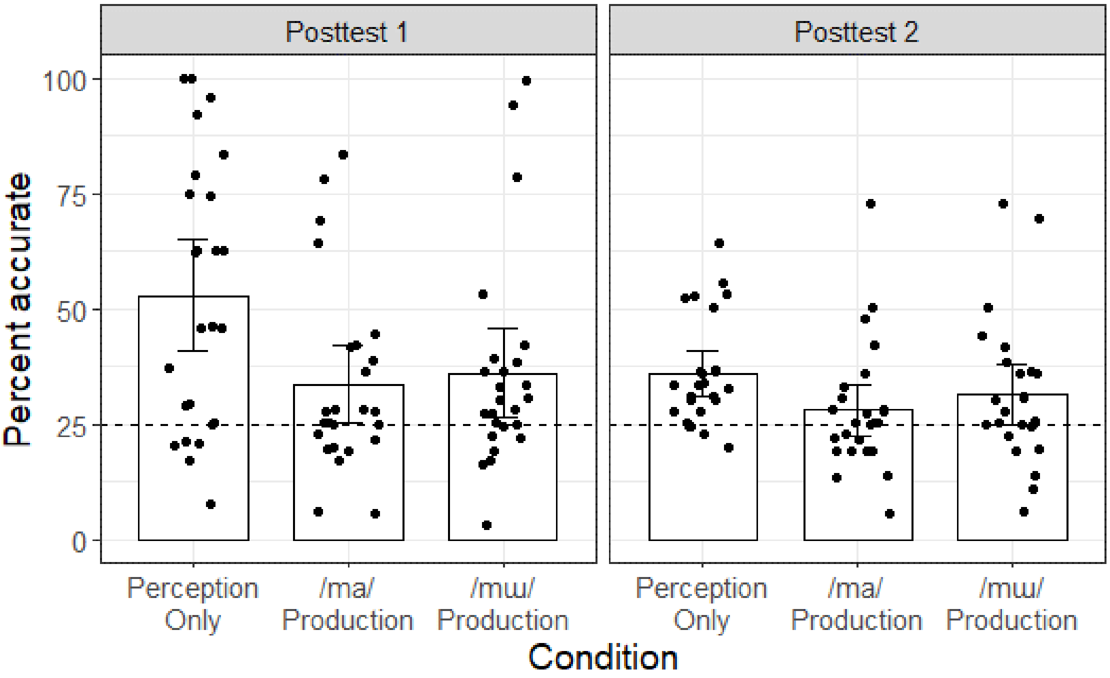

Figure 69. Mean proportion correct for all conditions on Posttest 1 and Posttest 2. Error bars represent 95% confidence intervals. The dashed line represents chance at 25%. The dots represent individual participants' proportion correct scores.

Overall, some participants in each of the three conditions displayed novel tone category acquisition on Posttest 1 and on Posttest 2, indicating that all conditions resulted in learning and that learning generalized to novel tokens on Posttest 1 and novel talkers on Posttest 2. However, a significantly larger proportion of participants learned in the Perception Only Condition compared to the two production conditions. Results from the /ma/ Production Condition and /mɯ/ Production Condition did not indicate distinct differences between the two conditions. It is clear that the /ma/ Production Condition did not result in learning above and beyond the /mɯ/ Production Condition, which contained the unfamiliar segment, /ɯ/. As the two production conditions produced equivalent results, the addition of the /mɯ/ Production Condition provides an internal replication of the results from the /ma/ Production Condition.

During training, greater learning was measured through reaction times becoming faster across training blocks. At test, greater learning was measured through higher accuracy scores. It was expected that faster reaction times at the end of training would correlate with higher accuracy scores at test for all conditions. Figure 70 illustrates the correlation between reaction times on block 4 and accuracy scores on Posttest 1, suggesting that the relationship between the two measures may be present in the production conditions.

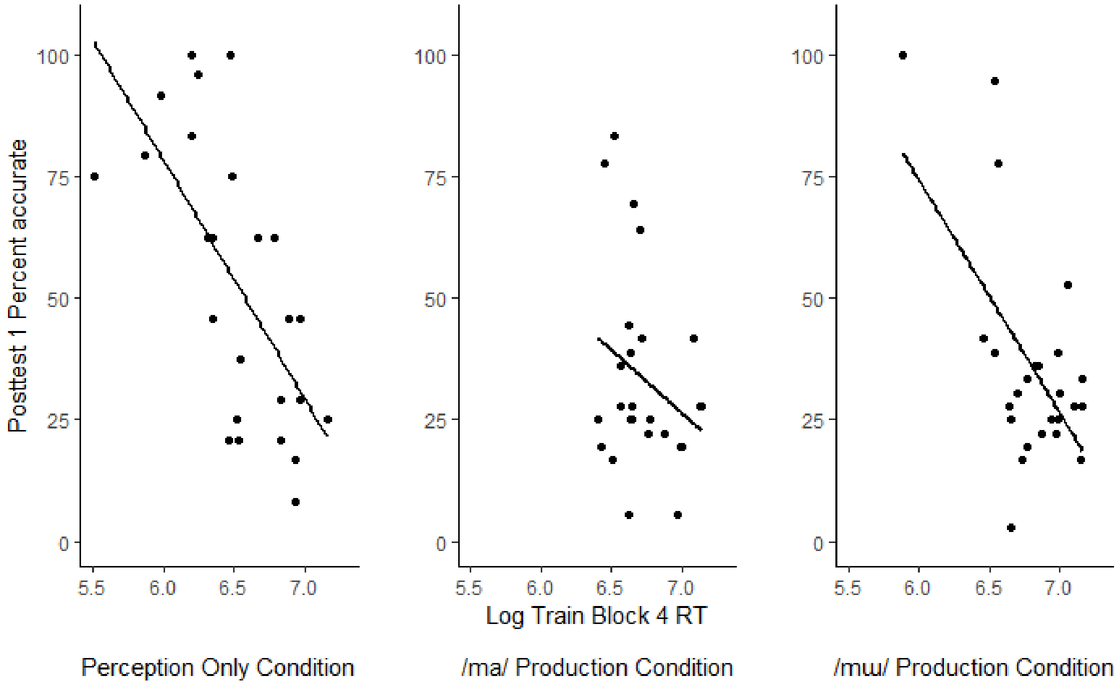

Figure 70. Relationship between two measures assessing category learning across conditions with log transformed reaction times on training block 4 on the x axis and accuracy scores on Posttest 1 on the y axis.

Spearman's rho correlation coefficient[47] was used to assess the relationship between reaction times on training block 4 and accuracy scores on Posttest 1. The relationship between the two measures was significant in the Perception Only Condition ($r$ = -.69, $p$ < .001). However, the relationship was not significant in the /ma/ Production Condition ($r$ = -.13, $p$ = .55) or in the /mɯ/ Production Condition ($r$ = -.33, $p$ = .11). The correlation between the two measures in the Perception Only Condition suggests that faster reaction times in training relates to better

[47] A Shapiro-Wilk normality test indicated the some of the data in the /ma/ Production Condition ($W$ = .87, $p$ = .003; $W$ = .92, $p$ = .06) and in the /mɯ/ Production Condition ($W$ = .79, $p$ < .001; $W$ = .90, $p$ = .02) were not normally distributed. Therefore, we conducted the non-parametric Spearman's test for all conditions. Although the data for the Perception Only Condition were normally distributed ($W$ = .93, $p$ = .08; $W$ = .96, $p$ = .37), Spearman's test was used for consistency. Pearson's correlation coefficient was also significant for the Perception Only Condition ($r$(23) = -.67, $p$ < .001).

accuracy on the generalization test and that both measures reliably assess category learning when there is only a perception component during training and the data are more evenly distributed among those that learned and those that didn't learn. In the production conditions, few participants showed signs of learning. However, almost all of those that scored above 50% accuracy were among the proportion of the participants that had the fastest reaction times. I hypothesize that if the number of participants that learned and didn't learn were more even, then there would be a correlation between reaction times during training and accuracy scores at test.

To test whether accuracy scores at test differed as a function of age for each condition, I compared models with and without age, and results indicated that accuracy scores did not significantly differ as a function of age in the Perception Only Condition on Posttest 1 ($X^2$ (1) = .44, *p* = .51) or on Posttest 2 ($X^2$ (1) = 3.62, *p* = .057). Accuracy scores did not significantly differ as a function of age in the /ma/ Production Condition on Posttest 1 ($X^2$ (1) = 1.70, *p* = .19), but accuracy scores did differ as a function of age on Posttest 2 ($X^2$ (1) = 5.15, *p* = .02). Further, accuracy scores did not significantly differ as a function of age in the /mɯ/ Production Condition on Posttest 1 ($X^2$ (1) = .52, *p* = .47) or on Posttest 2 ($X^2$ (1) = .28, *p* = .60).

accuracy ~ age + (1|participant)
accuracy ~ (1|participant)

Figure 71 illustrates accuracy scores on Posttest 1 and Posttest 2 as a function of age across conditions. The model comparison demonstrated that accuracy scores mostly did not differ as a function of age. Figure 71 suggests that the differences across age in the /ma/ Production Condition were largely driven by the oldest participants learning and very few younger participants learning. Overall, with few participants displaying learning in the production conditions, results are not as informative regarding the correlation between age and accuracy scores.

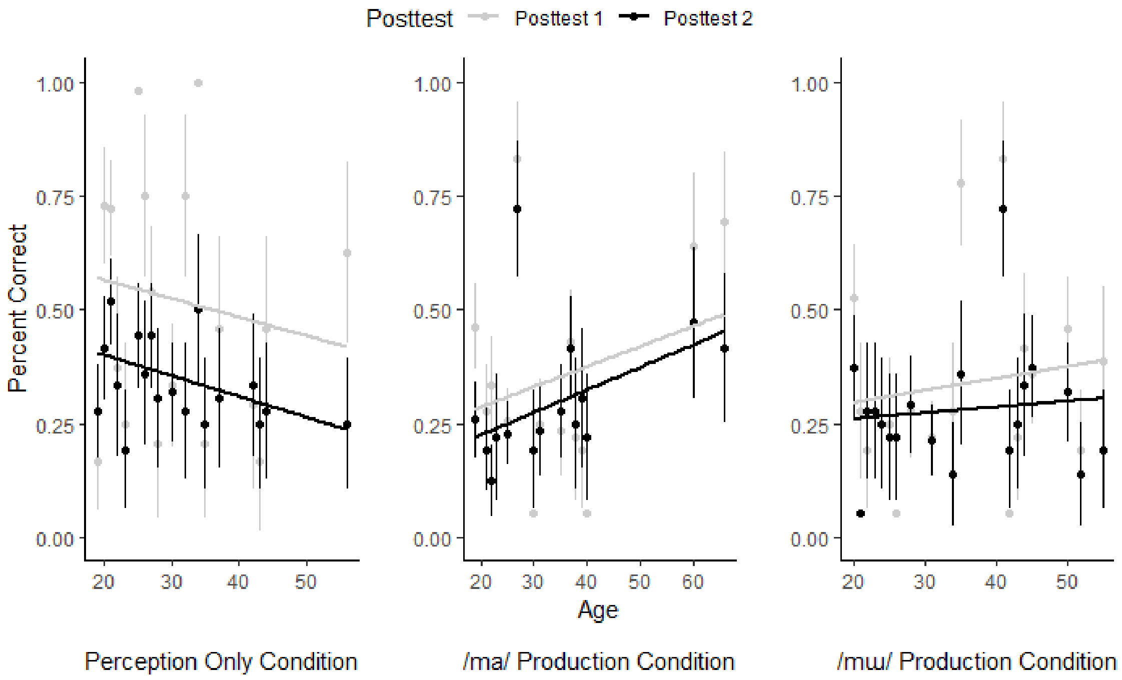

Figure 71. Accuracy scores on Posttest 1 and Posttest 2 across age in the Perception Only Condition, the /ma/ Production Condition, and the /mɯ/ Production Condition.

## 6.5 Discussion

In Experiment 4, I examined the impact of production during training on incidental perceptual learning of novel tone categories. I also examined the impact of segmental familiarity in the learners' productions during training. I compared three conditions. One condition was a Perception Only Condition that did not contain a production component. Two conditions contained tokens produced in different syllables. The /ma/ Production Condition was comprised of segments more familiar to the participants' language background experience. The /mɯ/ Production Condition contained a segment unfamiliar to the participants' language background experience.

It was expected that the two production conditions' results would differ from the Perception Only Condition. That is, the additional production by learners during perceptual learning would result in reduced learning compared to the Perception Only Condition. Further, it was expected that the lack of familiarity would negatively impact perceptual learning in the /mɯ/ Production Condition compared to the /ma/ Production Condition.

Results from training indicated that reaction times from participants in the Perception Only Condition became faster across training blocks. Participants in the /ma/ Production Condition also became faster across training blocks, but not as fast as participants in the Perception Only Condition. Participants in the /mɯ/ Production Condition did not become faster across training blocks.

Results from Posttest 1 indicated that participants in the Perception Only Condition were more accurate than the two production conditions. Further, as a whole, participants in the /ma/ Production Condition did not accurately identify the target categories above chance. Participants in the /mɯ/ Production Condition barely identified the target categories above chance. It is important to note that these results do not mean that the /ma/ Production Condition cannot result in perceptual learning. Some participants did show signs of learning. Instead, these results indicate that the two production conditions result in fewer participants successfully acquiring the novel tone categories than the Perception Only Condition.

Results from Posttest 2 indicated that generalizing to novel talkers was difficult for participants in all conditions. Specifically, the Perception Only Condition did not result in better generalization to novel talkers than the two production conditions.

Overall, the Perception Only Condition resulted in better perceptual learning than the production conditions, and the lack of segmental familiarity in the /mɯ/ Production Condition did not result in worse learning than the /ma/ Production Condition. However, it did result in slower reaction times than the /ma/ Production Condition. Below I discuss the implications of these results for categorization and perceptual learning.

### 6.5.1 The effect of production on perceptual learning

Production during perceptual learning hindered the formation of novel sound categories, replicating a finding from several previous studies (e.g., Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007, Baese-Berk & Samuel, under review). The production conditions contained elements that were not in the perception only condition. In the production conditions the participants produced the sound that they heard on each trial. They were also prompted to record themselves producing the sounds. As there were differences between the Perception Only Condition and the production conditions, it is important to consider whether differences between conditions might have resulted in decreased perceptual learning in the production conditions. To understand the potential differences between the conditions, it is important to take a closer look at what is occurring in the incidental learning paradigm.

Participants in incidental auditory learning paradigms (see Gabay et al., 2015) make little effort on each trial. They typically see the visual target on the screen and then they respond with the keyboard or mouse. They are not consciously trying to learn. Therefore, it may seem that learning in incidental paradigms is passive learning. However, learning is not passive. There is a feedback mechanism incorporated in the incidental learning paradigm (Schultz et al., 1993, 1997; Gabay et al. 2015; Ashby & Casale, 2003; Sutton & Barto, 2005; Lim et al., 2014, Reynolds & Wickens, 2002). In the incidental paradigm, learning occurs when the participant begins to use the auditory tokens as clues that predict the location of the visual targets. On a trial they hear the sounds and predict where the target will be when it appears. This provides implicit feedback telling them if they were right or wrong in their prediction. In this way the auditory-to-visuomotor correspondence reinforces learning by providing feedback on each trial. They use that feedback to refine their categorical judgments of the following auditory stimuli. As they become more confident in their predictions, they move the mouse cursor to the location where they think the visual target will appear. When it appears where they predicted, they are rewarded by being able to click on the visual target faster. If they are wrong in their prediction, they will have to move the cursor to the location of the visual target and their reaction time will be slower.

The learning reinforcement described here is a form of reinforcement learning, which is goal-directed, meaning that learning is driven by the participant's desire to achieve a goal (Sutton & Barto, 2005). In this case the goal is to minimize prediction errors (i.e., reward prediction error). Behavioral actions leading to rewards are reinforced, while behaviors leading to punishment become modified. Lim et al. (2014) argue that the learning reinforcement utilized in goal-directed learning has a neural basis that may not occur during passive exposure to stimuli or in explicit training paradigms. They argue that dopamine neurons in the basal ganglia can serve as a teaching signal to drive reinforcement learning. Dopamine neurons have been shown to be sensitive to reward prediction, firing when predictions are rewarded and depressed when predictions fail (Schultz et al., 1993, 1997). This process can lead to modulations in synaptic plasticity of cortico-striatal pathways (Reynolds & Wickens, 2002). However, this process may be time sensitive. Distractions during or immediately after the occurrence of the reinforcement learning mechanic may lead to disruptions in the process. Specifically, the process can be fragile, meaning that reinforcement signals must get to the right synapses at the right time to be effective (see Houk & Adams, 1995; Yagishita, 2014).

Thus, a key consideration in paradigms that implement reinforcement learning is the proximity of the stimuli to the learning reinforcement that occurs when the visual target appears and the participant responds to it (see Gabay et al., 2015 for discussion). If participants experience a disruption or delay between hearing the stimuli and seeing the visual target, reinforcement learning may be disrupted. There is a narrow window available for dopamine release to occur in order to strengthen synapses. Thus, it would be problematic for production to occur after exposure to the auditory stimuli and before the motor response due to the disruption to the striatal strengthening process. However, in the two production conditions, the proximity of the stimuli and visual target did not differ from the other conditions in the current study. Participants responded by producing the auditory target only after their motor response to the visual target. Therefore, temporally speaking, participants had a chance for striatal strengthening to occur.

A primary difference then between the production conditions and the Perception Only Condition could be that during the reinforcement period of the trial, participants in the production conditions were anticipating and perhaps preparing for the following production. This anticipation and preparation for producing the auditory target may have been a key factor in the disruption of the reinforcement learning. Thus, one possible explanation for the disruption of perceptual learning in the production condition was that participants were distracted by the need to produce the target during the window where the audio-to-visual reinforcement should have happened.

Results from previous studies indicate that participants can be distracted from the reinforcement learning found in the audio-to-visual correspondence. For example, Roark et al. (2020) also used an incidental paradigm to test multiple factors that might inhibit novel sound category acquisition and one of their conditions distracted participants from attending to the audio-to-video correspondence. In their Misalignment Condition, there was an audio-to-visual correspondence where the location of each visual target was predicted by the auditory tokens. However, the visual targets were different colors. The participants did not respond by pushing a button to select the location. Rather, they responded by pushing a button corresponding to the color and the auditory tokens did not predict which color was going to appear. Therefore, the task examined whether participants could still attend to the auditory-to-visual correspondence in the face of the distractor task. At test, participants were not successful at guessing the location of the visual targets. They were distracted from the salient cues that they needed to attend to. Thus, certain factors may distract participants from attending to the learning reinforcement used in incidental learning paradigms.

Research on category learning from cognitive psychology and neurobiology aid in the formation of hypotheses regarding the nature of the disruption of production to perceptual learning in the current experiments. Until the 1990s category learning research in cognitive psychology focused on single-system models of category learning. A single-system model postulates that there is a single structure or set of structures in the brain that is active during category learning. Initial single-system models included prototype, exemplar, and decision-bound models. Prototype models suggested that during category learning, category prototypes are developed and novel stimuli are compared to category prototypes during the categorization process (e.g., Reed, 1972; Homa et al., 1981; Posner & Petersen, 1990). Exemplar models proposed that unique instances of each category are stored for reference (e.g., Medin & Schaffer, 1978; Estes, 1986; Pierrehumbert, 2001), creating a cloud of stored perceptions (see Todd et al., 2019). When novel instances of that category are perceived, they are taken and merged with stored instances that are not noticeably different, forming exemplars. As more tokens are heard, they become integrated into existing exemplars as the existing exemplar is activated. Exemplars that are frequently activated increase in strength, exerting more force over new tokens (also see Goldinger, 1998). Decision-bound models propose that a stimulus is perceived as a point in multidimensional space. Category judgments are made by comparing the amount of perceptual distribution overlap of the stimuli. The greater degree of overlap in multidimensional space determines the degree of similarity and likelihood of the stimuli being judged to be in the same category (e.g., Ashby & Townsend, 1986; Ashby & Perrin, 1988).

The underlying hypothesis behind single-system models is that there is a single category learning system in the brain. The implication being that there is a single resulting impact on category learning should something happen to the neural structure that is activated during category learning. For example, if the structure is damaged, then category learning in general will be impacted, or if age-related plasticity degradation impacts the structure, then all category learning will be impacted. However, studies in the 1990s began to question this assumption (e.g., Maddox & Ashby, 1993; Nosofsky et al., 1994; Smith et al., 1996; Knowlton, 1999).

Studies examining participants with brain damage revealed that some types of categorization knowledge could be retained but other types of knowledge could not be retained. Amnesic patients that could not develop declarative memories could develop procedural memories (Hamann & Squire, 1997; Reed et al., 1997)). They had developed an implicit knowledge of categories even though they could not recall having seen the stimuli before or even the researcher (also see Squire & Knowlton, 1995; Knowlton, 1999). The conclusion from these studies was that categories could be developed implicitly even when there was no declarative knowledge of the individual exemplars or memory of learning at all.

Other evidence emerged supporting the idea that more than one category learning system may exist in the brain. Results from categorization studies indicated that category learning involving simple unidimensional rules qualitatively differed from category learning involving multidimensional rules (Maddox & Ashby, 1993; Ashby et al., 1998). When categories are presented to participants and the rules for category membership are clear and distinct, once the rule is learned, participants can easily categorize the stimuli and can state the rule for categorization. However, when the criteria for category membership are multidimensional and the features of categories overlap, participants demonstrate gradual learning and unable to state what the rules are for categorization, but rather categorize the stimuli based on a feeling of which category the stimuli should be in (also see Chandrasekaran, 2014a, 2014b). Corresponding research resulted in the proposal of multiple-system models with separate rule-based and exemplar-based category learning systems (Brooks, 1978; Allen & Brooks, 1991; Regehr & Brooks, 1993), followed by similar cognitive models (e.g., Squire, 1992; Nosofsky, 1994; Erickson & Kruschke, 1998). Later, insights developed by cognitive models were strengthened through findings from neurobiological research (Poldrack & Packard, 2003; Nomura et al., 2007).

One model that incorporated neural findings was the Competition between Verbal and Implicit Systems model (COVIS; Ashby et al., 1998, 2011; Chandrasekaran et al., 2014). The COVIS model proposes two learning systems, a reflective system and a reflexive system. These learning systems respond to two types of category information structures, rule-based structures and information integration structures (see Nomura et al., 2007 for review). The reflective learning system is activated during explicit category learning with rule-based category structures. The reflective system is rule-based in that rules are explicitly learnable and when learned are then applied to the categorization task. For example, we may have a visual classification task where all stimuli with a circle belong to one category and all stimuli with a square belong to another category. In this example, once the rule is understood, categorization will immediately become easier for the participant. Reflective learning engages executive attention and working memory, activating regions in the prefrontal cortex, anterior cingulate, and anterior caudate nucleus (e.g., Nomura et al., 2007; Chandrasekaran et al., 2014).

The reflexive learning system is activated during implicit category learning, which is typically used to learn information integration structures, which refers to category structures that contain two or more stimulus dimensions and cannot be described by simple rules. Rather, stimuli must be observed more holistically and categorization responses are selected based on the wider range of perceptual information (see Ashby & Gott, 1988; Ashby et al., 1998). Reflexive learning does not engage working memory or executive attention and activates the posterior caudate, putamen, basal ganglia, and the supplementary motor area (see Chandrasekaran et al., 2014; Lim et al., 2014 for review).

Most research investigating novel tone category acquisition has done so via reflective learning methodologies, where participants receive explicit instructions regarding the target tone categories and feedback on each trial. Evidence suggests that the reflective learning system is not the optimal system for speech category learning due to the multidimensional nature of speech sounds (Chandrasekaran et al., 2014). Results comparing category learning across reflective and reflexive learning systems suggest that speech categories are optimally learned through the reflexive learning system. Thus, in the current study we use a reflexive learning paradigm, incidental learning, to investigate factors known to impact novel speech category acquisition during reflective learning. However, it may be that producing the tokens during perceptual learning in the current experiments activates the reflective learning system, hindering the reflexive learning system engaged by the incidental learning paradigm. Evidence suggests that the two learning systems are distinct and competitive in visual and auditory domains (Ashby et al., 1998; Ashby & Ell, 2001; Ashby & Maddox, 2005; Lim et al., 2014).

A corresponding difference between the two systems is that reflective learning utilizes executive function and working memory and reflexive learning does not. This is a vital distinction between the two systems. The information learned through the reflexive learning system is not easily verbalizable. There are multiple dimensions that assign members to categories, making it impossible or extremely difficult to specify a single dimension necessary for category membership. For this reason, it is suggested that reflexive learning is optimal for learning speech sound categories (Chandrasekaran et al., 2014; Lim et al., 2014). Further, the effort to explicitly rationalize rules for category membership can be detrimental for reflexive learning (Ashby & Gott, 1988). Therefore, one possibility is that the need to produce the tokens leads participants to attempt to understand the rules for category membership and thus activates working memory and executive function, which hinders the reflexive learning system. When production or even sub-vocal rehearsal of production occurs with perception, an auditory-motor interface system that relies on working memory is activated (Hickok & Poeppel, 2000; Hickok et al., 2003).

Another potential explanation of the disruption that occurs from production comes from cognitive exemplar theory. Exemplar theory was initially incorporated into a single-system model of category learning, and later applied to multiple-systems models of category learning. More recently, exemplar theory has been reinterpreted in light of findings from neurobiological studies (Ashby & Rosedahl, 2017). The evolution of concepts around exemplar theory provide insight into the category learning process and potential hypotheses regarding the potential results of the current study, especially regarding the potential disruption to perceptual learning from producing the target speech sounds during learning.

Exemplar theory was originally introduced as a model of visual categorization in psychology and later extended to the auditory domain and the acquisition of speech sound categories (e.g., Johnson, 1997; Lacerda, 1997; Pierrehumbert, 2001). Exemplar theory posits that unique instances of each category are stored for reference, creating a cloud of stored perceptions, and when novel instances of that category are perceived, they are taken and merged with stored instances that are not noticeably different, forming exemplars. As more tokens are heard, they become integrated into existing exemplars as the existing exemplar is activated. Exemplars that are frequently activated by new exemplars increase in strength, exerting more force over new tokens.

Initial applications of exemplar theory to speech category learning incorporated concepts from usage-based phonology (Bybee, 2001), which proposes that sound categories are built up through experience with the language and exposure to sounds from the category over time and across contexts. During first language acquisition, as a child hears sounds in their language, they will store the individual instances of the input. The distribution around features such as voice onset time (VOT) will create an exemplar cloud that comprises the phonemic category (also see Todd et al., 2019). As the child hears tokens of that category, they will take them and merge them with other tokens that are not noticeably different, forming exemplars. As more tokens are heard, they become integrated into existing exemplars, activating that exemplar. Exemplars that are frequently activated increase in strength, exerting more force over new tokens (also see Goldinger, 1998).

During the acquisition of other languages, when a person hears sounds from novel categories, if the target categories are similar to sound categories in their L1, the strongest exemplars in their L1 categories will integrate the new token. This idea is reflected in Best's Perceptual Assimilation Model (Best et al., 1988; Best, 1995). PAM predicts that individual non-native sounds that are similar to established L1 sound categories are likely to be perceptually assimilated by naïve listeners to the most articulatorily-similar L1 category. Flege's Speech Learning Model (SLM) is similar, but it goes on to say that a new sound category can be established if the learner can discern enough acoustic differences between their closest L1 sound category and the novel sound (Flege, 1995). Similarly, if there are no L1 sound categories, new categories can be developed. However, the process of novel sound category formation can be disrupted.

From a cognitive perspective, the application of exemplar theory to disruptions during novel speech sound formation may lie in the interface between the phonological loop and the long-term memory system (Kaushanskaya and Yoo, 2011; Baddeley, 1986; 2000). Working memory is said to delineate the necessary processing and storage of auditory input for language comprehension and acquisition. The phonological loop is assumed to have two components, a phonological store and an articulatory control process. The phonological store holds the acoustic details of the input in short-term memory for one to two seconds before processing into long-term memory. Baddeley states that several things can affect this process, one of which is phonological familiarity, which strongly influences foreign language word learning (Baddeley, 1986; 2000). Production during perceptual learning may also disrupt the phonological loop if it occurs within the first one to two seconds after perception (also see Darwin et al., 1972; Baese-Berk & Samuel, 2016; Baese-Berk & Samuel, in press). Exemplar theory provides us with a hypothesis of what this disruption might look like. Upon initial exposure to an L2, L1 exemplars are very strong and are likely to incorporate L2 tokens into L1 categories. In order to create L2 sound categories that are similar to but differ from L1 phonemes, there would need to be a sufficient number of auditory tokens to create a new statistical distribution and create sufficient exemplar strength to incorporate L2 tokens into the new exemplar cloud. Thus, one hypothesis based on cognitive research is that production occurring immediately after exposure to auditory stimuli disrupts the phonological loop, keeping new tokens from being stored and a new cloud from being created. On the production side, exemplar theory states that a target is randomly selected from the existing exemplar cloud based on activation strength. So, the listener turned speaker extracts an exemplar from the L1 sound category to produce as the target, since the L2 sound category has not yet been created or does not yet have sufficient strength to stand as an exemplar. This causes the L1 sound category to be strengthened even further, and perceptual learning of the novel sound category does not occur.

This hypothesis could be tested by briefly delaying production after perception to see if that provides the needed time to store the auditory token (again see Darwin et al., 1972; Baese-Berk & Samuel, 2016; Baese-Berk & Samuel, under review). The expectation would be that perceptual learning would be higher for participants with a delayed production response than for participants that produce the token immediately after hearing it. However, there may still be interference from the activation of the L1 sound categories when they produce the targets. Therefore, initially it may be expected that they would not perform as well as those receiving perception only training (see Baese-Berk & Samuel, under review).

The neural interpretation of exemplar theory extends exemplar theory on the basis of findings from neurobiological studies on categorization and proposes that category learning occurs as synaptic connectivity between striatal neurons and neurons in sensory association cortex are altered (Ashby & Rosedahl, 2017). Rather than adding nodes to the exemplar network (see Kruschke, 1992; Nosofsky & Palmeri, 1997), learning occurs as the presentation of the stimuli strengthens existing cortical-striatal synapses or creates a new synapse. The neural exemplar model assumes that synaptic strengthening occurs in relation to the level of presynaptic activation, the resulting level of postsynaptic activation, and the level of dopamine present (see Ashby & Rosedahl, 2017). Neural exemplar theory aids in the understanding of reinforcement learning and by extension, incidental learning. It presents an understanding of the reflexive learning process and factors that inhibit or benefit the learning process.

As discussed above, dopamine neurons have been shown to be sensitive to reward prediction, firing when predictions are rewarded and depressed when predictions fail (Schultz et al., 1993, 1997). This process can lead to modulations in synaptic plasticity of cortico-striatal pathways (Reynolds & Wickens, 2002) and is thought to drive incidental learning through the reinforcement mechanism incorporated in the design of the learning paradigm (see Lim et al., 2014 for review). As participants make predictions based on the auditory stimuli and those predictions are validated upon appearance of the visual target, dopamine levels are increased resulting in synaptic plasticity at cortical-striatal synapses (Houk et al., 1995; Doya, 2000). This process provides a foundation for neural-based hypotheses regarding expected results for incidental learning. For example, it is expected that there is a narrow window of 0.3 to 2 seconds for reinforcement of predictions to occur (Yagishita, 2014). Further, the process can be fragile, meaning that reinforcement signals must get to the right synapses at the right time to be effective (Houk & Adams, 1995). Switching to production during this narrow window may inhibit the synaptic strengthening process.

A remaining issue is that some participants in the production conditions did learn the tone categories. Further, as illustrated in Figure 69, the few that learned had accuracy scores as high as those that learned in the Perception Only Condition. However, in the production conditions there is a clearer distinction between those that learned and those that didn’t learn. In the Perception Only Condition the results are less categorical. Therefore, we may conclude that due to the need to produce the tokens on each trial, participants that did not learn either did not attempt to make predictions regarding the location of the visual target or they tried to make predictions but were not very successful and ended up abandoning the effort to make predictions. In the first case we may conclude that participants were simply too distracted by producing the targets to attend to the primary requirement for learning in the task. In the second case, a lack of consistent reward may have resulted in dopamine depression and therefore a lack of synaptic strengthening and abandonment of the task. Conversely, participants that learned in the paradigm were able to make predictions regarding the location of the visual target and were successful enough that they received sufficient dopamine reward to continue making predictions. Thus, consistent strengthening of the synapses resulted in synaptic plasticity and acquisition of the categories.

The question remains though regarding how participants who learned were able to continue making predictions. The answer may have to do with attention and timing. Attention would first have to be given to the audio-to-visual correspondence. If all of the attention was on producing the tokens accurately, then they would not have been able to make predictions regarding the location of the visual target. Attention could be further investigated in future experiments by explicitly increasing attention to producing the targets. For example, participants could be instructed to produce the target as accurately as possible. It may be that exogenously orienting attention to producing targets in this way would further result in participants overlooking the audio-to-visual correspondence.

Regarding timing in the production conditions, participants that learned the tone categories may have initially directed their attention to making predictions on each trial. The supposition here is that only after making the predictions and clicking on the visual targets, they shifted their attention to preparation and execution of the productions. This hypothesis comes from the expectation that there is a narrow window of 0.3 to 2 seconds for reinforcement of predictions to occur (Yagishita, 2014). This could be tested by investigating the time it takes for successful learners to shift from clicking on the visual target to recording their production. If attention is only directed to production after the reflexive learning period occurs, then the duration of this period should be longer than those that do not learn the categories. Similarly, it is expected that if there is a longer delay between the perceptual reinforcement learning mechanism and production, perceptual learning would improve due to the separation of auditory, visual, and motor processes (see Baese-Berk & Samuel, under review). It is expected that distinct, non-overlapping processes will support learning (Forrin et al., 2012).

We might conclude that the distraction in the current experiments from producing the tokens is specific to incidental learning as it specifically disrupts the reinforcement learning mechanism, and that perhaps, producing tokens during perceptual learning may not impact explicit learning paradigms. However, there is growing evidence that the disruption of perceptual learning by efforts to produce the auditory targets extends beyond incidental paradigms. Production during perceptual learning of novel sound categories also impacts reflective learning (Baese-Berk & Samuel, 2016). However, when production is delayed for two or four seconds, perceptual learning is not hindered (Baese-Berk & Samuel, under review). Initially, the results from these experiments may seem to go against conventional understanding of the relationship between perception and production. Studies examining word learning found that production during word learning enhanced perceptual abilities (Gathercole & Conway; 1988). This led to the concept of the "production effect", where producing a word aloud during study improves retention of the word (MacLeod et al., 2010; Forrin et al., 2012), and resulted in the conclusion that production is needed during perceptual learning to establish a bidirectional link between the perception and production systems (Zamuner et al., 2016). However, studies that tested the impact of production during perceptual learning using novel words with unfamiliar segments or structures resulted in worse perceptual learning of the novel words (Leach and Samuel, 2007; Kaushanskaya and Yoo, 2011; Dahlen and Caldwell-Harris, 2013). The results from the current experiments support the conclusion that production during the perceptual formation of novel categories that are not familiar to the learner can hinder perceptual learning. The finding that a lack of familiarity can hinder perceptual learning, led to the expectation that a greater degree of unfamiliarity with the segments in the tokens may result in greater inhibition during perceptual learning.

### 6.5.2 Segmental familiarity and production during perceptual learning

It was expected that the lack of segmental familiarity in the /mɯ/ Production Condition would result in greater disruption of perceptual learning than the /ma/ Production Condition. The /mɯ/ Production Condition did result in slower reaction times during training than the /ma/ Production Condition, but there was little difference between conditions at test. Both conditions resulted in very few participants learning the novel tone categories. Combined accuracy scores for each condition were close to chance. This result from the current experiment is similar to the results from Experiment 3, where perceptual learning was equivalent in conditions with familiar segments and unfamiliar segments. An initial explanation for the results in Experiment 3 was that participants in the /mɯ/ Production Condition simply processed the unfamiliar segment as a familiar segment, such as /ə/. If they mapped /ɯ/ onto the acoustic space of an English vowel, they might avoid processing difficulties that may arise from the unfamiliar segment. However, the recordings from Experiment 4 indicate that participants regularly produced vowels unlike English vowels in an attempt to approximate /ɯ/.[48]

An expectation that the production of unfamiliar segments might result in a greater disruption to perceptual learning comes from findings indicative of processing differences when producing familiar and unfamiliar speech (Moser et al., 2009). The structures involved in speech motor control do not appear to differ when producing familiar and unfamiliar phonotactics. However, there is greater activation of the structures, both in extent and intensity, during the production of words with unfamiliar segments. Specifically, greater activity occurs bilaterally across the left anterior insula (aIns) and inferior frontal gyrus (IFG). Moser et al. (2009) suggest that these results indicate greater engagement across the entire motor speech system when producing unfamiliar segments. Thus, there are several potential explanations for the null results in Experiment 4. It may have been that the single unfamiliar vowel was not sufficient for greater activation of the motor speech system. In Moser et al. (2009) participants produced tri-syllabic non-words with various non-native consonants, a condition with much greater variability than the /mɯ/ Production Condition in Experiment 4. It may be that with a larger number of unfamiliar segments there would be greater disruption to perceptual learning. Another possibility is that the level of learning was too low to distinguish between the categories. That is, not enough participants learned the categories in the production conditions. To observe contrastive results, it may be necessary to increase the amount of learning, which may be possible to achieve by delaying production.

### 6.5.3 Learning differences as a function of age

In Experiment 4 participants' ages ranged from 19 to 66. In the Perception Only Condition and the /ma/ Production Condition, reaction times did not differ as a function of age. In the /mɯ/ Production Condition reaction times differed as a function of age, with older participants being slower than younger participants. These results more closely resembled the absence of learning in the Control Condition from Experiment 2. Accuracy scores in the two production conditions did not differ as a function of age. Overall, in the production conditions few participants learned. One observation that can be made is that a greater percentage of older participants in the

[48] This suggestion stems from a preliminary investigation of the acoustic data collected from participants in Experiment 4. Due to time constraints a deeper analysis of the data is not presented here.

production conditions learned compared to younger participants. This includes the 66-year-old participant, who had one of the highest scores in the /ma/ Production Condition. These results present an interesting hypothesis that arises from previous observations and hypotheses made. It is clear that older participants are slower at the task. Moving through the trial takes longer the older the person is. We hypothesized in Section 6.5.1 that a greater delay between the audio-to-visual correspondence and production would likely result in greater learning in a production condition. The longer reaction times by older participants indicate that older participants experienced a greater delay between the audio-to-visual correspondence and the productions. Therefore, it is possible that the effect of age resulted in greater distinctiveness between the processes engaged in the task and this resulted in higher perceptual learning (see Forrin et al., 2012).

## 6.6 Conclusion

In Experiment 4 we compared learning in a Perception Only Condition to learning in two production conditions to investigate the impact of production during perceptual learning. By examining production by participants immediately after perception and the corresponding motor response on each trial, we tested the impact that the anticipation of production during the audio-to-visual reinforcement had on perceptual learning. By including two production conditions, the /ma/ Production Condition and the /mɯ/ Production Condition, we also tested the additional impact that the lack of segmental familiarity during motor planning had on perceptual learning.

Experiment 4 demonstrated that production during the perceptual learning of speech sound categories in an incidental learning paradigm severely hinders perceptual learning. Specifically, if participants produce the token they hear on each trial during training, very few participants are able to acquire the novel tone categories compared to participants that do not produce the tokens. These results suggest that producing targets during incidental learning interrupts the learning that occurs in an incidental learning paradigm. A specific consideration is the timing of the production, which occurs directly after the learning reinforcement mechanism in the paradigm. We hypothesized that delaying production might reduce the inhibitory effect of production during incidental perceptual learning.

Experiment 4 also demonstrated that the inclusion of an unfamiliar segment in the stimuli did not result in greater interruption to perceptual learning than stimuli with familiar tokens. This result is similar to Experiment 3. However, in Experiment 4, participants were also producing the tokens on each trial. It was expected that the need to produce an unfamiliar segment would increase cognitive load due to higher levels of activation in motor planning structures, resulting in reduced learning compared to producing familiar segments. However, there was little difference between the conditions. Therefore, these results demonstrate that during incidental learning, learners are equally capable of attending to salient tone category features in tokens that contain familiar and unfamiliar segments.

# VII. CONCLUSION

This dissertation sought to examine the perceptual formation of novel tone categories with natural tokens through an incidental learning paradigm, where learning is driven by a reinforcement learning mechanism.  Further, we used incidental learning to investigate a number of factors known to impact the perceptual formation of novel sound categories during learning via explicit learning paradigms. In this chapter, we provide a summary of the main findings from the four experiments conducted in this study and illustrate the novel contributions of the study. We discuss future directions of the research focused on areas studied in the dissertation, and we present implications of the current study for second language pedagogy.

## 7.1 Summary of the current research

In Chapter 1 we provided an overview of the four experiments in the current study and the background of the research. Specifically, we discussed research on novel tone perception, including tone discrimination and novel tone category learning. We also discussed auditory perceptual learning of novel sound categories. In Chapter 2 we presented a characterization of the stimuli used in the four experiments, describing differences regarding duration and F0. We illustrated that the use of natural tokens from multiple talkers resulted in variability among the tokens that may have impacted perceptual learning in the four experiments. In Chapter 3 through Chapter 6 we presented four experimental studies performed to analyze different factors that impact the incidental formation of novel tone categories. Below, we summarize the main findings of the four experimental studies and the novel contributions of this work.

### 7.1.1 Main findings of the four studies

We found that native English participants from 19 years old to 66 years old with no prior experience with the target tone categories can use an incidental learning paradigm with natural tokens to form four novel tone categories after 30 minutes of training with very high, even perfect, accuracy. These findings extend the investigation of factors impacting incidental learning into natural speech sound categories, confirming hypotheses suggesting that incidental learning is an effective means of learning natural speech sound categories. Taken together, our results suggest that reflexive learning through an incidental paradigm is an effective and efficient means of category learning and provides an experimental foundation well suited for the examination of factors impacting novel sound category formation with natural tokens.

In the first experiment, we found that presenting five different tokens on each trial resulted in greater learning than presenting five identical tokens on each trial, indicating that high variability in close temporal proximity resulted in greater learning than when the variability was spread out across trials. In Experiment 2 we demonstrated that training on a single talker results in higher learning accuracy than training on multiple talkers. However, when trained on a single talker, accuracy was substantially reduced when generalizing to novel talkers. By contrast, when trained on multiple talkers, accuracy did not differ when generalizing to tokens from the same talkers and tokens from novel talkers. Further, in Experiment 2 we also demonstrated that participants can learn to categorize novel tone categories from passive exposure alone. In Experiment 3 we demonstrated that the presence of an unfamiliar vowel in the auditory stimuli did not impact the incidental formation of novel tone categories. That is, the additional complexity from processing unfamiliar segmental features did not result in reduced learning of the target tone categories. In Experiment 4 we demonstrated that production during the perceptual learning of speech sound categories in an incidental learning paradigm severely hinders perceptual learning. Specifically, if participants produce the token they hear on each trial during training, very few participants are able to acquire the novel tone categories compared to participants that do not produce the tokens. In Experiment 4 we also demonstrated that the inclusion of an unfamiliar segment in the stimuli did not result in greater interruption to perceptual learning than stimuli with familiar tokens.

In each experiment we considered age as a factor impacting learning. We demonstrated that age impacted reaction times during training. We also demonstrated that this was due to age affecting the visuomotor responses during the task rather. Age did not impact learning. That is, individuals across all ages were able to learn from the incidental paradigm and form the novel tone categories. However, it is important to note that the age of participants in the current study only went to 66. Studies investigating the impact of age on learning often have participants in their 70s and 80s. Further, there were not equal sample sizes across the range of ages. There were more younger participants than older participants. With equal sample sizes and a direct comparison of age groups, differences in category formation during incidental learning may be found.

### 7.1.2 Novel contributions of the current research

The four experiments provide novel contributions, informing novel tone category acquisition and auditory perceptual learning more broadly.

#### *7.1.2.1 Novel tone category acquisition*

The current study provides novel contributions that inform research investigating factors impacting the acquisition of novel tone categories. Specifically, we demonstrated that incidental learning provides an effective and efficient means of investigating the acquisition of novel tone categories by adults. The results from the current study add to our understanding of novel tone acquisition during incidental learning, permitting comparisons with research using explicit learning paradigms. It is important to note that we do not directly compare incidental learning to explicit learning in the current study. Comparisons are made to situate the current study in the wider literature and highlight differences such as the time course of learning.

A major contribution the current study provides to the conversation around novel tone category acquisition is that learning novel tone categories is not necessarily difficult. We demonstrate that adults with no experience with lexical tone, no significant language learning experience, and no specific motivation to learn can, indeed, quickly and easily learn novel tone categories. They do so without conscious effort through incidental learning. Further, we demonstrate that this learning ability is maintained across the lifespan. These results differ from the majority of the research on novel tone category learning that suggests otherwise (Kiriloff, 1969; Bluhme & Burr, 1971; Shen, 1989; Sun, 1998; Wang et al. 1999; Wayland & Guion, 2004; Reid et al., 2015). It may be that novel tone category learning has been difficult because of the learning paradigms typically used during training. My goal in stating this so explicitly is that future research will not cite the difficulty of learning novel tone categories without specifying the learning paradigm the difficulty occurs in, because as we demonstrate, learning novel tone categories is not difficult during incidental learning. The assumption that learning novel tone categories is difficult continues to permeate research that uses novel tone categories, but we demonstrate that this is not accurate and the ongoing assumption of difficulty may distract the field from studying differences that are meaningful.

The current research adds to the suggestion that the meaningful issues regarding novel tone category learning are the learning system used to acquire the categories, differences between the processes and mechanisms engaged through the learning systems, and how those processes and mechanisms change across the lifespan. As discussed in Section 6.5.1, there is evidence of multiple category learning systems and that different learning paradigms engage different learning systems. The reflexive/reflective distinction in the COVIS model is particularly applicable (Ashby et al., 1998, 2011; Chandrasekaran et al., 2014a). It is important to note that we do not directly test differences between learning systems in the current study and results from the current study are not evidence for learning systems. Further, we do not directly test differences between explicit learning paradigms and incidental learning paradigms. However, by discussing results from the current experiments in light of results from explicit learning studies and the wider discussion on cognitive and neural models, we situate our findings in the literature, which informs our understanding of our results.

In the current work we highlight similarities and differences between studies that use explicit learning paradigms that engage the reflective learning system and the current experiments, which use an incidental learning paradigm that engages the reflexive learning system. One difference we can note between the two systems is the time course of learning. Reflective tone category learning can take multiple sessions over the course of several weeks (Wang et al., 1999; Wong & Perrachione, 2007). By contrast, learning novel tone categories via reflexive learning may be much faster, with many participants achieving high levels of accurate categorization in a single session. One hypothesis regarding the difference in the time course of learning between the two systems is that reflexive learning is better suited for novel tone category acquisition due to the multidimensional nature of speech sound categories. Reflexive learning engages neural structures suited for the categorization of multidimensional stimuli, while reflective learning engages structures suited for unidimensional rule-based learning (see Chandrasekaran et al., 2014a, 2014b). Explicit training with explicit feedback may actually slow the sound category formation process.

As we compare novel tone category learning across the two systems, we discuss differences in learning between the two systems. For example, the current study illustrated that older adults can learn novel tone categories as well as younger adults during an incidental paradigm that engages the reflexive learning system. If older adults are worse than younger adults at learning novel tone categories through explicit learning paradigms, then older adults may learn novel tone categories better during reflexive learning than during reflective learning.

If so, it may be that age-related decline in working memory and the functioning of prefrontal structures impacts the reflexive learning system less than the reflective learning system (Daigneault & Braun, 1993; West, 1996; Clapp et al., 2011; Maddox et al., 2013; Chandrasekaran et al., 2014). Further research between age groups across the learning systems will provide insight into changes in neural plasticity across the lifespan (see Chandrasekaran & Kraus, 2010).

Although we do not directly compare learning paradigms and learning systems in the current work, we do discuss similarities in novel tone acquisition across learning paradigms. For example, token variability and talker variability appear to benefit learning in explicit learning paradigms and incidental learning paradigms. In the current study we demonstrated that token variability within trial aided in category formation during incidental learning (see Gabay et al., 2015). That is, learners that heard variable productions from the same talker learned better than those that heard identical tokens on each trial. Further, those that heard multiple talkers across trials were able to generalize learning to novel talkers at the same accuracy as generalization to novel tokens from the same talkers. When trained on only one talker, accuracy during generalization to novel talkers decreased drastically. This finding is similar to results from explicit learning paradigms for novel tone category learning (Wang et al., 1999) and novel segmental category learning (Logan et al., 1991). However, studies that use explicit learning paradigms tend to only include a single auditory token during categorization training. They play the sound and the participants respond with the category they think the sound belongs to. The results from the current study add to growing methodological considerations (see Gabay et al., 2015) by testing the impact of the composition of multiple tokens on a single trial, suggesting the possibility that explicit learning paradigms may also benefit from the inclusion of multiple tokens with high variability on each trial. Further, we examined talker variability across trials and concluded that learning in an incidental paradigm would also likely benefit from the inclusion of tokens from multiple talkers within trial, rather than across trials.

The current study also provides a novel contribution to tone category learning research by demonstrating that tone categories can be formed through passive exposure alone. In the Control Condition in Experiment 2 participants received no explicit instructions regarding the target categories and no feedback from the reinforcement mechanic in the incidental learning paradigm. It was expected that participants would not be able to consistently categorize the auditory stimuli in this condition, but they showed signs that they were able to form the tone categories. These results recommend a line of research investigating the extent to which participants can learn from passive exposure, directly comparing passive learning to learning that includes reinforcement from an audio-to-visual correspondence. Further, we note potential factors that resulted in successful category development during passive exposure to the stimuli. It is likely that the perceptual distinctiveness of each tone category in the current study aided in the passive formation of the novel tone categories (see Emberson et al., 2013). Therefore, success in forming novel sound categories from passive exposure alone may be moderated by the distinctiveness of the categories. Another factor that may benefit the formation of novel tone categories from passive exposure is the use of high-variability stimuli in close temporal proximity, which aids in the ability to identify the salient acoustic features of the category while learning to ignore the features that are not important for the category. Further, we present the hypothesis that the benefit of high-variability stimuli in close temporal proximity is not paradigm specific. Rather, this type of high-variability training may benefit category learning across paradigms, and it may be a key factor in the ability to form novel sound categories from passive exposure alone.

The current study demonstrates that an incidental learning paradigm can be used to study factors that also impact novel tone category formation during studies that use explicit learning paradigms. However, learning through incidental paradigms such as the one used in the current study can provide results in a single session, rather than over the course of weeks. Further, we demonstrate that incidental learning experiments can be run online rather than having to bring participants to a lab. Overall, we demonstrate the potential that incidental learning paradigms have for increasing our knowledge of factors impacting novel tone category acquisition.

#### *7.1.2.2 Auditory perceptual learning*

In the current study we demonstrated that naïve learners with no lexical tone experience were capable of learning four novel tone categories in a single session through an incidental learning paradigm. These results may provide support for the argument that natural sound categories are optimally learned through the reflexive learning system (Chandrasekaran et al., 2014a, Chandrasekaran et al., 2014b). Specifically, it is hypothesized that the multidimensionality of natural sound categories is best learned through reflexive learning paradigms such as the incidental paradigm used in the current study and in previous studies (Wade & Holt, 2005;

Chandrasekaran et al., 2014a; Lim Gabay et al., 2015; Roark et al., 2020). The results from the current study add to the growing body of research suggesting that there are differences in the processing of natural sound categories depending on the learning system engaged by the task. Therefore, we suggest that future work directed at natural sound category learning take into consideration the learning system being targeted by the training paradigm employed to study learning.

Results from the current study also provide novel contributions to research investigating factors that impact the acquisition of novel sound categories through the reflexive learning system. One factor considered that might impact reflexive learning was age. Participation in the current study was not limited by age. We found that age impacted the speed at which participants completed the task, but learning in each condition did not differ as a function of age. These results inform research on the extent to which age impacts novel sound category acquisition during reflexive learning. Some results from previous research indicate that age does impact category acquisition during reflexive learning (Maddox et al., 2013). Further, there is an expectation that auditory category learning should decline with age. The processing of nonlexical items is negatively impacted by age (Lima et al., 1991), and the neural processing of sounds decreases with age (Skoe et al., 2015). Results from the current study suggest that age may not always hinder auditory category learning. Further investigations of the effect of age on reflexive learning may aid in the understanding of the extent to which age might hinder learning.

Throughout the experiments in the current study, we noted potential differences between older and younger adults. It appeared that stimuli variability disproportionately impacted older and younger adults. Specifically, high-variability tokens within trial from the same talker seemed to help younger participants to learn the novel tone categories. However, high-variability tokens from multiple talkers across trials seemed to disproportionately hinder learning for younger participants, while older participants seemed to benefit from the greater variability of multiple talkers across trials. It is important to note that the current study did not directly test different groups of equal sample sizes at different ages. Rather, there was a range of ages from 19 to 66 and there were fewer older participants in the study than younger participants. Further study on the effects of talker variability across age groups during incidental learning may have important implications for understanding the underlying processes of perceptual categorization during incidental learning and how those processes change across the lifespan.

The current study also demonstrated that novel sound categories could be formed from passive exposure alone with no learning reinforcement. That is, Experiment 2 demonstrated that passive exposure without an audio-to-visual correspondence or a reinforcing motor response was sufficient for the formation of novel tone categories. These results present a contradiction to hypotheses made in the COVIS model regarding reflexive learning, which states that immediate feedback via the audio-to visual correspondence is critical to learning due to the reliance of the reflexive learning system on dopamine generated as participants make predictions and receive feedback (Chandrasekaran et al., 2014b). It is hypothesized that learning occurs due to the proximity of the token variability to the visuomotor association (Gabay et al., 2015) and that the audio-to-visual correspondence was necessary for reflexive learning to occur (Roark et al., 2020). Some previous research states that passive accumulation of acoustic input regularities is insufficient for learning (Roark et al., 2020), while others maintain that it may be possible but that reflexive learning would be much more effective with feedback (McClelland et al., 2002; Goudbeek et al., 2008; Chandrasekaran et al., 2014b). We suggest an investigation into the extent to which participants can learn from passive exposure, directly comparing category formation via passive exposure to learning that includes reinforcement from an audio-to-visual correspondence. We also hypothesize that the role of repeated variable tokens is important to category formation through passive exposure. It is likely that high-variability stimuli in close temporal proximity will benefit category learning in learning paradigms that do not contain audio-to-visual learning reinforcement. Further, we hypothesize that the benefit of high-variability stimuli in close temporal proximity is not paradigm specific. Rather, this type of high-variability training may benefit category learning across paradigms, and it may be a key factor in the ability to form novel sound categories from passive exposure alone.

The current study also demonstrated that token variability impacted learning results. We demonstrated that variable tokens from the same speaker within trial resulted in more robust learning than identical tokens within trial. These results replicate findings from Gabay et al. (2015), extending this finding from synthesized tokens to naturally produced tone categories. These findings suggest that close temporal proximity of variability is highly beneficial for categorization. Experiencing a full range of dimensional variability within trial is more effective than exposure to the full range of variability spread out across trials. We propose that these results also relate to talker variability.

We also demonstrated that talker variability impacts auditory sound category learning via reflexive learning. Exposure to a single talker during training resulted in a sharp decline in categorization accuracy when exposed to multiple new talkers. By contrast, training on multiple talkers resulted in the same ability to generalize to novel tokens from the same talkers and novel tokens from novel talkers. These results are similar to auditory category formation during reflective learning results suggesting that exposure to the full range of acoustic variability during training best prepares participants for generalization. However, we also demonstrated that fewer participants learned in the Multi-talker Condition than in the Single Talker Condition and those that learned did not achieve accuracy scores as high as the learners in the Single Talker Condition. Results from previous studies indicate that there can be a processing cost when listening to auditory tokens from multiple talkers (Wong et al., 2004); Kaganovich et al., 2006; Creel et al., 2008; Perrachione et al., 2011). When considering these results, we suggest an important consideration. In our Multi-talker Condition, each trial contained tokens from the same talker. Therefore, variability from talkers was spread out across trials. As pointed out in Perrachione et al. (2011), Reverse-Hierarchy Theory (RHT; Ahissar & Hochstein, 2004; Ahissar et al., 2009) posits that perceptual learning occurs when listeners identify the correct perceptual level (e.g., pitch contour) and attend to meaningful input. It may be that learners were not able to identify the correct perceptual level due to the exposure to various talkers being spread across trials. Therefore, participants may learn the target tone categories better by hearing multiple talkers' productions within trial (see Barcroft & Sommers, 2005). It is likely that talker variability within trial would train native English learners to ignore differences in pitch level across talkers and attend to pitch contours instead. Participants would hear the consistencies in the pitch contours, identify them as being salient to the category, and be trained to ignore differences in pitch height across talkers. When talker variability is only found across trials, this process is more difficult due to the temporal distance between the most variable exemplars of the category. If within-trial talker variability results in greater learning and greater ability to generalize to novel talkers, then it may indicate that the perceptual categorization mechanism that generalizes salient features of the category across the range of exemplars is most efficient when the full range of features found in the category exemplars occurs in close temporal proximity during training.

The current study also demonstrated that a lack of segmental familiarity did not negatively impact novel sound category learning during reflexive learning. Specifically, participants in were equally able to form novel sound categories when tokens were produced using familiar segments and unfamiliar segments. These results have implications for factors that contribute to effortful listening during novel sound category acquisition. It was expected that the segmental environment or the phonotactic environment of the target sound may inhibit attention to the target acoustic features (Guion & Pederson, 2007; Liu et al., 2011; Wright & Baese-Berk, under review). One hypothesis we present is that a lack of familiarity with the segmental or phonotactic structure may differentially increase the processing challenge depending on the learning paradigm and the learning system engaged by the task.

The current study also demonstrated that production during perceptual learning severely hindered the formation of novel sound categories during incidental learning. These results add to findings from several previous studies that suggest a disruption to perceptual learning can occur if the learners produce tokens on each trial (e.g., Baese-Berk, 2019; Baese-Berk & Samuel, 2016; Leach & Samuel, 2007, Baese-Berk & Samuel, under review). We presented several hypotheses that may account for the disruption of production to perceptual learning during the incidental acquisition of novel tone categories.

## 7.2 Future directions

### 7.2.1 Reflective learning, reflexive learning, and passive learning

In the current work, we situate discussions of the results in a wider understanding of auditory category learning. Specifically, we discuss differences between results from the current study and previous novel tone category studies in light of the multiple systems of learning presented in the COVIS model (Ashby et al., 1998, 2011; Chandrasekaran et al., 2014a). It is important to note that we do not directly test reflective and reflexive learning. A major difference between the current study and previous studies that use reflective learning methodologies is the time course of learning. Reflective methodologies typically require multiple sessions over the course of week to develop novel sound categories. In the current study participants were able to learn four new tone categories in a single session. Therefore, it may be, as suggested by the COVIS model (Ashby et al., 1998, 2011; Chandrasekaran et al., 2014a), that reflexive learning is better suited for learning natural sound categories. However, the time course of learning is only one measure of learning. There are other comparisons that could be made to test learning across systems. For example, besides behavioral mastery, retention is also an important measure of category learning. Further, sensory plasticity is a neural measure of category learning for tone categories. We suggest that studies examine novel tone category formation across reflexive, reflective, and passive learning systems. Participants could be trained to the point of behavioral mastery and the time course of learning could be measured (Wang et al., 1999), as well as retention after a set period of time past behavioral mastery (Reetzke et al., 2018). Also, learning could be measured across systems by examining the development of sensory plasticity, which is measured through the frequency-following response, a neurophonic potential encoding acoustic details along the early auditory pathway (see Reetzke et al., 2018). A study could measure the time course of the development of sensory plasticity across learning paradigms and investigate differences in retention after a set period of time.

### 7.2.2 Token variability

In the current study we demonstrated that token variability within trial aided in category formation during incidental learning. Participants that hear multiple variable productions learn better than those that hear identical tokens on each trial. The benefit of token variability within trial during incidental learning was first demonstrated in Gabay et al. (2015) and incidental learning studies that followed incorporated the idea into their methodology (Roark et al., 2020). However, there has not been much discussion on token variability within trial. The results from our Control Condition, which contained passive exposure to the stimuli and no learning reinforcement, lead us to hypothesize that token variability within trial may be very important to learning. We hypothesized that having variable tokens in close temporal proximity was particularly beneficial for participants in the Control Condition. By contrast, studies that use explicit learning paradigms tend to only include a single auditory token on each trial during categorization training. They play the sound and the participants respond with the category they think the sound belongs to. The results from the current study add to growing methodological considerations by testing the impact of the composition of multiple tokens on a single trial, suggesting the possibility that the benefit from the inclusion of multiple tokens with high variability on each trial may extend beyond the incidental learning paradigm.

### 7.2.3 Talker variability

The current research examined the effect of talker variability on the incidental acquisition of novel tone categories and found that training on multiple talkers hindered and benefited learning. Accuracy scores were lower for participants trained on multiple talkers. That is, participants trained on multiple talkers learned the novel categories much less accurately than participants trained on a single talker. However, scores for participants trained on multiple talkers did not decrease when generalizing to novel tokens from novel talkers, but there was a dramatic decrease for participants trained on a single talker. By demonstrating the difference between training on a single talker and on multiple talkers, we present a need for further investigation into the incidental acquisition of novel tone categories with multiple talkers. We suggest a line of research investigating the initial learning deficit found when trained on multiple talkers. We may expect to find an initial deficit when trained on multiple talkers due to the greater amount of variation in features across stimuli. However, there may be ways to moderate the impact the variability from multiple talkers has on learning. For example, it is likely that talker variability occurring within trials, rather than across trials would result in more robust learning because it would train the learner to ignore differences in pitch level across talkers and attend to pitch contours instead. Our hypothesis is that when talker variability is only found across trials, the categorization process is more difficult due to the temporal distance between the most variable exemplars of the category. If within-trial talker variability results in greater learning and greater ability to generalize to novel talkers, then it may indicate that the perceptual categorization mechanism that generalizes salient features of the category across the range of exemplars is most efficient when the full range of features found in the category exemplars occurs in close temporal proximity during training.

Another factor to consider with training on multiple talkers is that higher amounts of variability across the stimuli may requires more time to result in more robust learning (Goldinger, 1990; Goldinger et al., 1991; Magnuson & Nusbaum, 2007). In the present study, participants only heard about one thousand tokens over the course of thirty minutes. If participants in a single talker condition and a multiple talker condition were trained longer, it may be that accuracy on novel tokens from the same talker(s) would become move equivalent, and in that case, it would be expected that participants in the multiple talker condition would better generalize to novel talkers. If additional training led to improvements in the Multi-talker Condition over the Single Talker Condition, then it may suggest a more general rule that greater variability in the stimuli requires more time for category development to occur, which would suggest that the task of categorization becomes more difficult with the amount of variation. This is an area that the COVIS model does not fully clarify (Ashby et al., 1998, 2011; Chandrasekaran et al., 2014a). The COVIS model specifies that reflective learning is suited for learning unidimensional rule-based categories and that reflexive learning is suited for multidimensional information integration categories. However, the model simply posits a categorical learning model without specifying predictions regarding a range of dimensionality. Will multidimensional categories with less variability be acquired in the same way as multidimensional categories with higher amounts of variability? Questions regarding the time course of reflexive learning could be addressed in several ways. Participants in single talker and multiple talker conditions could be trained to behavioral mastery or there could be a set number of training sessions. Having a set number of training sessions would also permit an investigation into other training methods, such as increasing talker variability over the course of several training sessions.

### 7.2.4 Segmental familiarity and variability

The current work demonstrated that incidental learning was not hindered by a lack of segmental familiarity, indicating a potential robustness of resistance to distractors during incidental learning. We hypothesize that novel sound category formation in learning paradigms that engage the reflexive learning system may not be hindered in the same ways that learning is hindered in paradigms that engage the reflective learning system. Therefore, it is possible that the inclusion of unfamiliar segments during novel tone category learning could result in different impacts on the two types of learning. The lack of familiarity may increase the processing challenge for the neural systems that are engaged by working memory and executive attention during reflective learning but not increase the challenge for the systems engaged by reflexive learning tasks. However, in the current experiments only a single unfamiliar segment was used. The language acquisition process typically requires learners to attend to target features across multiple syllables containing familiar and unfamiliar segments and phonotactic structures. A greater range of segmental and phonotactic familiarity may impact reflexive and reflective learning systems to varying extents.

One area that we did not address in the current study is segmental and phonotactic variability. Only a single syllable structure was used across all experiments and segments did not vary within experiments. When learning languages, learners must attend to target features across a range of segments and phonotactic structures. Further, by increasing variability through the inclusion of variable phonotactic structures and segments produced by multiple talkers we could investigate details regarding the time course of category learning across the range of variability encountered by those seeking to acquire the target categories during language acquisition (Mullennix & Pisoni, 1999).

### 7.2.5 Production and perceptual learning

The current study demonstrated that producing the auditory token on each trial resulted in a disruption of perceptual learning. We hypothesized that during the reinforcement period of the trial, participants in the production conditions were anticipating and perhaps preparing for the following production. Further, this anticipation and preparation for producing the auditory target may have been a key factor in the disruption of the reinforcement learning. However, some participants learned regardless of the requirement to produce the tokens. Therefore, they were able to continue making predictions during the reinforcement learning section of each trial. Therefore, we hypothesized further that participants that learned were able to direct their attention appropriately. Attention would first have to be given to the audio-to-visual correspondence. If all of the attention was on producing the tokens accurately, then they would not have been able to make predictions regarding the location of the visual target. Therefore, we suggest that attention could be further investigated in future experiments by explicitly increasing attention to producing the targets. Participants could be instructed to produce the target as accurately as possible. It may be that exogenously orienting attention to producing targets in this way would further result in participants overlooking the audio-to-visual correspondence.

Attention that is directed towards production and is potentially disrupting perceptual learning could also be reduced in the learning paradigm. It might be expected that if there is a delay of two to four seconds between the perceptual reinforcement learning mechanism and production, perceptual learning would improve due to the separation of auditory, visual, and motor processes (see Baese-Berk & Samuel, under review). It is expected that distinct, non-overlapping processes will support learning (Forrin et al., 2012). However, if creating a delay between the motor response and the production of the auditory target improves learning, it may be difficult to determine what process benefits from the delay. We also hypothesized that if production is delayed for two to four seconds after perception, the learner may better acquire the perceptual categories due to the need to access the perceptual representation again before producing it. Under this hypothesis the perceptual representation of the word or sound is activated again. From a neural perspective this could suggest that further synaptic strengthening may occur with delayed production of the target. However, it may also simply provide time for the reinforcement learning process to occur. As discussed, there is a narrow window of 0.3 to 2 seconds for reinforcement of predictions to occur (Yagishita, 2014). Investigating differences in response time between successful learners and unsuccessful learners may help address this question.

### 7.2.6 Age and reflexive learning

As discussed 7.1.2, older participants learned as well as younger participants. However, there appeared to be differences in the impact of variability on older and younger participants. Specifically, high-variability tokens within trial from the same talker seemed to help younger participants to learn the novel tone categories. However, high-variability tokens from multiple talkers across trials seemed to disproportionately hinder learning for younger participants, while older participants seemed to benefit from the greater variability of multiple talkers across trials. However, results from the current study are limited by age range and sample size. Ages ranged from 19 to 66 and there were relatively few participants in their 50s and 60s compared with participants in their 20s. Further, many studies that test age as a factor in cognitive processing and learning include participants in their 70s and 80s. Thus, we propose that further study on the effects of talker variability across age groups during incidental learning may have important implications for understanding the underlying processes of perceptual categorization during incidental learning and how those processes change across the lifespan.

### 7.2.7 Further data analysis

Future analysis is also planned for the data collected in the current study. We intend to analyze participants responses at test to create a confusion matrix, which will provide insight into the tone categories that participants confused with each other. This analysis can provide details regarding tonal features that participants found to be salient. Further, we can compare results with reflective learning studies that measured tonal confusions to investigate confusions across learning systems (Francis et al., 2008; So & Best, 2010; Hao, 2012).

A primary reason for altering the methodology from Gabay et al. (2015), as described in Section 3.3.1, was to collect mouse tracking data. An analysis of the mouse tracking data will allow us to examine changes in the participant's decision space over the course of learning. Such analysis, which has not been done before, will provide precise details regarding the time course of learning. It also allows us to develop a dynamic confusion matrix, where we can see which sounds the participants confuse and the time course of the resolution of that confusion across training.

## 7.3 IMPLICATIONS FOR SECOND LANGUAGE ACQUISITION

Researchers involved in language acquisition may be interested in training programs that result in the greatest accuracy across potential variability in the shortest amount of time. Recent work investigating the application of incidental learning to real world language acquisition provides insight into this concern. Wiener et al. (2019) found that scaffolding learning by beginning with acoustically simpler categories in an incidental learning paradigm resulted in improved categorization and more native-like Mandarin tone productions than explicit speech training. We expect that further investigation of the effect of scaffolding from lower to higher levels of acoustic variability during incidental learning would prove beneficial for language acquisition pedagogy. For example, a study over several training periods that contains the Single Talker Condition and the Multi-talker Condition from Experiment 2, as well as a condition that begins with a single talker and increases the number of talkers over several days of training may show that increasing talker variability over time could result in a better ability to generalize to novel tokens from the same talkers and novel tokens from new talkers in the shortest amount of time. Further, we demonstrated that segmental familiarity in the tone bearing unit did not impact novel tone learning. However, we hypothesized that increasing the number of unfamiliar features may impact learning. For example, increasing the variability and lack of familiarity among segmental and phonotactic features may result in a slower time course of learning. These are features that can be examined in future research. It may prove beneficial for language acquisition researchers to have a more complete understanding of the time course of learning that includes a wider range of variability in both reflexive and reflective learning systems.

In Experiment 1 we demonstrated that variable tokens in close temporal proximity resulted in better novel tone category acquisition than hearing identical tokens. We hypothesized that hearing a range of acoustic variability in close temporal proximity helps the learner to extract the salient acoustic features of each tone category and ignore features that are unimportant. Further, we hypothesized that this benefit might extend beyond the incidental learning paradigm and benefit learning in other learning systems as well. For example, we hypothesized that this feature was likely what allowed learners that learned passively in the Control Condition to acquire the tone categories. This hypothesis can be tested in a variety of paradigms and should be easy to implement in language acquisition settings.

We also demonstrated that production during perceptual learning kept participants from learning the categories. Although it is likely that production is especially detrimental to reflexive learning due to the disruption to the reinforcement mechanic, the disruption that production causes to perceptual learning is also attested in other learning systems. Due to the speed at which learners can acquire novel sound categories in the incidental paradigm, it may be beneficial for those involved in language acquisition to seek to develop reflexive learning tools that will enable learners to perceptually form novel sound categories outside of settings where they will need to produce the categories. We hypothesize that initial perceptual training will allow learners to experience reduced inhibition from producing the target categories.

In the current study we do not directly test learning differences that occur between explicit and implicit learning methodologies. However, as the majority of previous research has focused on explicit learning methodologies, we situate our results in light of the similarities and differences between the implicit learning that occurred in our incidental paradigm with previous explicit learning. Further, we have attempted to discuss how efforts to learn language may differ depending on the type of learning, whether implicit or explicit, and the learning systems engaged by the learning paradigm[49]. Differences between implicit and explicit learning extend beyond novel sound category formation. For example, differences between the two types of learning are observed during the acquisition of grammar as well. Starting with Reber (1967), there has been a long history of the examination of the implicit learning of artificial grammar

[49] Throughout the paper, following the COVIS model, we refer to reflective and reflexive learning for explicit and implicit learning. However, research on grammar learning tends to use the terms, explicit and implicit.

(see Shanks, 2005 for review), with results suggesting that humans are able to abstract grammatical rules without conscious awareness and use those rules in making grammatical judgments. Further, evidence from studies comparing implicit and explicit grammar learning suggest, like the COVIS model, that different neural areas are activated during implicit and explicit grammar processing (Seger et al., 2000) and the areas activated during implicit processing coincide with the processing of abstract patterns, while the areas activated during explicit processing coincide with the processing of specific stimuli (Goldberg & Costa, 1981). Thus, similar to the learning that occurred in the current study, implicit learning procedures can result in the development of abstract rules that are not easily verbalized and these rules can then be applied to new stimuli. An area of interest for future study that will be meaningful for language acquisition research is the intersection of sound category learning and grammar learning. Specifically, does the formation of novel sound categories through implicit learning benefit grammar learning over the formation of novel sound categories through explicit learning?

## 7.4 CONCLUSION

In the four experiments in this dissertation, we assessed the acquisition of novel tone categories using natural tokens and an incidental learning paradigm that engages the reflexive learning system. Throughout the experiments we demonstrated that native English participants with no prior experience with the target tone categories, from 18 to 66 years old, can use an incidental learning paradigm with natural tokens to form four novel tone categories after 30 minutes of training with very high, even perfect, accuracy. These findings confirm hypotheses suggesting that incidental learning that engages the reflexive learning system is an effective means of learning sound categories, and we extend the investigation of factors impacting incidental learning into natural speech sound categories.

Across the four experiments we examined factors known to impact novel sound category acquisition. We demonstrated that high variability of tokens within trials resulted in greater learning than when the variability was spread out across trials. We also demonstrated that training on a single talker results in robust learning to novel tokens but a sharp decline when generalizing to novel talkers. By contrast, if participants are trained on multiple talkers during training, there is less learning, but there is little or no difference when generalizing learning to novel talkers. We also demonstrated that the presence of an unfamiliar vowel in the auditory stimuli did not impact the incidental formation of novel tone categories during perception only training. Further, we demonstrated that producing the tokens on each trial destroyed perceptual learning, and we presented multiple hypotheses regarding the nature of the disruption for future investigation. We also demonstrated that the presence of an unfamiliar vowel did not further disrupt perceptual learning over training with familiar segments. Thus, as a whole, this dissertation illustrated that learning paradigms that engage the reflexive learning system are effective and efficient means for learning novel tone categories. These paradigms can be used to investigate multiple factors known to impact novel sound category acquisition. Going forward, we propose that future research on novel speech perception consider the learning system engaged by the paradigm used in the study and interpret their findings accordingly. For example, when using a learning paradigm that engages the reflective learning system and findings suggest that learning novel tone categories is extremely difficult, one could state that learning tone categories is very difficult when attempting to learn the categories by engaging a learning system that is not well suited for the task. For, as we demonstrated, learning novel tone categories does not need to be difficult.

# REFERENCES CITED

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2008). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 285–299.

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*(1), 3.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience*, *32*(41), 14156–14164.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481. https://doi.org/10.1037/0033-295X.105.3.442

Ashby, F. G., & Casale, M. B. (2003). The cognitive neuroscience of implicit category learning. *Advances in Consciousness Research*, *48*, 109–142.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*(5), 204–210.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annu. Rev. Psychol.*, *56*, 149–178.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*(2), 83–89.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*(1), 124.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*(6), 1178–1199.

Ashby, F. G., & Rosedahl, L. (2017). A neural interpretation of exemplar theory. *Psychological Review*, *124*(4), 472.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154.

Baddeley, A. (1986). *Oxford psychology series, No. 11. Working memory. New York, NY, US*. Clarendon Press/Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*(4), 981–1005.

Baese-Berk, M. M., & Samuel, A. G. (n.d.). *Just give it time: Differential effects of disruption and delay on perceptual learning*.

Baese-Berk, M. M., & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, *89*, 23–36. https://doi.org/10.1016/j.jml.2015.10.008

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, 387–414.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217–239.

Best, C. T. (1995). *A direct realist view of cross-language speech perception. In. W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research (pp. 171–204)*. Baltimore: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, *1334*.

Bluhme, H., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics*, *22*, 51–57.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [computer program](2011). *Version*, *5*(3), 74.

Bradley, E. D. (2017). A Comparison of Stimulus Variability in Lexical Tone and Melody Perception. *Psychological Reports*, 0033294117734832. https://doi.org/10.1177/0033294117734832

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310.

Brooks, L. R. (1978). *Nonanalytic concept formation and memory for instances*.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*(2), 185–209.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. Wiley.

Burke, D., Mackay, D., & James, L. (2012). Theoretical approaches to language and aging. *Models of Cognitive Aging*.

Bybee, J. (2001). *Phonology and language use* (Vol. 94). Cambridge University Press.

Chandrasekaran, B., Koslov, S. R., & Maddox, W. T. (2014). Toward a dual-learning systems model of speech category learning. *Frontiers in Psychology*, *5*, 825.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236–246.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465.

Chandrasekaran, B., Yi, H.-G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review*, *21*(2), 488–495.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, *8 (45)*(30), 24–27. JSTOR.

Chen, J., Best, C., Antoniou, M., & Kasisopa, B. (2019). Cognitive factors in perception of Thai tones by naïve Mandarin listeners. *ICPHS*.

Chen, J., Best, C. T., Antoniou, M., & Kasisopa, B. (2018). Cross-language categorisation of monosyllabic Thai tones by Mandarin and Vietnamese speakers: L1 phonological and phonetic influences. *Proceedings of the Seventeenth Australasian International Conference on Speech Science and Technology, 4-7 December 2018, Sydney, Australia*, 169–172.

Chen, Y., & Pederson, E. (2017). Directing Attention during Perceptual Training: A Preliminary Study of Phonetic Learning in Southern Min by Mandarin Speakers. *Proc. Interspeech 2017*, 1770–1774.

Clapp, W. C., Rubens, M. T., Sabharwal, J., & Gazzaley, A. (2011). Deficit in switching between functional brain networks underlies the impact of multitasking on working memory in older adults. *Proceedings of the National Academy of Sciences*, *108*(17), 7212–7217.

Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings. *Hearing Research*, *264*(1–2), 48–55.

Colin, C., & Radeau, M. (2003). Les illusions McGurk dans la parole: 25 ans de recherches. *L'année Psychologique*, *103*(3), 497–542.

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131–138.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664.

Dahlen, K., & Caldwell–Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal*, *97*(4), 902–916.

Daigneault, S., & Braun, C. M. (1993). Working memory and the self-ordered pointing task: Further evidence of early prefrontal decline in normal aging. *Journal of Clinical and Experimental Neuropsychology*, *15*(6), 881–895.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*(2), 255–267.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annu. Rev. Psychol.*, *55*, 149–179.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, *10*(6), 732–739.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition*, *128*(1), 82–102.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107.

Escudero, P. R. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Netherlands Graduate School of Linguistics.

Estes, W. K. (1994). *Classification and cognition*. Oxford University Press.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *SPEECH PERCEPTION AND LINGUISTIC EXPERIENCE: ISSUES IN CROSS-LANGUAGE RESEARCH, Strange, Winifred [Ed], Timonium, MD: York Press, Inc, 1995, pp 233-277*. https://search.proquest.com/llba/docview/85604070/18E88132BF52464EPQ/1

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*(7), 1046–1055.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294.

Gabay, Y., Dick, F. K., Zevin, J. D., & Holt, L. L. (2015). Incidental auditory category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1124.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*(2), 110–119.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, *63*(1), 223–230.

Goldberg, E., & Costa, L. D. (1981). Hemisphere differences in the acquisition and use of descriptive systems. *Brain and Language*, *14*(1), 144–173.

Goldinger, S. D. (1990). Effects of talker variability on self-paced serial recall. *Research on Speech Perception Progress Report*, *16*.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251.

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.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds" L" and" R.". *Neuropsychologia*, 317–323.

Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication*, *50*(2), 109–125. https://doi.org/10.1016/j.specom.2007.07.003

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. *Language Experience in Second Language Speech Learning*, 57–77.

Hamann, S. B., & Squire, L. R. (1997). Intact perceptual memory in the absence of conscious memory. *Behavioral Neuroscience*, *111*(4), 850.

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. https://doi.org/10.1016/j.wocn.2011.11.001

Hao, Y.-C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, *61*(1), 135–152.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. https://doi.org/10.1162/jocn.2003.15.5.673

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218–1227.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418.

Houk, J. C., & Adams, J. L. (1995). A Model of How the Basal Ganglia Generate and Use Neural Signals That. *Models of Information Processing in the Basal Ganglia*, 249.

Houk, J. C., Davis, J. L., & Beiser, D. G. (1995). *Models of information processing in the basal ganglia*. MIT press.

Houtgast, T., & Steeneken, H. Jm. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acta Acustica United with Acustica*, *28*(1), 66–73.

Ioup, G., & Tansomboon, A. (1987). The acquisition of tone: A maturational perspective. *Texas Linguistic Forum*, 1–23.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English/ð/-/$þeta$/contrast by francophones. *Perception & Psychophysics*, *40*(4), 205–215.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *43*(1), 88.

Johnson, K. (1997). *Speech perception without speaker normalization: An exemplar model*. https://www.scinapse.io/papers/131022551

Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*(1), 121–148.

Ke, C., & Reed, D. J. (1995). An analysis of results from the ACTFL Oral Proficiency Interview and the Chinese Proficiency Test before and after intensive instruction in Chinese as a foreign language. *Foreign Language Annals*, *28*(2), 208–222.

Kiessling, J., Pichora-Fuller, M. K., Gatehouse, S., Stephens, D., Arlinger, S., Chisolm, T., Davis, A. C., Erber, N. P., Hickson, L., & Holmes, A. (2003). Candidature for and delivery of audiological services: Special needs of older people. *International Journal of Audiology*, *42*(sup2), 92–101.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*(2–4), 63–67.

Kluender, K. R., Lotto, A. J., & Holt, L. L. (2012). Contributions of nonhuman animal models to understanding human speech perception. In *Listening to Speech* (pp. 203–220). Psychology Press.

Knowlton, B. J. (1999). What can neuropsychology tell us about category learning? *Trends in Cognitive Sciences*, *3*(4), 123–124. https://doi.org/10.1016/S1364-6613(99)01292-9

Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology*, *42*(1), 277–303.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. *Handbook of Infant Perception*, *2*, 275–382.

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kuttruff, H. (2016). *Room acoustics*. Crc Press.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353.

LeBovidge, E. A. (2018). *Non-native tone production: Establishing a brain-behavior relationship* [Thesis]. The University of Texas at Austin.

Lee, D.-Y., & Baese-Berk, M. M. (n.d.). *Non-native English listeners' adaptation to native English speakers* [Paper submitted for publication].

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59–67.

Liberman, A. M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America*, *29*(1), 117–123.

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.

Lim, S., & Holt, L. L. (2011). Learning foreign sounds in an Alien World: Videogame training improves non-native speech categorization. *Cognitive Science*, *35*(7), 1390–1405.

Lim, S.-J., Fiez, J. A., & Holt, L. L. (2014). How may the basal ganglia contribute to auditory categorization and speech perception? *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00230

Lim, S.-J., Fiez, J. A., Wheeler, M. E., & Holt, L. L. (2013). Investigating the neural basis of video-game-based category learning. *Journal of Cognitive Neuroscience*.

Lima, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and Aging*, *6*(3), 416.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, *61*(4), 1119–1141.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 732.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English/r/and/l/by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369–390.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671.

Madden, D. J. (1988). Adult age differences in the effects of sentence context and stimulus degradation during visual word recognition. *Psychology and Aging*, *3*(2), 167.

Maddieson, I. (2013). Tone In: Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. *Leipzig: Max Planck Institute for Evolutionary Anthropology Available Online at Http://Wals. Info*.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H.-G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging*, *28*(4), 1042.

Maddox, W. T., Molis, M. R., & Diehl, R. L. (2002). Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels. *Perception & Psychophysics*, *64*(4), 584–597.

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*(2), 391.

Mama, Y., & Icht, M. (2018). Production on hold: Delaying vocal production enhances the production effect in free recall. *Memory*, *26*(5), 589–602.

McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior*, *77*(4), 657–662. https://doi.org/10.1016/S0031-9384(02)00916-2

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207.

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(4), 241.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of/b/and/w. *The Journal of the Acoustical Society of America*, *73*(5), 1751–1755.

Moser, D., Fridriksson, J., Bonilha, L., Healy, E. W., Baylis, G., Baker, J. M., & Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. *Neuroimage*, *46*(2), 549–557.

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

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*(1), 37–43.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. *Speech Perception, Production and Linguistic Structure*, 113–134.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, *5*(4), 291–303.

Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective & Behavioral Neuroscience*, *1*(4), 360–370. https://doi.org/10.3758/cabn.1.4.360

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, *127*(2), EL54–EL59. https://doi.org/10.1121/1.3292286

Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neuroscience Letters*, *352*(1), 25–28.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological Studies in Language*, *45*, 137–158.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42.

Qin, Z., & Zhang, C. (2020). How sleep-mediated memory consolidation modulates the generalization across talkers: Evidence from tone identification. *Age (Year)*, *24*(3.7), 23–3.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407.

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the Trajectory of Sensory Plasticity across Different Stages of Speech Learning in Adulthood. *Current Biology*, *28*(9), 1419-1427.e4. https://doi.org/10.1016/j.cub.2018.03.026

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General*, *122*(1), 92.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, *77*(2), 571–591.

Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, *15*(4–6), 507–521.

Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 75–94.

Roark, C. L., Lehet, M., Dick, F., & Holt, L. L. (2020). *Factors influencing incidental category learning*.

Salthouse, T. A. (1985). *Speed of behavior and its implications for cognition.*

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics*, *31*(4), 307–314.

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *Journal of Neuroscience*, *13*(3), 900–913.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Wadsworth Publishing Company.

Seger, C. A., Prabhakaran, V., Poldrack, R. A., & Gabrieli, J. D. E. (2000). Neural activity differs between explicit and implicit learning of artificial grammar strings: An fMRI study. *Psychobiology*, *28*(3), 283–292. https://doi.org/10.3758/BF03331987

Seitz, A. R., Protopapas, A., Tsushima, Y., Vlahou, E. L., Gori, S., Grossberg, S., & Watanabe, T. (2010). Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. *Cognition*, *115*(3), 435–443.

Shanks, D. R. (2005). Implicit learning. *Handbook of Cognition*, 202–220.

Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, *24*(3), 27–47.

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and Plasticity of Auditory Brainstem Function Across the Lifespan. *Cerebral Cortex*, *25*(6), 1415–1426. https://doi.org/10.1093/cercor/bht311

So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

Squire, L. R., & Knowlton, B. J. (1995). Learning about categories in the absence of memory. *Proceedings of the National Academy of Sciences*, *92*(26), 12470–12474.

Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. University of Hawaii Press.

Sutton, R., & Barto, A. (2005). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*. https://doi.org/10.1109/TNN.1998.712192

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, *185*, 1–20.

Tricomi, E., Delgado, M. R., McCandliss, B. D., McClelland, J. L., & Fiez, J. A. (2006). Performance feedback drives caudate activation in a phonological learning task. *Journal of Cognitive Neuroscience*, *18*(6), 1029–1043.

Vlahou, E. L., Protopapas, A., & Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. *Journal of Experimental Psychology: General*, *141*(2), 363.

Wade, T., & Holt, L. L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *The Journal of the Acoustical Society of America*, *118*(4), 2618–2633.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712.

Werker, J. F. (1989). Becoming a native listener. *American Scientist*, *77*(1), 54–59.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272.

Wiener, S., Murphy, T. K., Goel, A., Christel, M. G., & Holt, L. L. (2019). Incidental learning of non-speech auditory analogs scaffolds second language learners' perception and production of Mandarin lexical tones. *Proceedings of the International Congress of Phonetic Sciences*.

Wong, P. C. M., Nusbaum, H. C., & Small, S. L. (2004). Neural Bases of Talker Normalization. *Journal of Cognitive Neuroscience*, *16*(7), 1173–1184. https://doi.org/10.1162/0898929041920522

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585.

Wong Puisan & Lam Ka Yu. (2021). Characteristics of Effective Auditory Training: Implications From Two Training Programs That Successfully Trained Nonnative Cantonese Tone Identification in Monolingual Mandarin and Bilingual Mandarin–Taiwanese Tone Speakers. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2021_JSLHR-20-00436

Wright, J., & Baese-Berk, M. M. (n.d.). *The impact of phonotactic features on novel tone discrimination.* [Paper submitted for publication].

Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C., Urakubo, H., Ishii, S., & Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, *345*(6204), 1616–1620.

Zamuner, T. S., Morin-Lessard, E., Strahm, S., & Page, M. P. (2016). Spoken word recognition of novel words, either produced or only heard during learning. *Journal of Memory and Language*, *89*, 55–67.