December 2020

The Use of Musical and Perceptual Training in the Classroom for Teaching Tone to L2 Learners with a L1 Stress Language

Elizabeth Ann Hosking Elton
Syracuse University

ABSTRACT

In the study of teaching second languages, there has been limited research on the teaching of tonal languages as a second language (L2). Despite this, perceptual training and a background in musical training has been found to be useful for perception, discrimination, and identification of L2 tones. This study examined and compared the effects of two different training techniques, musical training (i.e., using musical concepts and/or instruments) and perceptual training (i.e., listening to targeted contrasts in tones), between musicians and nonmusicians on the learning of L2 tonal perception, discrimination, and identification (TPDI) accuracy.

A within-participants intervention research design was used, where each participant experienced both kinds of training, implemented in a counterbalanced order across training groups. The shelter-in-place mandate due to COVID-19 resulted in key changes to the planned methodology, principally an abrupt transition to online training and the reduction of training length from two days to one day. Extensive analyses of learner TPDI performance included in each training type at both the word and vowel level, as well as the ability to generalize to new tones and new tonal melodies, were conducted by individual participant as well as by group, including by level of musical background. Participant views of the training methods were also analyzed.

Perceptual training was found to be almost universally descriptively superior to the musical training, and at times also inferentially superior across all participants, and also within each group (i.e., musician vs. nonmusician). Between each group, the musicians descriptively outperformed the nonmusicians almost universally at the start and end of the study regardless of training. Perceptual training also enabled nonmusicians to narrow the performance gap to some extent between themselves and musicians. Regarding the ability of participants to generalize

their combined trainings, analyses revealed little if any effect on the ability to perceive, discriminate and identify new tones and tonal melodies. All above patterns were similar across word and vowel TPDI accuracy. In the post-training survey of attitudes, more than two thirds of all participants expressed a preference for the musical training compared to the perceptual training, citing that the musical training was more interactive. However, while the majority of musicians (six of seven musician participants) favored the musical training, only about half of the nonmusicians (five of nine nonmusician participants) favored the musical training as opposed to the perceptual training.

The Use of Musical and Perceptual Training in the Classroom for Teaching Tone to L2 Learners with a L1 Stress Language

by
Elizabeth Elton

B.A., Syracuse University, USA, 2018

Thesis
Submitted in partial fulfillment of the requirements for the degree of
Master of Arts in Linguistics Studies

Syracuse University
December 2020

Acknowledgements

First, I would like to express my profound appreciation for my advisor, Dr. Brown. Without her assistance and direction throughout the entirety of this process, this thesis would not have come to fruition. Thank you for your support. Thank you for your encouragement. Thank you for your guidance in every step of this process. Thank you.

Second, I would like to sincerely thank Dr. Green for his advice and support through several iterations of this thesis as I worked to present the information as clearly and concisely as possible. Without his feedback and patience, this thesis would not be what it is. Third, I would like to extend my sincere gratitude for the rest of my committee, Dr. Habib and Dr. Ticio Quesada for their support and insightful feedback to make this thesis stronger.

Fourth, I want to thank my fellow LSP peers, including (but by no means limited to) Ade, Chahinez, Nikkia, Kate, Liz, Bekka, Missi, Will, Xuan, and Yan. Thank you for listening to my struggles and for offering advice on how to stay sane through this process. Thank you for your persistent and continuous support and encouragement in my abilities.

Fifth, I want to thank my friends and family who have supported me and reassured me throughout my graduate studies. I want to thank them all for their patience with my limited time to give. Without you all, I would be lost. Specifically, thank you Sam, Zac, Mari, Ryanne, and my family.

Lastly, I have to thank my enduringly patient, kind, and generous partner, Zach. His assistance in taking care of household chores, caring for our fur-children, and general love has given me the strength and courage to keep going even as this study was deeply affected by COVID-19.

Thank you all.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In the study of teaching second languages (L2), there has been limited research on the teaching of tonal languages. Tonal languages are defined as languages "in which an indication of pitch enters into the lexical realization of at least some morphemes" (Hyman 2006, p. 229). Fundamentally, this refers to pitch being used to convey meaning. In tonal languages, high (H) tones have higher pitches relative to low (L) tones (which is illustrated in the pitch traces in (6a-c)), generally.[1] Orthographically, tones are represented with the examples below from Yoruba, a Kwa language spoken primarily in Nigeria with three phonological tones, H(igh), M(id), and L(ow) (Awobuluyi 1978; Good 2012; Ward 1952):

(1) bọ́
`to slip, to escape, to feed'

(2) bọ
`to worship'

(3) bọ̀
`to return, to come'

In (1), *bọ́*, the acute accent is used to depict a H tone while in (3), the grave accent on *bọ̀* is used to depict a L tone. In (2), though, the lack of an accent on *bọ* indicates a mid (M) tone. The only phonological difference between these three words is that (1) is produced H toned while (2) is produced M toned and (3) is produced L toned. Additionally, it should be clear from the gloss translation that each word has a different meaning. This is what is meant by "pitch

[1] Importantly, over the course of an utterance, however, the relative pitch differences between a H tone and a L tone can change. The phenomenon of downstep is an example of how, over an utterance, relative pitch differences may shift, such that a H tone will be produced lower that it generally would (i.e. as in isolation).

enter[ing] into the lexical realization;" it is only a difference in pitch that these two otherwise identical words differ in meaning. When pitch is used in this way, we can refer to it as tone. This thesis focuses on the learning of tones in a L2 that is a tonal language by learners whose first language (L1) does not make use of tone for lexical distinctions. For this study, the L1 of the learners is English, Spanish, or Albanian, which from the standpoint of prosodic typology are all stress languages[2].

A stress language is a language "in which there is an indication of word-level metrical structure" (Hyman 2006, p. 231). In other words, stress refers to the degree of prominence given to different syllables in a lexical word. Stress languages tend to have one primary prominence on every lexical word (Hyman 2006). An example from English can be used to illustrate this definition. Consider the following examples:

(4) *greenhouse*
[ˈgɹinˌhaʊs]

(5) *green house*
[ˈgɹinˈhaʊs]

These words differ in the relative prominence or strength between the syllables. The compound *greenhouse*, in (4), has primary stress (i.e., greater prominence) on its first syllable while *green house*, in (5), is marked for primary stress on both syllables, illustrating that each retains its status as a lexical word. The difference between tone languages and stress languages, then, is in how these types of languages make use of pitch and prominence. With tone, pitches relative to

[2] Prosodic phonology refers to the study of suprasegmentals. Alternatively, segmental phonology refers to the study of consonants and vowels (Gussenhoven 2015). A study of segmental phonology, then, would look at the processes and alternations of consonants and vowels while prosodic phonology is concerned with the processes and alternations of suprasegmentals, such as stress and tone.

one another is most important, but with stress, prominence or weight relative to the strength of the syllables around it is more important (Gussenhoven 2015).[3] This thesis explores how a musical training background, musical training techniques, and perceptual training techniques[4] affect the learning of a tonal language by learners of a L1 stress language.

It is pedagogically important to recognize the differences between perception, discrimination, identification, and production of a target feature in learning an L2. These terms are used widely in the literature, often without explicit definition. Perception refers to the L2 listener's awareness of a target feature's existence, such as tone. In other words, an L2 listener would consciously realize that tone is being used in the target language. Discrimination relates to the L2 listener's ability to differentiate a target feature's types (i.e., a H tone from a L tone). Identification refers to the L2 listener's ability to name a target feature's type (i.e., perceiving a H tone and identifying it as a H tone). Production, on the other hand, is the L2 listener's ability to produce a targeted feature and its types (i.e., being told to speak a word with a H tone, and the ability to produce that tone). Among the assumptions made in this thesis, one is that identification is a skill built from discrimination, and discrimination is built from perception. So, in order to refer to all three skills together as a whole, the acronym, PDI, will be used, generally.

---

[3] It is worth noting that these systems (tone and stress) can be used to typologize languages, where stress languages fall on one end of a cline and tonal languages fall on another. In the study of prosodic typology, languages that fall in the middle of this cline have been sometimes categorized as pitch-accent languages (Gussenhoven 2015; Hyman 2006), but more recently Hyman (2009) has called to dispense with the term "pitch-accent" when used to typologize a language. Instead, current literature is redefining these labels and focusing on the variability between stress and tone as a spectrum. This thesis focuses on stress language speakers attempting to learn tonal languages, and thus, it pays most attention to languages that fall at the ends of the spectrum rather than those in the middle.

[4] A musical training background, musical training techniques, and perceptual training techniques are defined later in this chapter.

However, to be more specific in regards to tonal PDI, this will be referred to as TPDI. This thesis seeks to test L1 stress language learners' ability to perceive, discriminate, and identify L2 tones. Since the effect of musical training techniques on L2 tonal production has been examined by Shi (2018), this thesis focuses on the effects of training on TPDI of learners' L2 tones.

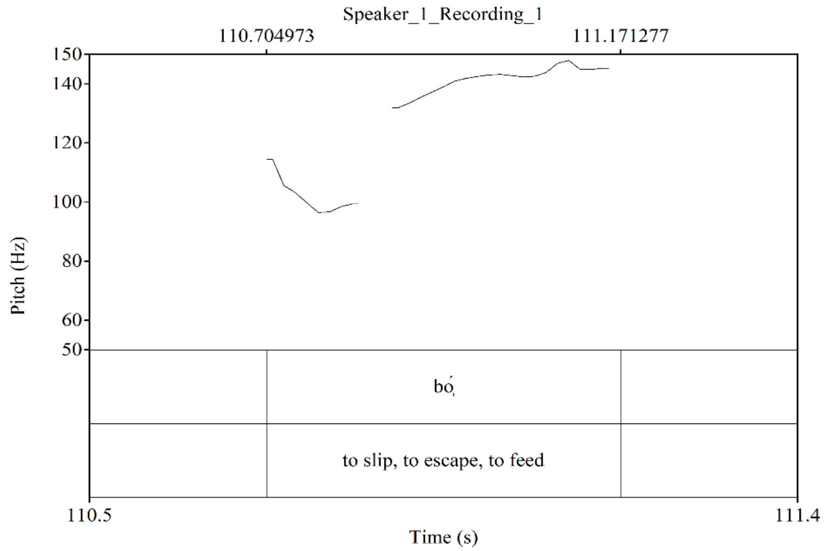
In the instruction of tonal languages, a learner's understanding of pitch height, direction, and slope impacts their TPDI and production of target tones. Pitch height refers to the acoustic frequency at which the pitch is produced while pitch direction refers to the fall or rise of the tone as it is produced. Pitch slope, on the other hand, relates to the change or lack thereof in movement of the acoustic frequency (Gandour & Harshman 1978). Relative changes in pitch slope or direction for one tone indicates a tonal contour or, in the absence of a change, a level tone.[5] Research has shown that languages employ pitch height, direction, and slope differently. For instance, Indo-European language learners perceive pitch height as more important in their perception of tonal language input while some tonal language speakers, like Mandarin and Thai speakers, are more likely to identify a change in pitch slope because these changes are lexically meaningful in their L1 (Li, Shao & Bao 2017; Mennen & Leeuw 2014). More specifically, Li, Shao and Bao's (2017) Indo-European language learners were all L1 stress language speakers, and they were more "influenced" by pitch height than pitch slope in their perception of the target tones (p. 120). This suggests that L1 stress language learners are listening for discrete pitch levels, and will potentially miss key changes in the pitch's movement (i.e., slope) that are

[5] To qualify, relative changes in pitch are only applicable to differences in linguistic tone when those variations are meaningful variations. Human produced pitch is not "pure." In other words, pitch produced by an instrument does not waver, but when produced by a human, pitch often does waver. Not all these pitch variations signal a meaningful contour.

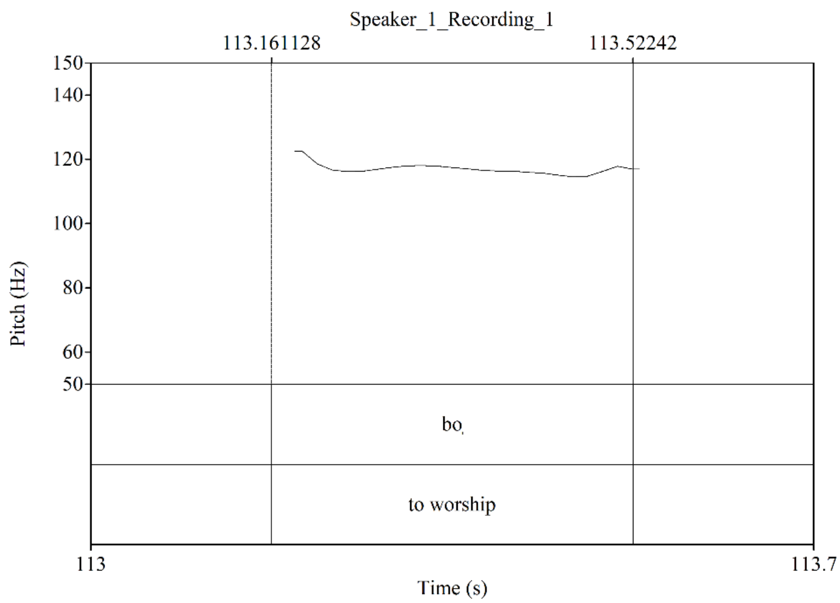
lexically critical in learning an L2 like Mandarin or Thai. Therefore, as an example, in a language that makes use of complex contours, the L1 stress language learner may pay most attention to the L2 tone's discrete pitch level, which may obscure the more lexically important change in slope. This is pedagogically important because teachers should be aware of their students' potential predispositions when perceiving, discriminating, and identifying L2 tones. More importantly for this study is that Yoruba tones are level tones, which indicates that the introductory L1 stress language learners in this study would likely focus on the lexically important difference in pitch levels of the L2 tones.

In terms of target language, this study explored the TPDI of Yoruba tones due to the distinct pitch ranges of each tone type (i.e., H, M, and L) in Yoruba, demonstrated with the pitch traces below in (6a-c), each tone's category is easily definable. Importantly, of the three tones in Yoruba, researchers have noted that M tone is unstable in the language, because in certain sentential environments, M alternates to a H tone (Akinlabi & Liberman 2000; Pulleyblank 2004). Since this thesis focuses on the introductory learning of monosyllabic and disyllabic words, not in sentential context, there was no concern of the M tone alternating to a H tone. Further regarding tonal processes that may impact TPDI, one such phonetic process generates a contour tone (either a rising or falling contour) on the final syllable of disyllabic words (Pulleyblank 2004). This process was taken into consideration during data collection.

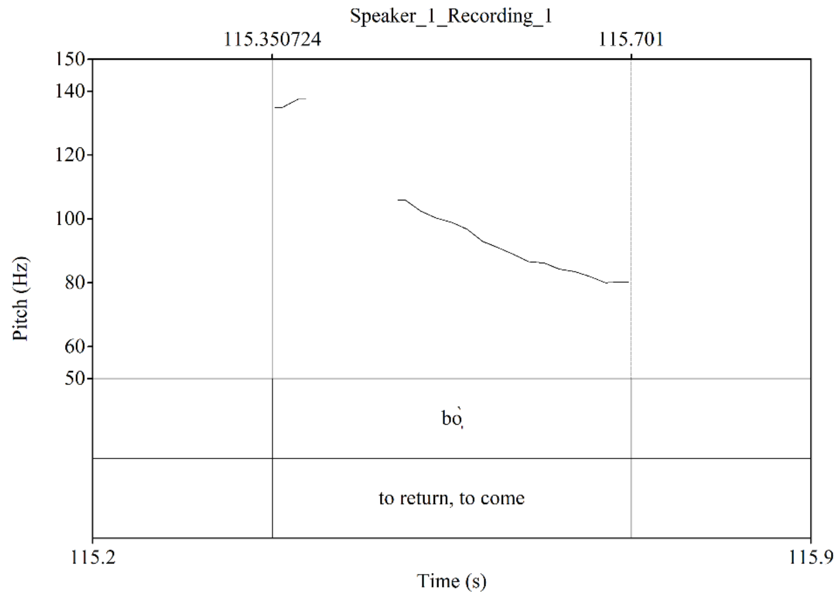
(6) a.


Speaker_1_Recording_1
110.704973
111.171277
150
140
120
100
80
60
50
Pitch (Hz)
bó̩
to slip, to escape, to feed
110.5
111.4
Time (s)

b.


Speaker_1_Recording_1
113.161128
113.52242
150
140
120
100
80
60
50
Pitch (Hz)
bo̩
to worship
113
113.7
Time (s)

c.



This thesis examines two different training procedures for teaching L2 tones: musical training and perceptual training. Musical training, for this thesis, has two related but distinct definitions that are crucial for understanding the present research. The first, notated as a "musical training background," refers to the study of music as a discipline. In prior research, participants' musical training background has been utilized to advantageously learn L2 phonology (Chobert & Besson 2013; Kirkham et al. 2011; Li & DeKeyser 2017; Pei et al. 2016; Perfors & Ong 2012; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007; Zhao & Kuhl 2015). This version of musical training will be referred to as a "musical training background." However, other research has examined the use of applying musical training into the L2 language classroom to apply to language learning, specifically for L2 phonology (Shi 2018). This second definition of musical training will simply be referred to as "musical training" in this thesis. On the other hand, perceptual training is a training procedure employed by researchers, most often in a laboratory, and rarely in a classroom setting. Perceptual training is defined as a learner's exposure to multiple speakers and/or different tokens of stimuli, which is focused on a specific and targeted contrast, as a way for the learner to gain awareness of some target L2 feature in the language (Zhang et al. 2013). Most of the reviewed literature in this thesis that uses perceptual training focuses on targeted contrasts between tones.

In the following chapters, a literature review on the effect of musical training and perceptual training on learning L2 tones will be examined in Chapter 2. The chapter will start by discussing how a musical training background is described in the literature as well as the efficacy of a musical training background when employed for the TPDI and tonal production in a tonal language by L1 stress speakers. Then, the chapter will shift focus to a review of the literature on perceptual training. This will cover its effectiveness for learners' TPDI and tonal production, reviewing specific types of perceptual training as well. In Chapter 3, the methodology is described, starting with how participants were recruited and what stimuli were obtained for the study. Next, this chapter will detail the various tests (pretest, posttest, and generalization test) and procedures that were employed for this study. In the following chapter, Chapter 4, results are presented. Chapter 5 reviews, discusses, and contextualizes the results of Chapter 4. Finally, Chapter 6 comprises the conclusion.

Chapter 2: Literature Review

This chapter will explore the literature related to the two trainings presented in Chapter 1: musical training and perceptual training. Specifically, a review of how effective a musical training background and musical training have been for tonal perception, discrimination, identification (TPDI), and tonal production will be explored in Section 2.1. Then in Section 2.2, the literature on perceptual training's effectiveness will be examined. The literature in Section 2.2 will also explore the effectiveness of different types of perceptual training on TPDI and tonal production. Lastly, Section 2.3 presents the research questions and hypotheses for the present study.

2.1. Musical Training

A background in musical training has been found to be useful for perception, discrimination, identification, and production of tones (Chobert & Besson 2013; Kirkham et al. 2011; Li & DeKeyser 2017; Pei et al. 2016; Perfors & Ong 2012; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007; Zhao & Kuhl 2015). However, in terms of tonal perception, discrimination, and identification, other research has challenged whether tonal perception, discrimination, and identification is truly aided by a prior musical training background (Wayland, Herrera & Kaan 2010; Zhao and Kuhl 2015). This section will also review whether a musician's superior ability in TPDI and production is due to aptitude or learned skill. Lastly, an examination of research that has applied musical training to the teaching of tones will be conducted. In all, this section of Chapter 2 seeks to synthesize the literature on musical training as well as a musical training background and its effects on learning suprasegmentals (i.e., tones, stress, intonation, etc.) by L1 stress language learners.

2.1.1. Perception, Discrimination, and Identification

A musical training background has been found to aid in pitch perception (Chobert & Besson 2013; Kirkham et al. 2011; Perfors & Ong 2012; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007), discrimination (Chobert & Besson 2013; Perfors & Ong 2012; Zhao and Kuhl 2015), and identification (Chobert & Besson 2013; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007). The landmark study by Wong and Perrachione (2007) set a precedent for later studies in the learning of TPDI, so this study will be examined at length. Additional research will be discussed below to elucidate what the literature has found since Wong and Perrachione's study.

Wong and Perrachione (2007) constructed a model in their methodology that later research followed. To begin with, they detailed what constitutes the definition of a musician for later literature. They identified a musician as a person who privately trained with their instrument for six years; additionally, they must have started lessons before the age of 10. On the other hand, a "nonmusician" was given its own definition as well. These individuals must not have had private training with an instrument or instruments for more than three years, regardless of their age when they began. They incorporated a perceptual training procedure into their methodology. Their stimuli for this training were licit English monosyllabic pseudowords (e.g. [pʰɛʃ], [dɹi], [nɛɹ], [vɛʃ], [nʌk], [fjut]), naturally produced by an L1 American English speaker. They, then, digitally altered the stimuli to superimpose Mandarin tonal melodies (Tone 1, Tone 2, and Tone 4) in Praat. Before training, they included a pretest that they entitled the "pitch pattern identification test" in order to test for participants' perception of linguistic pitch, generally. They recorded separate stimuli for this test. These stimuli were five Mandarin vowels produced by 4 L1 Mandarin speakers with Tone 1, the level tone. From these recordings, the researchers digitally altered the tones of these vowels by each speaker to include the other two tones, ending with a total of 60 stimuli. After each training session, they would quiz participants to test for accuracy on PDI of the stimuli.

Wong and Perrachione's (2007) findings indicate that a musical training background increases accuracy in tonal discrimination and identification. In analyzing their results, Wong and Perrachione divided their participants into "successful" and "less successful learners" (p. 573). Successful learners were defined as reaching 95% accuracy over two successive training sessions. Less successful learners were described as improving by 5% or less for four successive training sessions. Training sessions were not predetermined in this study. Rather, they continued training until their participants reached "their individual asymptotic performance" (p. 573). Nine of their learners were categorized as successful learners by the end of training while eight were categorized as less successful. They found that of the nine successful learners, seven were musicians, and only one musician ended training as a less successful learner. In fact, they found that a musical training background significantly predicted successful learning. This finding indicates that a musical training background is useful for L2 tonal teaching.

Perfors and Ong (2012) study differed in their methodology, but they did adapt their methodology from Wong and Perrachione (2007). While Perfors and Ong (2012) incorporated a training procedure into their methodology, they used distributional training instead. Distributional training differs from perceptual training in that it does not include multiple speakers for the stimuli and the stimuli are ordered in a continuum along some target feature. For Perfors and Ong, the target feature for the continuum was pitch. In other words, with seven different tokens of the vowel [i], the first token in the continuum is representative of Mandarin's Tone 1 while the seventh token is representative of Tone 2; the tokens in between are altered to be between the first and seventh token on a continuum, such that the second token is closer to sounding like Tone 1 than the third token. They repeated this continuum to their participants several times over 10 minutes, and participants were instructed only to listen. This constituted their training methodology. Their definition of a musician also slightly differed. In an adaption to Wong and Perrachione's (2007) study, Perfors and Ong (2012) defined musicians as individuals who began musical training before the age of 15 with 5 consecutive years of private music lessons. In an additional contrast, Perfors and Ong's nonmusicians were not given a unique definition as they were in Wong and Perrachione (2007). Also differing from Wong and Perrachione (2007), Perfors and Ong only had their participants complete posttests after training; no pretest was given. The test only assessed for discrimination abilities by having participants determine if one stimulus differed from the previous stimulus (in terms of the target feature). Identification tasks were not part of the testing.

Through this methodology, Perfors and Ong (2012) found that musicians exceeded nonmusicians in the discrimination tasks of the test. In fact, musicians were significantly more accurate in their test scores than nonmusicians. However, they found no significant effect of the training, indicating that distributional training is not as effective as perceptual training for teaching tones. Therefore, this study supports that a musical training background is helpful regardless of whether a training procedure is effective or not because musicians still outperformed nonmusicians. They also found that while a "total duration of musical training" background greatly correlated to higher scores on the posttest (p. 843), there was no significant effect on the scores due to length of training background among the musicians only. In other words, the varying length of an individual musician's prior musical background versus another's did not impact the musicians' within-group scores. Perfors and Ong discuss that this could actually indicate that musicians simply have better PDI abilities overall, despite length of training. It could also potentially indicate that limited musical training background is necessary to be advantageous in tonal discrimination. However, further research would need to test this. Nevertheless, this study corroborates Wong and Perrachione's (2007) study in perception and discrimination.

On the other hand, there are studies that challenge the results of the research outlined above (Wayland, Herrera & Kaan 2010; Zhou & Kuhl 2015). Wayland, Herrera and Kaan's (2010) study looked at learners' abilities to identify tonal contours categorically. In other words, they were testing to see if participants could identify that a specific change in pitch direction and slope was a distinct, meaningful unit despite the relative pitch changes of each token for one category (i.e. the rising tone as one category and the falling tone as another). They did this by exposing participants to the same minimal pair of words that only differed due to a difference in tone. For instance, one word would have a rising tone (the pitch rose in height) while the other would have a falling tone (the pitch fell in height). Participants were required to choose one of two visually presented tones[6] for each token of the minimal pair they heard. They specifically focused on contours because non-tonal L1 learners tend to only focus on a pitch's starting height rather than the slope or direction of the pitch contour (Li, Shao & Bao 2017), which would then interfere with the categorization of the tonal contour. Their methodology included a perceptual training procedure analogous to Wong & Perrachione's (2007) described above with a pretest

[6] They visually represented the tone by showing traces of the pitch contour: A line that started lower and rose to a higher position represented the rising tone while a line that started higher and fell to a lower position represented the falling tone.

and posttest. The pretest and the posttest included level tones[7] and contour tones. Their definition of musicians was comparable to Wong and Perrachione's (2007), though their musicians only needed six years of combined musical training background while their nonmusicians were defined as individuals with a maximum of two years of combined training.

Wayland, Herrera and Kaan (2010) found that their musicians outperformed nonmusicians in identification tasks. However, both nonmusicians and musicians improved at a comparable rate throughout their perceptual training sessions. In other words, musicians outperformed nonmusicians for level and contour tones in the pretests and posttests scores; with training, though, both nonmusicians and musicians improved in their identification abilities at about the same rate. In fact, the musician and nonmusician groups did not significantly differ in their scores on the pretest and posttest before and after training. Further, regardless of musical training background, Wayland, Herrera and Kaan found that perceptual training provides the ability to recognize the tonal contour as a linguistic category to some degree. This means that participants (musicians and nonmusicians) gained some ability to abstract these linguistic categories (i.e., the rising tone and falling tone) as a meaningful unit that can be applied to distinguish identical segmental input due to training. So, despite the capacity for musicians to outperform nonmusicians in identification for level tones, Wayland, Herrera and Kaan found that the two groups were about equal in their abilities to consider tones categorically. However, the authors noted a limitation that may have interfered with this finding. They explained that in their

[7] Level tones were incorporated into the testing because they noted that level tones are typically easier to perceive, discriminate, and identify than contour tones. So, this addition of level tones in their testing was used to assess participants' basic proficiency in TPDI.

pretest and posttest, participants were given a time limit, but during training, participants were not given a time limit. This is important because musicians did outperform nonmusicians in identification tasks during training sessions. So, this could have impacted results because musicians may have been able to outperform nonmusicians with additional time.

Similarly, Zhao and Kuhl (2015) examined how a musical training background does or does not affect a learner's categorical TPDI when learning tones for the first time. In their study, they had L1 English and L1 Mandarin speaking participants; only the L1 English speaking participants were further delineated by whether they were musicians or not. They included L1 Mandarin speaking participants as a way of comparing the categorical perception of L1 speakers to L2 learners of the tones. All of their participants completed discrimination and identification tasks for both level and tonal contours. Much like Wayland, Herrera and Kaan (2010), they incorporated perceptual training into their study as well. Also in line with Wayland, Herrera and Kaan (2010), Zhao and Kuhl (2015) found that perceptual training has a positive effect on discrimination of all tone types (level and contour) from pretest to posttest for both musicians and nonmusicians. Moreover, while musicians were found to improve in the identification tasks in their posttest scores, nonmusicians were not, which aligns with prior research as well. Importantly, though, despite musicians seeming superiority to identify tones after perceptual training, they discovered that their L2 learners, whether musicians or nonmusicians, did not perceive tonal categories as L1 speakers did even with the training. In other words, they indicated that English speaking participants used "different strategies in perceiving the tone" than L1 speakers (p. 1458). L2 learners perceived the pitch changes, but Zhao and Kuhl argue that their perception of the pitch change was not based on forming the different tones as linguistic categories. Rather, L2 learners viewed each stimuli's pitch change as separate from one another instead of, for instance, abstracting a falling pitch as always being part of the category of a falling tone. In other words, the L2 learners did not seem to recognize pitch as relative for a tonal category.

The results from Wayland, Herrera and Kaan (2010) and Zhao and Kuhl (2015) versus Wong and Perrachione (2007) and Perfors and Ong (2012) leave questions about how well musical training could be usefully adapted for teaching tones. Even though musicians appear to have better pitch PDI abilities, Wayland, Herrera and Kaan (2010) show that musicians and nonmusicians are about the same in their abilities to categorize relative pitch changes as a tonal category. Further, Zhao and Kuhl (2015) provide evidence that pitch is not being perceived linguistically by participants as it is by L1 tonal language speakers. If so, then musical training's benefits may very well be limited for the learning of L2 tones by L1 stress learners. For these reasons, Wayland, Herrera and Kaan's (2010) and Zhao and Kuhl's (2015) results leave questions about musical training's role as a training technique for teaching tone if a musical training background is unhelpful for learners to perceive the L2 tonal categories.

2.1.2. Production

A musical training background has also been found to aid in pitch production (Chobert & Besson 2013; Kirkham et al. 2011; Li & DeKeyser 2017; Pei et al. 2016). Li and DeKeyser's (2017) study provides an examination of musical ability on TPDI and production. In their study, their participants differed greatly from Wong and Perrachione's (2007). Rather than requiring musician participants to have a total of six years of formal training, where they must have begun lessons before the age of 10, Li and DeKeyser (2017) required that all their participants must not have more than three years of any kind of musical training background. In contrast to prior research, they identified musical ability through perceptive and productive musical ability tests.

Their training procedure was both similar to and different from the Wong and Perrachione (2007) study. It was similar in that monosyllabic words were chosen. Also, each of the words was presented to the participants with Mandarin tones used. It differed from Wong and Perrachione because Li and DeKeyser used 16 words, and each word was a real Mandarin word; additionally, they used all four Mandarin tones on each word, providing 64 distinct stimuli. Also, rather than recording the words and presenting them digitally, the authors opted to deliver them in real-time for the training in order to provide immediate feedback to learners on their productions.[8] They also had two different training conditions. Half of the participants were part of a perceptual training procedure while the other half participated in a production training.

Li and Dekeyser's (2017) results show that musicians in both training conditions were significantly rated more accurate and "nativelike" by L1 Mandarin speakers. However, they found that the production training condition resulted in increased "nativelike" pronunciation for all the participants in this condition (not only the musicians) as compared to the participants in the perceptual training condition. Additionally, in line with previous studies, they found that learners with higher musical ability outperformed those with less musical ability. Importantly, no participant had more than a combined three years of training. This could signal that limited training is needed to achieve the musical ability needed to apply to tonal learning. Alternatively, Li and Dekeyser noted that it could signal that a higher musical aptitude[9] is the reason for the higher rated scores of their participants with high musical abilities. This would indicate that

[8] One of the authors produced the stimuli in real-time while learners participated in the training.
[9] Aptitude, here, refers to an innate ability to perceive and produce pitch accurately.

training is not what provides musicians with greater accuracy in TPDI and production, but that musical aptitude is responsible for these greater gains. This is further reviewed below.

One question that previous research has asked is whether a background training in specific instruments would provide greater pitch PDI and productive abilities over other instruments (Wayland, Herrera & Kaan 2010; Alexander, Wong & Bradlow 2005). Current research has found that it does not matter whether a musician is a vocalist or an instrumentalist[10] (Kirkham et al. 2011); Kirkham et al. (2011) examined the differences between vocalists and instrumentalists in response to research that inquired whether vocalists, due to their extensive training with their vocal cords, would produce tones better than other instrumentalists (Alexander, Wong & Bradlow 2005). Kirkham et al. (2011) tested nonmusicians and an equal number of L1 English speaking vocalists to L1 English speaking instrumentalists. Their definition of a musician (either vocal or instrumental) was that they must have at least four years of formal training and still be playing their instruments. They found that vocalists did not significantly outperform instrumentalists in either the discrimination or production tests. Further, their research aligned with the literature previously described that musicians, generally, outperformed nonmusicians in discrimination and production tasks, as well. This shows that any kind of musical training background can be useful for TPDI and tonal production capabilities, regardless of whether a learner has experience in vocal training or instrumental training.

[10] An instrumentalist is a musician who uses a manmade instrument instead of their vocal cords.

2.1.3. Musical Aptitude vs. Learned Skill: The Effect on L2 Phonological Acquisition

Another important aspect of a musical training background that is important for teaching implications is whether musical experience or musical aptitude allows learners to achieve greater skills in L2 phonology. Talamini et al.'s (2018) study examined the difference between musical aptitude versus musical skill in perception and discrimination tasks. One limitation for the current thesis, however, is that this study did not focus on tonal learning or even stress learning. Their participants were L1 Italian (a stress language) speakers learning English (a stress language) segmental features. Also, distinct from previous studies, their participants were between the ages of 11 and 15,[11] and their musicians were defined as individuals who had been taking music lessons from 2-60 months. Though some of these distinctions pose limitations for the present thesis, the results shed light on the question of whether aptitude or skill is the reason for musician's greater phonological PDI and production. They used the Profile of Musical Perception Skills (PROMS) test for measuring aptitude in their study. Participants completed this test before taking an English Language Teaching (ELT) dictation test; this test required participants to listen to English words and identify them by spelling them on the testing sheet.

Talamini et al. (2018) found that musicians significantly outperformed nonmusicians in the dictation test. More importantly, though, they found that the scores of the PROMS test had no significant correlation on the results of the dictation test. This indicates that musicians' greater abilities in phonological PDI has little to do with aptitude. Additionally important is that 2-60 months was enough to set musicians apart from nonmusicians, and the musicians still

[11] The previous studies defined in this paper had adult participants.

outperformed nonmusicians. This potentially shows that as little as two months of a musical training background is enough to be useful in bolstering phonological perception. However, the limited relevance of this study to the current thesis cannot be overlooked. This study only examined segmental phonology, not suprasegmental phonology, which may demonstrate differences in results.

Further evidence suggests, though, that musical skill[12] would also aid in suprasegmental learning despite aptitude, yielding similar results to Talamini et al.'s (2018) study. Pei et al. (2016) examined tonal language speakers' musical experience versus aptitude to produce segmentals and suprasegmentals. They administered two musical aptitude tests prior to their primary testing, the Advanced Measures of Music Audiation (AMMA) test and a productive musical aptitude test adapted from prior research. They then completed a Foreign Language Imitation test. This test selected five sentences from four languages: French, German, Russian, and Japanese. Participants were to listen to the sentences then replicate one sentence from each language. They were tested on a five point scale to determine their accuracy for both segmentals and suprasegmentals (such as stress). Pei et al. (2016) found that musicians have a higher aptitude than nonmusicians for music, generally. While this may seem antithetical to Talamini et al.'s (2018) study, Pei et al. (2016) discovered that training could help increase musical aptitude,

---

[12] Skill, here, is used to delineate the use of training on suprasegemental PDI accuracy to innate ability (i.e., aptitude). In other words, the musical training background gave each learner the necessary skills to achieve higher PDI accuracy as opposed to the learner's innate ability in music. This is an important distinction because if training, alone, can aid learners in their phonological PDI accuracy, then this can be used in the classroom. If aptitude is responsible, musical training will not be beneficial to learners without musical aptitude.

which is important because it indicates that incorporating musical training into the classroom could help increase aptitude.

Fundamentally, the two studies, Talamini et al. (2018) and Pei et al. (2016), reveal that learned musical training skills alone may be enough to aid in suprasegmental PDI and production. However, while both studies are limited in scope for this thesis, as they did not focus on tone, both provide evidence that similar research applied to tones would yield similar results. This is an area that future research could expand upon, though, as there is still much to be accomplished in this area.

2.1.4. The Incorporation of Music in Teaching Tones

Despite the considerable research on the effect of a musical training background on phonological PDI and production, very little research exists on using musical training as a method for teaching tones. However, one dissertation has incorporated an aspect of musical training into a tonal training method. Shi (2018) drew from previous literature outlining a technique for Chinese Foreign Language teaching that involved musical training[13] (Duanmu 2007; Lin 1985; Yang 2014), but none of the prior literature had actually tested the technique.

Specifically, Shi's (2018) dissertation tested a lesson plan for teaching Mandarin tone that incorporated written musical scales to achieve greater accuracy in tonal production. The motivation behind this method is its focus on bringing tone to the level of awareness. She also used hand gestures (i.e., raising a hand high for a H tone and lowering a hand for a L tone) for this same purpose. Readers can find a lesson plan in the Data Collection and Analysis section of

---

[13] This technique was the incorporation of musical scales to teach Mandarin tones.

her dissertation. In the first procedure of the lesson plan: "Introduction to Chinese tones" (p. 84), tools listed are a PowerPoint and a video. These were used to bring explicit awareness to Chinese tones. Although the content of the video is not provided, Shi does provide the content of the PowerPoint. After the slides which discuss Chinese tone explicitly, there is follow-up content that reviews finding one's own pitch range and applying the tones to a written "tonal map;" this tonal map is equated to a visual musical scale which she then presented to students (whom she reported were mostly familiar with this aspect of musical theory). She used this to show pitch hierarchy of tones. Finally, she had the students practice what they learned. Ultimately, the research found that this method significantly increased the participants' accuracy in tonal production. Additionally, the students self-reported that bringing tone to the level of awareness aided their learning. Her results show that incorporating musical training may indeed help with tonal production accuracy.

However, it is worth noting that the incorporation of hand gestures with the use of musical scales makes it impossible to determine how the use of musical scales independently contributed to the increase in production accuracy. Despite this, some evidence from previous research into the use of hand gestures for L2 tonal learning shows how gestures may have potentially impacted Shi's (2018) results. Zheng, Hirata, and Kelly (2018) along with Baills et al. (2019) reviewed how body gestures affect tonal learning by L2 learners of a tonal language. Zheng, Hirata, and Kelly found mixed results. L2 learners were able to use body gestures to aid in their tonal learning of two Mandarin tones, but not the other two tones tested. Baills, on the other hand, found that body gestures improved their participants' tonal learning. Furthermore, they found that when learners produce the gestures themselves as opposed to simply observing the gestures, their tonal learning was improved. Overall, these results indicate that Shi's use of gestures in her methodology may have indeed been part of her participants' success, and not just the use of the musical scales.

2.1.5. Summary

Musical training and a musical training background may simply be useful in getting learners comfortable with pitch PDI as a relevant linguistic element. This may be due to areas in the brain that have been shown to link music and language together (Maess et al. 2001; Schon, Magne & Besson 2004; Patel 2011). For instance, Broca's area bridges music and language in the brain (Maess et al. 2001). It seems possible, then, that a transfer effect is what helps in the TPDI of tones. With this in mind, and with people being exposed to music generally, research could look even further into whether musical training can advantageously draw from this overlap in order to make connections from musical pitch to language pitch discrimination. This should additionally build from Wayland, Herrera and Kaan's (2010) and Zhao and Kuhl's (2015) studies in order to address the categorization of tones. Also, given that the research has shown important benefits of musical training on tonal learning, and the success of Shi (2018) in employing musical knowledge for L2 tonal production, the area of using musical training as a technique for teaching tone is one that needs further study. Since very little has been studied on musical training' effects in a classroom setting for L2 TPDI, this leaves many questions open as to how musical training can be implemented to help L2 tonal learners.

In all, research has found that musical training seems to enhance pitch PDI and production. It has been discovered that musicians seem to generally perceive, discriminate, identify, and produce tones better than nonmusicians. This may or may not be due to aptitude rather than learned skill. However, research has found that musicians are no better than nonmusicians at perceiving tones as a categorical linguistic element. While musicians can better identify discrete differences in pitch, neither nonmusicians nor musicians seem more able than the other to apply their identification of these discrete differences to the relative, but lexically crucial, pitch ranges that categorize each tone in the language. Furthermore, with training, musicians and nonmusicians begin to perceive tones categorically at about the same rate. Still, though, more research needs to be conducted in this area. Limited research has been administered on the application of musical training in the classroom. Shi's (2018) dissertation incorporated the use of musical scales as a visual representation, but other factors obscure whether the student's gains in tonal production were due to this use of musical scales or not. All in all, the incorporation of musical training in the classroom looks promising, so future research could explore filling this gap.

2.2. Perceptual Training

Perceptual training has many different names in the literature. Other than the name, "perceptual training," it has also been called "auditory training" and "phonetic training" (Inceoglu 2016; Pederson & Guion-Anderson's 2010). They all refer to the same kind of training as defined in the introduction, where learners are exposed to multiple speakers and/or multiple tokens by the same speaker many times to illuminate some targeted contrast. Studies have also found perceptual training to be useful for perception, discrimination, identification, and production of tones (Antoniou & Wong 2016; Godfroid, Lin & Ryu 2017; Li & DeKeyser 2019; Lu, Wayland & Kaan 2015; Perrachione et al. 2011; Wang 2013; Wang et al. 1999; Wang, Jongman & Sereno 2003; Wayland & Li 2008). Furthermore, the scoring of participants' tonal productions in perceptual training has been examined, and it has been found to have certain limitations, but also important benefits (Jiang 2017). Research has also found that in the use of perceptual training, explicit instruction of the target feature is necessary in order for learners to gain the benefits of training (Antoniou & Wong's 2016; Lu, Wayland and Kaan 2015; Pederson & Guion-Anderson's 2010). This literature review will also examine the varying types of perceptual training that have been employed in prior research. This will shed light on the benefits and limitations of certain forms of perceptual training for various types of learners. In all, this section seeks to synthesize the literature on perceptual training and its effects on learning tones by L1 stress language learners.

2.2.1. Perception, Discrimination, Identification, and Production

Perceptual training has been found to aid in pitch perception (Antoniou & Wong 2016; Godfroid, Lin & Ryu 2017; Li & DeKeyser 2019; Lu, Wayland & Kaan 2015; Perrachione et al. 2011; Wang 2013; Wang et al. 1999; Wang, Jongman & Sereno 2003; Wayland & Li 2008), discrimination (Antoniou & Wong 2016; Godfroid, Lin & Ryu 2017; Li & DeKeyser 2019; Lu, Wayland & Kaan 2015; Perrachione et al. 2011; Wang 2013; Wang et al. 1999; Wang, Jongman & Sereno 2003), identification (Antoniou & Wong 2016; Godfroid, Lin & Ryu 2017; Li & DeKeyser 2019; Perrachione et al. 2011; Wang 2013; Wang et al. 1999), and production (Li & Dekeyser 2019; Wang, Jongman & Sereno 2003). The landmark study by Wang et al. (1999) set a benchmark for the use of perceptual training on TPDI. Research conducted since Wang et al.'s (1999) study will also be discussed below to examine what current literature has found.

2.2.1.1. Perception, Discrimination, and Identification

Wang et al. (1999) was a landmark study in research for perceptual training of suprasegmentals, namely tone. They followed a general perceptual training method developed by

Logan et al. (1991). In this way, Wang et al. (1999) had trainees participate in a high variability[14] perceptual training. The training consisted of eight sessions that were 40 minutes long. 400 stimuli for the research were recorded by 6 Mandarin speakers; they recorded various syllabic structures for the stimuli, equally representing all four Mandarin tones. Their participants were 16 L1 American English speakers who had taken one or two semesters of Mandarin Chinese courses. However, only eight participated in the training, while the other eight were controls. They included a pretest and a posttest, which were used to determine the success of training in tone identification. They also included a retention test (conducted six months after training) and two wider application tests they entitled Generalization Test 1 and 2 which were all administered after the posttest. These tests required the participants to apply their knowledge from the training to new stimuli. The first generalization test assessed whether participants could apply their gained knowledge to new stimuli by the same speakers from their training stimuli. However, the second test, Generalization Test 2, tested for whether they could apply their training to both new stimuli and new speakers of the stimuli.

Wang et al (1999) found that perceptual training in high variability conditions showed important benefits in the experimental group for both short and long term retention. The experimental group greatly outperformed the control group. In fact, trainees showed significant improvement compared to the controls. While trainees improved by 21% from pretest to posttest, controls only improved by 3%. This shows that the training enhanced tonal identification

---

[14] Participants had to identify the four Mandarin tones in many different phonetic environments as well as by many different speakers, which constitutes this as high variability perceptual training. This kind of perceptual training is further examined below.

learning for the trainees. They also found that six months after training, the trainees largely retained their knowledge, scoring an average of 87% on the retention test while the controls scored an average of 68%. This reveals that perceptual training seems to have a long-term effect on gained skills in TPDI. However, the application of the tones to additional stimuli was far more difficult for the participants, indicating a challenge in categorical perception, which is corroborated by Zhao and Kuhl (2015) as well as Wayland, Herrera and Kaan (2010).

Wang (2013) found similar results in her study. This study primarily examined how learners with differing L1s (from stress, pitch accent, and tonal) would perceive L2 tones comparatively between each L1 group. More important for this thesis, Wang also looked at whether training would help to increase accuracy in TPDI. The participants for this study included primarily L1 American English, Japanese, and Hmong speakers. However, one L1 Spanish speaker, one L1 Khmer speaker, and one L1 Tagalog speaker were also included as participants. Additionally, none of the participants had experience with Mandarin prior to starting an introductory course during the semester they participated in the training. By the time they underwent the training, they were in their sixth week of the semester. By this time, they had completed their instruction of Mandarin sounds and tones. The stimuli for this experiment were produced by five L1 Mandarin speakers. The study used real Mandarin words of various syllabic combinations. A total list of 160 words (40 minimal quadruplets) for two types of training was created; each word was recorded by the 4 speakers, resulting in 640 tokens. For the pretest and posttest, additional productions were recorded by just one of the four speakers from the training. However, for their generalization test, additional stimuli from another speaker were recorded, constituting their five speakers overall. The participants were split into two groups for two different types of perceptual training procedures: auditory and production training. The auditory training itself required participants to practice identification skills. The researchers presented the stimuli and participants had to select which of the four tones a token corresponded to by pressing a button. Immediate feedback was given, whether correct or incorrectly answered. In the production training, Wang had participants listen to the stimuli and then record their own token. They could then see their production "visually" as a pitch trace. They could also replay their production in comparison to the target stimuli and visually compare their pitch trace to the stimuli's. They could then choose to rerecord or move onto the next stimuli. Additionally, and importantly, despite the fact that many participants completed the production training, all participants' completed an identification posttest. In other words, after the completion of training, no production test was administered for any group.

Wang (2013) found that both groups (the auditory and production training groups) were comparable to the control group (that did not receive training) in their pretest scores. However, The auditory and production training groups' posttest scores were much better than the control group's posttest scores. Their scores were also comparatively much better on the generalization test to the control group. She additionally found that there was no significant effect on scores due to participating in the auditory or production training. In other words, both training procedures showed about equal gains in posttest and generalization test scores. A limitation for this thesis, however, is that they did not determine whether either training procedure was more or less effective for their L1 English, Spanish, and Khmer speakers (stress learners). For this reason, it is difficult to determine how the L1 stress participants' results would have differed from other participants' (i.e., the Japanese, Tagalog, and Hmong[15] participants) had their contributions been considered separately.

2.2.1.2. Production

Wang, Jongman and Sereno's (2003) study was about whether perceptual training could enhance perception and discrimination, but they primarily looked at production of tones. They were testing for their participants' production before and after training. In their methodology, much like Wang et al (1999), Wang, Jongman and Sereno (2003) included 16 L1 American English speakers who had taken one or two semesters of Mandarin courses. Eight were trainees while the other eight were controls. They followed the perceptual training procedure of Wang et al. (1999). Unlike Wang et al. (1999), though, Wang, Jongman and Sereno (2003) had 82 L1 Mandarin speakers as judges. Additionally, Wang, Jongman and Sereno determined 80 stimuli would be used. Half was used for training while the other half was used for the pretest and posttest. The researchers recorded their participants' productions for the pretest and posttest, using the determined stimuli. A set of five judges assessed one stimulus at a time. They assessed the stimulus by writing down the word they perceived by the participants. Additionally, Wang, Jongman and Sereno (2003) also conducted an acoustic analysis,[16] comparing their participants' productions both before (in a pretest) and after (in a posttest) training to L1 Mandarin speaker's

[15] While Hmong would be on the far (tonal) end of the prosodic cline between stress and tone languages, Japanese and Tagalog are more in the middle of the cline. However it is worth noting that both would be closer to the tonal end of the cline than the stress end.

[16] An acoustic analysis compares a pitch trace of one production to another. This would then show the pitch trace's starting height, slope, and direction. This information can be used to compare the starting height, slope, and direction between the two pitch traces.

productions. They incorporated this acoustic analysis to analyze their participants' productions to a native speaker norm.

Wang, Jongman and Sereno (2003) found that perceptual training could help in TPDI and tonal production. Specifically, they found that the trainees significantly improved in their production scores from pretest to posttest while controls did not significantly improve. Further, they found that training had a significant effect on the trainees' improved production scores. In terms of the acoustic analysis, the researchers found that the participants' pretest productions did not as closely align with their native speaker's productions as the participants' posttest productions. In fact, they found that their participants' posttest productions significantly approximated their native speaker's productions better than their participants' pretest productions for all tonal categories.

One article, however, specifically set out to examine rating techniques employed in perceptual training. This revealed important implications for studies that use perceptual training to test for its impact on tonal production. Jiang (2017) examined how the rating methodology for tonal production scoring may be impacted by several different conditions. The stimuli for this study came from the pretest of another study. 35 nonnative speakers (NNS) of Mandarin in their first year of Mandarin coursework produced 20 Mandarin words. 20 participants of the original 35 were selected for this particular study. Each of the 20 produced recordings by the 20 participants was saved into short audio files of single, target words for this experiment. There were three conditions in which these recordings were presented to the raters. In Condition 1, segmental information was stripped from the recordings in PRAAT and the audio file was reduced to a hum. In Condition 2, the original recording with no other aid was given to raters to determine the speaker's tone category. In Condition 3, the original recordings were also used, but pinyin (a Romanized alphabet used to represent Mandarin words and their tones) was additionally provided without tonal symbols (i.e., tonal diacritics) given to mark the specific tone. Participants for Jiang's study were 4 L1 Mandarin speakers. In the experiment, the four participants were given the recordings and the pinyin sheet depending on which condition they were working under. They were required to complete their ratings under one condition in one day. They would then be required to not rate any recordings for one day in order to not impact the following condition(s). When rating, they were given a rating scale to select from: 0 for neutral, 1 for 1st tone, 2 for 2nd tone, 3 for 3rd tone, 4 for 4th tone, and 5 for none of them. Raters typically used 5 if they were unable to identify a speaker's tonal production as belonging to any of the other tonal categories. Additionally, Jiang also included an acoustic analysis of the speakers' productions as compared to native speaker's productions of the same words.

First, importantly, Jiang (2017) found that raters strongly agreed on their ratings of individual recordings, overall, showing that there was little discrepancy between judges in their ratings. In Condition 1, raters agreed that the learners produced the tones correctly about 54% of the time. In Condition 2, raters agreed that learners produced the tones correctly about 71% of the time while in Condition 3, they agreed that about 74% of the stimuli by learners were produced correctly. Jiang also found that from Condition 1 to Condition 2, the speakers' accuracy significantly increased. Further, from Condition 2 to Condition 3, accuracy also improved. This shows that if writing systems are employed as a method for identifying participants' tones in a study that this may lead to raters' greater accuracy in identifying the participants' tones. Further, only looking at suprasegmental data (i.e. just the pitch without the segmental information) impairs native speaker raters from determining the correct tone at times. This then implicates that lexical context is useful for listeners to identify the tones being used. In their acoustic analysis in

Condition 2, they found that participants' tokens marked incorrectly for tone were significantly different from participants' tokens marked correctly for tone. In other words, the participants' tones marked correctly by the raters were more closely aligned to a native speaker's tonal productions of the same word. Additionally, for Condition 3, they found that participants' tonal productions marked correctly or incorrectly by their raters did not significantly differ in their productions' alignment to the native speaker's production of the same words. Since Condition 3 additionally included a written representation of the words, the researchers explain how this could implicate that using written representations to rate tonal production accuracy might unduly bias raters' judgments about whether a tone was produced correctly or not.

2.2.2. Explicit vs. Implicit Instruction

Pederson and Guion-Anderson's (2010) study found important results about the use of explicit versus implicit instruction[17] in perceptual training. While their study did not focus on tones or other types of suprasegmentals, it is a crucial study on the importance of explicit instruction in perceptual training. Later articles, however, do examine the use of explicit vs. implicit instruction in tonal learning, which is examined further below. Pederson and Guion-Anderson (2010) used an identification perceptual training procedure on Hindi vowels and consonants. One goal of the study was to determine if explicit instruction showed any important gains in discrimination and identification of the target phonemes. They included an identical pretest and posttest, which assessed the participants' discrimination skills. They had 42

[17] Explicit instruction refers to explicitly orienting learners to a specific feature they will be learning. Implicit instruction means the instructor or researcher does not orient the learners to the feature they are supposed to be focusing on. In other words, the learners are unaware of what feature they are supposed to be learning.

participants, all monolingual English speakers. They were split into two groups for three training sessions. One group focused on Hindi consonants while the other focused on Hindi vowels. They produced two sets of stimuli, one for the pretest and posttest and the other for the training. The set of stimuli for the pretest and posttest was produced by one Hindi speaker. Target contrasts were obtained by recording minimal pairs for 8 consonants and 8 vowels. The training stimuli, on the other hand, was recorded by four additional Hindi speakers. They produced monosyllabic words that always began with one of the eight consonants and one of the eight vowels. No words were repeated between the separate sets of stimuli. They found that explicit instruction for the consonant group showed improved scores from pretest to posttest. However, for the vowel group, the researchers determined that training seemed to have no significant effect on scores from pretest to posttest. However, they point out that this may be due to the high pretest scores (97% accuracy) the group received prior to training. On the other hand, despite this, both groups showed general improvement throughout their three training sessions.

Antoniou and Wong’s (2016) study also looked at implicit versus explicit learning using perceptual training. They constructed a seven-day training for identifying tones and Voice Onset Time (VOT), which is described in their study as “the timing relationship between the start of vocal fold vibration relative to the release of a stop consonant” (p. 272). With four different training groups, they tested how implicit exposure or explicit instruction affects the learning of a target feature. They had 80 L1 American English speaking participants. Only 40 of the participants explicitly trained to identify tones, and the tones would vary in this training (i.e., participants would be exposed to any of the three tones throughout). These 40 were then split into 2 groups of 20 for separate training. The only difference between these two groups is that VOT would vary or was fixed. The other 40 were told to focus on VOT, and VOT would always vary. These participants were also split into 2 groups of 20; tone was implicitly presented in these groups, and in one training group tone varied while in the other, tones were fixed. They included 2 pretests (one for tones and one for VOT) with identical posttests for all participants. Participants also had to complete a generalization test in addition to the two posttests on the final day of training. They had 12 Hindi speakers produce 288 pseudoword stimuli, and then superimposed Mandarin-like level, rising, and falling tones onto these stimuli. They divided these stimuli into three groups, by speakers. Four of the speakers produced the stimuli for the pretests and posttests; another four produced the training stimuli while the last four produced the generalization test stimuli. During training, participants were presented with two words at a time and then repeated twice. After four of these rounds, participants were quizzed. They had to identify which of two pictures corresponded to the word presented. After 24 of these rounds, participants were presented a sheet with 24 pictures. One of the pseudowords was presented and participants had to match the word with one of the 24 pictures. Feedback was not part of the training.

Antoniou and Wong (2016) found that participants who explicitly trained to identify tones attained enhanced generalization of tone identification. They also found that when the feature implicitly presented was fixed while the feature explicitly presented was varied, greater learning gains were made. In other words, fixing the feature implicitly presented, but varying the feature explicitly presented allowed participants to generalize their knowledge better. As shown in prior research, Antoniou and Wong (2016) found that the training improved all participants’ scores from pretest to posttest in tonal identification. However, they found that only participants who participated in the tone training (where VOT was fixed) made significant improvement in tonal identification scores from pretest to posttest. They determined that explicit training is important for learning a nonnative target feature, such as tone for stress learners. They also elucidated that implicit exposure to a target feature is not likely to enhance learning. This study, as well as Pederson and Guion-Anderson's (2010), reveal that explicit instruction is important when using perceptual training in the classroom.

2.2.3. Perceptual Training Types

Several different types of perceptual training have been examined and studied in the literature. Specifically, literature has examined the following types: identification training (Wayland & Li 2008), discrimination training (Wayland & Li 2008), high variability training (Perrachione et al. 2011; Wang et al. 1999), and low variability training (Perrachione et al. 2011). The research has additionally examined how varying temporal distributions between trainings and final testing affects learning (Li & DeKeyser 2019). Lastly, the literature has explored the use of visual cues and how different types of cues impact learning (Godfroid, Lin & Ryu's 2017), which are all examined below.

2.2.3.1. Identification vs. Discrimination vs. Production Perceptual Training

Wayland and Li's (2008) study looked at two different types of perceptual training: identification (ID) and what they call "same/different discrimination" (SD) training, but SD merely refers to this thesis' definition of discrimination. They employed these two training types to determine if either was better than the other. They had two distinct sets of participants take part in the training of Thai tones. The first comprised 30 L1 Mandarin speakers, and the second comprised 21 L1 American English speakers. Both the Mandarin and English groups had never had experience with Thai. Each group was separated in half; they then made two new groups. Each new group included Mandarin and English speakers, and for each group, the researchers tested a different type of perceptual training method (either ID or SD). The training lasted for two days, one hour of training for each day. The first group participated in ID training while the other participated in SD training. Their methodology included both a pretest and a posttest, which were identical in content. For these tests, they were presented with a set of three tokens 80 times on a computer. Trainees were told to determine whether the second stimuli in the set was the same as the first or the last token, making this a discrimination test; they would then click a button to select their choice. No feedback was given during the testing. They had five L1 Thai speakers produce five minimal pairs for Thai's mid and low tones three times for their stimuli.

Wayland and Li (2008) ultimately found that neither the ID or SD training types were better than another in improving participants' discrimination skills. In fact, they found that both perceptual training types yielded significant improvements by English speakers from pretest to posttest. They noted that their results showed lesser gains than previous studies, specifically citing Wang et al. (1999). They explain that this is likely due to the shorter training time they instituted. However, despite this, improvements were made in only two days and in one-hour training sessions.

Lu, Wayland and Kaan (2015) looked at perceptual training and perceptual+production training to see if they differed from one another in a learner's ability to perceive and produce Mandarin-like tones both at the intentional (explicit) and unintentional (implicit) levels. Their participants were comprised of 22 L1 English speakers. Their training and testing only lasted three days, with testing on the first and last day and a one-hour training session on the second day. Their stimuli included eight monosyllabic minimal triplets produced by two American English speakers; the three tones (level, rising, and falling) were then superimposed on the tokens in Praat. In the perceptual (only) training, participants were asked to listen to one stimulus and then another. Afterward, they needed to decide if the first stimulus had the same tone or a different tone, making this a discrimination task. They were then presented with graphic representations of the first and second stimuli for both tones. The perception+production training followed a similar procedure. However, they additionally had to produce the stimuli as well. Importantly, neither training group was given feedback about whether they accurately discriminated or produced the tones. They found that both training types improved participants' discrimination abilities which follows the analogous procedure and results found in Wang (2013). Additionally, in line with previous research, they found that explicit instruction aided learning while implicit instruction did not. They also reported that neither training provided more improvement for participants over another, showing that the incorporation of production to improve tonal discrimination skills does not additionally aid discrimination abilities. Further, they found that while participants paid more attention to pitch height before training, they began to pay more attention to pitch direction after training.

2.2.3.2. High Variability Training and Low Variability Training

While Wang et al.'s (1999) study examined the use of high variability (HV) training on learners, they did not examine how HV vs. low variability (LV) training[18] impacted different learners. Perrachione et al.'s (2011) study reviews this difference. They also tested how these two different trainings affect different learners, based on their aptitudes for pitch PDI. They instituted these two different kinds of perceptual trainings to test for identification skills. Perrachione et al. recorded 18 productions for 18 vocabulary words for their HV training stimuli. Participants completed the training over eight sessions. Their stimuli, much like Wong and

[18] LV training is the use of only one speaker of the stimuli while still using multiple tokens to highlight targeted contrasts, which in Perrachione et al.'s case, are Mandarin tones.

Perrachione (2008), were pseudowords recorded originally by eight American English speakers. In Praat, they then superimposed three tones: level (Tone 1), rising (Tone 2), and falling (Tone 4). Four speakers' tokens were used for stimuli in training while the other four speakers' tokens were only used for post-training assessment. They had 64 L1 American English speakers as their training participants, and they administered a pretraining assessment to these participants. They included a "Pitch-Contour Perception Test" (PCPT) in order to determine their aptitude for pitch PDI ability. From this test, they split their participants into two types: High Aptitude Learners versus Low Aptitude Learners.[19] Their goal of the study was to determine if High Aptitude Learners versus Low Aptitude Learners learn better under high or low variability training. As such, they split their High and Low Aptitude Learners evenly into each training group. The methodology above constituted their first experimental conditions.

Perrachione et al. (2011), in experiment 1, found that both High Aptitude Learners and Low Aptitude Learners made greater gains in generalizing their tonal identification knowledge from training to the testing stimuli if they underwent either training. However, the researchers noted that despite the finding given above, Low Aptitude Learners in the High variability training were significantly impaired in their learning of the vocabulary. Further, they did not show as large an improvement as the Low Aptitude Learners in the low variability training. In fact, Low Aptitude Learners in the Low Variability Training showed more comparable improvement to High Aptitude Learners in either training than to Low Aptitude Learners in High variability training.

[19] They described their High Aptitude Learners as learners whom were likely to accurately learn the vocabulary. They described their Low Aptitude Learners as learners whom were unlikely to accurately learn the vocabulary.

Alternatively, High Aptitude Learners in the LV Training and the HV training reached comparable accuracy by the end of the eight training sessions. Overall, though, the researchers noted that despite certain impairments based on aptitude, either training provided greater accuracy for both types of learners. These results instigated the researchers' second experiment.

Perrachione et al. (2011) conducted a second experiment as well. This experiment tested for specific types of HV training: Blocked High Variability (HV-B), Repeated High Variability (HV-R), and Blocked and Repeated High Variability (HV-BR). HV-B training constituted the same training conditions as HV training from their first experiment, except that the stimuli were organized by speaker and presented to participants with all of one speakers' stimuli before moving onto the next speakers'. In the HV-R training condition, they listened to one speaker's productions of the 18 vocabulary items used four times (i.e., the participants in this training heard 72 tokens from one speaker), and they listened to all four speakers' 72 tokens in one training session. This is different from the HV-B training because participants in HV-B training only listened to one token for each of the 18 vocabulary items used by each speaker. However, in the HV-R training, all 288 (72 X 4) tokens were not organized by speaker. In the HV-BR training, the researchers presented the 288 tokens to participants as organized by talkers. The researchers recruited new participants for their second experiment. They were able to find 61 new participants; these participants were also tested for aptitude. 30 were High Aptitude Learners while 31 were Low Aptitude Learners.

Perrachione et al.'s (2011) second experiment showed that HV-B and HV-BR training conditions resulted in significantly increased learning rate for High and Low Aptitude Learners. The -R condition (repeating the stimuli) resulted in significantly increased learning as well. However, they found that combining the -B and -R conditions did not increase learning rate when put together (HV-BR) than when either of these conditions was used separately (HV-B, HV-R). They also found that the -B condition was more successful than the -R condition for Low Aptitude Learners while the High Aptitude Learners benefited from all training conditions. In all, the researchers suggested that if any particular type of perceptual training should be administered for tonal identification when aptitudes are unknown, the HV-B training conditions are best used – as this is beneficial to all kinds of learners based on their aptitudes.

A later study also supported that high variability training would not benefit certain learners. Chang and Bowles (2015) tested the relationship between variability in the pitch changes of the tones themselves and in the phonetic contexts in which they are presented. Their participants included 160 American English native speakers whose responses in the study were analyzed. All 160 were naïve learners to Mandarin or any other tone language. Before these participants completed the tonal training, they participated in several tasks that tested for pitch PDI abilities, language learning aptitude, and general cognitive ability. In these pre-training tasks, the stimuli used for testing pitch PDI comprised monosyllabic words (as opposed to the disyllabic and monosyllabic contexts contrasted in the training). They used two different speakers for this stimuli. One recorded stimuli for a discrimination task while the other recorded stimuli for an identification task. For the training stimuli, the researchers recruited six native Mandarin speakers. Additionally, the stimuli constituted 24 pseudowords (four disyllabic quadruplets and 2 monosyllabic quadruplets). These pseudowords were paired with drawings that matched the meanings the researchers assigned. The participants completed 6 trainings over the course of two weeks in which they were training to identify and discriminate the stimuli. They completed a variety of tasks during this training which included selecting the tone they believed they heard after listening to a stimuli, listening to two stimuli spoken by different speakers to determine if the word was the same or not, and listening to stimuli and choosing which correlative drawing matched the stimuli's meaning.

Chang and Bowles (2015) found through an acoustic analysis that the pitch variability of the tones was greater across contexts (within speakers) than it was across speakers. They also found that disyllabic words were more difficult to learn than monosyllabic words. Additionally, they discovered that whether the contrastive tone for the tasks was on the penultimate syllable or the final syllable (for disyllabic words) made no significant difference to the participants' accuracy in TPDI, overall. They did find differences in accuracy for individual tones being contrasted on the penultimate or final syllables. However, they inferred that these results likely meant that the difficulty in identifying and discriminating the stimuli had little to do with the segmental part of the word; rather, they explain that the difficulty arises from the tones themselves. Ultimately, they concluded that contextual variation is not supportive for new learners of a tonal language, and in fact, it could be inhibitory to learning the tones. As opposed to previous studies' testing of more advanced learners, this provides evidence for reduced contextual variability in the onset of learning.

2.2.3.3. Temporal Distribution of Training and Retention

In Li and DeKeyser's (2019) study, they examined the effects of temporal distribution on training sessions. In other words, they examined how varying lengths between training sessions and posttests would affect learning. They had a total of 80 L1 English speaking monolinguals without prior experience with a tonal language. However, only 68 were included for hypothesis testing for various reasons.[20] The recorded stimuli for this study were comprised of real Mandarin words. Participants completed three trainings with one training session occurring only once in a day. All tests were completed on separate days either before or after training days. Training days were spaced out according to the experimental conditions, and the post-training retention test was completed one or four weeks later, depending on the training condition. There were four training conditions: (1) Condition A had a one-day Intersession Interval[21] (ISI) with a one-week Retention Interval[22] (RI); (2) Condition B had a one-day ISI with a four-week RI; (3) Condition C had a one-week ISI with a one-week RI; and (4) Condition D had a one-week ISI with a four-week RI. The training itself consisted of explicit instruction given on Mandarin tones. Then, they practiced their knowledge by listening to one of the pre-recorded stimuli. They were asked to identify its tone. In the next task of training, participants were asked to listen to one of the stimuli and identify the correct, corresponding picture. Feedback was provided at the time of both practice tasks. Participants also participated in production training. They were exposed to a stimulus (either pinyin or a picture) and asked to immediately produce it. Feedback was given, and they could play the model over again if desired.

Li and Dekeyser (2019) found that the two groups with one-day ISI (Condition A and B), significantly outperformed those who trained with seven days between trainings on pre-training

---

[20] One of these primary reasons for multiple exclusions included participants practicing between days when they were told not to because this would impact the results. Another participant was excluded from hypothesis testing because they claimed this participant was not an actively engaged learner as the other participants were.

[21] ISI refers to the time between training sessions. A one day ISI would mean only one day is given between two sessions.

[22] RI refers to when the retention test was administered after the final training. A one-week RI would mean that participants would take the retention test one week after the final training session.

session quizzes given on their second and third training days. They also found that in their stimuli-picture identification task, RI had an important effect. Specifically, they found that the four-week long RI groups performed at a lower accuracy rate on their retention test in comparison to those who had only one week before the retention test. They found that the experimental conditions did not affect the ability of participants' identification in matching stimuli to word. In other words, no condition outperformed another in this identification ability. They did find, however, that in tonal production accuracy, the differing ISIs had different effects on accuracy. In other words, training groups that had a one-day ISI outperformed those with a one-week ISI in tonal production accuracy. RI was shown to not effect tonal production accuracy.

2.2.3.4. The Effect of Visual Cues

Godfroid, Lin and Ryu's (2017) study aimed to determine the effectiveness of associating colors, symbols, and numbers as visual representations of tones in 5 trainings over 10 days. They had 303 L1 English speaking participants who were assigned to either one of 5 experimental training groups or a control group. However, it is worth noting that 60 of these participants had taken phonology or phonetics classes while 144 reported having played a musical instrument. They included a pretest, posttest (which was completed immediately after training), and a delayed posttest (which was completed one week after training), which involved matching the given tone to the same tone in the response options. The stimuli for the testing were recorded by two L1 Mandarin speakers. These speakers recorded at least 225 stimuli for all three tests. The stimuli for the training were recorded by two different L1 Mandarin speakers who recorded 200 tokens of words with Mandarin tones.

There were five different training groups for the five different training conditions in Godfroid, Lin and Ryu's (2017) study: (1) Number,[23] (2) Symbol,[24] (3) Color,[25] (4) NumberColor,[26] and (5) SymbolColor.[27] The training, itself, was conducted online through a web-based platform. In the training, an instructional video bringing tones to the level of awareness was the first task completed by participants. For all experimental conditions, the first part of the video was the same. However, in the end, the narrator of the video would explicitly associate the Mandarin tones to one of the visual representations or visual representation combinations (i.e., one of the experimental conditions). The procedure was presented as a game, where participants were required to listen to two stimuli, and then identify which tone it was by clicking the corresponding button.

Godfroid, Lin and Ryu (2017) found that all three single visual cue training groups (i.e., only Number, Symbol, and Color) significantly improved from pretest to immediate posttest scores as compared to the control group. Importantly, though, they found that training group 3 (Color) did not keep these improved scores in their delayed posttest while the first two groups (Number and Symbol) did benefit from training in their delayed posttest scores. Also, they found that the last two groups (NumberColor and SymbolColor) did not show any enhanced benefits as compared to the single visual cue training groups. In fact, they found that overall the Number and Symbol training groups had greater gains as compared to the other experimental training groups.

[23] Tones were only associated as a number, like Tone 1, Tone 2, etc.
[24] Tones were only associated as a symbol, like -, /, \, etc.
[25] Tones were only associated as a color, like hearing a high and level pitch and associating that as yellow
[26] Tones were associated with number and color.
[27] Tones were associated with symbol and color

Importantly, they also found that all participants' abilities to generalize their tone training after only a week greatly diminished. In all, though, they concluded that the conventional single cues (either Number or Symbol) provided greater benefits for training, but all training groups showed improvement from pretest to posttests regardless.

2.2.4. Summary

Perceptual training has been found to aid in perception, discrimination, and identification with little dispute in the literature. In fact, limitations that have been pointed out are typically due to shortcomings of specific types, such as the use of color as a visual cue in Godfroid, Lin and Ryu (2017). One important limitation is the use of implicit instruction to learners when engaged in perceptual training. The training must be explicit to show enhanced accuracy post-training (Antoniou & Wong's 2016; Lu, Wayland & Kaan 2015; Pederson & Guion-Anderson's 2010). The literature finds that perceptual training, when used explicitly, generally increases TPDI. Furthermore, perceptual training has also been shown to aid in production (Li & DeKeyser 2019; Wang, Jongman & Sereno's 2003). Lastly, different types of perceptual training have been found to be more effective for specific learners. In Perrachione et al.'s study, HV training was found to impede Lower Aptitude Learners, but when certain conditions were employed onto HV training, namely blocking (-B), both High and Low Aptitude Learners were found to make important gains in tonal learning. This has important implications for the use of perceptual training in the classroom.

In all, research has found that perceptual training seems to enhance pitch PDI and production. More research needs to be done, though, on the use of perceptual training in a classroom setting. Most of the studies in this literature review are conducted in a laboratory setting. While this is useful for showing the benefits and limitations of the training methodology, more research can be done on how implementation in a classroom can benefit learners. Furthermore, TPDI and tonal production have been found to be enhanced when perceptual training and a musical training background are combined. Together, this leaves a gap in the literature of how to bring the enhanced learning capabilities that musical training brings into use for learners who do not have a musical training background. Incorporating musical training with perceptual training in the classroom seems like an effective way to enhance learners' TPDI and tonal production.

2.3 The Present Study

The present study aims to fill a gap in the literature. As has been shown thus far in Chapter 2, musical training has rarely been employed in the classroom in order to teach tones. Further, while many studies have reviewed the effectiveness of perceptual training, and even the use of perceptual training combined with a musical training background, a gap exists of whether musical training[28] applied in the classroom would be useful for L2 TPDI and if musical training or perceptual training are more or less beneficial as compared to each other. Importantly, while Shi (2018) focused on musical training's effect on production, her dissertation did not assess gains in TPDI. For this reason, the current thesis reviews the use of musical training on L2 TPDI. Also, the scope of this thesis prevented the inclusion of production, which is the second reason why production is not tested in this study. Through the use of the training methodology described in Chapter 3, this thesis has four research questions.

---

[28] This term is used to mean the use of instruments in the classroom (in the case of this study, a digital piano) and the training of pitch as a musical element, called "notes," to apply to the understanding of linguistic pitch use.

2.3.1. Research Questions and Hypotheses

This thesis' research questions are as follows:

1. In tonal learning, is the use of musical training more or less beneficial than perceptual training, and is there a difference between learners with and without musical training backgrounds?
2. How do the combined musical and perceptual trainings affect the ability to generalize novel words, tones (i.e., M tone), and tone melodies[29] (i.e., HH, MH) for musicians versus nonmusicians?

The first question aims to determine if the use of musical training as applied to tonal learning is more beneficial than perceptual training. Or is perceptual training more effective, still, than this application of musical training? For this question, the hypothesis for this thesis was that musical training would be more beneficial than perceptual training. In the literature, perceptual training has indeed been shown to increase PDI for nonmusicians, but musicians were still able to outperform nonmusicians when all participants were given perceptual training. This seems to indicate that the use of musical training should more beneficially improve participants' L2 TPDI.

The second question seeks to determine whether the combined trainings (i.e., musical and perceptual training) will allow trainees to generalize their knowledge after training to new words, tones (i.e., M tone), and tone melodies (i.e., HH, MH). In other words, will unknown words with unknown tones be perceived, discriminated, and identified by trainees, despite having no

[29] In this thesis, "tone melody" or "tonal melody" refers to a sequence of tones across the span of a word as opposed to the more complex definition referring to lexically contrastive tonal melodies mapped across different words as seen in, for example, Mende (Leben 1971).

exposure to those items previously? The hypothesis for this question is that the combined trainings will result in the trainees' abilities to generalize their knowledge to novel stimuli.

Chapter 3: Methodology

3.1. Participants

16 trainees were recruited to participate in this study, using a flyer with relevant contact information and pertinent information about the study (see Appendix A). Participants comprised L1 stress language speaker trainees, with both musicians and nonmusicians represented (following Wayland, Herrera, & Kaan 2010).[30] This study mostly followed Wong & Perrachione's (2007) definition of "musician," (i.e., a person who privately trained with their instrument for six years; additionally, they must have started lessons before the age of 10) but an alteration to the definition was made. Rather than require a specific age for the learners to have begun playing their instrument, no age requirement was specified. This restriction was removed in accordance with Wayland, Herrera, and Kaan (2010), who found that a musical training background without this requirement still aided participants labeled "musicians." In other words, Wayland, Herrera, and Kaan found similar results (i.e., that musicians outperformed nonmusicians in TPDI accuracy) with their definition of "musician" as previous studies with stricter age requirements found with their definitions. For this reason, an age requirement for the definition of "musician" for this study was not included.

[30] One exception was made to this definition. Despite noting only 4 years of private music training, Participant 14 also noted years of additional training through participating in bands and other musical exploits. He specifically stated that these years also afforded a kind of "training." These additional years were seen as enough to qualify him as having at least 6 years of a musical training background.

3.2 Materials

A description of stimuli, the pretests and posttests, the generalization test, and the survey are presented below.

3.2.1. Stimuli

A group of eight native speakers of the tonal language, Yoruba, were recruited to record productions of stimuli for the training and testing of the trainees in order to increase the variability of the testing stimuli, particularly for the perceptual training (e.g., Perrachione et al. 2011; Wang et al. 1999). Yoruba speakers were recruited to record each stimulus item three times, following Wayland and Li (2008). 15 minimal monosyllabic pairs and eight minimal or near minimal pairs of disyllabic words were recorded for the training tasks and pretests/posttests. One Yoruba speaker produced another 28 additional words, monosyllabic and disyllabic, for the generalization test. A summary of how each word was used in each task or test can be found in Table 3.1, below. Importantly, all stimuli used for a task or test were used uniformly across groups and participants.

3.2.2. Pretests and Posttests

The pretests and posttests were used to test for progress in tonal perception, discrimination, and identification (TPDI) skills before the training began and after each training ended (following Antoniou and Wong 2016; Wang et al. 1999; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007). This resulted in two pretests and two posttests. A set of audio stimuli (i.e. the stimuli outlined in Table 3.1) was played for participants. On a worksheet, they marked each word's tone(s) by indicating either H or L for each vowel (See Appendix B). The worksheets for the pre- and posttests included identical words with the only differences being the order in which the words were presented and which native speaker's rendition was used in order to minimize repeated testing bias.[31] [32] Additionally, while the generalization test contained entirely novel words, some words used in training were also used in the pre- and posttests. Importantly, feedback specifying the correct tones for words shared in the training tasks and the pretests/posttests was never provided in order for participants to not rely on their knowledge of these shared words that would have been explicitly identified through feedback.

3.2.3. Generalization Test

The generalization test consisted of randomly presented audio stimuli not previously included in the training nor in the pretests or posttests. Specifically, this generalization test required learners to generalize their knowledge to new words, a new speaker, a new tone (M tone), and two new tonal melodies (MH and HH disyllabic words) (See Appendix C). Audio stimuli for this test was separately recorded and only used for the purposes of this test (following Antoniou & Wong 2016; Wang et al. 1999; Wang 2013). The generalization test was given after the final posttest, and tested whether participants' TPDI skills could be applied to novel stimuli. A summary of tonal categories and words used for the stimuli is presented below in Table 3.1.

---

[31] Selection of the test stimuli was done through a random number generator. First, selection of the position of the 16 words was chosen through this method. Then, for the pretests and posttests, the researcher assigned each speaker/token pair a number. Starting with Pretest 1's first word, the generator picked the speaker/token that would be used for that word. This continued until the 16th word on Pretest 1. Then, this same method was employed for each of the following pretest and posttests.

[32] Different renditions of native speaker's stimuli was also used to more closely simulate, and therefore test, how learners would perceive, discriminate, and identify tones in a more naturalistic context.

Table 3. 1: Summary of Stimuli Used for Testing and Training[33]

Word Type	Musical Training Task 1	Musical Training Task 2	Perceptual Training Task 1	Perceptual Training Task 2	Pretests/ Posttests	Generalization Test	Power Point
Monosyllabic H	bí, kí, rá,	fá, **yá,** ṣú	yọ́, lọ́, ṣú	bú, **kọ́, lú,**	bá, fọ́, **kọ́, lú,** ró	bọ́, lé	**yá**
Monosyllabic L	bì, kì, rà,	fà, **yà,** ṣù	yọ̀, lọ̀, ṣù,	bù, **kọ̀, lù,**	bà, fọ̀, **kọ̀, lù,** rò	bọ̀, lè	**yà**
Monosyllabic M	X	X	X	X	X	bi, bọ, ki, kọ, le, lọ, lu, ra, ro, ṣu, ya, yọ	X
Disyllabic HH	X	X	X	X	X	rárá, kóró, kéré, lábẹ́, kúrú	X
Disyllabic HL	yálà, jálẹ̀	kúrò, **rárà**	**rárà,** kúkù	**kókò, búlù**	**búlù,** gúsù, **kókò**	dọ́là	X
Disyllabic LH	bàbá, bàjẹ́	bùbá, **ràrá**	**ràrá**, yàrá	**kòkó, jùjú**	kẹ̀kẹ́, **kòkó, jùjú**	gègé, fùfú	X
Disyllabic MH	X	X	X	X	X	ṣeré, kọjá, ṣubú, burú	X

3.2.4. Survey

A survey was given to participants in order to identify each participant's musical training background, tonal language learning background, and training preference (i.e. perceptual training versus musical training) (See Appendix D).

3.3. Procedure

The training consisted of two 1-hour sessions completed on the same day with a 10-15 minute break given between each. One of the training sessions was a musical training while the other was a perceptual training. Both trainings were organized into a Presentation Practice Production (PPP) format due to the PPP format's widespread use in language classrooms and its positive effects found in the literature (Criado 2013; Swan 2005). Further, interactive elements

[33] Bolded words in Table 3.1 reveal words that are used in more than one task/test/training.

were incorporated into both trainings in line with Communicative Language Teaching (CLT) principles and characteristics as a method of avoiding bias. At the beginning of the first training session (whether it was the musical or perceptual training), a PowerPoint presentation (See Appendix E) was given to bring tones to the level of awareness, which in the literature is a term used to refer to participants' conscious and explicit perception of tones, based on the findings that explicit instruction better benefits learners (Antoniou & Wong 2016; Lu, Wayland & Kaan 2015; Pederson & Guion-Anderson 2010). Participants were separated into two groups: a Musical+Perceptual (MP) training group and a Perceptual+Musical (PM) training group; in these two groups, there were both musicians and nonmusicians. In the MP training group, participants took the 1-hour musical training first. Then, they took the 1-hour perceptual training after a short 10-15 minute break.[34] The PM training group's order of trainings was reversed. By giving half the participants the MP training order and the other half the PM training order, the training order delivery was counterbalanced (following Inceoglu 2015 and Wayland & Li 2008). Given the demonstrated efficacy of perceptual training in the literature (Antoniou & Wong 2016; Godfroid, Lin & Ryu 2017; Li & DeKeyser 2019; Lu, Wayland & Kaan 2015; Perrachione et al. 2011; Wang 2013; Wang et al. 1999; Wang, Jongman & Sereno 2003; Wayland & Li 2008), perceptual training was used as the baseline method with which to compare the new musical training and a separate control group with no training was not included. In the descriptions of the musical training and perceptual training procedures below, it should be noted that when participants

[34] Importantly, while each training was only scheduled to be one hour each, technical difficulties across groups that arose resulted in the trainings lasting longer than an hour at times in order to address the technical issues.

within their MP or PM group were paired off or put into groups of three, these smaller groupings always included a mix of musicians and nonmusicians to control for musicianship background in the relevant training tasks.

After the training and all testing, the survey on musical and language background and training preference was administered. Participants returned the survey no later than 24 hours after completing the generalization test.[35] Additionally, lesson plans were designed for both trainings and can be found Appendix F. A summary of the research design is found below, in Table 3.2.

Table 3. 2: Summary of Research Design

MP Group	PM Group
-Pretest 1 -Musical training -Posttest 1	-Pretest 1 -Perceptual Training -Posttest 1
10-15 Minute Break	
-Pretest 2 -Perceptual Training -Posttest 2 -Generalization Test -Survey	-Pretest 2 -Musical training -Posttest 2 -Generalization Test -Survey

The procedure is thus summarized: a short five minute pretest (i.e., Pretest 1) before the first 1-hour training; a five minute posttest (Posttest 1) following the first training; a 10-15 minutes break before the second five minute pretest (Pretest 2), which precedes the second 1-

[35] One participant, however, is an exception. Participant 13 returned his survey within a week due to personal reasons.

hour training; a second five minute posttest (Posttest 2) followed immediately by the generalization test and then the participant survey.

3.3.1. Musical training Session Procedure

In the musical training, participants first watched the PowerPoint presentation if they were part of the MP group, bringing Yoruba's tones to the level of awareness. Then, different "notes" on a digital piano were related to different pitches, and these notes were related to linguistic tone: H and L. In other words, using one, specific pitch on the piano (which was explained as representing a specific tone in the target language) does not equate to all tonal language speakers using that specific pitch every time for the same tone. Rather, the relativeness of one tone to another in one utterance reveals the tonal category. Importantly, in order to encourage learners to construct tonal categories regardless of individual speakers, it was pointed out that tone is relative to speaker. By showing the pitches on the piano, it was explained that the importance of pitch to linguistic tone is in the differences between one tone's height in comparison to another's. Additionally, it was explained how combining two notes can make a pitch rise or fall and how in Yoruba, this rise or fall can happen over a single vowel (i.e., a contour tone).

Participants were then given a set of words, and some were marked for tone with tonal symbols in the form of an information gap activity in line with CLT principles and characteristics:[36] participant pairs or triplets were given two or three versions of a worksheet. In

[36] The principles are grounded in communication with some freedom of language learning exploration (Richards & Rodgers 2001). Particularly, student-student interaction and goal-oriented communication are key characteristics of this approach, making the information gap activity in this training more aligned with CLT principles that can be adapted for L2 tonal language teachers.

one version, a word was given with the tonal symbol. In the other versions, the corresponding word with its tonal symbol was not shown on the worksheet. For example, Participant A and B had a tonally unmarked word. They would ask what tone this word had to Participant C who had this particular word marked for tone on their worksheet. Participant C answered by playing a corresponding note on the piano or by using their voice if they could not download an app for a digital piano. Then, Participants A and B marked the tone down on their worksheet. In this example, Participant C would also have unmarked tones and would inquire about the tone's identification to whichever participant had the tone marked on their worksheet. They would fill in their worksheets until both (or three if in a triplet) completed their own versions. Next, participants were given a new sheet with different words that were unmarked for tone, and they worked with the same partner(s). They listened to the audio stimuli created for training that corresponded with the order of the unmarked words on the worksheet. Together, they needed to determine what each vowel's tone was. When they decided, they marked it by indicating the H or L letters on their worksheets. All worksheets were submitted electronically by the end of the training day. A summary of this procedure is found below in Table 3.3.

Table 3. 3: Summary of Musical training Procedure

Time	Lesson Activity	Activity Description
10 minutes	PowerPoint (if musical training was first)	Brought tones to the level of awareness and explicit instruction of tones (Omitted if participants were part of a PM group)
15 minutes	Piano Presentation	Different notes on a digital piano were related to different pitches (level and contour).
15 minutes	Information Gap	Participant pairs were given two/three versions of a worksheet. In one/two version(s), a word is given without the tonal symbol. In the other version, the corresponding word with its tonal symbol were given. Pairs/triplets filled in each other's unmarked words by using the piano or their voices.
15 minutes	Partner Listening	Partners listened to audio stimuli to determine what each unmarked word's tone was on a worksheet.

3.3.2. Perceptual Training Session Procedure

In the perceptual training, participants were given the introductory PowerPoint presentation if they were part of the PM training groups. The perceptual training was then started. Much like Wayland and Li (2008), stimuli for this training were presented in pairs or sets of targeted contrasts. Participants marked their identification of the tone on a worksheet. Immediate feedback was given about which tone was spoken in the audio stimuli. This continued for 10 minutes. After this task ended, participants continued the training with new words for another 10 minutes. However, during this second 10 minute portion, immediate feedback was not given because this test included shared words with the testing stimuli. Participants were then paired together and told to review their answers to each of the words for the second training period. They were allowed to listen to any of the stimuli again, but feedback was not provided. All materials were turned into the researcher once the training session ended. A summary of this procedure is found below in Table 3.4.

Table 3. 4: Summary of Perceptual Training Procedure

Time	Lesson Activity	Activity Description
10 minutes	PowerPoint (if perceptual training was first)	Brought tones to the level of awareness and explicit instruction of tones (Omitted if participants was part of an MP training group)
15 minutes	1st 10 minute Training	Participants marked their identification of the tone on a worksheet. Immediate feedback was given by the researcher about which tone was spoken in the audio stimuli. Five additional minutes at the beginning of this activity was used for directions given to the participants.
15 minutes	2nd 10 minute Training	Participants marked their identification of the tone on a worksheet. Immediate feedback was **not** given by the researcher. Five additional minutes at the beginning of this activity was used for directions given to the participants.
15 minutes	Paired Review	Participants were then paired together and told to review their answers to each of the words for the second post-break training period.

3.4 Method of Analysis

In answering the research questions outlined at the end of Chapter 2, the data was analyzed in various ways. Participant test answers were assessed through each word's TPDI accuracy and through each vowel's TPDI accuracy, which are defined below. Additionally, participant accuracy scores were compared, first, via descriptive statistic methods, and second through inferential statistics. A review of each of these methods of analysis is discussed at length in Chapter 4. Important to note, meaning was not assessed in this study. While meaning is crucially important, the focus of this study was on the PDI accuracy of the tones themselves, which actually aids in meaning because a difference in tones can indicate a difference in word meaning.

Regarding the review of test answers, a word's TPDI accuracy score was calculated by assessing each test for TPDI accuracy over the entire word. A word's TPDI accuracy was considered correctly identified by the participant if the tones were chosen accurately on each vowel in the word. Incorrect identifications (between H and L for the pretests and posttests, and between H, M, and L for the generalization test) and the absence of a choice for even one tone on one vowel in a word were assessed as inaccurate. In the case of monosyllabic words, if the word was H toned, and the option for H was not selected on the test, this was assessed as an incorrect answer. In the case of disyllabic words, if just one tone on one vowel in the word was incorrectly identified by the participants, then the whole word was marked as inaccurately identified. For example, if a HL word was marked as HH by the participant, this was counted as an incorrect response.

Further regarding the review of test answers, a vowel's TPDI accuracy score was calculated by looking at each vowel's tone individually, and independent of word shape (i.e., monosyllabic or disyllabic). In this way, each vowel's tone was assessed for accuracy. If the vowel's tone was identified correctly, it was assessed as accurate. If the vowel's tone was marked incorrectly or if the vowel was left unmarked, it was assessed as incorrectly identified.

The TPDI accuracy of each word and vowel over each test was reviewed for inaccurately identified tones by each participant. It was found that many of the same participants consistently marked the same words/vowels incorrectly in Posttest 1, Pretest 2, and/or Posttest 2. For some words/vowels even though a participant incorrectly identified a tone in Posttest 1, Pretest 2, and Posttest 2, they oddly marked the word/vowel accurately in Pretest 1. This indicated a pattern in the results. Participants had a 50% chance of accurately guessing. If they scored correctly on an earlier test and then incorrectly on a later test, their initial, correct response could be considered a lucky guess as opposed to an accurate reflection of PDI. As a result, a revised scoring procedure was instituted, the details of which follow.

The following condition was utilized when revising the participants' pretest and posttest scores: the word/vowel had to be marked incorrectly in at least two following tests in order to re-mark a participant's word/vowel as inaccurate in an earlier test.[37] Notably, only Pretest 1 and Posttest 1 were able to be revised due to this condition because only these two tests had at least two following tests. The condition requires two following incorrect identifications for a principled reason. One following incorrect identification may have been a mistake by the participant. However, two or three following incorrect identifications were assumed to indicate a pattern that revealed the participant's lack of knowledge. The reason that the identifications needed to follow (as opposed to precede) the test being revised was because participants could have made gains due to trainings. It would be inappropriate to assume that a word or vowel incorrectly identified in two previous tests meant the participant had not learned the word/vowel by the third or fourth test.[38] For this reason, two following incorrect identifications of the word/vowel for the condition were required. The data presented in Chapter 4 will use the revised scores.[39]

3.5 Methodological Changes due to COVID 19

The onset of COVID-19 prompted rapid, necessary changes to the original training methodology. Prior to this global health crisis, this study was planned to take place face-to-face

---

[37] In other words, take for example, the word bá. For this example, Participant X incorrectly identified bá in Pretest 2, and Posttest 2 while Participant Y incorrectly identified this word in Pretest 1 and Pretest 2. In revising Posttest 1 scores, the researcher would review these participants' incorrect identifications in the tests following Posttest 1. In this example, both incorrectly identified the word in two tests, but only Participant X incorrectly identified the word in two *following* tests to Posttest 1. For this reason, only Participant X meets the revision protocol condition.

[38] In fact, of course, it is the expectation that training would improve PDI accuracy over time.

[39] While the revised scores are the ones presented in Chapter 4, Appendix G contains non-revised results for comparison.

(f2f) and over the course of two days. However, due to New York State's social distancing mandate, this study was adapted to a synchronous computer-mediated communication (SCMC) training. A discussion of how this change in modality may have impacted results is considered in Section 5.1, but the current section, Section 3.5, will describe how the change in modality prompted technical challenges to the study.

First, regarding participants, of the 16 participants, Participant 2 and 10 were excluded from analysis because technical issues resulted in unsaved pretests/posttests, which made the independent contributions to their progress indeterminable. Concerning the stimuli, while eight Yoruba speakers participated in this study, only five speakers' productions were used for various reasons, including too much audio interference, unsigned consent forms, and late submissions. Additionally, despite asking all the speakers to record the stimuli three times, only one complied with this instruction and the remainder recorded each item only once. Pertaining to the procedure, the study's online delivery resulted in participants being able to independently listen to the stimuli in the audio tracks for the tests, which could not be entirely controlled. In order to offset prior review, the audio stimuli for the tests was sent to the participants just before they would take the test. With the first training group (Participants 1-5), the participants were not instructed to limit their listening of the track to one playthrough. In the second, third, and fourth training groups (Participants 6-16), however, participants were instructed to listen to the audio track for each test only once, unless they experienced internet or other technical issues which prevented them from hearing the word clearly. Despite this, it is worth noting that Participants 1, 3, and 5 from the first training group followed these guidelines despite not being instructed. Participants 2 and 4, however, listened through each audio track twice. In the first training group, as well, it is important to note that the explanation of contour tones was missed. Further regarding the procedure, it is worth noting that despite attempts to adhere strictly to the one-hour timing of the training schedule, various technical issues often arose during each group's trainings. Therefore, while the times outlined in Table 3.3 and 3.4 roughly estimate the amount of time spent proportionally on each activity, some activities were potentially halted by technical challenges. One last effect on the study's procedure due to the technical issues arising from a SCMC modality was the inclusion of immediate feedback in the perceptual training procedure. In the first training group (Participants 1-5), feedback was not provided in the perceptual training's first training task. Due to technical errors, immediate feedback could not be provided after each set of targeted contrasting words. Instead, participants listened to the audio stimuli on their own. Then, feedback was given by providing all the correct answers after all the participants had finished their worksheet.

Importantly, when this study was originally planned to take place over two days, a posttest after the training on the first day and a pretest before the training on the second day was included. In rapidly adapting to an online modality, the study was restructured to take place on one day with only a short 15 minute break in between both trainings. Posttest 1 and Pretest 2 were not *both* needed with such a short break between, but due to the circumstances, both tests remained part of the study's procedure. Additionally, as explained above, Pretest 2 was used as one of the tests for the revision protocol condition described in Section 3.4. However, Pretest 2 and Posttest 2 could not be revised themselves because an insufficient number of tests followed them in following the condition of the revision protocol. Since there was only a 10-15 minute break between training sessions 1 and 2, and because Posttest 1 could be revised while Pretest 2 could not, Pretest 2 is not represented in the results presented below in Chapter 4. Therefore, importantly, Pretest 2’s only contribution to this study is its use in revision through the protocol condition.

Chapter 4: Results

This chapter reports on the revised[40] results of the present study outlined in Chapter 3. In Chapter 2, two research questions and a hypothesis for each were proposed based on a review of the published literature. Each research question is repeated below for convenience:

1) In tonal learning, is the use of musical training more or less beneficial than perceptual training, and is there a difference between learners with and without musical training backgrounds?
2) How do the combined musical and perceptual trainings affect the ability to generalize the training to novel words, tones (i.e., M tone), and tone melodies (i.e., HH, MH) for musicians versus nonmusicians?

In the subsections below, the data were examined through descriptive and inferential statistics. Section 4.1 presents the individual scores for each training and test. Section 4.2 shows the results relevant to answering research question 1 while Section 4.3 shows the results relevant to answering research question 2.

4.1 Results of Individuals

This section presents individuals' results for each training and each test. Important to remember in the following sections is the counterbalanced training methodology this study employed. Half the participants took the musical training first while the other half took the perceptual training first. Importantly, then, when reference to musical/perceptual pretests/posttests are made, this refers to the relevant pretest or posttest (i.e., Pretest 1, Posttest

[40] Here, "revised" refers to the revision protocol discussed in 3.4.

1[41] and 2) the participant took for each training.[42] Section 4.1.1 examines individual results related to the musical training. Section 4.1.2 examines individual results related to perceptual training. Section 4.1.3 explores the percent change scores within individuals, and Section 4.1.4 examines the individuals' generalization test scores in comparison to their training scores. In each of these sections, both the word's and the vowel's tonal perception, descrimination, and identification (TPDI) accuracy are reviewed separately.

4.1.1. Musical Training Individual Results

The below subsections, 4.1.1.1 and 4.1.1.2 show the results of individual performance in the musical training.

4.1.1.1 The Word's TPDI Accuracy

Figure 4.1 provides an overview of each participant's musical training-relevant pretest and posttest scores in the word's TPDI, with Figure 4.2 showing individual percentage gains between tests. As can be seen in both graphs below, some participants made gains, though three lost accuracy, and of the four who made no gains, two were already at ceiling. Of those who made gains, four were nonmusicians while three were musicians.

---

[41] This posttest was used as a pretest for the second training each participant took due to the revision protocol described in Chapter 3

[42] In other words, for example, Participant 1 took musical training first. So Pretest 1 and Posttest 1 would be the relevant pretest and posttest for his musical training. However, Participant 6 took perceptual training first, so Posttest 1 and Posttest 2 would be the relevant pretest and posttest for her musical training.

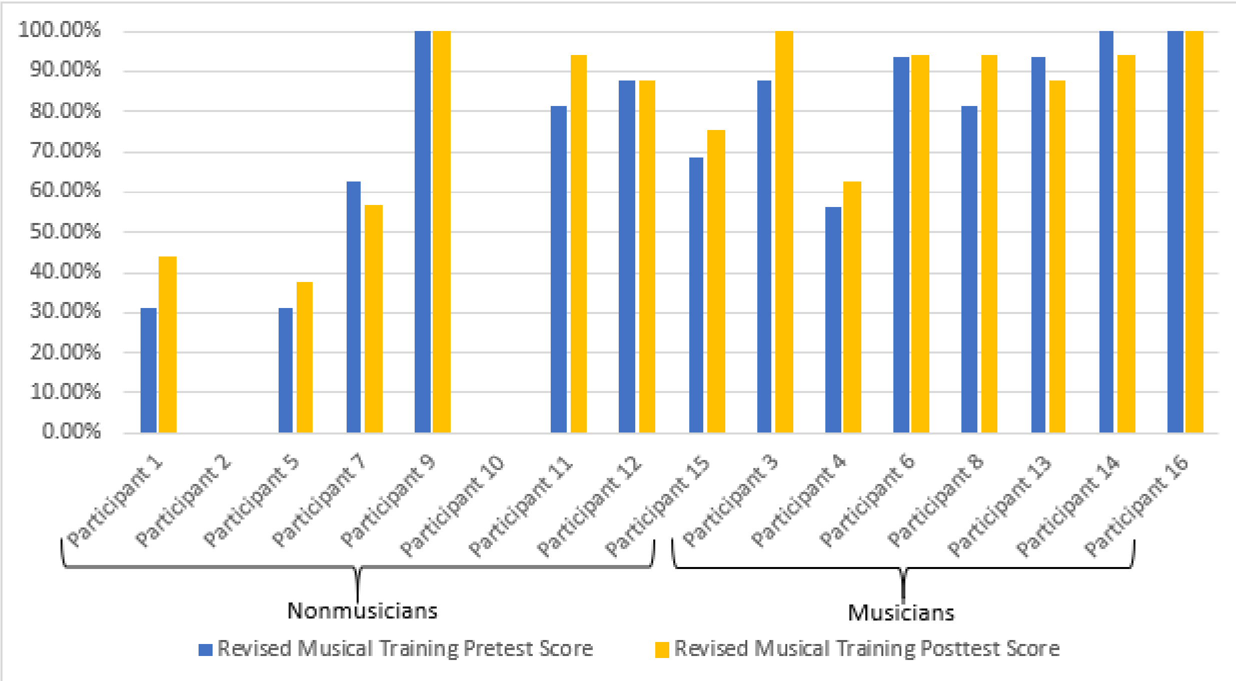


Figure 4.1: Musical Training Pretest and Posttest Scores of the Word's TPDI by Participant[43]

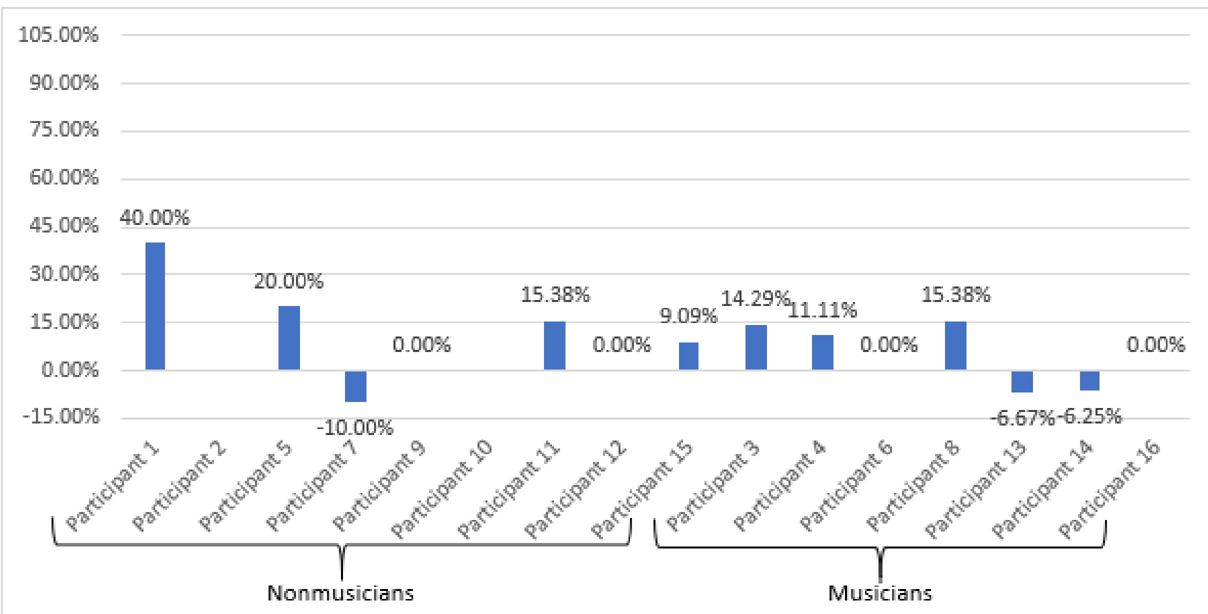


Figure 4.2: Musical Training Percent Change for the Word's TPDI Accuracy between Pretest and Posttest by Participant

---

[43] Participants 2 and 10 were excluded from analysis because both submitted one or more empty pretests/posttests.

4.1.1.2 The Vowel's TPDI Accuracy

Figure 4.3 provides an overview of each participant's musical training-relevant pretest and posttest scores, with Figure 4.4 showing percentage gains between tests each participant made from pretest to posttest in the vowel's TPDI accuracy for the musical training. Similar to the analysis of words, some participants made gains, though three lost accuracy, and two with no gains were already at ceiling. Even while three participants accuracy decreased for both the word's and the vowel's TPDI accuracy, it is interesting to note that these participants' were not exactly the same. For the word's TPDI accuracy, Participants 7, 13, and 14 showed decreased accuracy while for the vowel's TPDI accuracy, it was Participants 13, 14, and 15. Of the five who made gains, three were nonmusicians while two were musicians.

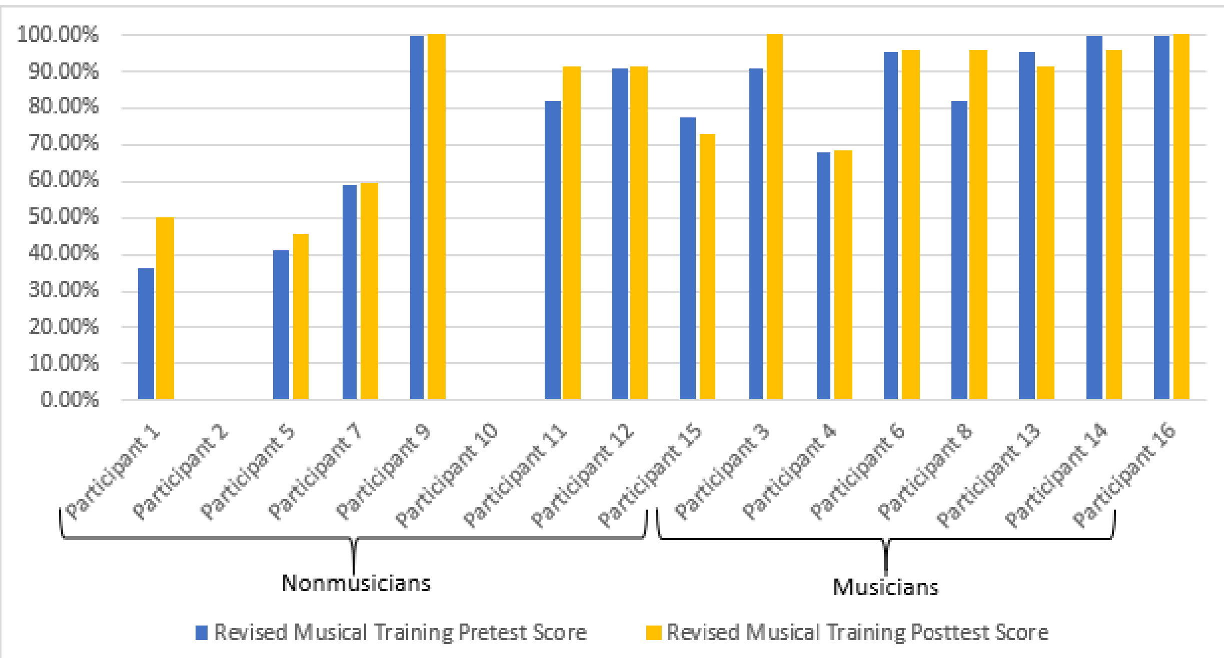


Figure 4.3: Musical Training Pretest and Posttest for the Vowel's TPDI Scores by Participant[44]

[44] Participants 2 and 10 were excluded from analysis because both submitted one or more empty pretests/posttests. From this point on, all graphs examining individuals will exclude Participants 2 and 10.

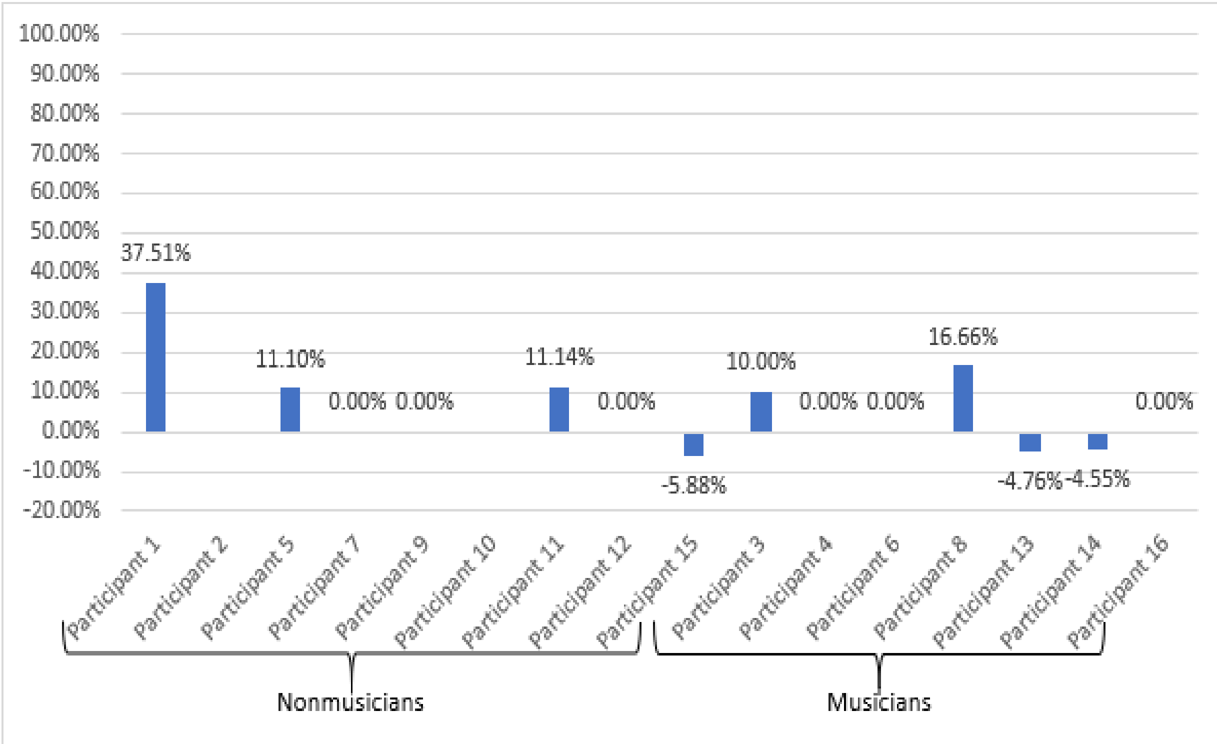


Figure 4.4: Musical Training Percent Change between Pretest and Posttest for the Vowel's TPDI Scores by Participant

4.1.2 Perceptual Training Individual Results

The below subsections, 4.1.2.1 and 4.1.2.2 show the results of individual performance in the perceptual training.

4.1.2.1 The Word's TPDI Accuracy

Figure 4.5 provides an overview of each participant's perceptual training-relevant pretest and posttest scores, with Figure 4.6 showing percentage gains between tests. As can be seen in both of these graphs, most participants made gains after perceptual training. Though four participants made no gains at all, two of those were already at ceiling at the start of training. Of those 10 who made gains, 6 were nonmusicians while 4 were musicians.

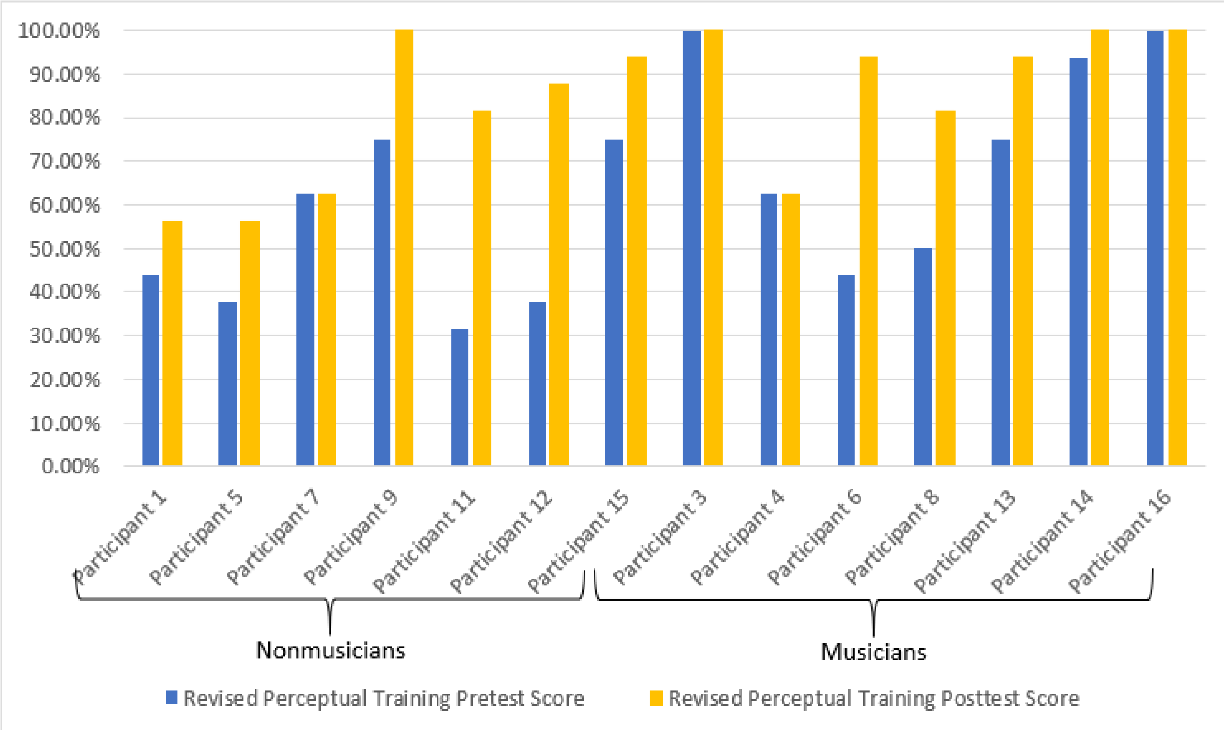


Figure 4.5: Perceptual Training Pretest and Posttest Scores of the Word's TPDI by Participant

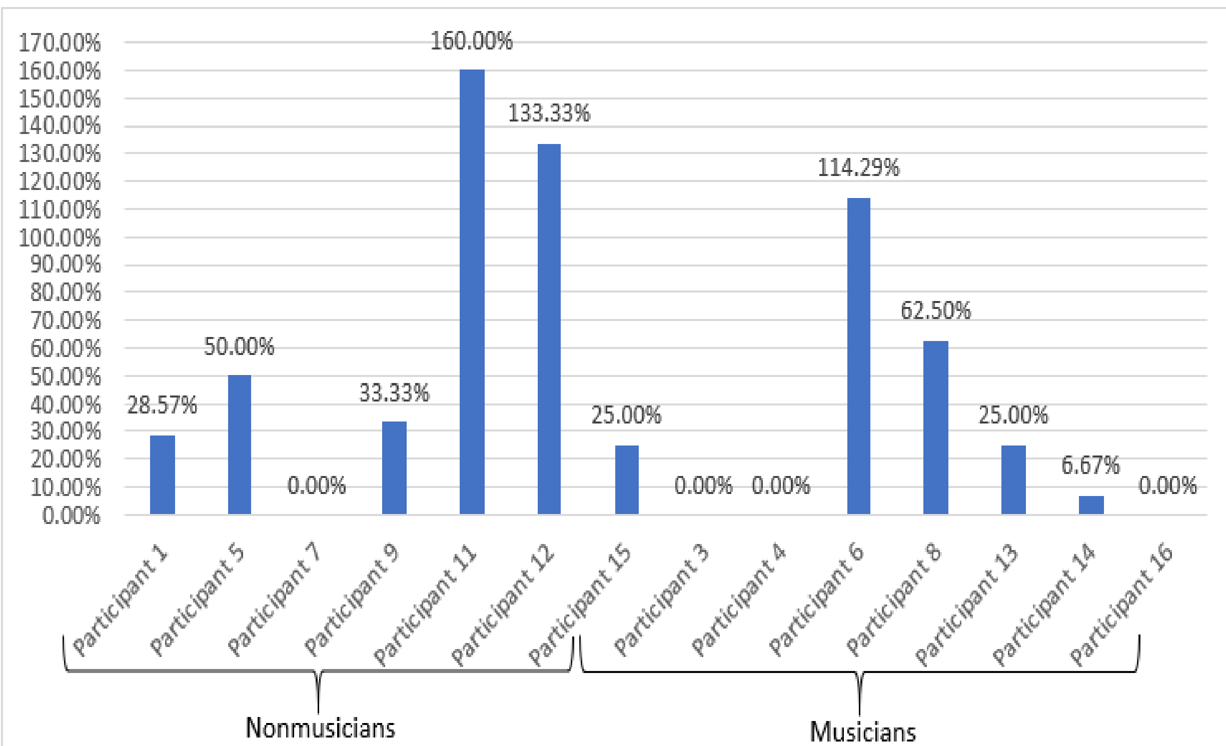


Figure 4.6: Perceptual Training Percent Change of the Word's TPDI between Pretest and Posttest by Participant

4.1.2.2 The Vowel's TPDI Accuracy

Figure 4.7 provides an overview of each participant's perceptual training-relevant pretest and posttest scores, with Figure 4.8 showing percentage gains between tests. Similar to the analysis of the word's TPDI accuracy, most participants made gains, though one lost accuracy, and two with no gains were already at ceiling. Of those nine who made gains, six were nonmusicians while four were musicians.

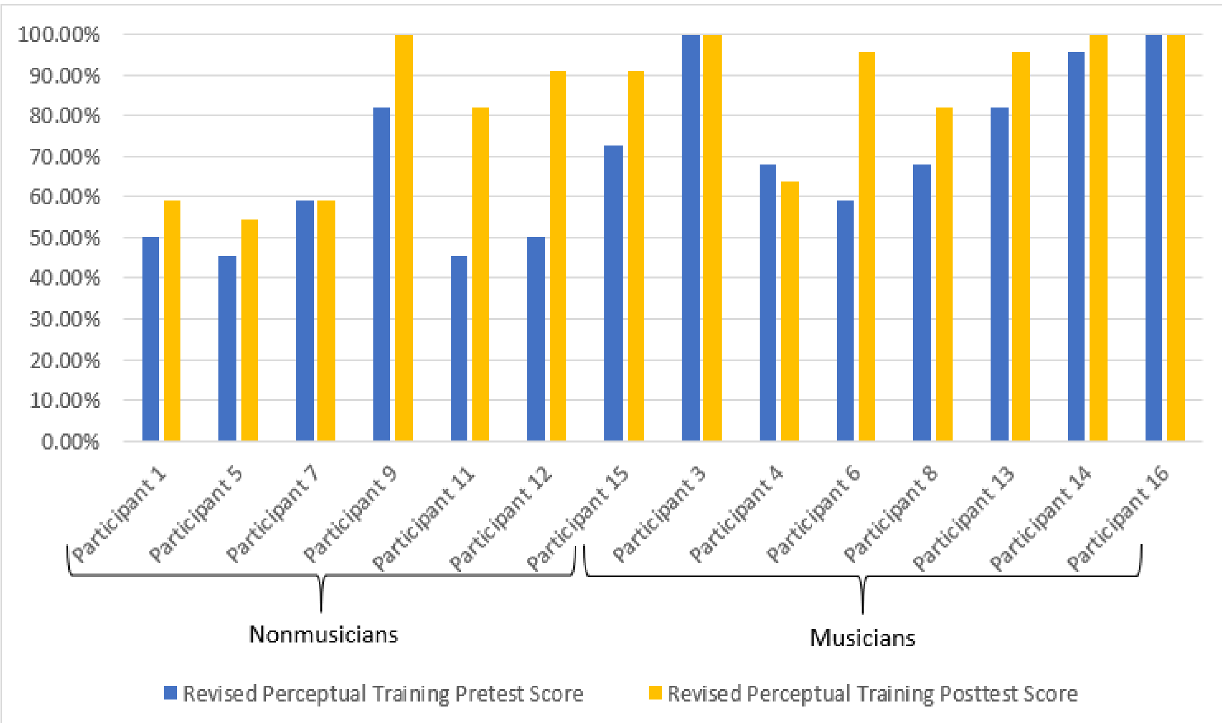


Figure 4.7: Perceptual Training Pretest and Posttest for the Vowel's TPDI Accuracy Scores by Participant

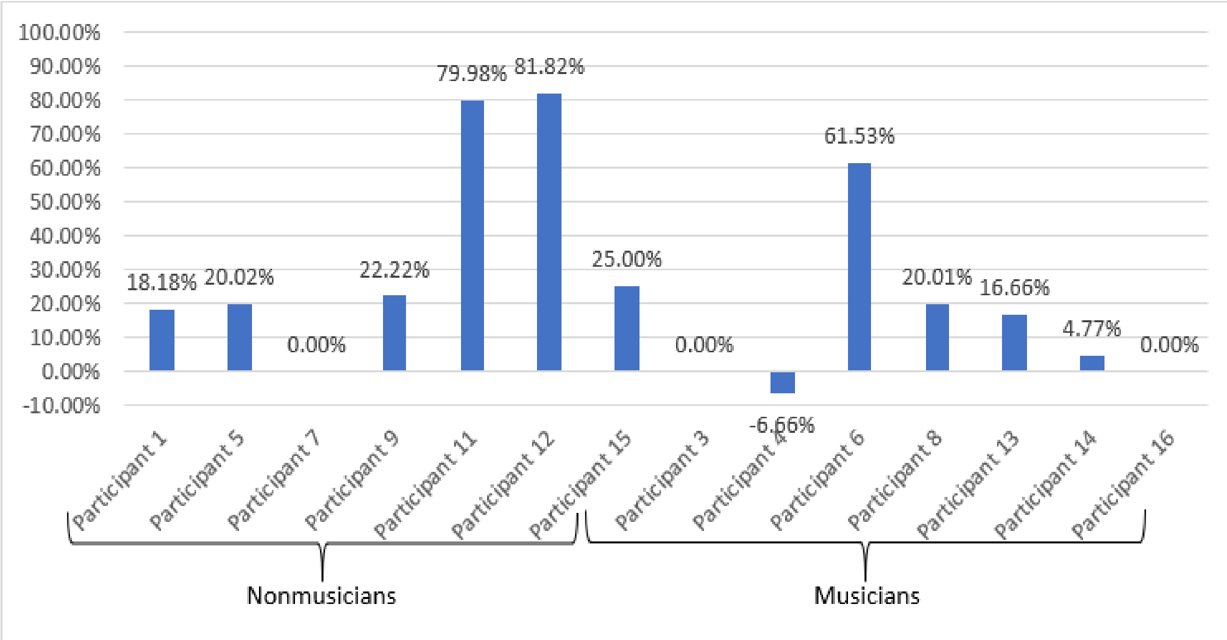


Figure 4.8: Perceptual Training Percent Change Scores of the Vowel's TPDI between Pretest and Posttest by Participant

4.1.3 Musical vs. Perceptual Training Individual Results

The below subsections, 4.1.3.1 and 4.1.3.2, show the results of individual performance between musical and perceptual training.

4.1.3.1 The Word's TPDI Accuracy

An overview of the percent changes from pretest to posttest for each training – musical and perceptual – by each participant can be seen in Figure 4.9 below. As can be seen in the graph, the gains made from the perceptual training are visibly greater than those made from the musical training, which is an observation that will be statistically explored later in Section 4.2. In fact, three participants' percent changes are even negative changes after the musical training. While this could be due to many factors, this discussion will be reserved for Chapter 5. Only three participants' musical training percent change scores were higher than the perceptual training percent change scores. In the case of Participant 3, in her first pretest, she was already near ceiling with an 87% score. After taking the musical training as her first training, her posttest score was 100%. She simply could not achieve a higher percent change score with the perceptual training as her second training. Moreover, these three participants took the musical training first, which is further explored below.

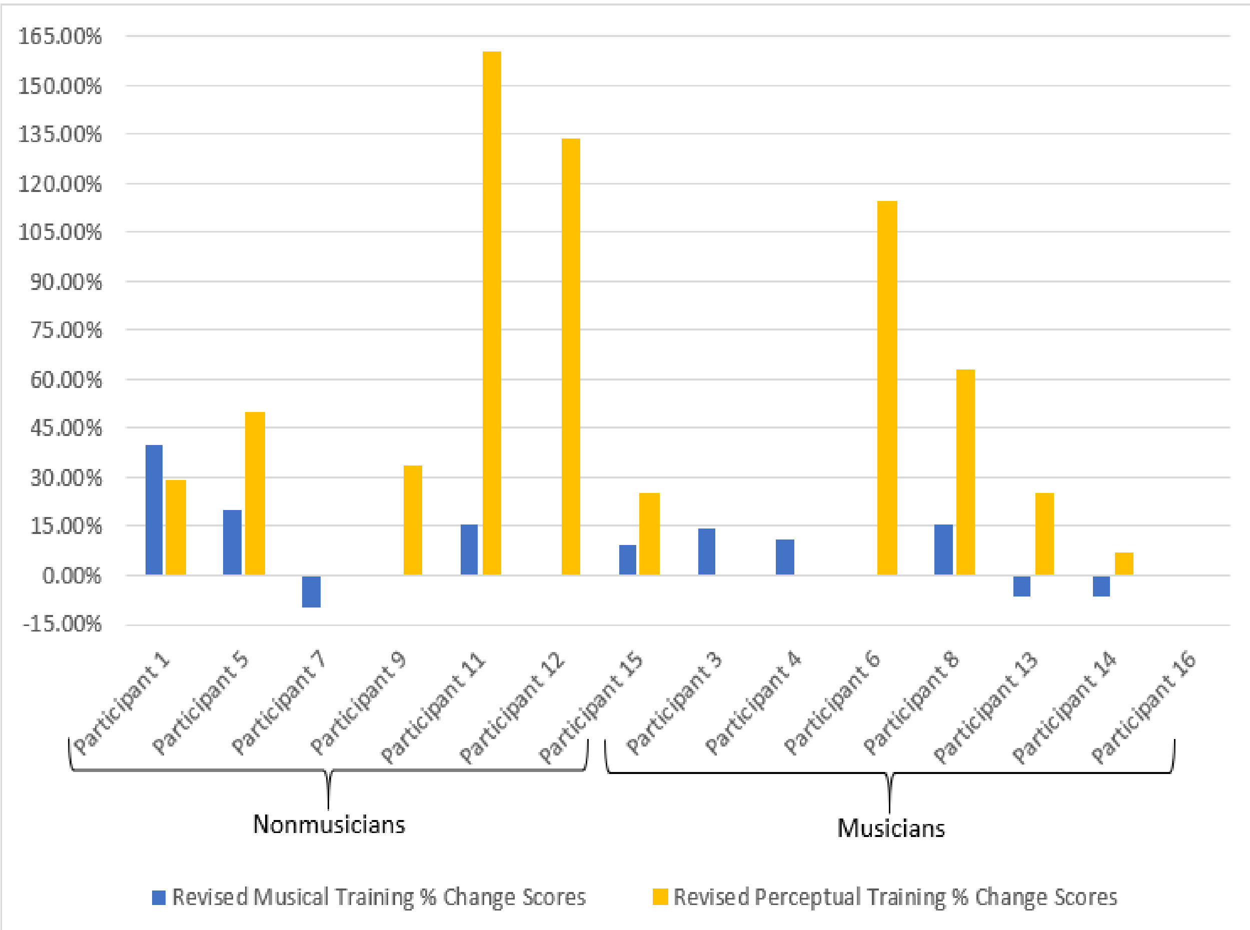


Figure 4.9: Musical vs. Perceptual Training Accuracy Percent Change Scores the Word's TPDI by Participant

Table 4.1 and 4.2, below, address the effects of the order of training. As a methodological control, training order was counterbalanced, with half the participants receiving the musical training first and the other half, the perceptual training. In terms of the word's TPDI accuracy raw scores, musicians were often close to ceiling with seemingly little effect of training order. In terms of the word's TPDI accuracy percent change, musicians in the Musical+Perceptual (MP) training group displayed much lower percent changes than those in the Perceptual+Musical (PM) training group. This may have been because having the more beneficial perceptual training first, additionally and positively affected the subsequent musical training posttest scores, whereas having the less beneficial musical training first had less additional and positive effect on the subsequent perceptual training TPDI accuracy post test scores.

The effects of training order on the nonmusicians' TPDI accuracy raw and percent change scores seem to be more variable, but may be at least in part explained again by having the potentially more beneficial perceptual training first. Furthermore, even while the musical training did not have a high impact on most participants, it still aided most nonmusicians to a limited degree. Additional factors include technical difficulties due to the online modality (lack of immediate instructional feedback, missing information on the language's rising and falling tones on disyllabic words, audio issues) which affected Participants 1-7.[45]

The presence of some outliers contribute to some high standard deviations in the data in Tables 4.1 and 4.2, but overall, the musicians patterned together fairly uniformly while the nonmusicians also pattern together to some extent. Thus, for the remainder of the analyses, scores from each training type are collapsed regardless of its presentation order.

---

[45] About half of the participants had scores above 80% after taking the perceptual training, the other half was about 60% or lower. In the first training group (Participants 1-5), technical difficulties prevented immediate feedback in the perceptual training's first task; instead, the feedback was delayed. Similarly, in the second training group (Participants 6-10), technical difficulties with the audio were particularly difficult for Participants 6 and 7, but Participant 6 likely compensated for these audio issues due to her previous training in music. Furthermore, Participants 11 and 12's much higher percent change scores in the perceptual training are likely due to their low Pretest 1 scores and the perceptual training being their first training. It is also worth noting that the higher scores attained by Participants 9 and 15 are potentially due to their previous exposure to languages that use pitch contrastively (i.e., Zambian Tonga for one semester and Japanese for 1 year, respectively).

Table 4.1: Individual Posttest of the Word's TPDI Accuracy Scores for Each Training by Each Participant

Musicians		
MP Musicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 3	100.00%	100.00%
Participant 4	62.50%	62.50%
Participant 14	93.75%	100.00%
Participant 16	100.00%	100.00%
PM Musicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 6	93.75%	93.75%
Participant 8	93.75%	81.25%
Participant 13	87.50%	93.75%
Nonmusicians		
MP Nonmusicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 1	43.75%	56.25%
Participant 5	37.50%	56.25%
Participant 15	75.00%	93.75%
PM Nonmusicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 7	56.25%	62.50%
Participant 9	100.00%	100.00%
Participant 11	93.75%	81.25%
Participant 12	87.50%	87.50%

Table 4.2: Individual TPDI of the Word Accuracy Percent Change Scores for Each Training by Each Participant

Musicians		
MP Musicians	Musical Training % Change Score	Perceptual Training % Change Score
Participant 3	14.29%	0.00%
Participant 4	11.11%	0.00%
Participant 14	-6.25%	6.67%
Participant 16	0.00%	0.00%
PM Musicians	Musical Training % Change Score	Perceptual Training % Change Score
Participant 6	0.00%	114.29%
Participant 8	15.38%	62.50%
Participant 13	-6.67%	25.00%
Nonmusicians		
MP Nonmusicians	Musical Training % Change Score	Perceptual Training % Change Score
Participant 1	40.00%	28.57%
Participant 5	20.00%	50.00%
Participant 15	9.09%	25.00%
PM Nonmusicians	Musical Training % Change Score	Perceptual Training % Change Score
Participant 7	-10.00%	0.00%
Participant 9	0.00%	33.33%
Participant 11	15.38%	160.00%
Participant 12	0.00%	133.33%

4.1.3.2 The Vowel's TPDI Accuracy

An overview of the percent changes from pretest to posttest for each training – musical and perceptual – by each participant can be seen in Figure 4.10 below. Similar to the word's TPDI accuracy results above, the gains made from the perceptual training are visibly greater than those made in the musical training for the vowel's TPDI accuracy. As before, some of the percent changes are negative changes for the musical training. Again, discussion will be reserved for Chapter 5. Table 4.3 and 4.4 below resemble Tables 4.1 and 4.2 with the same participants and similar patterns seen in both sets.

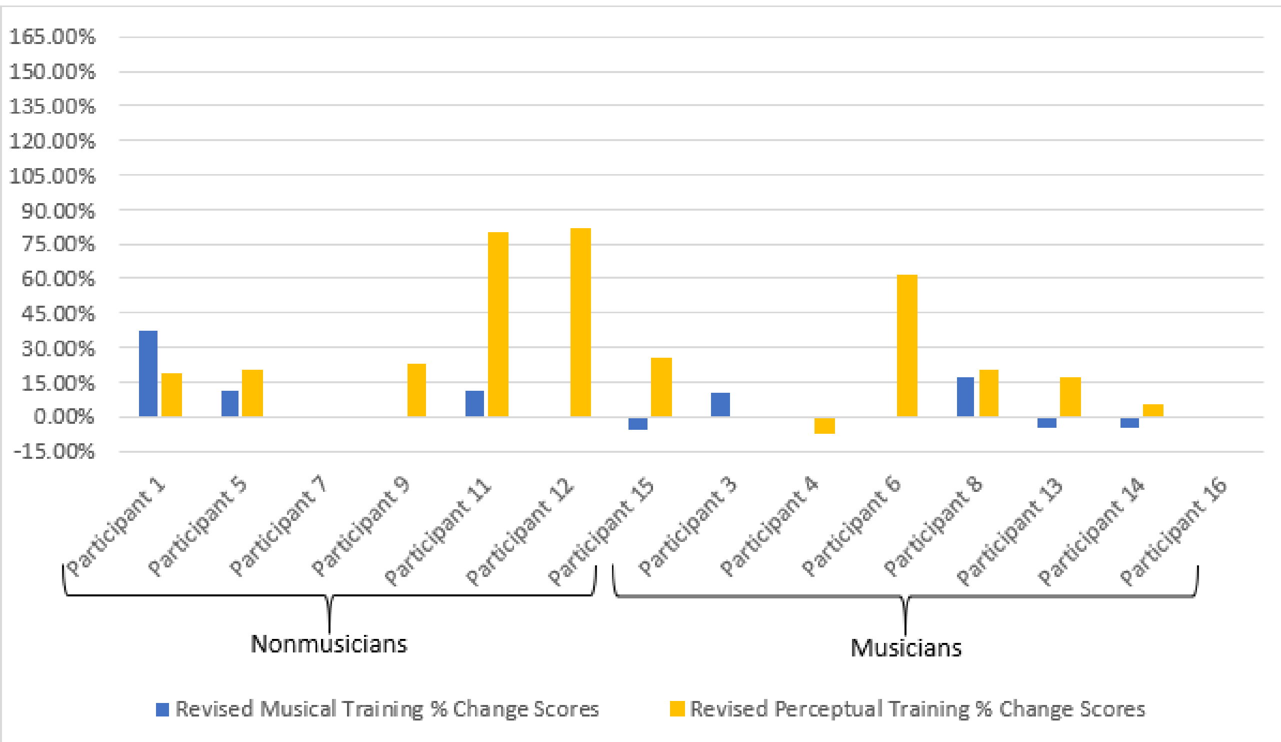


Figure 4.10: Musical vs. Perceptual Training for the Vowel's TPDI Accuracy Percent Change Scores by Participant

Table 4.3: Individual Posttest For the Vowel's TPDI Accuracy Scores for Each Training by Each Participant

Musicians		
MP Musicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 3	100.00%	100.00%
Participant 4	68.18%	63.64%
Participant 14	95.45%	100.00%
Participant 16	100.00%	100.00%
PM Musicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 6	95.45%	95.45%
Participant 8	95.45%	81.82%
Participant 13	90.91%	95.45%
Nonmusicians		
MP Nonmusicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 1	50.00%	59.09%
Participant 5	45.45%	54.55%
Participant 15	72.73%	90.91%
PM Nonmusicians	*Musical Training Posttest Score*	*Perceptual Training Posttest Score*
Participant 7	59.09%	59.09%
Participant 9	100.00%	100.00%
Participant 11	90.91%	81.80%
Participant 12	90.91%	90.91%

Table 4.4: Individual For the Vowel's TPDI Accuracy Percent Change Scores for Each Training by Each Participant

Musicians		
MP Musicians	*Musical Training % Change Score*	*Perceptual Training % Change Score*
Participant 3	10.00%	0.00%
Participant 4	0.00%	-6.66%
Participant 14	-4.55%	4.77%
Participant 16	0.00%	0.00%
PM Musicians	*Musical Training % Change Score*	*Perceptual Training % Change Score*
Participant 6	0.00%	61.53%
Participant 8	16.66%	20.01%
Participant 13	-4.76%	16.66%
Nonmusicians		
MP Nonmusicians	*Musical Training % Change Score*	*Perceptual Training % Change Score*
Participant 1	37.51%	18.18%
Participant 5	11.10%	20.02%
Participant 15	-5.88%	25.00%
PM Nonmusicians	*Musical Training % Change Score*	*Perceptual Training % Change Score*
Participant 7	0.00%	0.00%
Participant 9	0.00%	22.22%
Participant 11	11.14%	79.98%
Participant 12	0.00%	81.82%

4.1.4 Individual Results of Generalizability

After the completion of both types of training and pre- and post-testing of perception, discrimination, and identification of L and H tones, participants were tested on their ability to generalize to a new category – that of M tones. Additionally, they were tested on their ability to generalize to new words and new tonal melodies (i.e., HH and MH). Results from these tests follow, with the word's TPDI accuracy distinguished from the vowel's TPDI accuracy.

4.1.4.1 The Word's TPDI Accuracy

An overview of the scores from each participants' Pretest 1, Posttest 1, and Posttest 2 to the generalization test by each participant can be seen in Figure 4.11 below.

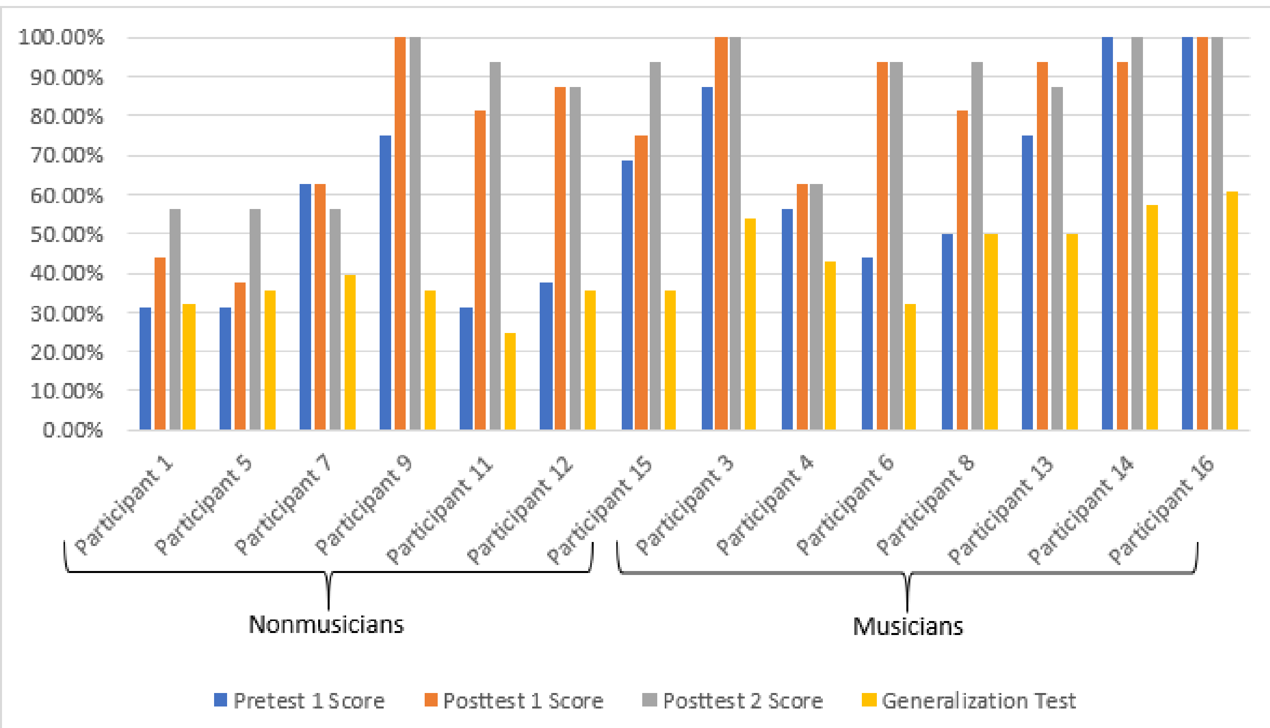


Figure 4.11: Pre- and Posttest Scores of the Word's TPDI to the Generalization Test by Participants

As can be seen in the graph, the generalization test scores are generally lower than the pretest and posttest scores by participants. In only two cases, are the generalization scores better than previous scores: Participants 1 and 5's Pretest 1 scores. This rather surprising finding will be discussed in Chapter 5.

4.1.4.2 The Vowel's TPDI Accuracy

An overview of the scores from each training's pretest and posttest to the generalization test by each participant can be seen in Figure 4.12 below.

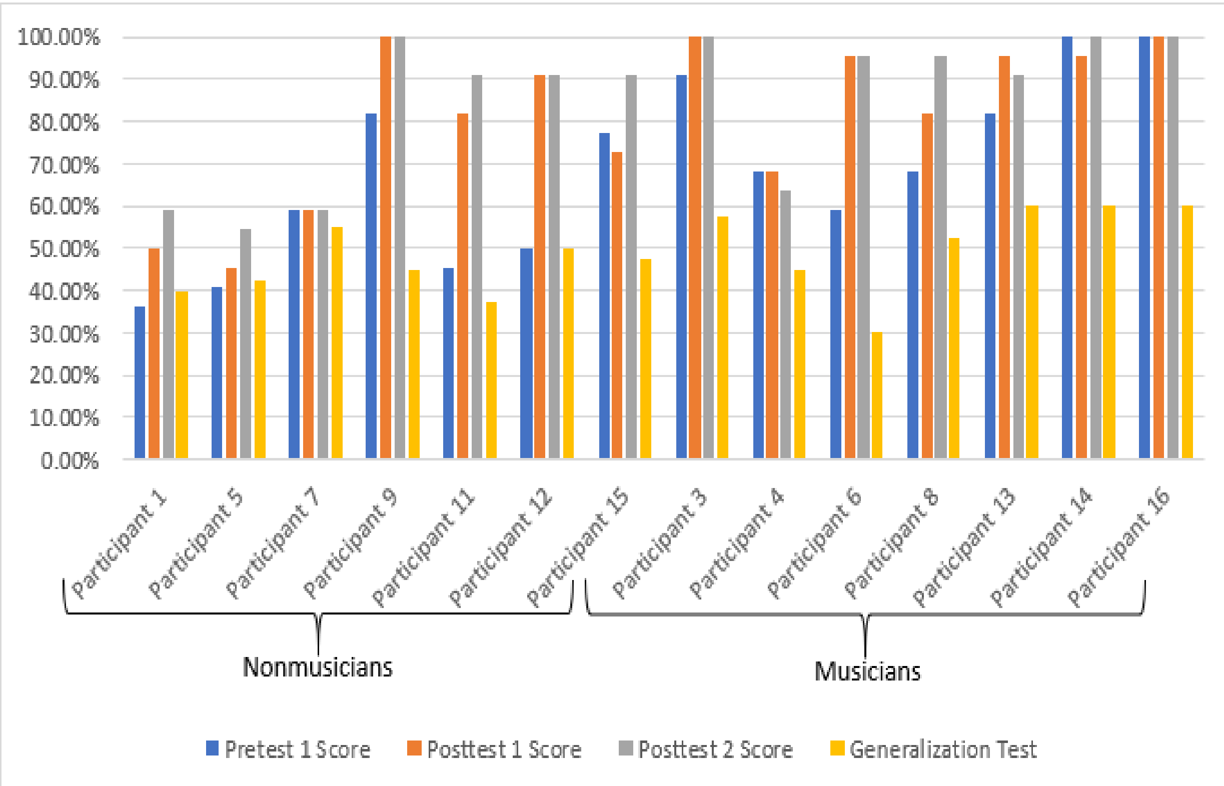


Figure 4.12: Pre- and Posttest Scores to Generalization Test Scores of the Vowel's TPDI by Participants

As can be seen in the graph, the generalization test scores are generally lower than all the pretest and posttest scores by participants as was the case with the word's TPDI accuracy scores. In only three cases, were the generalization scores equal to or better than the pretest scores: Participants 1 and 5's Pretest 1 scores were lower than their Generalization Test scores while Participant 12's Pretest 1 score was equal to his Generalization Test score.

4.1.5 Individual Survey Results

The results of the survey revealed that 11 of the 16 (69%) participants who partook in the study preferred the musical training (See Figure 4.13). Of the 11 who preferred the musical training, 5 were in the MP group while 6 were in the PM group. Alternatively, of the four who preferred the perceptual training, three were in the MP group and one was in the PM group. Further, as can be seen in Table 4.5, while most musicians (i.e., 86%) preferred the musical training because of their familiarity with this kind of training, more than half of the nonmusicians (i.e., 56%) also preferred the musical training. Moreover, three nonmusicians (Participant 2, 10, and 11) pointed out that the musical training provided a greater foundational context; importantly, Participant 2 was part of the MP group while Participants 10 and 11 were part of the PM group, showing that ordering did not affect this observation. Additionally notable, both participants with some tonal language experience expressed that the perceptual training was their preferred training.

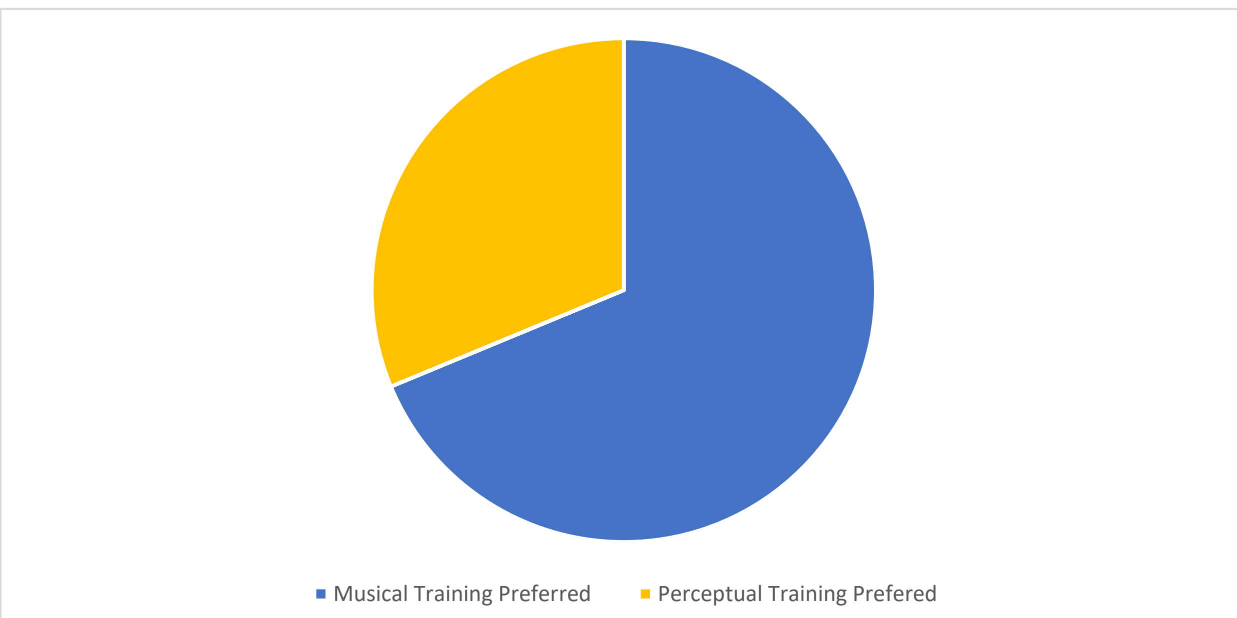


Figure 4.13: Percentage Breakdown of Preferred Training Across All Participants

Table 4.5: Survey Results by Each Participant Showing Experience, Preferred Training, and Reason

Participant	Musician vs. Nonmusician	MP vs PM Group	Preferred Training	Reason	Tonal Language Experience; Tonal Language Learned
1	Nonmusician	MP	Musical	Without much/any real experience in tonal languages the musical component more directly translated to my understanding. Do Re Mi Fa So La Ti Do and so-forth.	No
2	Nonmusician	MP	Musical	Because it was the basis for repetition training and listening for tone seems natural while communicating	No
3	Musician	MP	Musical	It was more interactive	No
4	Musician	MP	Perceptual	music is intimidating	No
5	Nonmusician	MP	Perceptual	each person had a different tonal range, so the repetition helped me hear the tones better.	No
6	Musician	PM	Musical	I am more used to that kind of training so I enjoy it more.	No
7	Nonmusician	PM	Musical	It was easier to follow. there were fewer audio mistakes. There was a range so it was challenging but easier (idk i felt it better)	No
8	Musician	PM	Musical	Musical Training because I am used to that but I also liked the repetition training as well	No
9	Nonmusician	PM	Perceptual	I think that this type of training sticks better with me specifically, the more I hear something the more it becomes second nature to me to understand it	Yes; Zambian Tonga
10	Nonmusician	PM	Musical	It had a more defined example for what we were looking for when listening to the sounds	No
11	Nonmusician	PM	Musical	I liked hearing the different tones expressed through the musical context, it helped me understand	No
12	Nonmusician	PM	Perceptual	I am able to gather more from hearing/experiencing the language more than the application of music.	No

13	Musician	PM	Musical	Because it relates to something I understand. However, using the examples of different languages having the meanings of words change based on tones gave perspective.	No
14	Musician	MP	Musical	The initial explanation of relative pitches was something I was able to relate to existing knowledge, although... that actually makes matters a bit confusing (trying to cross-reference what I was learning with what I already knew.)	No
15	Nonmusician	MP	Perceptual	Not really familiar with musical training. It's something I do enjoy but I cannot pick up the distinctions as quickly as someone with possibly a bit more music experience.	Yes; Japanese
16	Musician	MP	Musical	Once I learned the gist of it auditorily it was very easy for me from that point on. However maybe if I had the repetition first it would have been easy too	No

4.2 Results to RQ1: Musical vs. perceptual training effectiveness

To test if using musical versus perceptual training in the classroom is effective for increasing TPDI accuracy, average scores for the word's and vowel's TPDI accuracies are presented, with differentiation between musicians and nonmusicians. Section 4.2.1 presents results for musical training with 4.2.1.1 presenting the word's TPDI accuracy and 4.2.1.2, results of the vowel's TPDI accuracy. Likewise, Section 4.2.2 presents results for perceptual training with 4.2.2.1 presenting the word's TPDI accuracy and 4.2.2.2, results of the vowel's TPDI accuracy. Section 4.2.3 directly compares the effectiveness of musical versus perceptual training with 4.2.3.1 presenting the word's TPDI accuracy and 4.2.3.2, results of the vowel's TPDI accuracy.

4.2.1 Musical Training

The below subsections, 4.2.1.1 and 4.2.1.2, show results across participants in the musical training.

4.2.1.1 The Word's TPDI Accuracy Results

In looking at the average scores across participants, in Figure 4.14 below, it can be seen that there is almost a modest 5% of TPDI accuracy gain from musical training across participants.

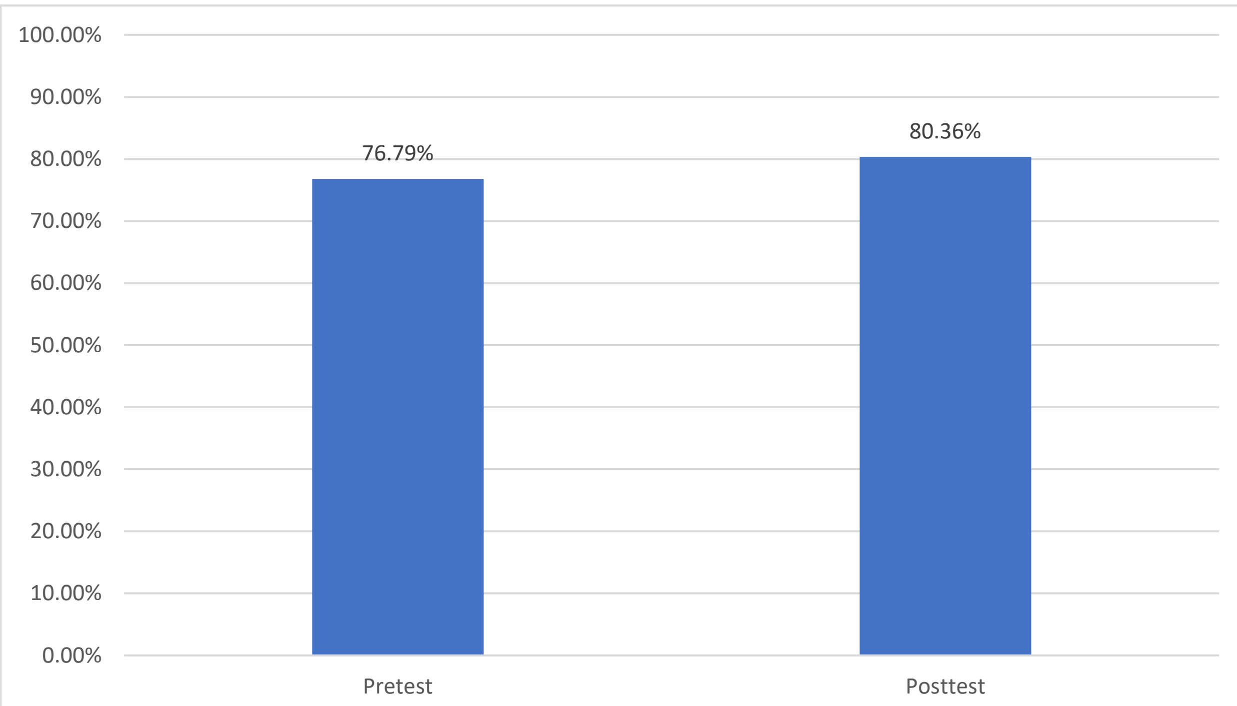


Figure 4.14: Musical Training Pretest and Posttest Average Scores of the Word's TPDI Across Participants

In order to determine whether the difference between pretest and posttest scores for the musical training was significant, a paired-samples t-test was conducted to evaluate the effect of the training on the word's TPDI accuracy scores. There was not a statistically significant increase in accuracy scores from pretest (M=76.79, SD=23.70) to posttest (M=80.38, SD=21.63), t(13)=-1.847, p=.088. However, the eta squared statistic (.21) indicated a large effect size.

Summaries of test results are presented in Tables 4.6 and 4.7. In other words, the musical training will not be likely to improve the word's TPDI accuracy scores in a general population.

Results before and after the musical training were also examined based on musical training background. Below, Figure 4.15 indicates that neither the musicians nor nonmusicians benefitted more from the training. Both only show an average increase of about 2-3%.

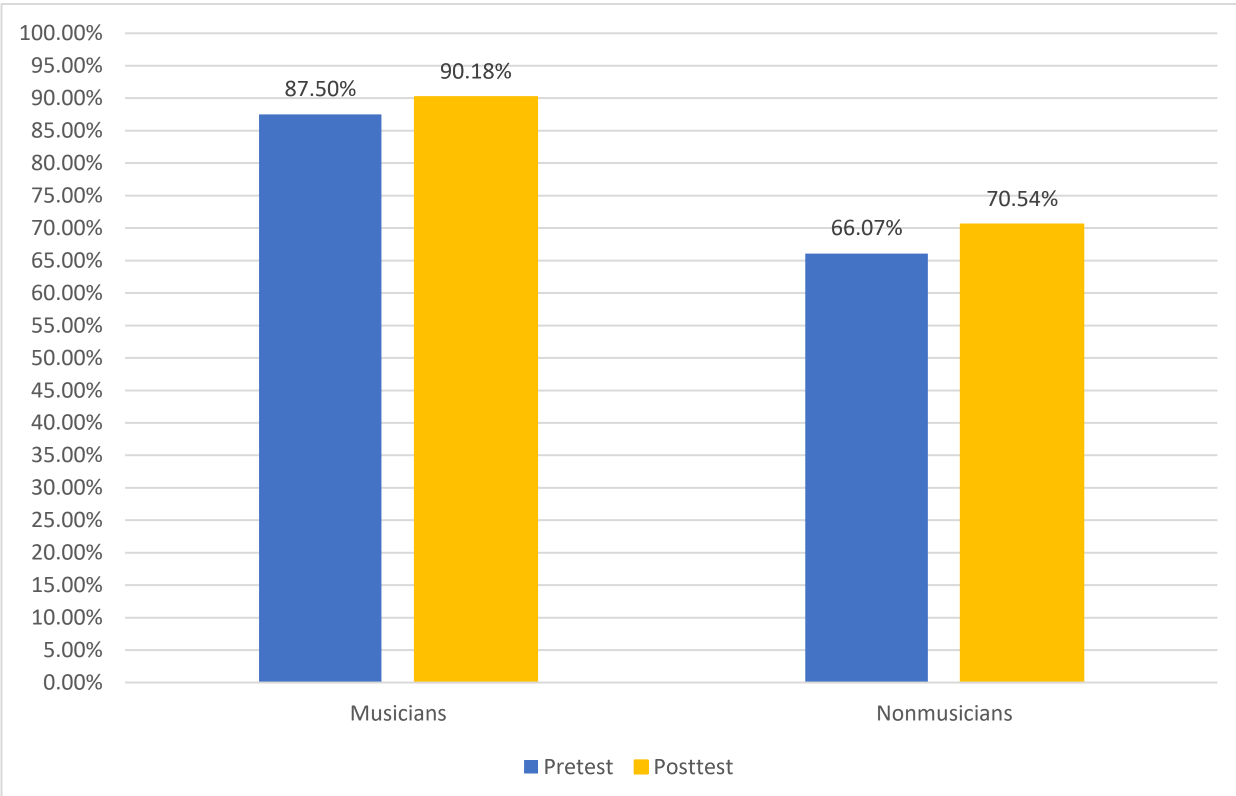


Figure 4.15: Average Improvement of the Word's TPDI from Musical Training Pretest to Posttest Within the Musician and Nonmusician Groups

In order to determine whether there was any interaction between the differences in scores and a musical training background, a mixed between-within subjects analysis of variance (see Tables 4.6, 4.8, and Figure 4.16) was conducted to assess the impact of the musical training on musicians' and nonmusicians' TPDI accuracy scores on the word across two time periods (pretest and posttest). There was no significant main effect for time, Wilks' Lambda=.79, F (1, 12)=3.200, p=.099, partial eta squared=.211. The main effect comparing the two types of musical backgrounds was not significant F (1, 12)=3.511, p=.086, partial eta squared=.226, suggesting that the descriptive difference may not be statistically reliable. There was no significant interaction between a musical training background and time, Wilks' Lambda=.98, F (1, 12)=.200, p=.663, partial eta squared=.016. In other words, no groups improved over time and the descriptive difference between musicians and nonmusicians was not statistically robust.

Table 4.6: Descriptive Statistics for the Word's TPDI Accuracy of Musical Training Pretest and Posttest Scores

		Mean	N	Std. Deviation
All Learners	Musical Training Pretest Score of the Word's TPDI	76.7857%	14	23.79535
All Learners	Musical Training Posttest Score of the Word's TPDI	80.3571%	14	21.63080
Musician	Musical Training Pretest Score of the Word's TPDI	87.5000%	7	15.30931
Musician	Musical Training Posttest Score of the Word's TPDI	90.1786%	7	12.93873
Nonmusician	Musical Training Pretest Score of the Word's TPDI	66.0714%	7	26.72612
Nonmusician	Musical Training Posttest Score of the Word's TPDI	70.5357%	7	24.92548

Table 4.7: Paired Samples Test of Musical Training Pretest and Posttest Scores for the Word's TPDI

	Paired Difference			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Musical Pretest-Posttest	-3.57143%	7.23668	1.93408	-7.74976	.60690	-1.847	13	.088

Table 4.8: Mixed Between-Within Analysis of Variance Statistics Tests for the Word's TPDI Accuracy Scores of Musical Training Pretest and Posttest Scores

	Tests of Within-Subjects Effects				
Effect	Value	F	Hypothesis df	Error df	Sig.
Time (Wilks' Lambda)	.789	3.200^{b}	1.000	12.000	.099
Time * Musician (Wilks' Lambda)	.984	.200^{b}	1.000	12.000	.663
	Tests of Between-Subjects Effects				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Musician	2952.009	1	2952.009	3.511	.086
Error	10089.286	12	840.774		

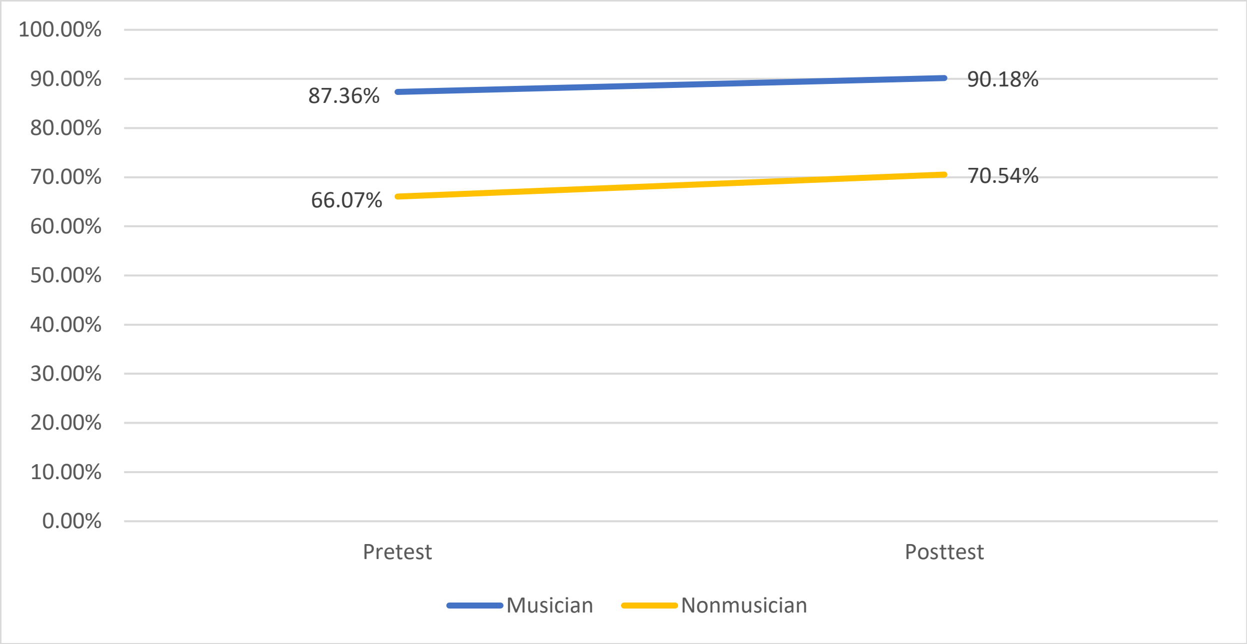


Figure 4.16: Mixed Between-Within Estimated Marginal Mean Plot of Musical Training the Word's TPDI Pretest and Posttest Scores between Musicians and Nonmusicians

4.2.1.2 The Vowel's TPDI Accuracy Results

In looking at the average scores across participants, in Figure 4.17 below, it can be seen that there is less gain (less than 3%) in the vowel's TPDI accuracy as compared to the 5% gained when looking at the word's TPDI accuracy.

In order to determine whether the difference between pretest and posttest scores for the musical training was significant, a paired-samples t-test was conducted to evaluate the effect of the training on the vowel's TPDI accuracy scores. There was not a statistically significant increase in accuracy scores from pretest (M=79.87, SD=21.37) to posttest (M=82.47, SD=19.44), t(13)=-1.529, p=.150. However, the eta squared statistic (.15) indicated a large effect size. Summaries of test results are presented in Tables 4.9 and 4.10. In other words, the musical training will not be likely to improve the vowel's TPDI accuracy scores in a general population.

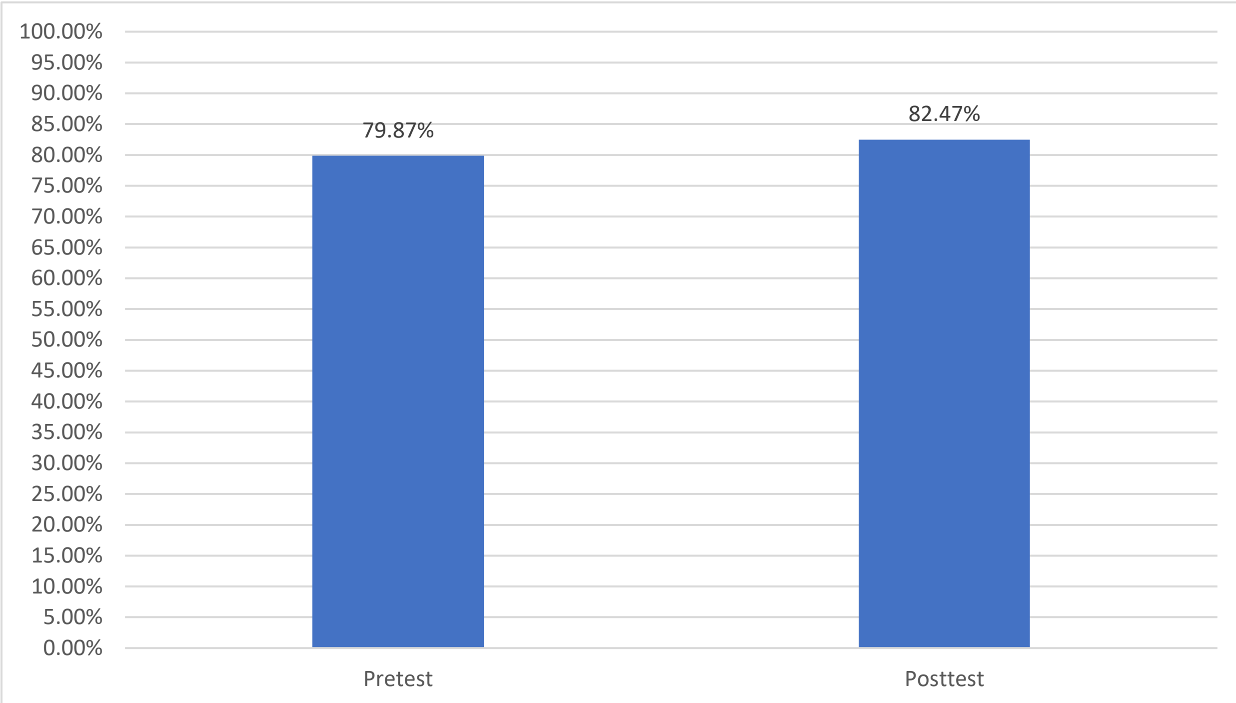


Figure 4.17: Musical Training Pretest and Posttest Average Scores of the Vowel's TPDI

Results before and after the musical training were also examined based on musical background. Figure 4.18 below indicates that neither the musicians nor nonmusicians benefitted more from the training in terms of the vowel's TPDI accuracy. Both only show an average increase of about 2-3%.

In order to determine whether there was any interaction between the differences in scores and a musical training background, a mixed between-within subjects analysis of variance (see Tables 4.9, 4.11, and Figure 4.19) was conducted to assess the impact of the musical training on musicians' and nonmusicians' TPDI accuracy scores of the word across two time periods (pretest and posttest). There was no substantial main effect for time, Wilks' Lambda = .85, $F(1, 12) = 2.184$, $p = .165$, partial eta squared = .154. The main effect comparing the two types of musical backgrounds approached but did not reach statistical significance, $F(1, 12) = 4.392$, $p = .058$, partial eta squared = .011. There was no significant interaction between a musical training background and time, Wilks' Lambda = .99, F (1, 12) = .137, p = .718, partial eta squared = .011. In other words, neither group improved over time, but musicians descriptively and almost statistically outperformed nonmusicians in pre- and post-tests.

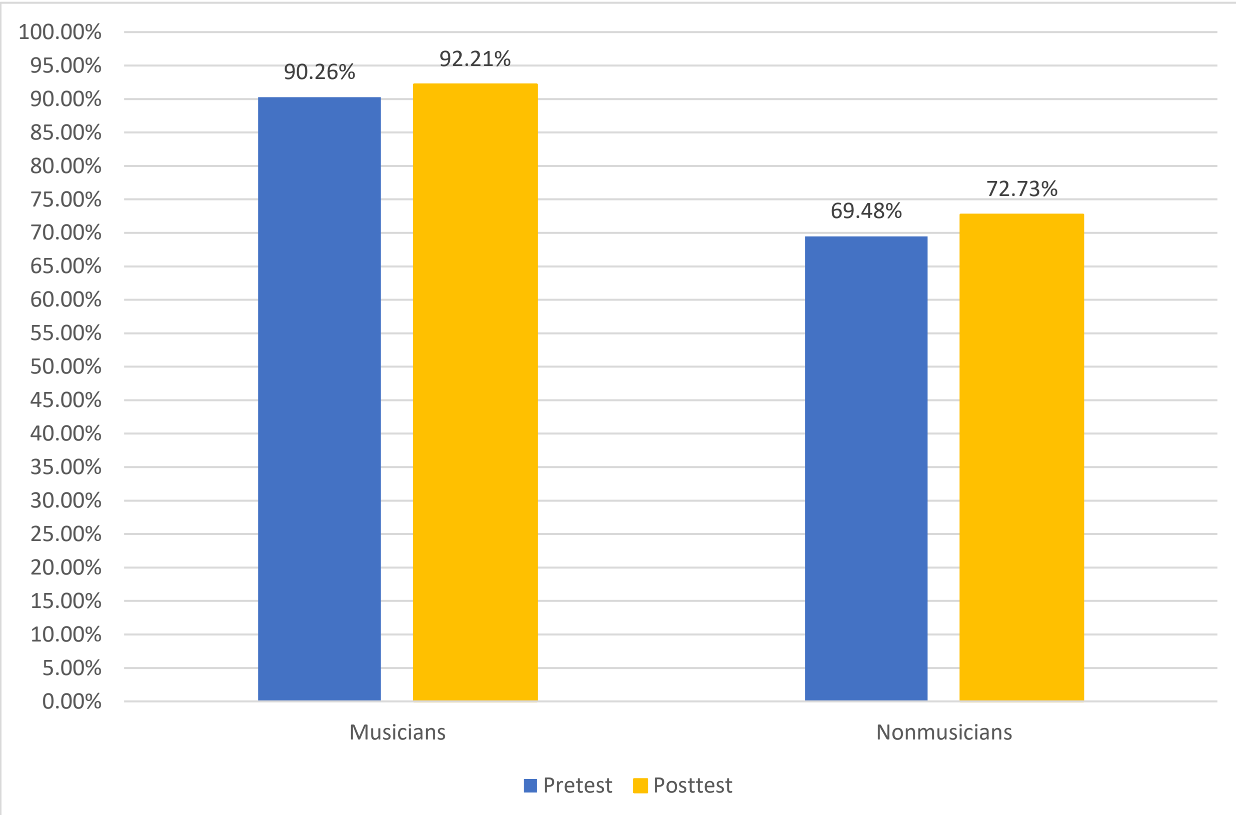


Figure 4.18: Average Improvement from Musical Training Pretest to Posttest Scores of the Vowel's TPDI Within the Musician and Nonmusician Groups

Table 4.9: Descriptive Statistics for the Vowel's TPDI Accuracy of Musical Training Pretest and Posttest Scores

		Mean	N	Std. Deviation
All Learners	Musical Training Pretest Score of the Vowel's TPDI	79.8679%	14	21.37331
All Learners	Musical Training Posttest Score of the Vowel's TPDI	82.4664%	14	19.43732
Musicians	Musical Training Pretest Score of the Vowel's TPDI	90.2586%	7	26.72612
Musicians	Musical Training Posttest Score of the Vowel's TPDI	92.2057%	7	11.56712
Nonmusicians	Musical Training Pretest Score of the Vowel's TPDI	69.4771%	7	24.57747
Nonmusicians	Musical Training Posttest Score of the Vowel's TPDI	72.7271%	7	21.80053

Table 4.10: Paired Samples Test of Musical Training Pretest and Posttest Scores for Scores of the Vowel's TPDI

	Paired Difference			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Musical Pretest-Posttest	-2.59857%	6.35760	1.69914	-6.26934	1.07220	-1.529	13	.150

Table 4.11: Mixed Between-Within Statistics Tests for Scores of the Vowel's TPDI of Musical Training Pretest and Posttest Scores

	Tests of Within-Subjects Effects				
Effect	Value	F	Hypothesis df	Error df	Sig.
Time (Wilks' Lambda)	.846	2.184[b]	1.000	12.000	.165
Time * Musician (Wilks' Lambda)	.989	.137[b]	1.000	12.000	.718
	Tests of Between-Subjects Effects				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Musician	2836.518	1	2836.518	4.392	.058
Error	7750.920	12	645.910		

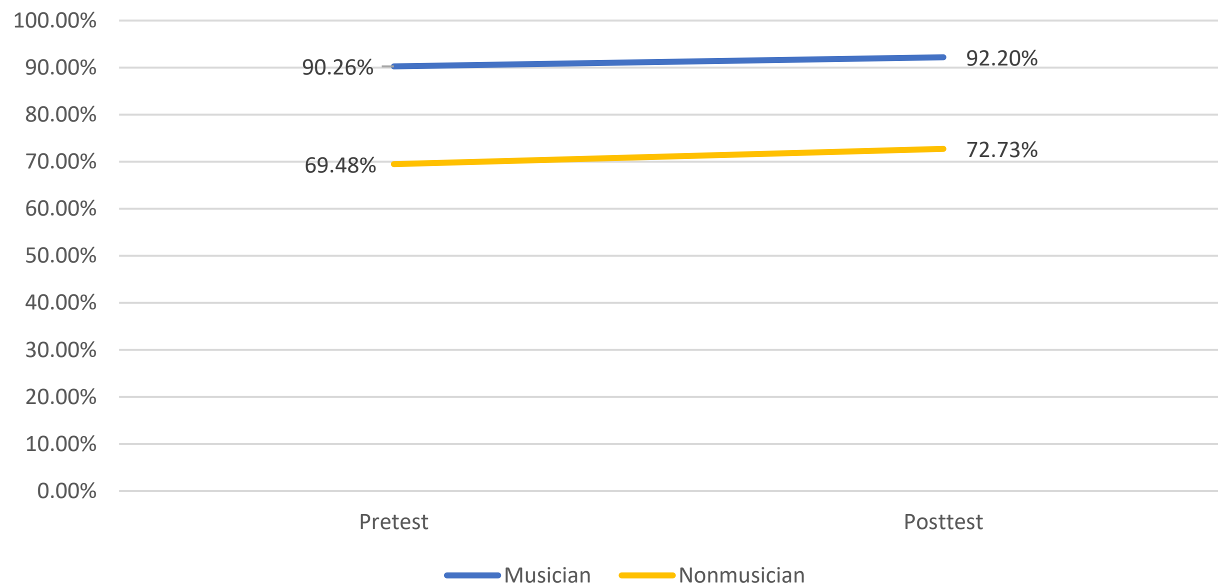


Figure 4.19: Mixed Between-Within Estimated Marginal Mean Plot Pretest and Posttest Scores of the Vowel's TPDI between Musicians and Nonmusicians

4.2.2 Perceptual Training

The results below in subsections 4.2.2.1 and 4.2.2.2 show the performance across participants in the perceptual training.

4.2.2.1 The Word's TPDI Accuracy

In looking at the average scores across participants, in Figure 4.20 below, it can be seen that there is more than a 20% TPDI accuracy gain. In order to determine whether the difference between pretest and posttest scores for the perceptual training was significant across all participants, a paired-samples t-test was conducted to evaluate the effect of the training on the word's TPDI accuracy scores. There was a statistically significant increase in accuracy scores with the pretest (M=63.39, SD=23.75) for each participant's perceptual training compared to the posttest (M=83.48, SD=17.09), t(13)=-3.953, p=.002. The eta squared statistic (.55) indicated a large effect size. Summaries of test results are presented in Tables 4.12 and 4.13, below. In other words, the perceptual training will be likely to improve the word's TPDI accuracy scores in a general population.

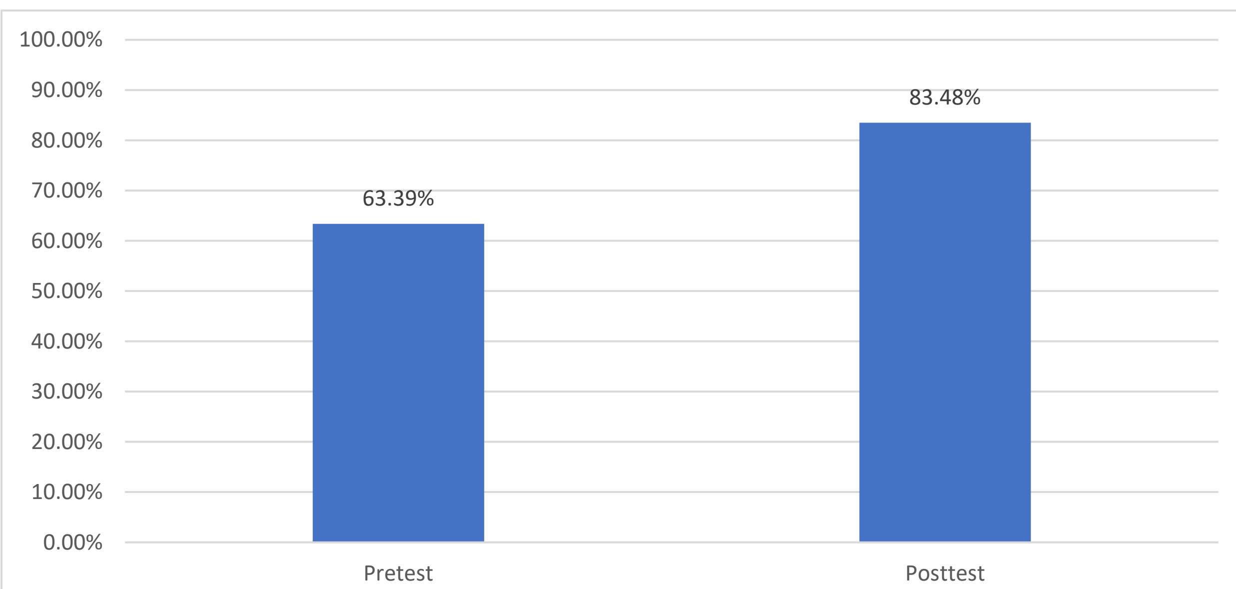


Figure 4.20: Perceptual Training Pretest and Posttest Average Scores of the Word's TPDI

Results before and after the perceptual training were also examined based on musical background. Figure 4.21 below indicates that the nonmusicians in this dataset benefitted more from the training than the musicians. While the musicians show about a 15% increase in the word's TPDI accuracy after the training, the nonmusicians increased their accuracy by 25%.

A mixed between-within subjects analysis of variance (see Tables 4.12, 4.14, and Figure 4.22) assessed the impact of the perceptual training on musicians' and nonmusicians' TPDI accuracy scores across two time periods (pretest and posttest). There was a substantial main effect for time, Wilks' Lambda = .44, F (1, 12) = 15.537, p = .002, partial eta squared = .564, with both groups showing an increase in the word's TPDI accuracy scores across the two time periods. The main effect comparing the two types of musical backgrounds approached but did not reach statistical significance, F (1, 12) = 4.374, p = .058, partial eta squared = .267, suggesting some difference between the two types of participants (i.e., musician vs. nonmusician). The interaction was not statistically significant between a musical training background and time, Wilks' Lambda = .93, F (1, 12) = .928, p = .354, partial eta squared = .072,. In other words, all the groups improved over time and musicians generally outperformed non musicians and this indicates that groups will improve after training. Further, musicians will likely always have higher scores before and after training, but nonmusicians could potentially narrow the gap between themselves and the musicians.

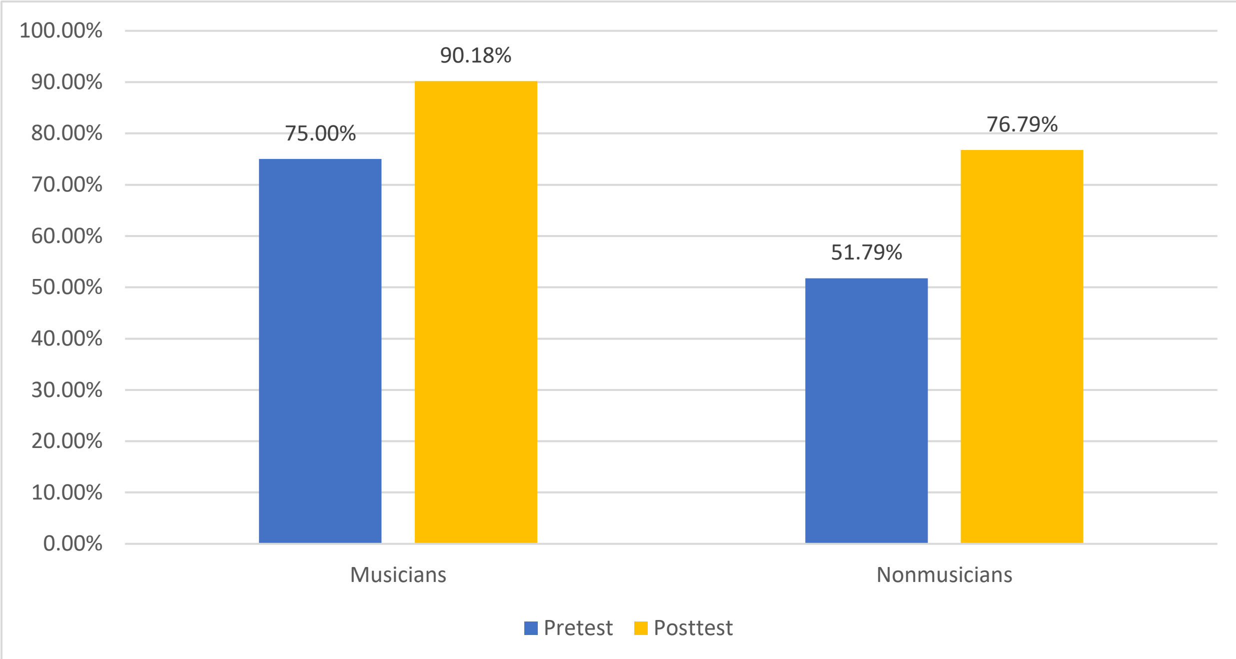


Figure 4.21: Average Improvement with the Word's TPDI Accuracy Scores from Perceptual Training Pretest to Posttest Within the Musician and Nonmusician Groups

Table 4.12: Descriptive Statistics for Perceptual Training's Pretest and Posttest Scores of the Word's TPDI

		Mean	N	Std. Deviation
All Learners	Perceptual Training Pretest Score of the Word's TPDI	63.3929%	14	23.74964
All Learners	Perceptual Training Posttest Score of the Word's TPDI	83.4821%	14	17.09123
Musician	Perceptual Training Pretest Score of the Word's TPDI	75.0000%	7	23.66212
Musician	Perceptual Training Posttest Score of the Word's TPDI	90.1786%	7	13.90872
Nonmusicians	Perceptual Training Pretest Score of the Word's TPDI	51.7857%	7	18.65053
Nonmusicians	Perceptual Training Posttest Score of the Word's TPDI	76.7857%	7	18.29813

Table 4.13: Paired Samples Statistics for Perceptual Training's Pretest and Posttest Scores of the Word's TPDI

	Paired Difference			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Perceptual Pretest-Posttest	-20.08929%	19.01719	5.08253	-31.06943	-9.10914	-3.953	13	.002

Table 4.14: Mixed Between-Within Statistics Tests for Perceptual Training Pretest and Posttest Scores of the Word's TPDI

	Tests of Within-Subjects Effects				
Effect	Value	F	Hypothesis df	Error df	Sig.
Time (Wilks' Lambda)	.436	15.537[b]	1.000	12.000	.002
Time * Musician (Wilks' Lambda)	.928	.928[b]	1.000	12.000	.354
	Tests of Between-Subjects Effects				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Musician	2345.145	1	2345.145	4.374	.058
Error	6434	12	536.179		

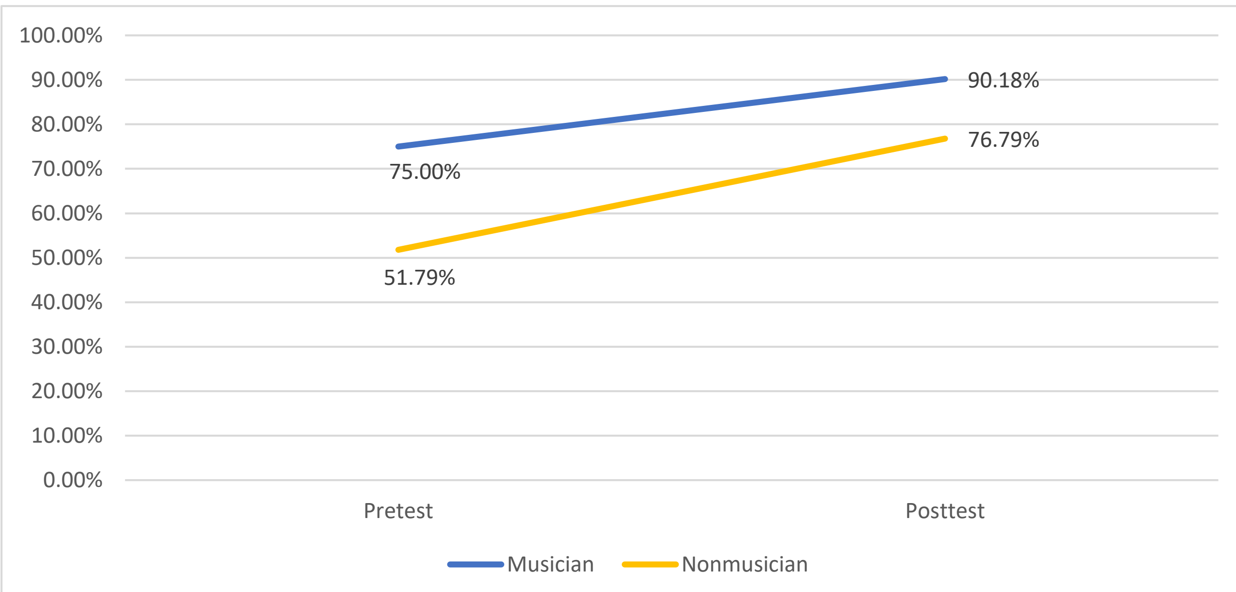


Figure 4.22: Mixed Between-Within Estimated Marginal Mean Plot of Perceptual Trainings' Pretest and Posttest Scores of the Word's TPDI between Musicians and Nonmusicians

4.2.2.2 The Vowel's TPDI Accuracy

In looking at the average scores across participants, in Figure 4.23 below, it can be seen that there is less gain (about 14%) in the vowel's TPDI accuracy as compared to the 20% gained when looking at the word's TPDI accuracy. In order to determine whether the difference between pretest and posttest scores for the perceptual training was significant, a paired-samples t-test was conducted to evaluate the effect of the training on the vowel's TPDI accuracy scores.

There was a statistically significant increase in accuracy scores with the pretest (M=69.80, SD=19.58) for each participant's perceptual training compared to the posttest (M=83.77, SD=17.35), t(13)=-3.542, p=.004, with the eta squared statistic (.49) indicating a large effect size. Summaries of test results are presented in Tables 4.15 and 4.16, below. In other words, the perceptual training will be likely to improve the vowel's TPDI accuracy scores in a general population.

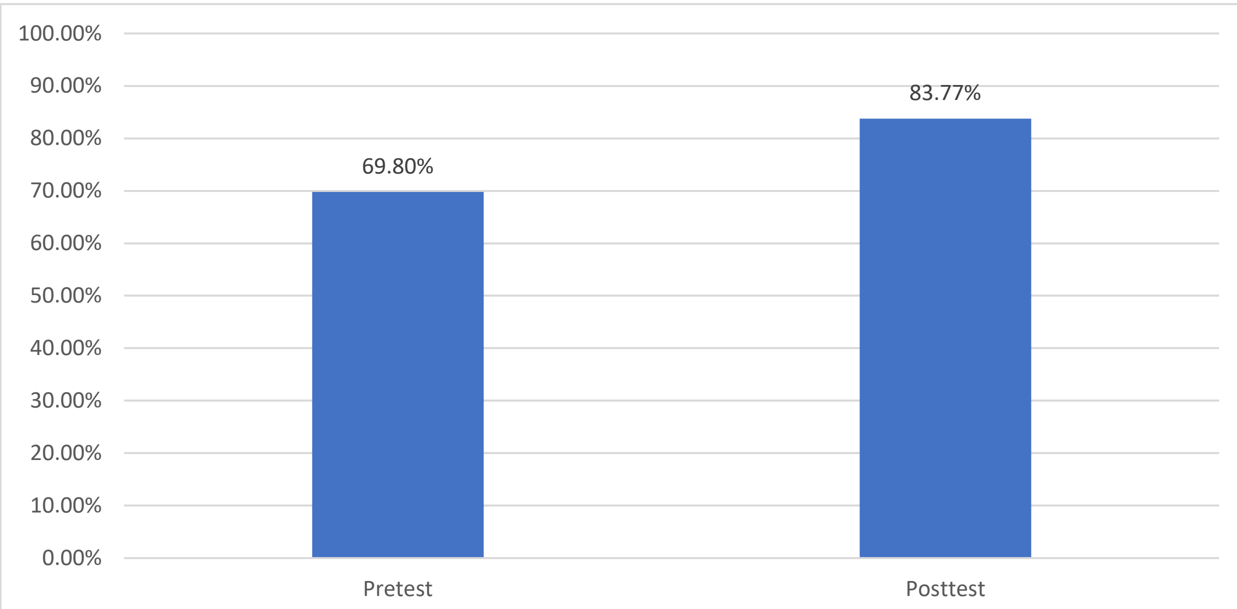


Figure 4.23: Perceptual Training Pretest and Posttest Average Scores of the Vowel's TPDI

Results before and after the musical training were also examined based on musical background. Figure 4.24 below, indicates that the nonmusicians benefitted more from the training. While the musicians show a nearly 9% gain from pretest to posttest, the nonmusicians show a nearly 20% gain from pretest to posttest.

A mixed between-within subjects analysis of variance (see Tables 4.15, 4.17, and Figure 4.25) was conducted to assess the impact of the perceptual training on musicians' and nonmusicians' TPDI accuracy scores on the vowels across two time periods (pretest and posttest). There was a substantial main effect for time, Wilks' Lambda = .48, F (1, 12) = 13.118, p = .004, partial eta squared = .522, with both groups showing an increase in the word's TPDI accuracy scores across the two time periods. The main effect comparing the two types of musical backgrounds was significant, F (1, 12) = 6.27, p = .028, partial eta squared = .343, suggesting a difference between the two types of participants (i.e., musician vs. nonmusician). There was no significant interaction between a musical training background and time, Wilks' Lambda = .88, F (1, 12) = 1.596, p = .230, partial eta squared = .117. In other words, all the groups improved over time and musicians generally outperformed non musicians and this indicates that groups will improve after training. Further, musicians will likely always have higher scores before and after training.

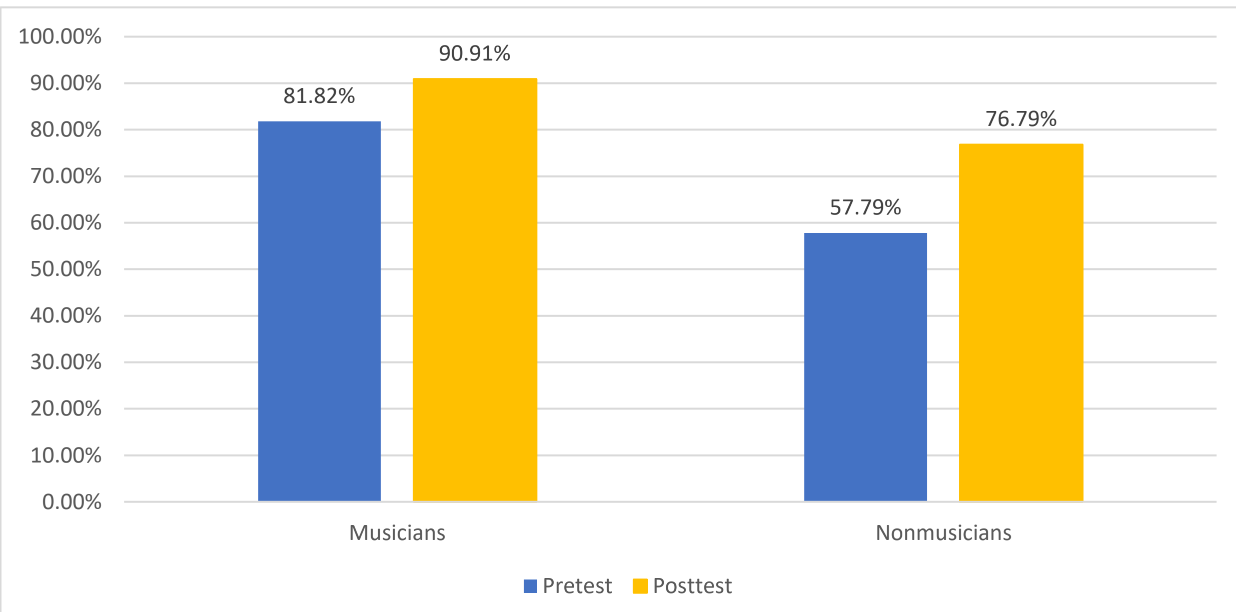


Figure 4.24: Average Improvement with the Vowel's TPDI Accuracy Scores from Perceptual Training Pretest to Posttest Within the Musician and Nonmusician Groups

Table 4.15: Descriptive Statistics for Perceptual Training's Pretest and Posttest Scores of the Vowel's TPDI

		Mean	N	Std. Deviation
All Learners	Perceptual Training Pretest Word Score	69.8043%	14	19.58047
All Learners	Perceptual Training Posttest Word Score	83.7650%	14	17.35017
Musician	Perceptual Training Pretest Word Score	81.8171%	7	17.00761
Musician	Perceptual Training Posttest Word Score	90.9086%	7	13.63444
Nonmusicians	Perceptual Training Pretest Word Score	57.7914%	7	14.30765
Nonmusicians	Perceptual Training Posttest Word Score	76.6214%	7	18.63488

Table 4.16: Paired Samples Statistics for Perceptual Training's Pretest and Posttest Scores of the Vowel's TPDI

	Paired Difference			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Perceptual Pretest-Posttest	-13.96071%	14.74914	3.94187	-22.47661	-5.44482	-3.542	13	.004

Table 4.17: Mixed Between-Within Statistics Tests for Perceptual Training Pretest and Posttest Scores of the Vowel's TPDI

	Tests of Within-Subjects Effects				
Effect	Value	F	Hypothesis df	Error df	Sig.
Time (Wilks' Lambda)	.478	13.118[b]	1.000	12.000	.004
Time * Musician (Wilks' Lambda)	.883	1.596[b]	1.000	12.000	.230
	Tests of Between-Subjects Effects				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Musician	2568.781	1	2568.781	6.272	.028
Error	4914.726	12	409.561		

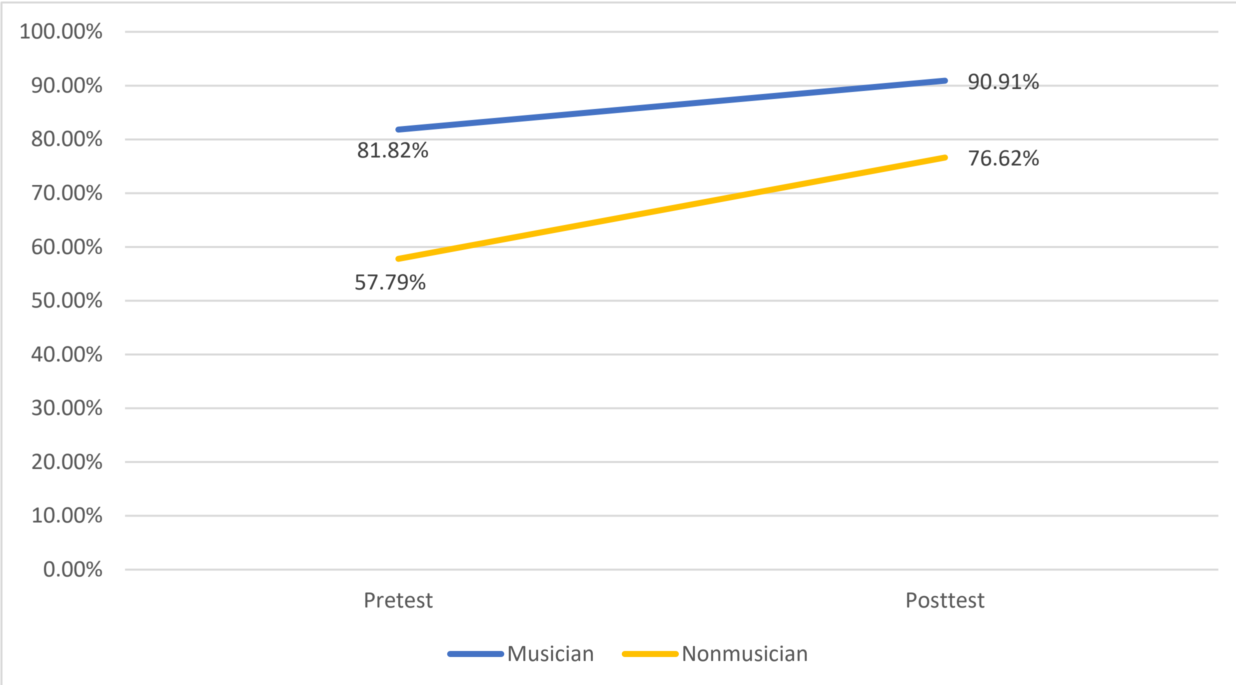


Figure 4.25: Mixed Between-Within Estimated Marginal Mean Plot of Perceptual Training's Pretest and Posttest Scores of the Vowel's TPDI between Musicians and Nonmusicians

4.2.3 Musical vs. Perceptual Training

The subsections below, 4.2.3.1 and 4.2.3.2, compare musical and perceptual training directly across participants.

4.2.3.1 The Word's TPDI Accuracy

As part of the methodology for this thesis, half the trainees received the musical training first while the other half received the perceptual training first, as explained earlier in this chapter and in Chapter 3. For this reason, in order to more easily interpret the results, percent change score have been derived from Pretest 1 to Posttest 1 and from Posttest 1 to Posttest 2[46] with each

[46] This process was described in Section 3.4 of Chapter 3.

participant's respective training.[47] When these gains are averaged by training, the averages show that perceptual training far outperformed musical training by almost 40% for the word's TPDI accuracy. This can be seen in Figure 4.26 below.

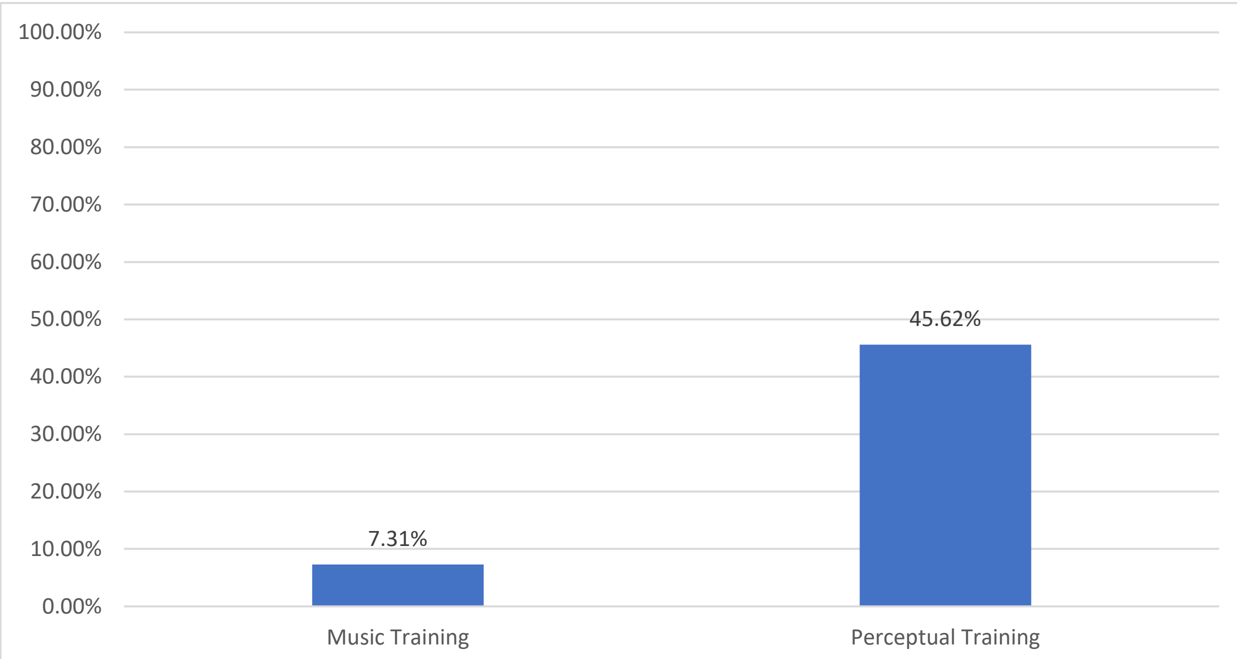


Figure 4.26: Average Increased Percent Change Scores of the Word's TPDI for Musical and Perceptual Training

In order to determine whether this difference was significant, a paired-samples t-test was conducted to evaluate each training's effectiveness on the word's TPDI accuracy. Percentage gains in accuracy after perceptual training (M=45.62, SD=53.30) were significantly higher than gains in accuracy after musical training (M=7.31, SD=13.36), t(13)=2.67, p=.019. The eta squared statistic (.35) indicated a large effect size. In Tables 4.18 and 4.19, below, the results of

[47] Percent change was calculated with the following calculation: (X-Y)/Y. X is the relevant training's posttest score and Y is the relevant training's pretest score.

the this test are shown. This reveals that the perceptual training was much more effective in the word's TPDI accuracy than the musical training across participants.

Results before and after both the musical and perceptual training were also examined based on musical background and are shown in Figure 4.27. While musicians' average improvement after the musical training was almost 4%, their perceptual training improvement was nearly 30%, revealing an almost 25% difference between the two trainings in absolute terms; in relative terms, their gains in perceptual training was 7.5 times more than the musical training. Nonmusicians had about a 10% improvement after the musical training while the perceptual training afforded a nearly 60% improvement. Between the two trainings, that shows a more than 50% difference in absolute terms; in relative terms, their gains in perceptual training were 6 times more than the musical training. With this in mind, it is clear that the nonmusicians benefitted more from the perceptual training in absolute terms, though musicians benefitted more from the perceptual training in relative terms.

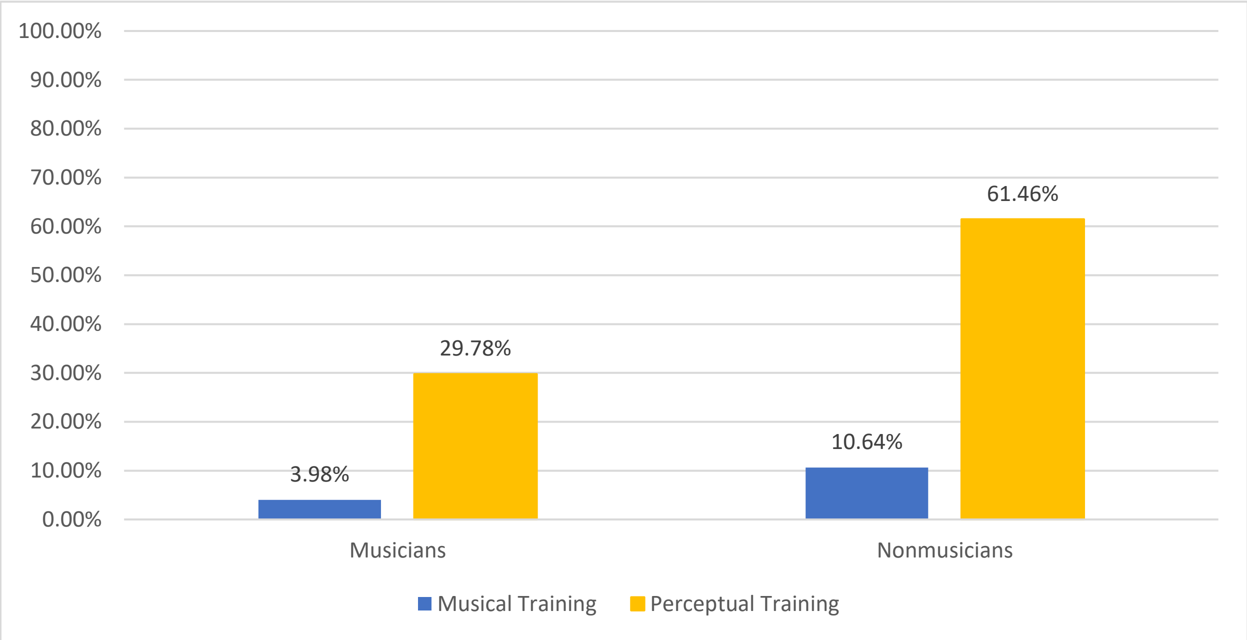


Figure 4.27: Average Improvement by Training the Word's TPDI Accuracy Percent Change Within the Musician and Nonmusician Groups

Two paired-samples t-tests evaluated the differences in accuracy gains from musical training versus perceptual training for musicians and nonmusicians separately (see Tables 4.18 and 4.19). Despite the descriptive difference in accuracy gains for musicians, there was no statistically significant difference in the effectiveness of perceptual training (M=29.78, SD=43.65) compared to musical training (M=3.98, SD=9.46), t(6)=1.519, p=.180, though eta squared (.28) indicated a large effect size. For nonmusicians, the large descriptive difference between the effectiveness of perceptual training (M=61.46, SD=60.53) versus musical training approached but did not reach statistical significance (M=10.64, SD=16.47), t(6)=2.167, p=.073, with eta squared (.44) indicating a large effect size. In other words, these results indicate that the perceptual training may not generally be more impactful than musical training on the word's TPDI accuracy gains for musicians, but they may lead to higher gains for nonmusicians. These nonsignificant results are surprising given the descriptive differences, but should be interpreted alongside the small sample sizes and the large standard deviations, which indicated considerable individual variability in the data.

Table 4.18: Descriptive Statistics for the Word's TPDI Accuracy of the Difference Between Pretest and Posttest for Each Training

		Mean	N	Std. Deviation	Std. Error Mean
All Learners	Perceptual Training Pre-Post Difference	45.6207%	14	53.29849	14.24462
All Learners	Musical Training Pre-Post Difference	7.3093%	14	13.35591	3.56952
Musicians	Perceptual Training Pre-Post Difference	29.7800%	7	43.6431	16.49676
Musicians	Musical Training Pre-Post Difference	3.9800%	7	9.45912	3.57521
Nonmusicians	Perceptual Training Pre-Post Difference	61.4614%	7	60.53439	22.87985
Nonmusicians	Musical Training Pre-Post Difference	10.638%	7	16.46672	6.22384

Table 4.19: Paired Samples Test for Each Training on All Learners, Musicians, and Nonmusicians for the Word's TPDI Gains of the Difference Between Pretest and Posttest for Each Training

	Paired Difference			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
All Learners Perceptual %Change – Musical %Change	38.31143%	53.64974	14.33850	7.33499	69.28787	2.672	13	.019
Musician Perceptual %Change – Musical %Change	25.80000%	44.94547	16.98779	-15.76762	67.36762	1.519	6	.180
Nonmusician Perceptual %Change – Musical %Change	50.82286%	62.05610	23.45500	-6.56946	108.21518	2.167	6	.073

Finally two independent-samples t-tests were conducted to compare the perceptual and musical training accuracy gains for musicians versus nonmusicians. Again, surprisingly given the descriptive differences, no significant difference in percentage gains were found after perceptual training for musicians (M=29.78, SD=43.65), versus nonmusicians (M.=61.46, SD=60.53; t(12)=1.123, p=.283), though the magnitude of the differences in the means was moderate (eta squared=.095). No significant difference in gains were found after musical training for musicians (M=3.98, SD=9.46), versus nonmusicians (M.=10.64, SD=16.47; t(12)=.928, p=.372). The magnitude of the differences in the means was moderate (eta squared=.067). Below, Table 4.20 summarizes the output of the analyses. In other words, the difference in gains between musicians and nonmusicians was comparable for both trainings. Note that these analyses of percentage gains present slightly different results from those e.g. Section 4.2.1.2 which analyze raw scores and show a stronger distinction between musicians and nonmusicians.

Table 4.20: Independent Samples Test for Each Training Between Musicians and Nonmusicians for the Word's TPDI Gains of the Difference Between Pretest and Posttest for Each Training

	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Perceptual % Change with Word Gains	1.081	.319	1.123	12	.283	31.68143
Musical % Change with Word Gains	1.204	.294	.928	12	.372	6.65857

4.2.3.2 The Vowel's TPDI Accuracy

Again, to start, when the percent change scores are averaged by training, the averages show that perceptual training still outperforms musical training. However, as opposed to the nearly 40% difference between the two when looking at the word's TPDI accuracy, the difference between the two for the vowel's TPDI accuracy is nearly 20%. This can be seen in Figure 4.28 below.

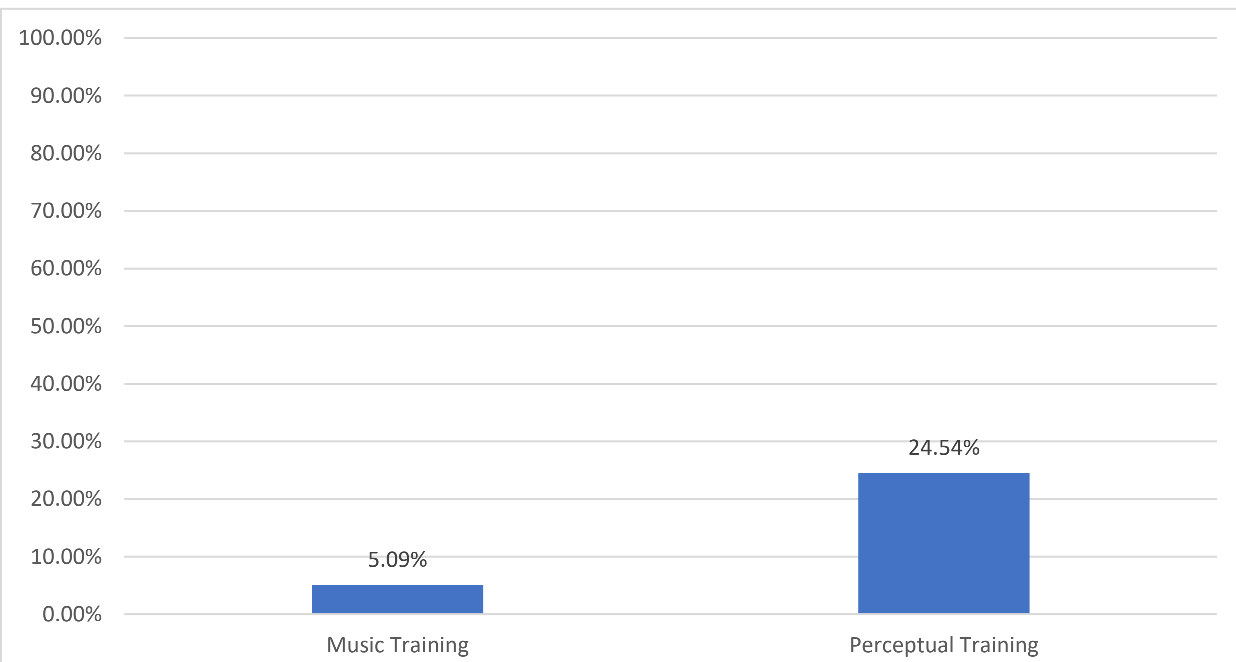


Figure 4.28: Average Increased Percent Change Gains of the Vowel's TPDI for Musical and Perceptual Training

In order to determine whether this difference was significant, a paired-samples t-test was conducted to evaluate each training's effectiveness on the vowel's TPDI accuracy for all participants combined. Percentage gains in accuracy after perceptual training (M=24.54, SD=29.15) were significantly higher than gains in accuracy after musical training (M=5.087, SD=11.59), $t(13)=2.345$, $p=.036$. The eta squared statistic (.297) indicated a large effect size. In Tables 4.21 and 4.22, below, the results of the this test are shown. In other words, perceptual training led to the higher TPDI accuracy gains as compared to the musical training, across participants.

Results before and after both the musical and perceptual training were also examined based on musical background. Figure 4.29 below indicates similar patterns as found with the word's TPDI accuracy in the last section. While musicians' average improvement in the musical training was almost 2.5%, their perceptual training improvement was nearly 14%, revealing about a 10% difference between the two trainings in absolute terms; in relative terms, the gains made in perceptual training was about 5.5 times more than the musical training. Nonmusicians had about an 8% improvement after the musical training while the perceptual training afforded a more than 35% improvement. Between the two trainings, in absolute terms, this shows a more than 25% difference; in relative terms, nonmusicians' gains in perceptual training were more than 4 times more than their gains in musical training. While the differences are not as dramatic when looking at the vowel's TPDI accuracy as opposed to the word's TPDI accuracy, it is clear that the nonmusicians benefitted more in absolute terms from the perceptual training, though musicians benefitted more in relative terms.

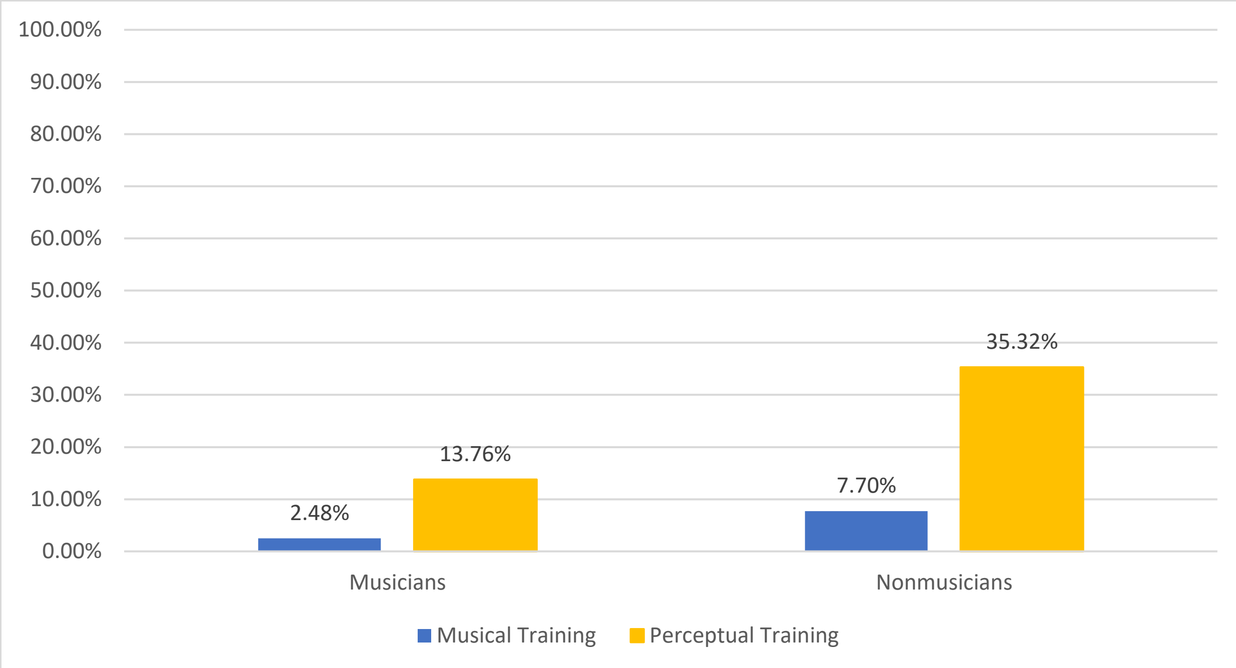


Figure 4.29:Average Improvement by Training of the Vowel's TPDI Accuracy Percent Change Scores Within the Musician and Nonmusician Groups

Two paired-samples t-tests evaluated the differences in accuracy gains from musical training versus perceptual training for musicians and nonmusicians separately. Despite the descriptive difference in accuracy gains for musicians, again there was no statistically significant difference in the effectiveness of perceptual training (M=13.76, SD=23.12) compared to the musical training (M=2.48, SD=7.94), t(6)=1.213, p=.269, with the eta squared statistic (.20) indicating a large effect size. A final paired-samples t-test was conducted to evaluate each training's effectiveness on nonmusicians' TPDI accuracy on the vowels. In opposition to the word's TPDI accuracy, for nonmusicians, the large descriptive difference between the effectiveness of perceptual training (M=35.3171, SD=32.17) compared to the musical training was not significant (M=7.70, SD=14.57), t(6)=2.002, p=.092, though the eta squared statistic (.40) indicated a large effect size. In other words, within the musician and nonmusician groups, both trainings were comparable in increasing the vowel's TPDI accuracy. Again these nonsignificant results are surprising, but should again be interpreted alongside the small sample sizes and large standard deviations, which are noted in Section 4.1. Table 4.21 and Table 4.22 below show the results of each paired samples t-test.

Table 4.21: Descriptive Statistics for The Vowel's TPDI Accuracy of the Difference Between Pretest and Posttest for Each Training

		Mean	N	Std. Deviation	Std. Error Mean
All Learners	Perceptual Training Pre-Post Difference	24.5379	14	29.14535	7.78942
All Learners	Musical Training Pre-Post Difference	5.0871	14	11.59408	3.09865
Musicians	Perceptual Training Pre-Post Difference	13.7586	7	23.11535	8.73678
Musicians	Musical Training Pre-Post Difference	2.4786	7	7.93634	2.99965
Nonmusicians	Perceptual Training Pre-Post Difference	35.3171	7	32.17206	12.15990
Nonmusicians	Musical Training Pre-Post Difference	7.6957	7	14.57347	5.50825

Table 4.22: Paired Samples Test for Each Training on All Learners, Musicians, and Nonmusicians for Gains of the Vowel's TPDI of the Difference Between Pretest and Posttest for Each Training

	Paired Difference			95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
All Learners Perceptual %Change – Musical %Change	19.45071%	31.03912	8.29555	1.52926	37.37217	2.345	13	.036
Musician Perceptual %Change – Musical %Change	11.28000%	24.48309	9.25374	-11.36308	33.92308	1.219	6	.269
Nonmusician Perceptual %Change – Musical %Change	27.62143%	36.49971	13.79559	-6.13517	61.37803	2.002	6	.092

Finally two independent-samples t-tests were conducted to compare the perceptual and musical training accuracy gains for musicians versus nonmusicians. No significant difference in gains were found after perceptual training for musicians (M=13.76, SD=23.12), versus nonmusicians [M.=35.32, SD=32.17; t(12)=1.440, p=.175], though the magnitude of the differences in the means was large (eta squared=.147). No significant difference in gains were found after musical training for musicians (M=2.48, SD=7.94), versus nonmusicians [M.=7.70, SD=14.57; t(12)=.832, p=.422], even while the magnitude of the differences in the means was moderate (eta squared=.055). Table 4.23 summarizes the data below. In other words, the difference in gains between musicians and nonmusicians was comparable for each training. Once again, these analyses of percentage gains present slightly different results from those (e.g. Section 4.2.2.1) which analyze raw scores and show a stronger distinction between musicians and nonmusicians.

Table 4.23: Independent Samples Test for Each Training between Musicians and Nonmusicians for Gains of the Vowel's TPDI of the Difference Between Pretest and Posttest for Each Training

	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Perceptual % Change with Word Gains	1.428	.255	1.440	12	.175	21.55857
Musical % Change with Word Gains	1.248	.286	.832	12	.422	5.21714

4.3 Results to RQ2: The effect of the trainings on generalizability

Research Question 3 sought to test how both trainings, when combined, impacted musicians' versus nonmusicians' capabilities to generalize knowledge to novel words, tones (i.e., M tone), and tone melodies (i.e., HH, MH) . As with the previous sections in this chapter, TPDI accuracy scores will be distinguished by the word's TPDI accuracy (see Section 4.3.1) versus the vowel's TPDI accuracy (see Section 4.3.2). Within each of these subsections, all participants' and musicians' versus nonmusicians' accuracy scores and averages in each test[48] will be compared, descriptively and through statistical tests.

[48] Different trainings were given to different participants at each time period for each test, but when comparing Pretest 1 to Posttest 1, and Posttest 1 to Posttest 2, trainings are collapsed within each test.

4.3.1 The Word's TPDI Accuracy Results

In order to determine if the differences from test to test were significant, a mixed between-within subjects analysis of variance was conducted to assess the impact of the trainings on musicians' and nonmusicians' TPDI accuracy scores across four time periods (Pretest 1, Posttest 1, Posttest 2, and Generalization Test). There was a substantial main effect for time, Wilks' Lambda = .120, F (3, 10) = 24.337, p < .001, partial eta squared = .880. In post-hoc analyses, the results for each test were considered separately, using a Bonferroni adjusted alpha level of .013. Results from the full set of pairwise comparisons are available in Appendix H, and to expedite exposition, only the post-hoc analysis results from immediately consecutive tests are reported here. Descriptively, both groups showed an increase in the word's TPDI accuracy scores across the first three time periods, and then a steep drop from Posttest 2 to the Generalization Test (see Table 4.24 and Figure 4.30). The increase between Pretest 1 and Posttest 1 was nonsignificant (p = .028), the increase between Posttest 1 and Posttest 2 was nonsignificant (p = .274), and the decrease between Posttest 2 and the generalization test was significant (F (3, 10) = 24.337, p < .001, partial eta squared = .880, see Table 4.24).

The main effect comparing the two types of musical backgrounds was significant (F (1, 12)=6.704, p= .024, partial eta squared=.358), suggesting a difference between the two types of participants (i.e., musician vs. nonmusician). For this reason, further post-hoc comparisons using independent samples t-tests indicated that the difference between groups in mean scores at Pretest 1 approached significance (t(12)=-2.154, p=.052, eta squared=.28), at Posttest 1 was nonsignificant (t(12)=-1.957, p=.074, eta squared=.24]), at Posttest 2 was nonsignificant (t(10.389)=-1.452, p=.176, eta squared=.15]), and in the generalization test was significant (t(12)=-3.831, p=.002, eta squared=.55, see Table 4.25). The interaction effect for time and musical background was nonsignificant (F (1, 12)=.710, p=.568). In other words, all the groups lost accuracy from the end of training to the Generalization Test, and musicians descriptively outperformed nonmusicians, a difference which was significant at specific time periods. These results indicate then that learners will not likely improve generalization to novel words with a novel tone (i.e., M tone) or tone melodies (i.e., HH, MH) after trainings on L and H tones.

Table 4.24: Mixed Between-Within Analysis of Variance Statistics for Scores of the Word's TPDI of Pretest 1, Both Posttests, and the Generalization Test

		Descriptive Statistics		
		Mean	Std. Deviation	N
All Learners	Pretest 1 Score of the Word's TPDI	60.7143%	24.56699	14
All Learners	Posttest 1 Score of the Word's TPDI	79.4643%	20.71855	14
All Learners	Posttest 2 Score of the Word's TPDI	84.3750%	17.97267	14
All Learners	Generalization Test Score of the Word's TPDI	41.8350%	10.70782	14
Musicians	Pretest 1 Score of the Word's TPDI	73.2143%	23.58338	7
Musicians	Posttest 1 Score of the Word's TPDI	89.2857%	13.36306	7
Musicians	Posttest 2 Score of the Word's TPDI	91.0714%	13.43248	7
Musicians	Generalization Test Score of the Word's TPDI	49.4886%	9.54426	7
Nonmusicians	Pretest 1 Score of the Word's TPDI	48.2143%	19.66989	7
Nonmusicians	Posttest 1 Score of the Word's TPDI	69.6429%	22.94371	7
Nonmusicians	Posttest 2 Score of the Word's TPDI	77.6786%	20.36680	7
Nonmusicians	Generalization Test Score of the Word's TPDI	34.1814%	9.54426	7

	Tests of Within-Subjects Effects				
Effect	Value	F	Hypothesis df	Error df	Sig.
Time (Wilks' Lambda)	.120	24.337[b]	3.000	10.000	.000
Time * Musician (Wilks' Lambda)	.824	.710[b]	3.00	10.000	.568

	Tests of Between-Subjects Effects				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Musician	4706.778	1	4706.778	6.704	.024
Error	8425.411	12	702.118		

Table 4.25: Post-hoc Analysis Statistics for Scores of the Word's TPDI of Pretest 1, Both Posttests, and the Generalization Test for Musicians versus Nonmusicians

	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Pretest 1 Scores of the Word's TPDI	.308	.589	-2.154	12	.052	-25.00000
Posttest 1 Scores of the Word's TPDI	2.848	.117	-1.957	12	.074	-19.64286
Posttest 2 Scores of the Word's TPDI	5.718	.034	-1.452	10.389	.176	-13.39286
Generalization Test Scores of the Word's TPDI	2.066	.176	-3.831	12	.002	-15.30714

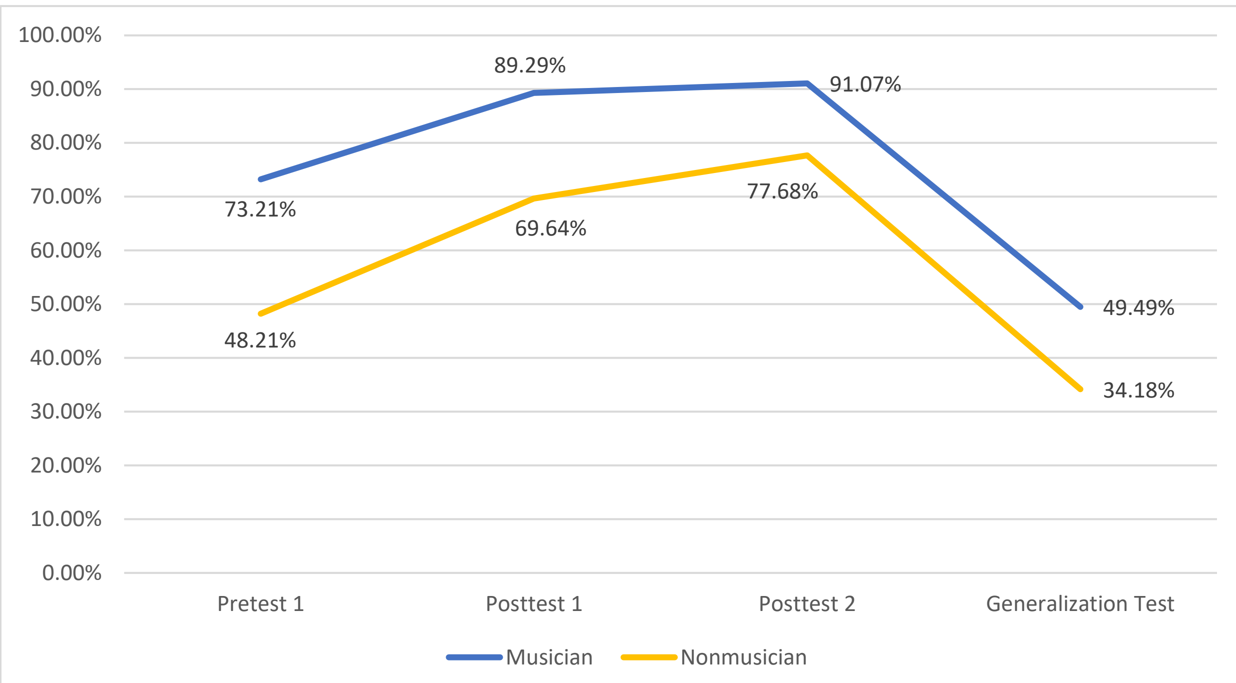


Figure 4.30: Mixed Between-Within Estimated Marginal Mean Plot of Each Test's TPDI Accuracy Scores of the Word between Musicians and Nonmusicians

4.3.2 The Vowel's TPDI Accuracy Results

Finally in order to determine if the differences from test to test were significant, a mixed between-within subjects analysis of variance was conducted to assess the impact of the trainings on musicians' and nonmusicians' TPDI accuracy scores across four time periods (Pretest 1, Posttest 1, Posttest 2, and Generalization Test). There was a substantial main effect for time, Wilks' Lambda=.158, $F(3, 10)=17.736$, $p < .001$, partial eta squared=.842. In post-hoc analyses, the results for each test were considered separately, using a Bonferroni adjusted alpha level of .013. Results from the full set of pairwise comparisons are available in Appendix H, and to expedite exposition, only the post-hoc analysis results from consecutive tests are reported here. Descriptively, both groups showed an increase in the word's TPDI accuracy scores across the first three time periods, and then a steep drop from Posttest 2 to the Generalization Test, which was the same pattern seen in Section 4.3.1 (see Table 4.26 and Figure 4.31). The increase between Pretest 1 and Posttest 1 approached significance ($p=.066$), the increase between Posttest 1 and Posttest 2 was nonsignificant ($p=.280$), and the decrease between Posttest 2 and the generalization test was significant ($F(3, 10)=24.337$, $p < .001$, partial eta squared=.842, see Table 4.20).

Table 4.26: Mixed Between-Within Analysis of Variance Statistics Test for Scores of the Vowel's TPDI of Pretest 1, Both Posttests, and the Generalization Test

		Descriptive Statistics		
		Mean	Std. Deviation	N
All Learners	Pretest 1 Score of the Vowel's TPDI	68.5057%	21.05580	14
All Learners	Posttest 1 Score of the Vowel's TPDI	81.1664%	19.10726	14
All Learners	Posttest 2 Score of the Vowel's TPDI	85.0650%	17.50641	14
All Learners	Generalization Test Score of the Vowel's TPDI	48.7500%	9.34129	14
Musicians	Pretest 1 Score of the Vowel's TPDI	81.1686%	16.47945	7
Musicians	Posttest 1 Score of the Vowel's TPDI	90.9071%	11.73579	7
Musicians	Posttest 2 Score of the Vowel's TPDI	92.2071%	13.04468	7
Musicians	Generalization Test Score of the Vowel's TPDI	52.1429%	11.22020	7
Nonmusicians	Pretest 1 Score of the Vowel's TPDI	55.8429%	17.74449	7
Nonmusicians	Posttest 1 Score of the Vowel's TPDI	71.4257%	20.78230	7
Nonmusicians	Posttest 2 Score of the Vowel's TPDI	77.9229%	19.36079	7
Nonmusicians	Generalization Test Score of the Vowel's TPDI	45.3571%	6.02574	7

	Tests of Within-Subjects Effects				
Effect	Value	F	Hypothesis df	Error df	Sig.
Time (Wilks' Lambda)	.158	17.736[b]	3.000	10.000	.000
Time * Musician (Wilks' Lambda)	.563	2.592[b]	3.000	10.000	.111

	Tests of Between-Subjects Effects				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Musician	3797.323	1	3797.323	6.354	.027
Error	7171.921	12	597.660		

The main effect comparing the two types of musical backgrounds was significant, F (1, 12)=6.354, p=.027, partial eta squared=.346, suggesting a difference between the two types of participants (i.e., musician vs. nonmusician). For this reason, further post-hoc comparisons using independent samples t-tests indicated that the difference between groups in mean scores at Pretest 1 was significant (t(12)=-2.767, p=.017, eta squared=.39), at Posttest 1 approached significance (t(12)=-2.160, p=.052, eta squared=.28), at Posttest 2 was nonsignificant (t(10.517)=-1.619, p=.135, eta squared=.18), and in the generalization test was surprisingly nonsignificant (t(12)=-1.410, p=.184, eta squared=.14) as compared to Section 4.3.1 (see Table 4.27). The interaction effect for time and musical background was nonsignificant (F (1, 12)=2.592, p=.111). These results indicate that these groups will not likely improve generalization after training.

Table 4.27: Post-hoc Analysis Statistics for Scores of the Vowel's TPDI of Pretest 1, Both Posttests, and the Generalization Test for Musicians versus Nonmusicians

	Levene's Test for Equality of Variances		t-test for Equality of Means			
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Pretest 1 Scores of the Vowel's TPDI	.032	.862	-2.767	12	.017	-25.32571
Posttest 1 Scores of the Vowel's TPDI	3.340	.093	-2.160	12	.052	-19.48143
Posttest 2 Scores of the Vowel's TPDI	5.277	.040	-1.619	10.517	.135	-14.28429
Generalization Test Scores of the Vowel's TPDI	1.723	.214	-1.410	12	.184	-6.78571

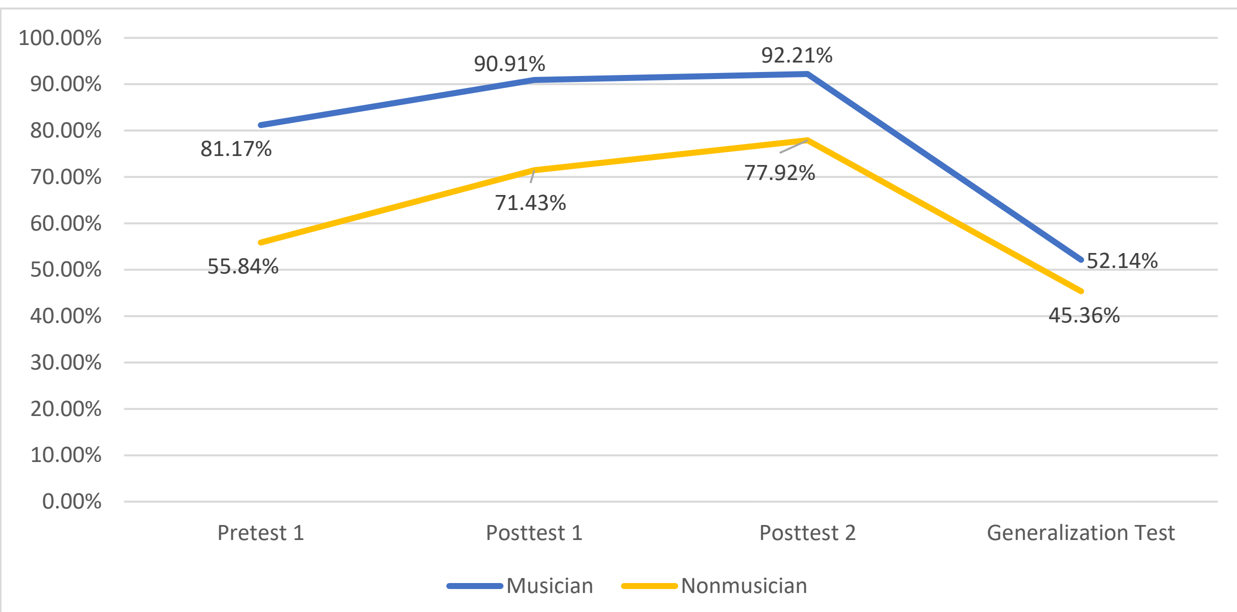


Figure 4.31: Mixed Between-Within Estimated Marginal Mean Plot of Each Tests' Pretest and Posttest Scores of the Vowel's TPDI between Musicians and Nonmusicians

Chapter 5. Discussion

In this chapter, the primary focus is the discussion and interpretation of the results. However, Section 5.1 considers the change in method required by the COVID-19 global health crisis that necessitated the use of a quickly-adapted online modality as opposed to a face-to-face modality. This change is discussed here as a way of interpreting potential limitations of the findings due to this change in modality. In the following sections, each training is examined and compared; the results are discussed against the literature surveyed in Chapter 2. Then, pedagogical implications are discussed, and relevant limitations are considered. Section 5.2 examines the effects of the musical training. Section 5.3 examines the effects of the perceptual training while 5.4 compares and contrasts the musical training and perceptual training. Section 5.5 examines the effects of the combined trainings on generalizability. Section 5.6 is a summary of findings.

5.1 The Effect of an Online Modality on the Trainings

The implications of the use of an online modality would be better understood if research existed that examined the differing effects of a face-to-face (f2f) training vs a synchronous computer-mediated communication (SCMC) training on perception, discrimination, and identification (PDI) accuracy. However, a study of this kind has yet to be explored. One study in the literature did examine the effects of course delivery modality methods on students' abilities to achieve the American Council on the Teaching of Foreign Languages (ACTFL) benchmarks in overall proficiency, pronunciation, fluency, sentence formation, and vocabulary (Moneypenny & Aldrich 2018). Specifically, this study examined L2 Spanish learners taking university Spanish classes in either a primarily asynchronous online modality or a f2f modality. They found that course modality did not significantly predict test scores. Further, they explain that this indicates that oral proficiency was not impacted by the kind of course students were enrolled in. Of course, their study has limited application to the potential effects that the SCMC modality of the present study may have had. Moneypenny and Aldrich's study is an indication, however, that had the research been conducted f2f, scores may not have been much different.

Alternatively, another study does look at the differences between SCMC (as opposed to the asynchronous modality of Moneypenny and Aldrich's (2018) study) vs f2f settings. Kim's (2014) research investigated these settings on their effectiveness in collaborative communicative interaction and learning strategies. The only difference related to pronunciation showed that learners used avoidance strategies for linguistic purposes (including difficulties in pronunciation and production) more often with the SCMC method than with the f2f method. This has implications for online teaching of pronunciation, and suggests that learners may avoid difficult pronunciation more frequently through an online modality than with a f2f modality. However, this avoidance in pronunciation and production may simply be due to poor audio quality of the SCMC setting (Guan 2014). This could indicate that PDI skills are more challenging to apply in a SCMC setting, where audio quality can be poorer, but this would be a question for future research to explore.

Another area of the literature to examine to determine how the current study was impacted by being delivered through a computer is computer-assisted pronunciation training (CAPT). Limited research has explored the use of CAPT as an effective tool for pronunciation (see Levis 2007; Luo 2016; Tsai 2019). Only one article has examined the effect of CAPT in increasing accuracy in the pronunciation of the target language (Luo 2016). Luo's (2016) article found that when CAPT was assigned to one of two groups of students where both groups were currently enrolled in a primary f2f class setting, the group with the additional CAPT training lead to greater pronunciation production accuracy. While this methodology has limitations for the current study because of its focus on production, on segmental phonological learning, and on English as the target language, Luo's study does reveal that online learning can be beneficial to second language pronunciation learning.

A similar study, Tsai (2019), found that L2 learners paid more attention to suprasegmentals (i.e., pitch as applied to intonation by L1 tonal language speakers learning English as an L2) when using CAPT and found a positive impact of the CAPT training on raising awareness of prosodic production in the target language.[49] The positive impacts shed light on the potential effect of the use of online learning for the current study. However, a limitation for extension of findings to the current study is posed by Tsai's methodology. Similar to Luo (2016), Tsai's (2019) participants were all in f2f classes while they took the CAPT training. These two articles do indicate that the online modality employed in the current study may have had only a limited impact on the learning of TPDI accuracy than would have been observed in the originally planned f2f training. However, Tsai's focus on the added benefit of CAPT training as opposed to a direct comparison of the effect of CAPT versus f2f training on the teaching of pronunciation ultimately makes it difficult to determine the independent contributions of each. Future research could replicate the present study through a f2f modality to determine if this change in modality played a role in the effect of each training.

---

[49] There were also negative impacts noted by the learners and the author, but these impacts were isolated to production training scoring procedures of the software in particular, not the training methodology.

In all, limited research in the area of SCMC vs f2f mediated instruction and their effects on L2 pronunciation training makes it difficult to determine how the sudden adaptation of the study to a synchronous online modality impacted the results. The studies discussed above positively indicate that there may have been no or limited effects due to the modality of the trainings. However, as Guan (2014) and Kim's (2014) research indicate, it is entirely possible that poor audio quality due to the nature of the SCMC setting could have led to avoidance strategies or further challenges for learners in their developing PDI skills in the target language. These mixed results leave questions that create a limitation to the current study, but also provide an opportunity for future research.

5.2 The Effect of Musical Training

The first research question asked: In tonal learning, is the use of musical training more or less beneficial than perceptual training, and is there a difference between learners with and without musical training backgrounds? In Chapter 4, results showed that the musical training was beneficial for 7 of the 14 participants that were included in the analysis for the word's TPDI accuracy and detrimental for 3 participants, and 4 showed no gains (see Table 4.1 and 4.2). Of those seven who made gains, four were nonmusicians while three were musicians, indicating little difference between musicians and nonmusicians. In fact, with the musician and nonmusician groups, the average increase within each group was about the same, a nearly 3% increase for musicians and over 4% for nonmusicians. Between each group, there was about a 20% difference in scores for both the pretests and the posttests relevant to each participant's musical training order (see Figure 4.15). Generalizing, this increase in scores from pretest to posttest between musicians and nonmusicians was not statistically significant across participants nor was it statistically significant within either the musician and nonmusician groups. While the descriptive results show a gain with a large effect size between musician and nonmusician scores, this was likely not statistically significant because of the number of participants.

For the vowel's TPDI accuracy, the results were very similar. Five of the 14 participants showed improvement from pretest to posttest while 3, as with the word's TPDI accuracy, showed a decrease in scores, and 6 showed no gains (see Table 4.3 and 4.4). Again, within the two groups (i.e., musicians and nonmusicians), both showed only about a 2-3% gain in accuracy across participants in each group. The musician group gained about 2% accuracy from pretest to posttest while the nonmusician group gained a little more than 3% accuracy, which was similar to the results for the word's TPDI accuracy. As with the word's TPDI accuracy, between each group, there was about a 20% difference in scores for both the pretests and the posttests relevant to each participant's musical training order (see Figure 4.18). Generalizing, the accuracy gains made across participants was not significant. Within each group, the musicians' scores from pretest to posttest were not significant and neither were the nonmusician's scores from pretest to posttest. However, the difference in scores between the musicians and nonmusicians approached statistical significance, suggesting again the influence of sample size.

These results have several implications to the literature surrounding the benefits of musicianship in the second language learning of phonology. As was described in Chapter 2, a musical training background tended to result in greater TPDI and tonal production accuracy (Chobert & Besson 2013; Kirkham et al. 2011; Li & DeKeyser 2017; Pei et al. 2016; Perfors & Ong 2012; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007; Zhao & Kuhl 2015). However, the various studies only looked at the differences in baseline abilities between musicians and nonmusicians. In actually testing pedagogical techniques in the L2 classroom, though without focusing on TPDI accuracy, Shi's (2018) dissertation investigated the application of musical training for developing tonal production skills in Chinese. She found that her musical training procedure increased tonal production accuracy (though it is uncertain how the use of musical scales specifically and independently contributed to the results over and above the additional use of hand gestures). In the studies on TPDI accuracy, however, musicians were ubiquitously better at pitch perception, discrimination, and identification of tones even while musicians and nonmusicians were typically equal in their categorization of relative pitch changes to tonal categories of the target language (Wayland, Herrera & Kaan 2010; Zhao & Kuhl 2015). Indeed, despite the non-statistically significant impact of the musical training method applied in this study, the results corroborate previous findings that musicians are greater at TPDI than nonmusicians as evident from the finding that the musicians outperformed the nonmusicians.

There are further implications of this study in regards to second language acquisition. In Chapter 2, a musical training background and aptitude were explored. Particularly, recall that Talamini et al.'s (2018) study found that musicians outperformed nonmusicians (all between the ages of 11-15) in a dictation test that targeted segmental phonology, but that the scores of the musical aptitude test, PROMS, had no significant correlation with the results of the dictation test. This indicates that higher musical aptitude has little to do with musicians' superior phonological PDI accuracy. Pei et al. (2016), however, found that their musically trained participants made more gains in production of suprasegmentals and they found that training in music seemed to help increase musical aptitude for their adult participants as well, which could potentially increase the suprasegmental production abilities of participants who start musical training at any age. In the current study, musical training did not narrow the gap between musicians and nonmusicians for TPDI accuracy. The present study's findings more closely align with Talamini et al.'s study, which found that musical aptitude had little to do with their participants' scores on the phonological dictation test. This study's findings contribute to the literature on musical training and aptitude by indicating that whether musical training increases musical aptitude or not, musical training to teach tones does not allow nonmusicians to narrow the gap between themselves and nonmusicians in TPDI accuracy.

Further, in the studies from Chapter 2 on how a musical training background affects phonological PDI accuracy, the age when a musician started lessons with their instrument(s) was crucial to some definitions of a "musician" (Perfors and Ong 2012; Wong & Perrachione 2007; Zhao & Kuhl 2015). For Wayland, Herrera, and Kaan (2010), however, this was not a factor in their definition. As explained in Chapter 3, this lack of an age requirement did not change that they found similar results in that musicians outperformed nonmusicians in TPDI accuracy. Similarly, the current study's musicians outperformed nonmusicians, revealing the same trends found in previous studies. This indicates that the age at which a musician begins privately studying and practicing their instrument does not affect the positive benefits to L2 tonal learning. This is an important insight of the current study, especially considering that Granena and Long (2012) found evidence that phonology is likely to be the first aspect of language to close in the critical period for language acquisition. Given this evidence from Granena and Long and considering the prior research reviewed in Chapter 2, while L2 phonology may be more challenging to acquire with age, a musical training background can aid learners in their L2 phonological PDI accuracy. Further, the present study indicates that regardless of age, training in music can benefit learners of any age in L2 phonological PDI accuracy.

Although these results show that musical training had little if any robust effect on TPDI accuracy for participants, this training might have been more effective under different circumstances. As previously explained, the fact that these trainings were mediated through a synchronous online modality as opposed to the f2f modality initially proposed (and because this change had to be applied rapidly) is one reason why future research should explore musical training pedagogy further. Despite prior research findings that the difference in modalities (i.e., online vs f2f) has little if any effect on pronunciation instruction (Moneypenny & Aldrich 2018), the training undertaken in the current study unexpectedly required the use of digital instruments, and collaboration with these instruments would have been much easier for participants to manipulate had they been in a f2f setting. For this reason, the results of the musical training may have been impacted by the online setting. Additional training sessions may also impact the results. Talamini et al.'s (2018) study included participants coded as musicians who had been taking music lessons for only two months. Their study showed that even with only two months of a musical training background, the musician participants outperformed nonmusicians in PDI accuracy of an English dictation test. While these participants were between the ages of 11-15 and while the target features were not tonal or even prosodic, a limited musical training background was still beneficial. Thus, future research should examine whether application of the musical training procedure over a longer time period, e.g. two months, would prompt significant gains in TPDI accuracy.

Should teachers, then, consider incorporating musical training into the classroom for introductory learners? Three out of 14 participants' TPDI scores decreased after the musical training. At least in one case, with Participant 13, this could have been due to internal factors. This participant took the musical training as his last training. He was in a rush to finish Posttest 2 in order to arrive at a virtual class on time. In regards to the individual circumstances of the other participants, it is not as clear what may have led to the decrease. It is worth noting, though, that 66% of the participants (i.e., Participants 13 and 14) whose scores decreased after their musical training posttest expressed that the musical training was their favorite training (see Table 4.5). In all, though, it is difficult to determine if musical training would be beneficial to incorporate in the classroom since only 50% of the participants increased their accuracy of the word's TPDI and only 36% of participants increased their accuracy of the vowel's TPDI after taking the musical training and neither of these increases in TPDI accuracy were statistically significant. Perhaps, until further research can resolve the questions raised above, teachers could cautiously implement musical training into the L2 language classroom as long as they also incorporate perceptual training and monitor the effects of each on their specific group of learners. As will be discussed and as can be garnered from the results in Chapter 4, the perceptual training was far more effective and no matter the order of musical training and perceptual trainings, participants still made gains, showing that musical training (at least when) in coordination with perceptual training can be beneficial to learners.

5.3 The Effect of Perceptual Training

In continuing to discuss findings related to the first research question, results showed that the perceptual training was beneficial for 10 of the 14 participants for the word's TPDI accuracy and detrimental for none while 4 participants showed no gains (see Table 4.1 and 4.2). For 64% of the participants, then, the perceptual training was beneficial and increased their accuracy. Of those 10 who made gains, 6 were nonmusicians while 4 were musicians, indicating the trainings impacted both groups almost equally. However, with the musicians, the average accuracy increase of the word's TPDI was over 15% while with the nonmusicians, the average accuracy increase of the word's TPDI was 25% (see Figure 4.21). Comparing musicians and nonmusicians, there was about a 25% difference in the perceptual training-relevant pretest scores and a nearly 15% difference in the posttest scores. In generalizing, the increase in scores from pretest to posttest was statistically significant across all participants. However, the differences between the musician and nonmusician groups' pretests and posttest scores were not significant, which is likely due to the limited number of participants and higher standard deviations, considering the large effect size.

For the vowel's TPDI accuracy, the results were very similar. In comparison to the word's TPDI accuracy, 10 of the 14 participants showed improvement from pretest to posttest for the vowel's TPDI accuracy. Unlike with the word's TPDI accuracy, however, one participant (Participant 4) showed a decrease in scores, and three showed no gains, which were the same participants who made no gains in the word's TPDI accuracy. Of the 10 who made gains, 6 were nonmusicians while 4 were musicians, indicating little difference again (see Table 4.3 and 4.4). The musicians' average accuracy increase of the vowel's TPDI was less than 10% while the nonmusicians' average accuracy increase of the vowel's TPDI was nearly 20% (see Figure 4.24). Between each group, there was about a 25% difference in their perceptual training-relevant pretest scores and a nearly 15% difference in the posttest scores. In generalizing these results, it was found that the increase in scores from pretest to posttest was statistically significant across participants, and the differences between the groups' (i.e., musicians vs. nonmusician) pretests and posttest scores were also significant.

By returning to Figures 4.6 and 4.8, it can be seen that the first training group (Participants 1-5) seemed to achieve much less progress with the perceptual training than the following groups. It is important to note that perhaps the lower perceptual training scores in this group may be due to the technical audio challenges experienced. Since this group was the first, many technical issues arose despite abundant preparation and tests of the system prior to this group's training day. For this reason, feedback given in the first task of the perceptual training, which was supposed to be immediate feedback, was not given immediately, but instead after each set of 10 words the participants listened to. As this group was the only group who did not receive immediate feedback, perhaps the perceptual training was not optimally effective. Feedback has been shown to be an important and effective tool in L2 learning and instruction (Li 2010; Saito & Lyster 2012; Spada & Lightbown 1993). Future research may want to explore the effect of immediate versus non-immediate feedback for tonal training to determine if this was the cause of the lower impact of perceptual training on the initial participants.

These results have several implications to the literature surrounding the benefits of perceptual training in the learning of L2 tones. Previous literature has tested perceptual training in a laboratory context, as explained in Chapter 2 (Antoniou & Wong 2016; Godfroid, Lin & Ryu 2017; Lu, Wayland & Kaan 2015; Perrachione et al. 2011; Wang 2013; Wang et al. 1999; Wang, Jongman & Sereno 2003; Wayland & Li 2008). Only Li & Dekeyser's (2017) study tested perceptual training in a classroom-like context, with immediate feedback given by one of the authors on learner productions. In the present study, the presentation-practice-production (PPP) pedagogical method was applied to both trainings. This application did not alter the basic elements of perceptual training, found to be effective in prior literature, but it did alter their organization and presentation. Even with such an alteration, as can be gleaned from the results, perceptual training is still beneficial to learners. The results in Chapter 4 clearly show that the word's and vowel's TPDI increase from pretest to posttest were significant, revealing that perceptual training delivered in a PPP format in the teaching of tones was successful.

This has further pedagogical implications. Specifically, perceptual training in its laboratory context does not entirely fit one of the most common approaches to language teaching currently established: Communicative Language Teaching (CLT). CLT is an approach characterized by its grounding in communication, with (some) freedom of exploration in language learning (Richards & Rodgers 2001). Its widespread use comes from its practical theoretical underpinnings that learners need to accomplish simple communicative tasks in order to mediate their environment (Littlewood 2011). For this reason, it is important to be able to apply CLT principles of communicative competency to classroom methods and techniques, which includes trainings like the musical training and perceptual training done in the current study. Previous studies have not structured their perceptual training to adhere to CLT principles.[50] In the current thesis, though, the perceptual training incorporated the use of student-student discussion, which reflects several of the characteristics of CLT (e.g. student-student interaction, goal oriented communication). While this discussion was not conducted in the target language due to the scope of this current research, the model for this training procedure certainly could utilize the target language and incorporate further tenets of CLT (grammatical, discourse, and strategic competence). Importantly, the current research shows that an application of perceptual training that adheres to CLT principles can be effective and that teachers can adapt the technique to their own CLT-based classrooms.

More research needs to be conducted on this training method. While this training was adapted to a PPP structure and incorporated student-student interaction and collaboration in the production part of the training, the interaction was teacher-student and/or student-isolated in the presentation and practice portions of the PPP format. Perhaps future research could examine how

[50] CLT principles primarily rely on the focus on developing communicative competence, which entail several sub-competencies: grammatical, sociolinguistic, and discourse competencies (Savignon 1983).

incorporating multiple speakers and stimuli focused on a targeted tonal contrast could be applied in dialogues or other communicative materials. Furthermore, only about one third of participants preferred this training to the musical training. However, while the musician group almost universally preferred the musical training (i.e., only Participant 4 preferred perceptual training), the nonmusician group were nearly split. Five of the nonmusicians preferred musical training, but four preferred the perceptual training. This is important information to a teacher's choice of methods and techniques as a way of motivating students, but in a classroom context, teacher's may not know who, in their class, is a musician or nonmusician. However, since most of the participants (regardless of musical training background) preferred musical training, citing that the musical training was more "interactive," "easier to follow," and "enjoy[able]"(see Table 4.5), musical training can still be useful as a motivational tool. Further, such views indicate that the perceptual training was not as engaging which can result in learners' lack of interest. Therefore, additional research should explore methods of adapting perceptual training to be more communicative and more interactive while retaining its effectiveness.

5.4 The Effect of Musical Training vs. Perceptual Training

In Chapter 4, results showed that perceptual training far outperformed musical training. With the word's TPDI accuracy, perceptual training outperformed musical training by nearly 40% (see Figure 4.26), which was statistically significant. In looking at individual score differences between trainings, only three participants' musical training percent change scores were higher than the perceptual training percent change scores, and each of these participants were from the first training group (which was discussed above). Looking deeper, within the musician and nonmusician groups, on average, the musicians' improvement contrasting musical versus perceptual training was about 25%, generally, and about 7.5 times, relatively (with perceptual training's percent change scores being higher). For nonmusicians, though, the improvement between trainings was about 50%, generally, and about 6 times, relatively (see Figure 4.27). Despite the lack of statistically significant differences in percentage change scores either between trainings (but within the musician versus nonmusician groups) or between groups for each training, descriptively the higher perceptual training percent change scores compared to the musical training percent change scores for the group of nonmusicians were striking.

With the vowel TPDI accuracy, perceptual training generally outperformed musical training by nearly 20% (see Figure 4.28), which was statistically significant. Within the musician and nonmusician groups, on average, the percentage change difference for musicians between musical and perceptual training was about 10%, generally, and about 5 times, relatively (with perceptual training percent change scores being higher). For nonmusicians, though, the difference between trainings was a little more than 25%, generally, and 4 times, relatively (see Figure 4.29). Again, the lack of statistically significant differences in percentage change scores either between trainings (but within the musician versus nonmusician groups) or between groups for each training is striking. However, the descriptively higher perceptual training percent change scores compared to the musical training percent change scores for the group of nonmusicians together with the significantly higher percentage change scores for perceptual training overall indicated the impact of that technique.

It is curious that while the difference between trainings was statistically significant across all participants with both the word's and vowel's TPDI accuracy, the trainings were not statistically different within and between each group. Perhaps, this is due to the high standard deviation for these scores or the limited number of participants. Within the musician and nonmusician groups, the standard deviation was often between about 10-30%, with the musicians often having a higher standard deviation than that of nonmusicians. This wide standard deviation comes from the wide ranging differences in percentage change scores. In the musician group, a percent change score could be as low as -6% and as high as 114%. With the nonmusicians, a percent change score was as low as -10% and as high as 160%. Both groups had individuals with both high and low percent change scores for both trainings. When together, the standard deviation evened out with additional participants. Perhaps, for this reason, the trainings showed no statistically significant differences within each group while there were statistical differences across all participants. Additionally, in cutting the sample size in half to examine the results within each group, statistical power is reduced.

These results have several implications to the literature surrounding the effect of perceptual training on musicians vs. nonmusicians. In Chapter 2, research was discussed which examined the TPDI and tonal production effect of perceptual training on musicians vs. nonmusicians (Li & Dekeyser 2017; Wayland, Herrera & Kaan 2010; Wong & Perrachione 2007; Zhao & Kuhl 2015). As may be recalled, these studies showed that both musicians and nonmusicians improved their scores at about the same rate. In other words, even with the perceptual training, musicians and nonmusicians retained the same gap in accuracy scores from pretest to posttest in these studies. In the present study, results were different. In the mixed between-within subjects analysis of variance of the word’s TPDI accuracy scores before and after perceptual training (see Table 4.14), the difference between musicians and nonmusicians approached statistical significance, while in the analysis of the vowel TPDI accuracy scores, a main effect of group was found (see Table 4.17). This indicates that there was a greater impact on TPDI accuracy scores for nonmusicians than musicians after taking the perceptual training; in other words, perceptual training seemed to help close the gap between musicians and nonmusicians. Future research should explore this possibility further by examining if more training sessions would close the gap further. Nevertheless, this is a clear indication that incorporating perceptual training procedures into the classroom would be beneficial for all learners, regardless of musicianship background. Moreover, this result indicates that perceptual training is beneficial in an online modality for tonal training, which is important to the growing field of CAPT and SCMC language teaching.

Turning to musical training, the difference between musicians and nonmusicians' TPDI accuracy scores of the word and vowel from the musical training's pretests to posttests were not significant, indicating that the musical training had little if any effect in raising nonmusicians' abilities to the level of musicians (see Tables 4.8 and 4.11). Of course, as mentioned in section 5.2, this could certainly have been due to various external factors outside the control of the present study. For this reason, more research should certainly be conducted on the efficacy of this training, particularly over more sessions with more time and in a f2f setting as opposed to a SCMS setting. However, the results imply, with the current state of the literature and this current study's contribution, that – again – instructors should think about adding musical training techniques with caution and perhaps always with the addition of perceptual training if they do decide to incorporate this training technique. In fact, given that participants as a whole did not prefer the perceptual training, teachers may consider ordering the less engaging perceptual training first to possibly increase the overall potential for learning, but then follow with the musical training in order to sustain motivation. Despite the fact that nonmusicians were nearly equally split in regards to their preferences, the majority of nonmusicians still favored musical training. Given that most musicians and more than half of nonmusicians prefer the musical training, this training can still act as a motivational tool to retain engagement in tonal learning.

Overall, then, in discussing the results of each training and in comparing the two trainings' effectiveness to each other, it seems clear that perceptual training is far superior. Even while perceptual training does not seem to significantly affect the word's and vowel's TPDI accuracy scores in some instances, overall we can conclude that it is a much more effective training than musical training in instruction of L2 tones.

5.5 The Effect of the Trainings on New Word and Tone Generalizability

The second research question asked: How do the combined musical and perceptual trainings affect the ability to generalize the training to novel words, tones (i.e., M tone), and tone melodies (i.e., HH, MH) for musicians versus nonmusicians? In examining the effect of the combined trainings on new word and tone generalizability, results from Chapter 4 revealed that participants' were largely unable to generalize their training to a new tone (M tone), new tone melodies (MH and HH), and new words. In the mixed between-within analysis of variance tests for the word's and vowel's TPDI accuracy for the generalization test (see Figures 4.30 and 4.31), there is a visibly steep decline from Posttest 2 to the generalization test. Post hoc analyses revealed that this decline was statistically significant for the word's and vowel's TPDI accuracy across participants. Even more strikingly, most participants' generalization test scores were even lower that their Pretest 1 scores. Indeed, for both the word's and vowel's TPDI accuracy, there was a statistically significant difference between Pretest 1 and the generalization test (see Appendix H), with Pretest 1 having a higher mean score than the generalization test. Although scores in the generalization test were not zero and participants identified some new words, tones and melodies correctly, the above results, and especially the comparison of Pretest 1 to the generalization test, indicate that participants did not appear to develop the ability to generalize the knowledge gained from training to new words, tones, and melodies.

One possible reason to account for the poor performance in the generalization test is that the latter was more difficult than the training tests. First, an entirely new third tone was added, so participants had to identify one of three tones as opposed to one of two tones. In addition, they had to identify two new tonal melodies. A second possibility relates to the relative distinction between the tones. The difference in relative pitch categories between the H and the L tone is quite large (see pitch traces in Chapter 1), which may have been easier to identify by participants even before training began, while differences between H and M and M and L are much smaller. Thus, the generalization test may have been too challenging. In the future, research should consider using the M and the L tone in training and introducing the H tone as a generalizable tone. This would make the post-training tests more difficult than the generalization test and potentially reveal some ability to generalize knowledge and skills.

Further, in looking at participants' Pretest 1 and generalization TPDI accuracy scores, individually, of the word and vowels, only two participants (Participants 1 and 5) of the 14 subjected to analysis had lower Pretest 1 scores than the generalization test scores. Importantly, though, these participants had the lowest Pretest 1 scores among all the participants.[51] Their higher generalization test scores, then, are likely due to the low baseline set by their low Pretest 1 scores. The previous studies that employed generalization tests yielded different findings (Antoniou & Wong 2016; Wang et al 1999; Wang 2013). Their participants universally did

---

[51] Participant 11 shared their same TPDI accuracy score of the word on Pretest 1, but she did much better on the vowel's TPDI accuracy on this test. Oddly, though, her generalization test word and syllable PDI scores were much lower. Perhaps this lower score was due to her inability to understand the tonal categories, which has been shown to be difficult for other L1 stress language learners of L2 tonal languages (Wayland, Herrera & Kaan 2010; Zhao & Kuhl 2015).

better on their generalization tests as compared to the first pretest. However, these studies did not test whether their participants could generalize their skills to novel tones or tonal melodies and only tested whether the participants could generalize to novel speakers and novel words (with the same tones used). This further suggests that generalization to the M tone in the current study may have been too challenging. Future research should further investigate the extent of ability to generalize TPDI skills and knowledge.

Between musicians and nonmusicians, a post hoc independent samples analysis revealed no significant difference in the word's and vowel's TPDI accuracy scores in Posttest 2. However, the post hoc test did reveal a significant difference between musicians and nonmusicians in the word's TPDI accuracy scores for the generalization test, with musicians' mean scores being higher than nonmusicians, though the statistical difference did not extend to the vowel TPDI accuracy. For the word's TPDI accuracy, this shows that the gap between musicians and nonmusicians by Posttest 2 for the word's TPDI accuracy was not as large as it was in the generalization test. This could indicate that even though nonmusicians raised their accuracy of the word's TPDI much closer to musicians with the trainings, these trainings were not useful in generalizing to novel words, tones, and tonal melodies. Alternatively, we may interpret the results such that the stark contrast between the H and L tone did not provide sufficient skills to differentiate between two tones with less contrast. On the other hand, the lack of statistical difference in the vowel's TPDI accuracy scores between musicians and nonmusicians could reveal, alternatively, that nonmusicians were better at identifying individual tones, but worse at identifying both tones of a disyllabic word correctly. In all, these results may indicate that nonmusicians were able to apply some attained skills to identify new words and tones but were not able to retain the closing gap between themselves and the musicians.

These results have several implications to the literature surrounding the effect of perceptual training on generalizability. Previous studies that tested perceptual training's effects on generalizability to new words and speakers showed that perceptual training was effective for learners to more easily generalize to novel vocabulary (Antoniou & Wong 2016; Wang et al 1999; Wang 2013). The current study adds to this literature by challenging whether perceptual training really enhances generalization skills. It is worth noting that in Wang et al. (1999) and Wang (2013), score comparisons were between two different sets of participants. They compared a control group, who experienced no training, to a treatment group, who participated in a perceptual training procedure. The treatment group had significantly higher scores compared to the controls. It is possible, then, that including a control group who experienced no training in the current study might have yielded a lower generalization scores than those from the two training groups. Thus, future research should consider adding a control group to a replication of this present study.

In all, these results indicate that perceptual and musical trainings as implemented in the current study may not aid leaners enough to generalize their knowledge to new words, tones, and tone melodies. In terms of pedagogical implications, instructors of tonal languages should include all tones in the language for their materials and trainings. If teachers want to introduce fewer tones than what the language actually displays, it might be advisable to use tones with a smaller gap between relative pitch ranges at first. This may help learners to better generalize their knowledge to novel vocabulary and tones with a larger gap in relative pitch. Of course, future study is needed to examine this hypothesis. In the meantime, instructors should not rely on tonal training of limited tones in the language with the expectation of generalized skills and should provide perceptual training on all tones in the language at some point.

5. 6 Summary of Findings

In all, results revealed that the perceptual training far outperformed the musical training in the pretests and posttests across participants. Within each group, there were similar results. Both musicians and nonmusicians progressed their TPDI accuracy more with the perceptual training than the musical training. Between each group, the musicians always outperformed the nonmusicians on average, regardless of training. However, with perceptual training, specifically, nonmusicians were able to narrow the gap between their accuracy scores and the musicians' scores, and slightly less than half of the nonmusician group preferred the perceptual training. On the other hand, more than two thirds of all participants preferred the musical training to the perceptual training, citing that the musical training was more interactive and fun. In terms of generalizability, combined, both trainings appeared to have little if any effect on the ability to generalize new words and tones.

In terms of pedagogical implications, the above research can offer some insights. As explained above, the current research shows that perceptual training is far more effective for learners to increase TPDI accuracy as compared to musical training. However, musical training needs to be further researched due to the sudden change in modalities that the COVID-19 global health crisis required. Despite this, or perhaps in light of this change, perceptual training in this format was shown to be effective through a SCMC modality, which is informative to this growing field, especially as COVID-19 continues to impact instruction of all kinds. However, given the lower scores in perceptual training of the first training group, it is possible that immediate feedback would be necessary to keep in this perceptual training procedure or it would not be as effective. As noted, though, this needs further research to confirm. In the meantime, instructors can include perceptual training procedures and musical training (as long as perceptual training is also included, potentially first), to increase TPDI accuracy in the CLT-informed, L2 classroom.

Chapter 6. Conclusion

This study examined and compared the effects of two different training techniques (i.e., musical and perceptual training) between musicians and nonmusicians on the learning of tonal perception, discrimination, and identification (TPDI) accuracy. Musical training involved the use of an instrument (i.e., a digital piano or voice) while perceptual training involved listening to a targeted set of tonal contrasts to teach Yoruba tones. A within-participants intervention research design was used, where each participant experienced both kinds of training, implemented in a counterbalanced order across training groups. The onset of a shelter-in-place mandate due to COVID-19 caused key changes to the planned methodology, principally an abrupt transition to online training and the reduction of training length from two days to one day. Extensive analyses of learner TPDI performance included in each training type at both the level of the word and vowel, as well as the ability to generalize to new tones and new tonal melodies, were conducted by individual participants, including an analysis of the effects of training order, as well as by group, including by level of musical training background. Participant views of the training methods were also analyzed. This chapter provides a brief summary of the results, weaknesses of the study, suggestions for future research, and implications for pedagogy.

6.1 Summary of Results

Results of the study revealed considerable individual differences, which is expected in any educational context, including language learning. The counterbalanced training methodology was applied to help reduce the statistical effect of this variation. However, the data were also analyzed by the training order to determine if this counterbalanced training affected the results. It was found that musicians' scores patterned together and nonmusicians' scores patterned together regardless of training order, revealing little effect on the scores due to the counterbalanced training methodology.

Despite individual differences, the perceptual training was found to be almost universally descriptively superior to the musical training, and at times also inferentially superior across all participants, and also within each group (i.e., musician vs. nonmusician). Between each group and in line with prior research, the musicians descriptively outperformed the nonmusicians almost universally at the start and end of the study with large effect sizes, regardless of training, although between-group differences did not always reach statistical significance likely due to sample size and standard deviations. Strikingly, perceptual training enabled nonmusicians to narrow the performance gap to some extent between themselves and musicians. In terms of training order, a slight advantage was found when perceptual training, the more effective training type, was experienced first. Regarding the ability of participants to generalize their combined trainings, analyses revealed little if any effect on the ability to perceive, discriminate and identify new tones and tone melodies. All above patterns were similar across word and vowel TPDI accuracy. Finally, both trainings were successfully incorporated into a classroom setting and learners overall exhibited learning, but in the post-training survey of attitudes, more than two thirds of all participants expressed a preference for the musical training compared to the perceptual training, citing that the musical training was more interactive. However, musical training was only favored by slightly more than half of the participants in the nonmusician group while it was favored by all but one participant of the musician group.

6.2 Summary of Weaknesses

Limitations to the current study included the number of participants, the rapid change to an online modality, and technical issues due to the online transition. A total of 16 participants, with eight musicians and eight nonmusicians, was the maximum feasible number of participants for the scope of this thesis. Small group sizes likely impacted statistical power, yielding evidence of trends that in some cases did not reach statistical significance. However, the large effect sizes observed indicate large descriptive differences in the data collected. Combined with the results that were found to be statistically significant and approaching significance, the study provides some key findings as well as areas important for future research to explore.

The online delivery of the training, caused by the COVID-19 related shelter-in-place order was another weakness. Little research has been conducted on the effects of an online modality on the development of pronunciation in general, with none specifically on TPDI accuracy, making it difficult to determine how the online modality may have impacted results. Although the existing limited research suggests that the online modality may not have had much of an effect on the results, the use of musical instruments through a digital medium without the advantage of being face-to-face may have altered the results in particular for the musical training. Additionally, the abruptness of the change to an online delivery presented technical challenges that seemed to impact the results of the study, particularly with the first group (Participants 1-5) who did not receive immediate feedback in the first perceptual training task, and the participants (2 and 10) who failed to save their test output.

6.3 Suggestions for Future Research

A number of opportunities for future research arise from the current study. More participants overall would help to reduce the statistical noise from the individual variation. Additionally, while perceptual training, which has been shown to be effective, was used as a baseline against which to measure musical training, and both trainings were found to generate learning, with perceptual training being superior, future research could add a non-training control group of 16 new participants (musicians and nonmusicians). This addition would facilitate assessment of the effectiveness of the trainings in a classroom setting as well as the effectiveness of musical training versus no training. The addition of more training sessions is also an important area for future research to explore. Given that students generally did not prefer the perceptual training, it is critical to test if perceptual training would maintain its effectiveness over time or lose effectiveness due to lack of learner engagement or demotivation.

Relatedly, future research should also consider adapting perceptual training to align more closely with the principles of Communicative Language Teaching (CLT) to offset learner criticisms. More trainings could also test whether additional musical training sessions would increase the effectiveness of the musical training overall. In terms of the potential for generalization of learning, future research could also use tones with a smaller gap in relative pitch ranges for training, and tones with a wider gap in relative pitch ranges for generalization testing. Lastly, the rapid change in teaching modality was a weakness that provides an opportunity for future research. The study should be replicated in a face-to-face (f2f) modality, particularly due to the challenge of using digital instruments through an online medium.

6.4 Implications for Pedagogy

Results from this study demonstrate that both perceptual and musical training can be incorporated into a (loosely defined) CLT class and can yield learning in perception, discrimination and identification of tones. While teachers will not necessarily know who in their classroom is a musician and who is a nonmusician, a mix of techniques is advisable especially given that although perceptual training is more effective, most learners preferred musical training (and, specifically, more than half of nonmusicians preferred musical training) and research has found at least a musical training background to be helpful in pronunciation learning. In such a mix, it might be helpful for teachers to order the less engaging perceptual training first, which may increase overall potential for learning, and then follow with the more engaging musical training as a way of sustaining motivation. Importantly, given the lower scores in perceptual training of the first training group, immediate feedback should be consistently incorporated in the trainings. Overall, this thesis provides teachers with additional techniques that they can use in their classroom for the teaching of a language's tones.

Appendices

Appendix A

Flyer for trainee participants

SYRACUSE UNIVERSITY

College of Arts & Sciences

Languages, Literatures, and Linguistics

Consent to Participate

(Study of Teaching Pronunciation in Tonal Languages)

My name is Elizabeth Elton and I am a graduate student in the Linguistic Studies program at Syracuse University. I am interested in researching effective methods of teaching pronunciation in a second language and I am inviting you to take part in a research project on this. The training will be two 1-hour sessions completed on different days. In addition to the training, initial surveys will be conducted, which will take 5 mins. Also, after training a retention test will be administered, which will take 15 mins. Finally, training will be recorded. Recordings provided during the training will be manually transcribed, de-identified, and stored on a computer.

Please note that if you consent to participate in this study, data from the study may be stored on a secure internet storage site and correspondences online about the study may take place over email. Whenever one works with e-mail or the internet there is always the risk of compromising privacy, confidentiality and/or anonymity. Your confidentiality will be maintained to the degree permitted by the technology being used. It is important for you to understand that no guarantees can be made regarding the interception of data sent via the internet by third parties.

Participants must be 18 years of age or older. Participation is absolutely voluntary and you can opt-out at any time. Any questions you may have can be directed to me or my faculty supervisor, Dr. Amanda Brown. My email address is eaelton@syr.edu and my faculty supervisor's email address is abrown08@syr.edu.

Please check the following as appropriate:

I am 18 years or older ☐

I agree to participate in this study as described. ☐

I consent to be audio recorded. ☐

Printed name of participant	**Signature of participant**	**Date**
Printed name of researcher	**Signature of researcher**	**Date**

Flyer for Yoruba speakers to record stimuli

SYRACUSE UNIVERSITY

College of Arts & Sciences

Languages, Literatures, and Linguistics

Consent to Participate

(Study of Teaching Pronunciation in Tonal Languages)

My name is Elizabeth Elton and I am a graduate student in the Linguistic Studies program at Syracuse University. I am interested in researching effective methods of teaching tone to speakers of stress-accent languages. A tone language is a language that uses pitch to mark differences in words. For instance, in Mandarin Chinese (a tone language), the word *ma* can have a high, level pitch or a falling pitch. The use of one over another indicates a different meaning. *Ma* with a high, level pitch means "mother" while *ma* with a falling pitch means "scold." On the other hand a stress-accent language uses stress (i.e., placing greater emphasis on one part of a word) on words. This research will be examining the use of music training and perceptual training (a kind of training that exposes a learner to various spoken words in the language being learned) on the teaching of tones. In order to administer this training, I require recordings of words spoken in your tonal language. For this research, I would ask to record you speaking a set of words to be used for the training mentioned above. This will take 30 minutes. Recordings provided during the research will be manually transcribed, de-identified, and stored on a computer.

Please note that if you consent to participate in this study, data from the study may be stored on a secure internet storage site and correspondences online about the study may take place over email. Whenever one works with e-mail or the internet there is always the risk of compromising privacy, confidentiality and/or anonymity. Your confidentiality will be maintained to the degree permitted by the technology being used. It is important for you to understand that no guarantees can be made regarding the interception of data sent via the internet by third parties.

Participants must be 18 years of age or older. Participation is absolutely voluntary and you can opt-out at any time. Any questions you may have can be directed to me or my faculty supervisor, Dr. Amanda Brown. My email address is eaelton@syr.edu and my faculty supervisor's email address is abrown08@syr.edu.

Please check the following as appropriate:

I am 18 years or older ☐

I agree to participate in this study as described. ☐

I consent to be audio recorded. ☐

________________________	________________________	________________
Printed name of participant	**Signature of participant**	**Date**
________________________	________________________	________________
Printed name of researcher	**Signature of researcher**	**Date**

Appendix B

Pretest 1

Directions: Listen to the audio track. The words below are listed in the order you will hear them. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. There will be about six seconds between each word (HINT: H tone is á; L tone is à).

1. juju: first tone: H L second tone: H L
2. ba: H L
3. gusu: first tone: H L second tone: H L
4. fọ: H L
5. koko: first tone: H L second tone: H L
6. fọ: H L
7. lu: H L
8. ro: H L
9. bulu: first tone: H L second tone: H L
10. lu: H L
11. koko: first tone: H L second tone: H L
12. kọ: H L
13. kẹkẹ: first tone: H L second tone: H L
14. kọ: H L
15. ba: H L
16. ro: H L

Posttest 1

Directions: Listen to the audio track. The words below are listed in the order you will hear them. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. There will be about six seconds between each word (HINT: H tone is á; L tone is à).

1. kọ: H L
2. koko: first tone: H L second tone: H L
3. ba: H L
4. koko: first tone: H L second tone: H L
5. kẹkẹ: first tone: H L second tone: H L
6. fọ: H L
7. kọ: H L
8. fọ: H L
9. lu: H L
10. gusu: first tone: H L second tone: H L
11. ba: H L
12. juju: first tone: H L second tone: H L
13. ro: H L
14. lu: H L
15. bulu: first tone: H L second tone: H L
16. ro: H L

Pretest 2

Directions: Listen to the audio track. The words below are listed in the order you will hear them. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. There will be about six seconds between each word (HINT: H tone is á; L tone is à).

1. gusu: first tone: H L second tone: H L
2. fọ: H L
3. kẹkẹ: first tone: H L second tone: H L
4. kọ: H L
5. lu: H L
6. juju: first tone: H L second tone: H L
7. ba: H L
8. ro: H L
9. lu: H L
10. ba: H L
11. koko: first tone: H L second tone: H L
12. fọ: H L
13. ro: H L
14. koko: first tone: H L second tone: H L
15. bulu: first tone: H L second tone: H L
16. kọ: H L

Posttest 2

Directions: Listen to the audio track. The words below are listed in the order you will hear them. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. There will be about six seconds between each word (HINT: H tone is á; L tone is à).

1. lu: H L
2. juju: first tone: H L second tone: H L
3. lu: H L
4. ba: H L
5. koko: first tone: H L second tone: H L
6. fọ: H L
7. koko: first tone: H L second tone: H L
8. fọ: H L
9. gusu: first tone: H L second tone: H L
10. kọ: H L
11. ba: H L
12. ro: H L
13. kẹkẹ: first tone: H L second tone: H L
14. kọ: H L
15. ro: H L
16. bulu: first tone: H L second tone: H L

Appendix C

Generalization Test

Directions: Listen to the audio track. The words below are listed in the order you will hear them. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. There will be about six seconds between each word (HINT: H tone is á; L tone is à; M tone is a).

1. kọ: H M L
2. kere: first tone: H M L second tone: H M L
3. kuru: first tone: H M L second tone: H M L
4. le: H M L
5. yọ: H M L
6. bọ: H M L
7. le: H M L
8. bi: H M L
9. ya: H M L
10. bọ: H M L
11. koro: first tone: H M L second tone: H M L
12. lọ: H M L
13. gege: first tone: H M L second tone: H M L
14. le: H M L
15. ra: H M L
16. dọla: first tone: H M L second tone: H M L
17. rara: first tone: H M L second tone: H M L
18. ki: H M L
19. labẹ: first tone: H M L second tone: H M L
20. ṣu: H M L
21. fufu: first tone: H M L second tone: H M L
22. buru: first tone: H M L second tone: H M L
23. lu: H M L
24. ro: H M L
25. ṣubu: first tone: H M L second tone: H M L
26. ṣere: first tone: H M L second tone: H M L
27. kọja: first tone: H M L second tone: H M L
28. bọ: H M L

Appendix D[52]

Participant Survey

1. Do you have musical training experience (either formal or informal)? In other words, have you trained yourself or been trained by someone else to use an instrument (i.e., piano, guitar, voice, etc.)

   Yes **No**

2. If you answered "yes" to question one, what instrument did you train with?

   ____
   ____

3. If you answered "yes" to question one, how many years of training have you had?

   ____
   ____

4. If you answered "yes" to question one, are you self-taught or have you trained with a teacher?

   ____
   ____

5. Do you have any experience learning a tonal language? A tone language is a language that uses pitch to mark differences in words. For instance, in Mandarin Chinese (a tone language), the word "ma" can have a high, level pitch or a falling pitch. The use of one over another indicates a different meaning. "Ma" with a high, level pitch means "mother" while "ma" with a falling pitch means "scold." On the other hand a stress-accent language uses stress (i.e., placing greater emphasis on one part of a word) on words. Examples of tone languages include: Mandarin Chinese, Cantonese, Thai, Vietnamese, Punjabi, Yoruba, Igbo, Ewe, Zulu, etc.

   Yes **No**

   Please indicate the language you studied:______________________________

6. If you answered "yes" to question four, how long did you study the language?

   ____

[52] Note that "repetition" refers to "perceptual" training, but "repetition" was used to refer to the perceptual training with the participants

____

____

7. If you answered "yes" to question four, would you consider yourself proficient or fluent in the language?

____

____

8. Which training type did you like the most?

    Musical Training   Repetition Training

9. Why did you like the training type you circled in Question 8 best?

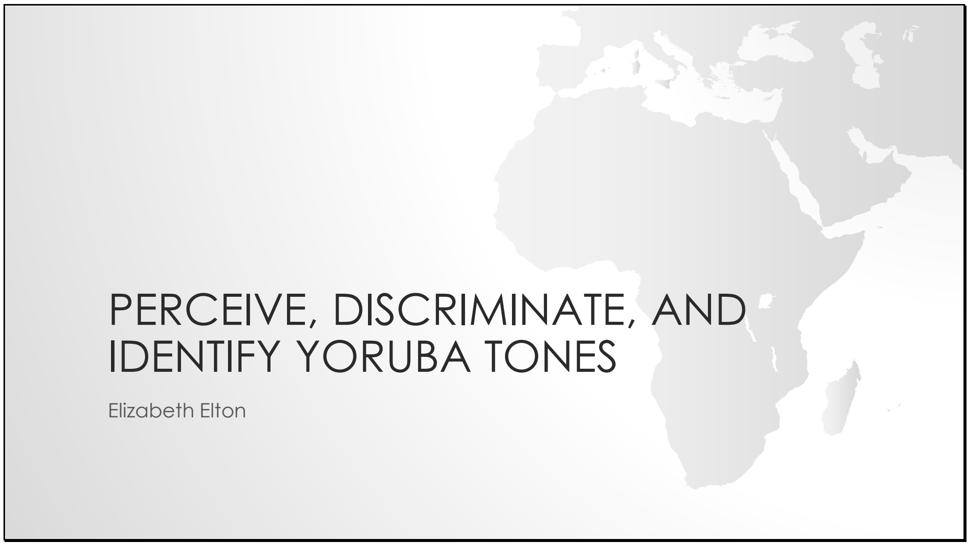
____

____

Appendix E[53]

General Information Powerpoint

Slide 1

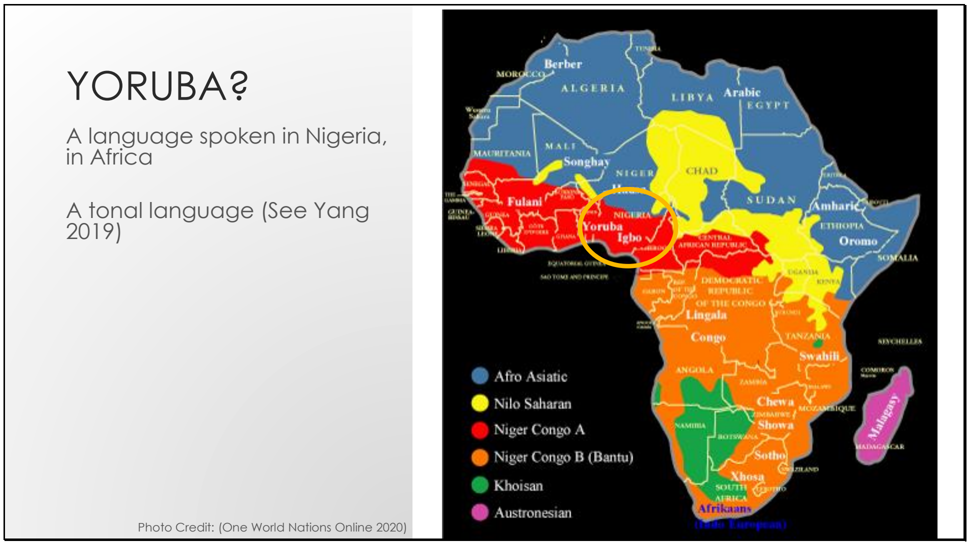


Link: https://drive.google.com/file/d/15QsmlKMxqs4XWnzbs2IMn0LUhDfEN-R8/view?usp=sharing

Piano: https://drive.google.com/file/d/18zgbwnXVIetzkgb92tq5Otm9jKQALHVc/view?usp=sharing

[53] Text under slides were the notes drawn upon during the trainings

Slide 2

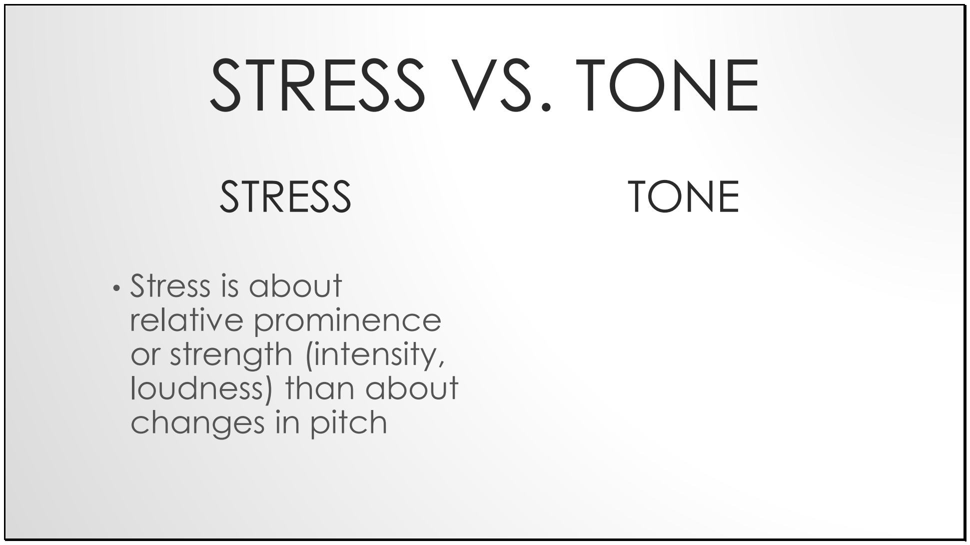


We’re going to be learning the language, Yoruba, today. Yoruba is a language spoken in Nigeria, which is in the western part of Africa [point to where Nigeria is].

Specifically, Yoruba is a tonal language. Unlike English, which is a stress language, Yoruba uses tones instead of stress.

https://www.nationsonline.org/oneworld/african_languages.htm

Slide 3

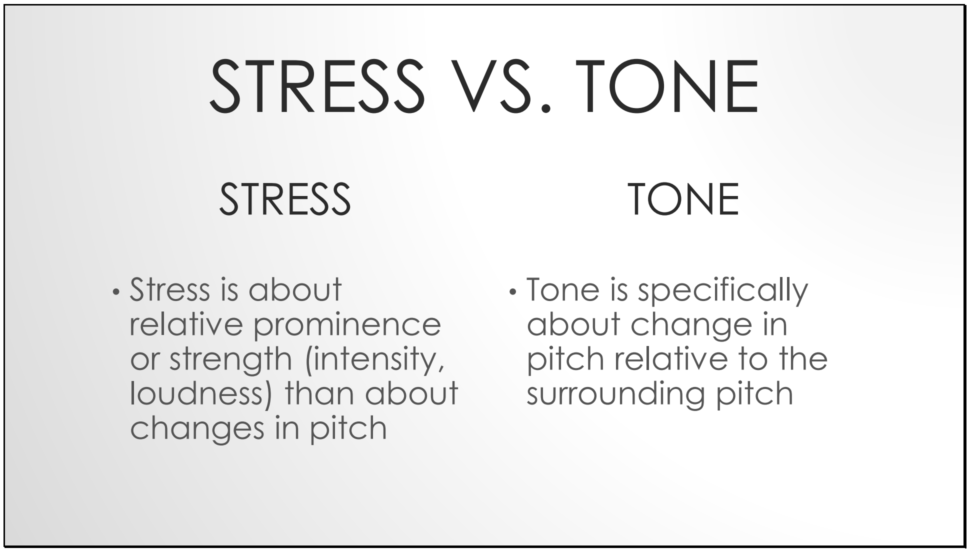


So, what is stress and tone? Stress is about prominence, which we can think of as intensity or loudness on a syllable rather than about changes in pitch. Tone, on the other hand, is about pitch. As speakers of English all of us are users of stress.

Slide 4

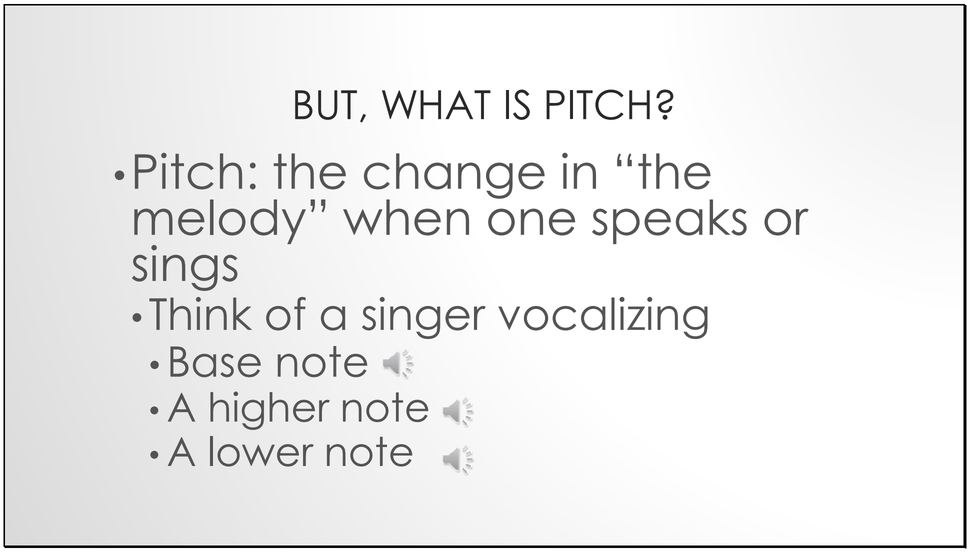
Stress is more about intensity than the change in pitch

Slide 5



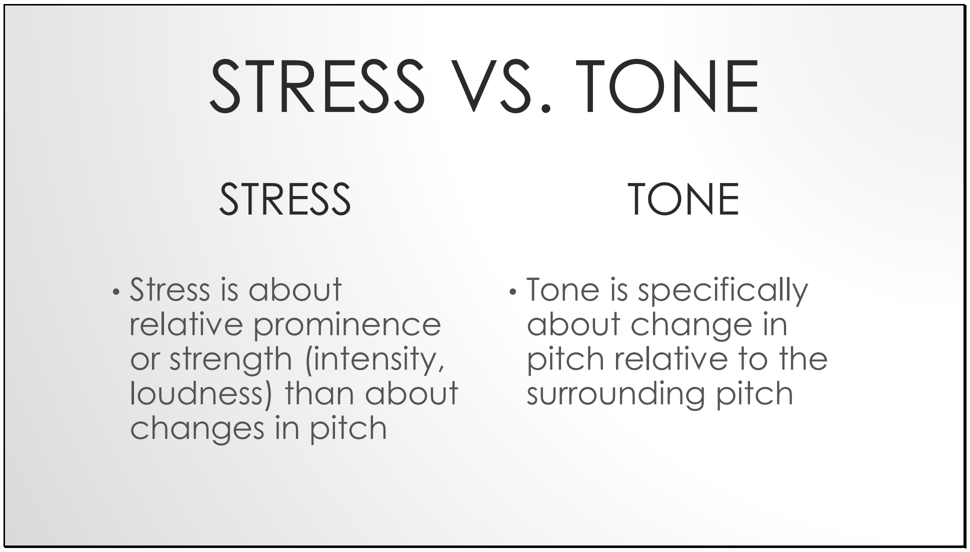
So, what is stress and tone? Stress is about prominence, which we can think of as intensity or loudness on a syllable rather than about changes in pitch. Tone, on the other hand, is about pitch. As speakers of English all of us are users of stress.

Slide 6



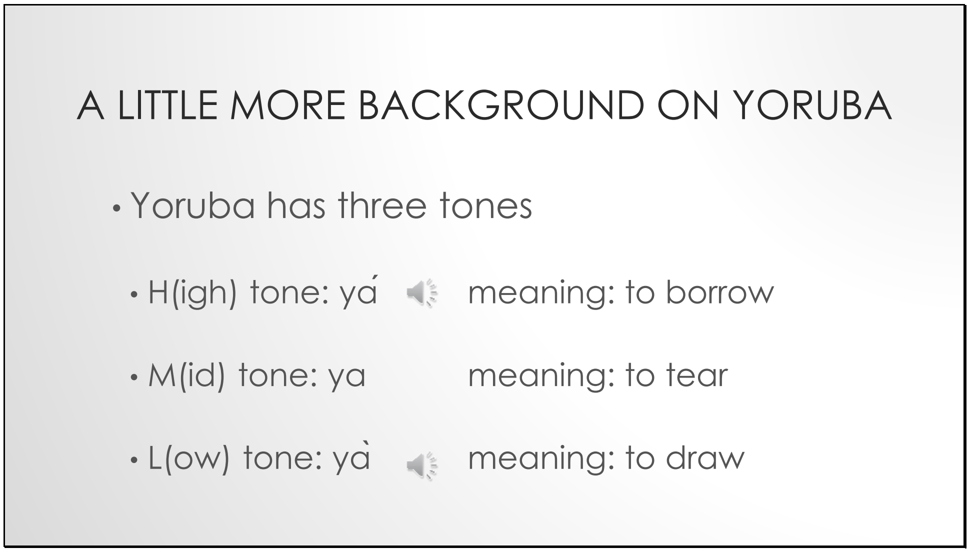
But what is pitch exactly?

Slide 7



So, what is stress and tone? Stress is about prominence, which we can think of as intensity or loudness on a syllable rather than about changes in pitch. Tone, on the other hand, is about pitch. As speakers of English all of us are users of stress. In tonal languages, pitch is used to show meaning.

Slide 8

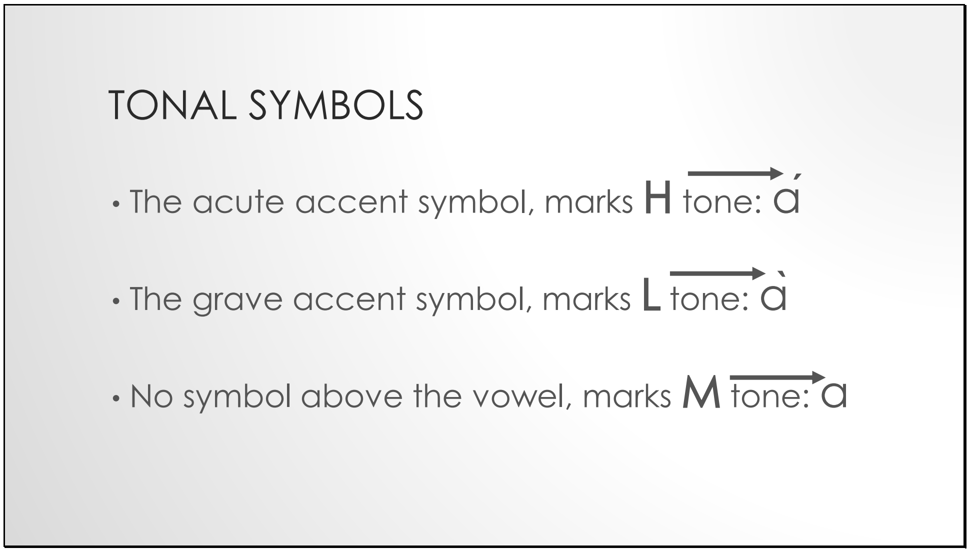


So, Yoruba has three tones: H, M, and L.

Explain the words and meanings.

Make sure to tell the participants that we will not be training with the M tone, but it may be on a test, so keep it in mind.
Make sure to mention noticing the symbols that mark the tones…

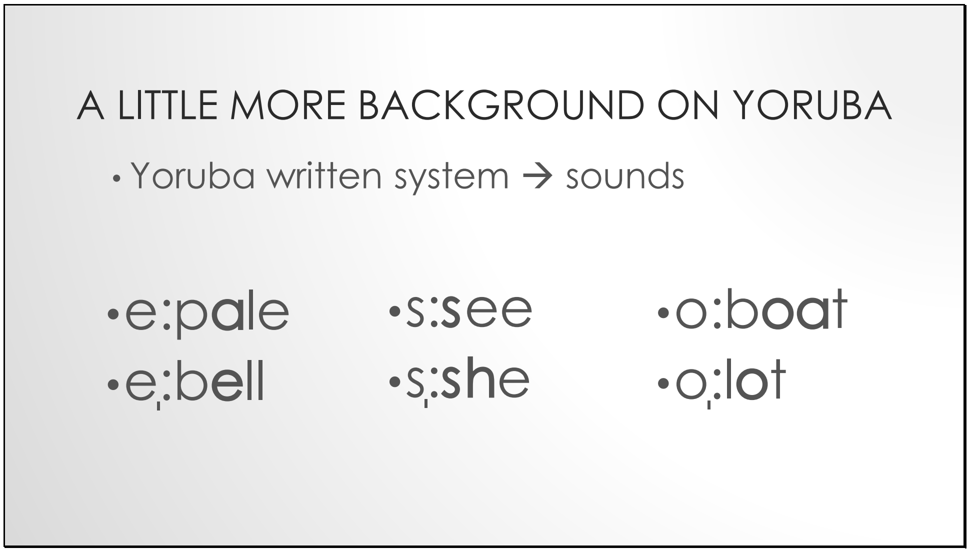
Slide 9



Also, I just want to make you all aware of how H and L tones are denoted.

Explain and read

Slide 10

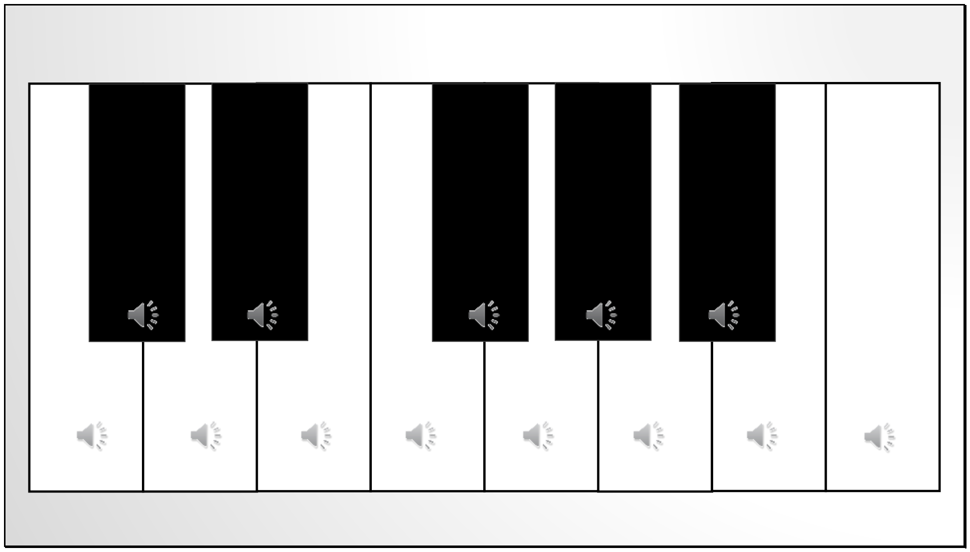


You are NOT going to be tested on this. I'm just showing you this, so you know the sounds we will be using. But this is not going to be tested.

Musical Training Related PowerPoint

Slide 1

Slide 2



Using this piano, when I click on a key or a note, it produces a pitch. If I go to the left of that note and click another, the pitch becomes lower. If I click on a note to the right of it, the pitch becomes higher. Do you hear that. [pick on a person]: which note produces a higher pitch [click on two notes]. Using your fingers to pick, which one was a higher pitch, the first or the second. [Repeat until everyone understands]. Good!

Tones in a tonal language use pitch. Its just rather than using the pitch to create melodies for music, tone uses pitch to convey different meanings. So, I could have a word, like [ya].

https://drive.google.com/open?id=18zgbwnXVIetzkgb92tq5Otm9jKQALHVc

Slide 3

This word: ya

Slide 4

It can be a high tone. If it is this word means to borrow

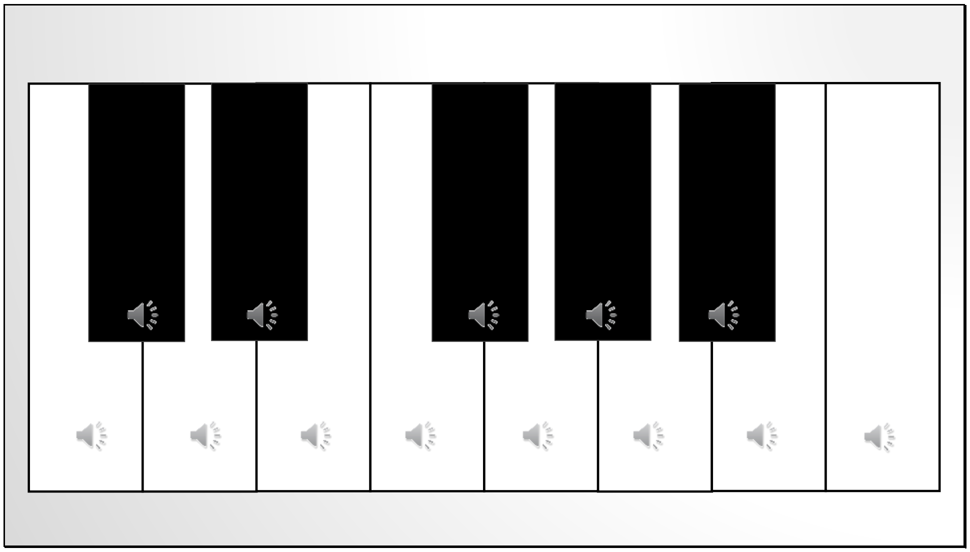
Slide 5

It could be a low tone, and if it is, it means to draw.

Slide 6

Both words have the same sounds, the only difference is the pitch. This difference in pitch creates meaning.

Slide 7



So, for [ya] if my L tone is pitched somewhere in here [play the bottom two notes], then this would be the word that means to draw. And if my H tone is pitched somewhere in here [play some top two notes], then this would be the word meaning to borrow.

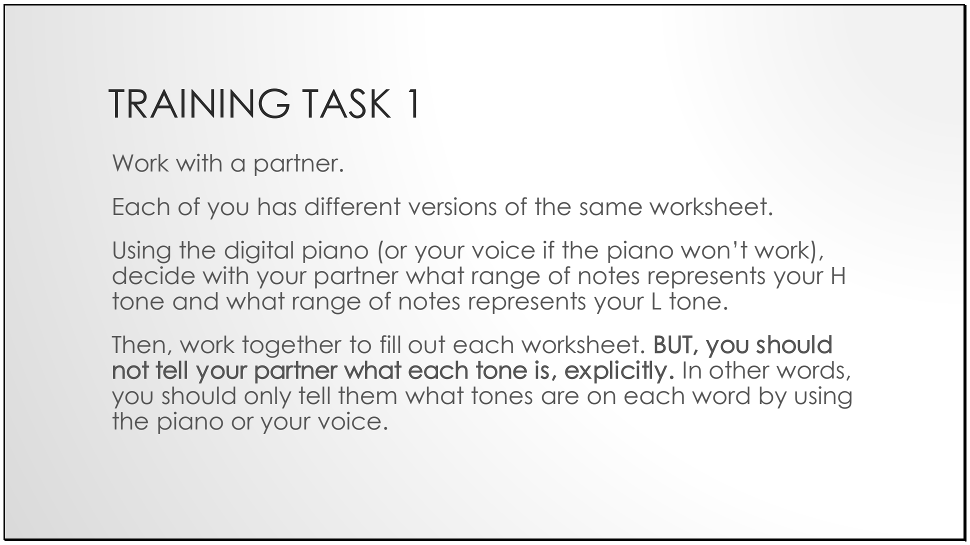
Now in tone languages, different people can have different pitch ranges. So one person may have a pitch range for their H tone around here [demonstrate with top two notes], but another person may have a pitch range for their H tone around here [demonstrate with lower two notes]. Just because two people have different pitch ranges doesn't mean they aren't both using H tones nor does it mean that these two people would not understand each other as using H tones. It just means that tone is relative. It is relative to the person using it.

Also, in Yoruba, with two syllable words, sometimes the pitch changes on one syllable. So, it would sound like this. This will happen on the last syllable of a two syllable word. If it's a HL word it will sound like this, but if it is a LH word it will sound like this.

So, I want us to practice this with an activity.

https://drive.google.com/open?id=18zgbwnXVIetzkgb92tq5Otm9jKQALHVc

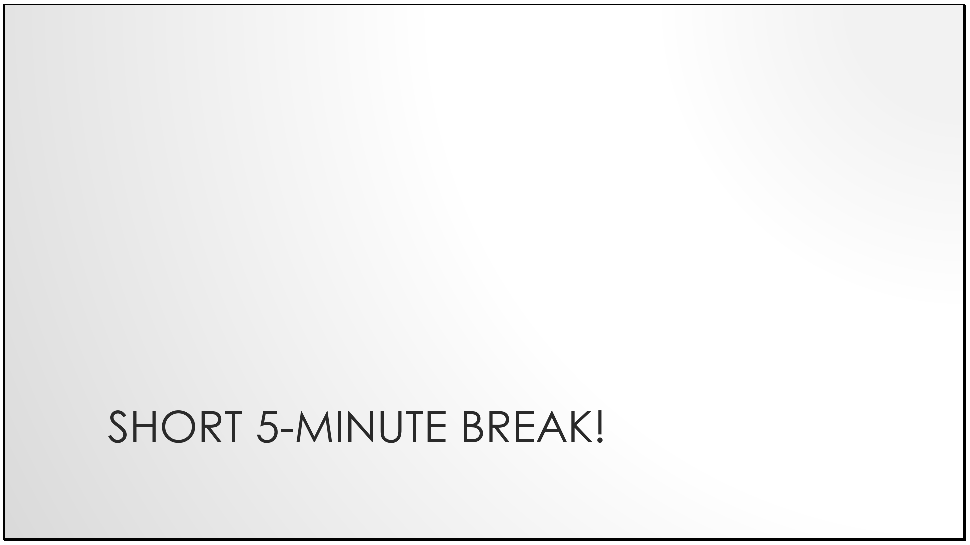
Slide 8



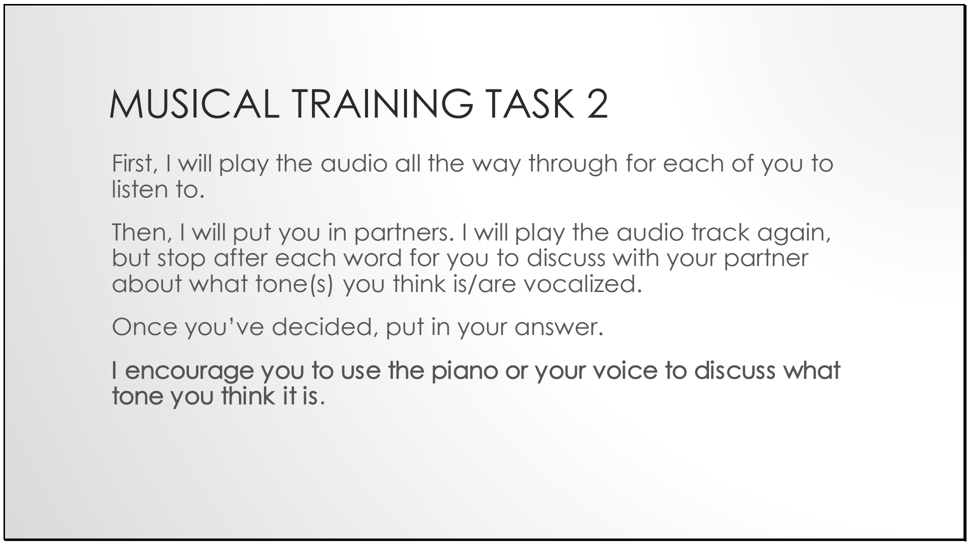
Version A:
https://drive.google.com/open?id=1kd7XtSHDBmLMrAZU3U6auFSQ8gUkm16p
Version B: https://drive.google.com/open?id=10CyxfwC6IsDUJd-Q4msyRqV5Yn7arwRA

Version C:
https://drive.google.com/open?id=1_YT4mobt7qtjtjakNrUIuTSHnPEtVaua
Version D:
https://drive.google.com/file/d/1uYwDZ73fMGMGFONPhBt_NoyD_2Zratk1/view?usp=sharing
Version E: https://drive.google.com/open?id=1KjZfaEplztkxQRuLy1g76-LC6IA4Dpv7

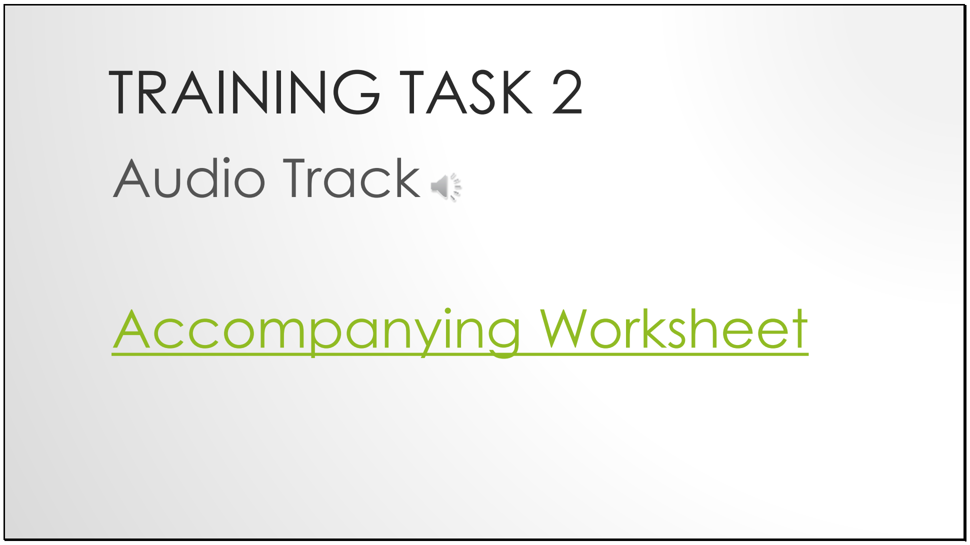
Slide 9



Slide 10


MUSICAL TRAINING TASK 2
First, I will play the audio all the way through for each of you to listen to.
Then, I will put you in partners. I will play the audio track again, but stop after each word for you to discuss with your partner about what tone(s) you think is/are vocalized.
Once you've decided, put in your answer.
I encourage you to use the piano or your voice to discuss what tone you think it is.

Slide 11



Audio Track Link: https://drive.google.com/file/d/1A-6yi69e8bL4WkALIcFLBzfKYFluQAWo/view?usp=sharing

Perceptual Training Related PowerPoint

Slide 1

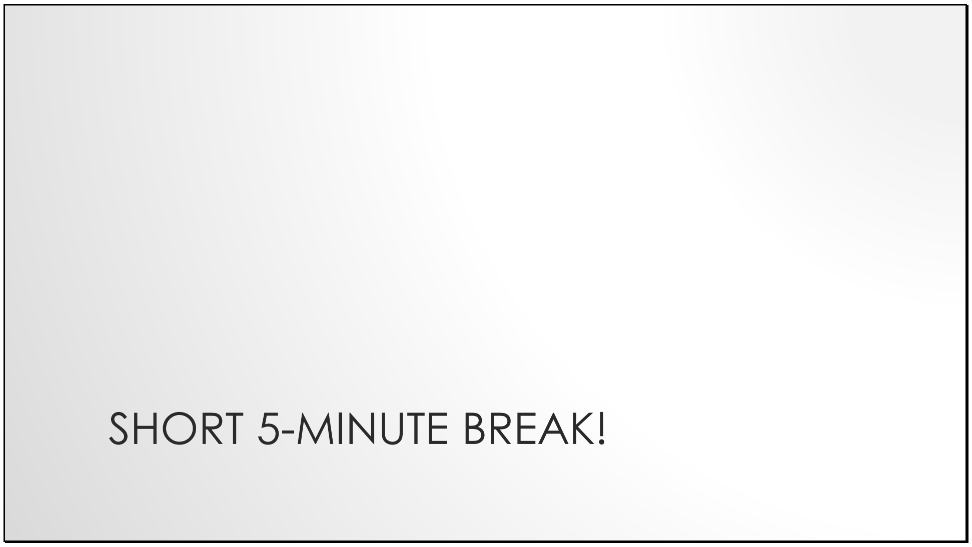
Slide 2

Audio Track Link:
https://drive.google.com/file/d/1XDeJpe8IT2ZHLWIPPQg3X2OJwf-IFSbY/view?usp=sharing

Slide 3

Audio Track Link:
https://drive.google.com/file/d/1si6F5cRQIybvDTRw5nIlrCNvwpw1cBOt/view?usp=sharing

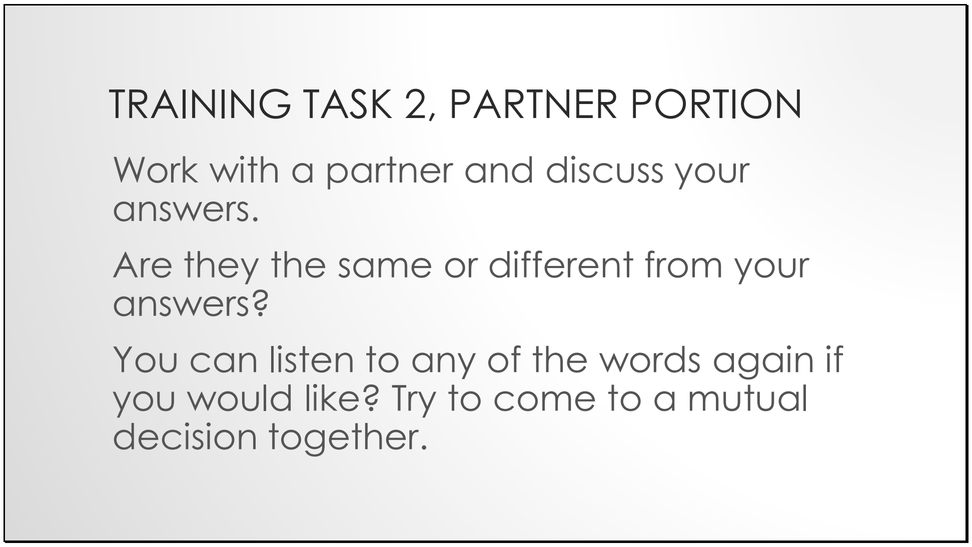
Slide 4



Slide 5

Audio Track Link: https://drive.google.com/file/d/1yu4q-IMnanpep-d4Z635K9DLSN_eQo72/view?usp=sharing

Slide 6


TRAINING TASK 2, PARTNER PORTION
Work with a partner and discuss your answers.
Are they the same or different from your answers?
You can listen to any of the words again if you would like? Try to come to a mutual decision together.

Appendix F

<table>
<tr><th colspan="3">Musical Training Lesson Plan</th></tr>
<tr><td colspan="3">Teacher Name: Elizabeth Elton</td></tr>
<tr><td colspan="3">Course: Musical Training Lesson</td></tr>
<tr><th colspan="3">Preliminary Info</th></tr>
<tr><td>Class Level</td><td colspan="2">All participants will have had limited or no exposure to a tone language</td></tr>
<tr><td>Time/Length</td><td colspan="2">60 minutes</td></tr>
<tr><td>Class Profile</td><td colspan="2">There will be musicians and nonmusicians in this lesson</td></tr>
<tr><td>Lesson Objectives</td><td colspan="2">To explicitly bring the tones of Yoruba to the attention of the participants through musical training techniques</td></tr>
<tr><td>Language Analysis</td><td colspan="2">See Appendix</td></tr>
<tr><td>Assumed Knowledge</td><td colspan="2">No assumed knowledge of the language</td></tr>
<tr><td>Materials</td><td colspan="2">• PowerPoint Presentation on Tones
• A digital piano
• Information Gap set of worksheets (2 versions)
• Audio stimuli
    ○ Worksheet that goes with the audio stimuli</td></tr>
<tr><th colspan="3">Procedure</th></tr>
<tr><td>Stage: Presentation</td><td>Instructions:</td><td>Part 1:
The students/participants will listen to a PowerPoint Presentation given by the instructor/researcher. This PowerPoint will bring tones to the level of awareness and explicitly instruct on tones.

Part 2: The researcher will show how different “notes” on a digital piano correspond to different pitches. It will be shown that some pitches are higher while some lower. The researcher will explain how combining two notes can make a pitch rise or fall. Subsequently, the researcher will show how notes on</td></tr>
</table>

		a piano can correspond to pitches that represent linguistic tone. Importantly, the researcher will point out in this presentation that tone is relative to speaker, in order to encourage the categorization of tones. By showing the pitches on the piano, the researcher will explain the importance of pitch to linguistic tone is in the differences between one tone's height, slope, and direction, in comparison to another's. In other words, using one, specific pitch on the piano (which the researcher will correspond to a specific tone in the target language) does not equate to all tonal language users using that specific pitch every time for the same tone. Rather, the relativeness of one tone to another in one utterance reveals the tonal category.
	Aims:	To bring the tones to the level of awareness (Antoniou & Wong's 2016; Lu, Wayland & Kaan 2015; Pederson & Guion-Anderson's 2010); to introduce the tonal symbols being used to represent the tones in writing (Godfroid, Lin and Ryu 2017)
	Types of interaction:	S-T Class work
	Timing	**Part 1:**10 minutes **Part 2:** 15 minutes
	Whiteboard Use:	Might be used to supplement the use of tonal symbols in writing to correlate to the tones. They can be put on the board to specifically use as a reference
Stage: Practice	Instructions:	Participants will then be given a set of words, and some will be marked for tone with tonal symbols in the form of an information gap activity. In this activity, participant pairs will be given two versions of a worksheet. In one version (Version A), a word is given without the tonal symbol. In the other version (Version B), the corresponding word with its tonal symbol is given[54]. The participant with Version A will ask what tone their word has. The participant with Version B will answer by playing a corresponding note on the piano. The participant with Version A will mark the tone down on their worksheet. The participant with Version B can also have an unmarked tone and inquire about it to the participant with Version A. They will continue to fill in their sheet until they both complete their own versions. Directions: Now, I will hand out a pair of worksheets that require you to have a partner. So, with the person sitting next to you, one of you will get Version A of the worksheet, and the other will get a Version B. Explain to your partner before you start filling in the gaps what relative pitch range on the piano you

[54] With three partners, this is split into three worksheets (Version C, D, and E)

		want to use. So, is this note (play note) going to represent the H tone or the M tone. Now, if this note is the H tone, then what note would have to be the M tone? So, decide this and then tell your partner. Once you start filling in the blanks, you can't tell your partner which note represents which tone for you. Then, find your first word with a tonal symbol, and tell your partner which number the word is. Then play the tone on the piano according to your relative pitch range that you explained to your partner.
	Aims:	To have participants start practicing with relative pitch in correspondence to tones. To use musical training to allow participants to learn the grammar of tones.
	Type of Interaction	S-S
	Timing	15 minutes
	Whiteboard Use:	Might be used to supplement the use of tonal symbols in writing to correlate to the tones. They can be put on the board to specifically use as a reference
Stage: Production	Instructions:	**Part 1:** Then participants will be given a short break. **Part 2:** Next, participants will be given a new sheet with different words that are unmarked for tone, and they will work with the same partner. They will listen to the audio stimuli created for training that corresponds with the order the unmarked words are listed on the sheet. Together, they will need to determine what each word's tone is. When they decide, they will mark it with a tonal symbol. All worksheets will be handed to the researcher before the training ends. Directions: Now, with your same partner, I am handing out one worksheet for each of you. I will play an audio file for you all to hear. With your partner, decide what tone you think you hear. Then, once you've agreed, mark the tone on your worksheet with the corresponding tonal symbol.
	Aims:	To have participants listen to input and process the tones being used.
	Type of Interaction	S-S
	Timing	**Part 1:** 5 minutes **Part 2:** 15 minutes
	Whiteboard Use:	Might be used to supplement the use of tonal symbols in writing to correlate to the tones. They can be put on the board to specifically use as a reference

Appendix

Practice Task Worksheet Versions

Directions

Objective: **Don't show your partner your paper at any time during this activity.** You and your partner each have a different version of this worksheet. You have 5 words marked with tone and 5 words not marked with tone. You need to know what tones each word has.

Directions: Using what we learned with the digital piano, decide what keys/notes on the piano will correspond to your H tone and your L tone. Then, check the box that designates the tone you think it is. **Remember**: pitch is relative. This means you could have a few keys/notes represent your H tone and a few keys/notes represent your L tone. Just be sure that you are clear with your partner about which keys/notes correspond to which tones.

Then, start at the top of the list. If your first word is unmarked with tone, ask your partner what tone it is. If your first word is marked for tone, **do not answer your partner by telling them which tone it is**. You should play one of the keys/notes on the digital piano to demonstrate which tone is marked on the word. Continue until every word in the list is marked with tone (**HINT: H tone is á; L tone is à**).

Version A:

1. fa: H L
2. fà
3. ya: H L
4. ya: H L
5. ṣú
6. ṣù

Directions: For this second part, when you need to demonstrate the tones with the piano, play both keys/notes to represent the tones.

7. ràrá
8. rárà
9. kuro: first tone: H L second tone: H L
10. buba: first tone: H L second tone: H L

Version B

1. fá
2. fa: H L
3. yà
4. yá

5. ṣu: H L
6. ṣu: H L

Directions: For this second part, when you need to demonstrate the tones with the piano, play both keys/notes to represent the tones.

7. rara: first tone: H L second tone: H L
8. rara: first tone: H L second tone: H L
9. kúrò
10. bùbá

Version C

1. fa: H L
2. fa: H L
3. ya: H L
4. ya: H L
5. ṣú
6. ṣù

Directions: For this second part, when you need to demonstrate the tones with the piano, play both keys/notes to represent the tones.

7. ràrá
8. rárà
9. kuro: first tone: H L second tone: H L
10. buba: first tone: H L second tone: H L

Version D

1. fá
2. fà
3. ya: H L
4. ya: H L
5. ṣu: H L
6. ṣu: H L

Directions: For this second part, when you need to demonstrate the tones with the piano, play both keys/notes to represent the tones.

7. rara: first tone: H L second tone: H L
8. rara: first tone: H L second tone: H L
9. kúrò:
10. buba: first tone: H L second tone: H L

Version E

	1. fa: H L 2. fa: H L 3. yà 4. yá 5. ṣu: H L 6. ṣu: H L Directions: For this second part, when you need to demonstrate the tones with the piano, play both keys/notes to represent the tones. 7. rara: first tone: H L second tone: H L 8. rara: first tone: H L second tone: H L 9. kuro: first tone: H L second tone: H L 10. bùbá
Production Task Worksheet	Directions: Listen to the audio track. The words listed below are in the order you will hear them. However, they do not have the tone symbols marked. Work with your partner to determine the tones for each word, and check the box that designates the tone you think it is. The instructor will play the track once all the way through. Then, the instructor will play it again, stopping each time to allow you to discuss with your partner. Once everyone thinks they have figured out the tones, the instructor will play it once through again (HINT: H tone is á; L tone is à). 1. bi: H L 2. ki: H L 3. bi: H L 4. ra: H L 5. ki: H L 6. ra: H L 7. baba: first tone: H L second tone: H L 8. bajẹ: first tone: H L second tone: H L 9. yala: first tone: H L second tone: H L 10. jalẹ: first tone: H L second tone: H L

Perceptual Training Lesson Plan
Teacher Name: Elizabeth Elton

Course: Perceptual Training Lesson		
Preliminary Info		
Class Level	All participants will have had limited or no exposure to a tone language	
Time/Length	60 minutes	
Class Profile	There will be musicians and nonmusicians in this lesson	
Lesson Objectives	To explicitly bring the tones of Yoruba to the attention of the participants through perceptual training techniques	
Language Analysis	See Appendix	
Assumed Knowledge	No assumed knowledge of the language	

Materials	• PowerPoint Presentation on Tones
	• Audio stimuli
	    ○ Worksheets
	        ▪ 1st Training Worksheet
	        ▪ 2nd Training Worksheet	

Procedure		
Stage: Presentation	Instructions:	**Part 1:** The students/participants will listen to a PowerPoint Presentation given by the instructor/researcher. This PowerPoint will bring tones to the level of awareness and explicitly instruct on tones.
	Aims:	To bring the tones to the level of awareness (Antoniou & Wong’s 2016; Lu, Wayland & Kaan 2015; Pederson & Guion-Anderson’s 2010); to introduce the tonal symbols being used to represent the tones in writing (Godfroid, Lin and Ryu 2017)
	Types of interaction:	S-T Class work
	Timing	**Part 1:** 10 minutes

Stage: Practice	Instructions:	**Part 1:** The researcher will then start the perceptual training. Much like Wang (2013), stimuli for this training will be presented one at a time. Participants will mark their identification of the tone on a worksheet. Immediate feedback will be given by the researcher about which tone was spoken in the audio stimuli. **Part 2:** Then, participants will be given a short break. Directions: I am passing out a worksheet. This worksheet has the words in the order you will hear them; you should notice that the tones are not marked on this worksheet. I will play the corresponding audio, and you will mark the tones you hear with the tonal symbols we learned. After you have marked down the tones you heard, I will tell you the correct answer.
	Aims:	To have participants practice listening to different phonetic contexts and pitch variability corresponding to tones.
	Type of Interaction	S-T
	Timing	**Part 1:** 5-7 minutes **Part 2:** 5-7 minutes **Part 2:** 5 minutes
Stage: Production	Instructions:	**Part 1:** When they return, they will continue the training for another 5-10 minutes with new words. They will hear the words by additional speakers and more times than in the Practice stage. During this second portion, immediate feedback will not be given. However, the researcher will play the audio track again, so the participants can double check their first answers. Directions: This will be similar to the last task. This worksheet also has all the words in the order you will hear them, and the tones are not marked on this worksheet either. I will play the corresponding audio, and you will mark the tones you hear with the tonal symbols we learned. After you have marked down the tones you heard, however, I will not be telling you the correct answer. **Part 2:** Participants will then be paired together and told to review their answers to each of the words for the second post-break training period. They will be allowed to listen to any of the stimuli again, and are able to change their answers. All worksheets and materials will be turned into the researcher once the training session has ended.

		Directions: Now, find a partner and look over your answers. You can listen to any audio file if you disagree on an answer. After you have reviewed together and come to a conclusion, I will give the answers. You can then review the audio files again.
	Aims:	To have participants listen to input and process the tones being used.
	Type of Interaction	S-S S-T
	Timing	**Part 1:** 15 minutes **Part 2:** 15 minutes
	Whiteboard Use:	Might be used to supplement the use of tonal symbols in writing to correlate to the tones. They can be put on the board to specifically use as a reference
	Areas of Flexibility/ Comments:	
Appendix		
Practice Task Worksheet	Directions: Listen to the audio track. The words are listed in the order they are spoken. There will be three sets. Set 1 is first, Set 2 is second, and Set 3 is third (which are on the last page). Each blank line between words within sets denotes about six seconds before the instructor will give the answer. Between each set is about ten seconds. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. **Set 1:** 1. yọ: H L 2. yọ: H L 3. lọ: H L 4. lọ: H L 5. ṣu: H L 6. ṣu: H L	

7. yara: first tone: H L second tone: H L
8. rara: first tone: H L second tone: H L
9. rara: first tone: H L second tone: H L
10. kuku: first tone: H L second tone: H L

Set 2:

1. yọ: H L
2. yọ: H L

3. lọ: H L
4. lọ: H L

5. ṣu: H L
6. ṣu: H L

1. kuku: first tone: H L second tone: H L
2. rara: first tone: H L second tone: H L
3. rara: first tone: H L second tone: H L
4. yara: first tone: H L second tone: H L

Set 3:

1. yọ: H L
2. yọ: H L

3. lọ: H L
4. lọ: H L

5. ṣu: H L
6. ṣu: H L

	7. yara: first tone: H L second tone: H L 8. rara: first tone: H L second tone: H L 9. rara: first tone: H L second tone: H L 10. kuku: first tone: H L second tone: H L
Production Task Worksheet	Directions: Listen to the audio track. The words are listed in the order they are spoken. There will be five sets. Set 1 is first, Set 2 is second, Set 3 is third, etc. Each blank line between words within sets denotes about six seconds before the next word will be spoken. Between each set is about ten seconds. As you listen to the audio track, determine the tones for each word, and check the box that designates the tone you think it is. The track will be played twice. The instructor will not give the answers to these words. **Set 1:** 1. bu: H L 2. bu: H L 3. lu: H L 4. lu: H L 5. kọ: H L 6. kọ: H L 7. koko: first tone: H L second tone: H L 8. koko: first tone: H L second tone: H L 9. bulu: first tone: H L second tone: H L 10. juju: first tone: H L second tone: H L **Set 2:** 1. bu: H L 2. bu: H L

	3. lu: H L 4. lu: H L 5. kọ: H L 6. kọ: H L 7. koko: first tone: H L second tone: H L 8. koko: first tone: H L second tone: H L 9. juju: first tone: H L second tone: H L 10. bulu: first tone: H L second tone: H L **Set 3:** 1. bu: H L 2. bu: H L 3. lu: H L 4. lu: H L 5. kọ: H L 6. kọ: H L 7. koko: first tone: H L second tone: H L 8. koko: first tone: H L second tone: H L 9. bulu: first tone: H L second tone: H L 10. juju: first tone: H L second tone: H L **Set 4:** 1. bu: H L 2. bu: H L

3. lu: H L
4. lu: H L

5. kọ: H L
6. kọ: H L

7. koko: first tone: H L second tone: H L
8. koko: first tone: H L second tone: H L

9. bulu: first tone: H L second tone: H L
10. juju: first tone: H L second tone: H L

Set 5:

1. bu: H L
2. bu: H L

3. lu: H L
4. lu: H L

5. kọ: H L
6. kọ: H L

7. koko: first tone: H L second tone: H L
8. koko: first tone: H L second tone: H L

9. juju: first tone: H L second tone: H L
10. bulu: first tone: H L second tone: H L

Appendix G

```
 SAVE OUTFILE='/Users/abrown08/Amanda/Courses/Independent Studies/2019 '+
    'fall/Elizabeth/Analyses/Elizabeth analyses original.sav'
  /COMPRESSED.
T-TEST PAIRS=OriginalWordsPercentChangePerceptualTraining
    MusiciansOriginalWordsPercentChangePerceptualTraining
    NonMusiciansOriginalWordsPercentChangePerceptualTraining
    OriginalTonesPercentChangePerceptualTraining
    MusiciansOriginalTonesPercent
ChangePerceptualTraining
    NonMusiciansOriginalTonesPercentChangePerceptualTraining WITH
    OriginalWordsPercentChangeMusicTrainingMusiciansOriginalWordsPercen
    tChang
eMusicTraining
    NonMusiciansOriginalWordsPercentChangeMusicTraining
OriginalTonesPercentCh angeMusicTraining
    MusiciansOriginalTonesPercentChangeMusicTraining
    NonMusiciansOriginalTonesPercentChangeMusicTraining (PAIRED)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS.
```

T-Test

Notes

Output Created		**23-APR-2020 10:20:...**
Comments		
Input	**Data**	**/Users/abrown08/Amanda/Courses/Independen t Studies/2019 fall/Elizabeth/Analyses/ Elizabeth analyses original.sav**
	Active Dataset	**DataSet2**
	Filter	**<none>**
	Weight	**<none>**
	Split File	**<none>**
	N of Rows in Working Data File	**16**

Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.

Notes

Syntax		T-TEST PAIRS=OriginalWordsPercentChangePerceptualTraining MusiciansOriginalWordsPercentChangePerceptualTraining NonMusiciansOriginalWordsPercentChangePerceptualTraining OriginalTonesPercentChangePerceptualTraining MusiciansOriginalTonesPercentChangePerceptualTraining NonMusiciansOriginalTonesPercentChangePerceptualTraining WITH OriginalWordsPercentChangeMusicTraining MusiciansOriginalWordsPercentChangeMusicTraining NonMusiciansOriginalWordsPercentChangeMusicTraining OriginalTonesPercentChangeMusicTraining MusiciansOriginalTonesPercentChangeMusicTraining NonMusiciansOriginalTonesPercentChangeMusicTraining (PAIRED) /CRITERIA=CI(.9500) /MISSING=ANALYSIS.
Resources	Processor Time	00:00:00.01
	Elapsed Time	00:00:00.00

```
[DataSet2] /Users/abrown08/Amanda/Courses/Independent Studies/2019
fall/Elizab eth/Analyses/Elizabeth analyses original.sav
```

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	OriginalWordsPercentChangePerceptualTraining	20.9729	14	39.86330	10.65392
	OriginalWordsPercentChangeMusicTraining	-2.2836	14	10.47231	2.79884
Pair 2	MusiciansOriginalWordsPercentChangePerceptualTraining	14.7586	7	35.20101	13.30473
	MusiciansOriginalWordsPercentChangeMusicTraining	-2.9729	7	7.26269	2.74504
Pair 3	NonMusiciansOriginalWordsPercentChangePerceptualTraining	27.1871	7	45.97602	17.37730
	NonMusiciansOriginalWordsPercentChangeMusicTraining	-1.5943	7	13.55587	5.12364
Pair 4	OriginalTonesPercentChangePerceptualTraining	15.3379	14	27.81384	7.43356
	OriginalTonesPercentChangeMusicTraining	-3.2300	14	13.53285	3.61681
Pair 5	MusiciansOriginalTonesPercentChangePerceptualTraining	11.5243	7	30.17681	11.40576
	MusiciansOriginalTonesPercentChangeMusicTraining	-2.8500	7	7.13528	2.69688
Pair 6	NonMusiciansOriginalTonesPercentChangePerceptualTraining	19.1514	7	27.04775	10.22309
	NonMusiciansOriginalTonesPercentChangeMusicTraining	-3.6100	7	18.58897	7.02597

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	OriginalWordsPercentChangePerceptualTraining & OriginalWordsPercentChangeMusicTraining	14	.723	.003
Pair 2	MusiciansOriginalWordsPercentChangePerceptualTraining & MusiciansOriginalWordsPercentChangeMusicTraining	7	.292	.525
Pair 3	NonMusiciansOriginalWordsPercentChangePerceptualTraining & NonMusiciansOriginalWordsPercentChangeMusicTraining	7	.914	.004
Pair 4	OriginalTonesPercentChangePerceptualTraining & OriginalTonesPercentChangeMusicTraining	14	.511	.062
Pair 5	MusiciansOriginalTonesPercentChangePerceptualTraining & MusiciansOriginalTonesPercentChangeMusicTraining	7	.300	.513
Pair 6	NonMusiciansOriginalTonesPercentChangePerceptualTraining & NonMusiciansOriginalTonesPercentChangeMusicTraining	7	.707	.076

Paired Samples Test

		Paired Differences			
		Mean	Std. Deviation	Std. Error Mean	95% Confidence ... Lower
Pair 1	OriginalWordsPercentChangePerceptualTraining - OriginalWordsPercentChangeMusicTraining	23.25643	33.08920	8.84346	4.15129
Pair 2	MusiciansOriginalWordsPercentChangePerceptualTraining - MusiciansOriginalWordsPercentChangeMusicTraining	17.73143	33.79981	12.77513	-13.52818
Pair 3	NonMusiciansOriginalWordsPercentChangePerceptualTraining - NonMusiciansOriginalWordsPercentChangeMusicTraining	28.78143	34.03851	12.86535	-2.69894
Pair 4	OriginalTonesPercentChangePerceptualTraining - OriginalTonesPercentChangeMusicTraining	18.56786	23.92057	6.39304	4.75653
Pair 5	MusiciansOriginalTonesPercentChangePerceptualTraining - MusiciansOriginalTonesPercentChangeMusicTraining	14.37429	28.84863	10.90376	-12.30625
Pair 6	NonMusiciansOriginalTonesPercentChangePerceptualTraining - NonMusiciansOriginalTonesPercentChangeMusicTraining	22.76143	19.14354	7.23558	5.05661

Paired Samples Test

		Paired ... 95% Confidence Interval of the .. Upper	t	df	Sig. (2-tailed)
Pair 1	OriginalWordsPercentChangePerceptualTraining - OriginalWordsPercentChangeMusicTraining	42.36157	2.630	13	.021
Pair 2	rcentChangePerceptualTraining - rcentChangeMusicTraining	48.99104	1.388	6	.214
Pair 3	NonMusiciansOriginalWordsPercentChangePerceptualTraining - NonMusiciansOriginalWordsPercentChangeMusicTraining	60.26180	2.237	6	.067
Pair 4	OriginalTonesPercentChangePerceptualTraining - OriginalTonesPercentChangeMusicTraining	32.37918	2.904	13	.012
Pair 5	MusiciansOriginalTonesPercentChangePerceptualTraining - MusiciansOriginalTonesPercentChangeMusicTraining	41.05482	1.318	6	.235
Pair 6	NonMusiciansOriginalTonesPercentChangePerceptualTraining - NonMusiciansOriginalTonesPercentChangeMusicTraining	40.46625	3.146	6	.020

```
T-TEST GROUPS=Musician(0 1)
  /MISSING=ANALYSIS
  /VARIABLES=OriginalWordsPercentChangePerceptualTraining OriginalWordsPercent ChangeMusicTraining
    OriginalTonesPercentChangePerceptualTraining OriginalTonesPercentChange Mus icTraining
  /CRITERIA=CI(.95).
```

T-Test

Notes

Output Created		23-APR-2020 10:26:...
Comments		
Input	Data	/Users/abrown08/Amanda/Courses/Independen t Studies/2019 fall/Elizabeth/Analyses/ Elizabeth analyses original.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	16
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST GROUPS=Musician(0 1) /MISSING=ANALYSIS /VARIABLES=OriginalWordsPercentChangePerceptualTraining OriginalWordsPercentChangeMusicTraining OriginalTonesPercentChangePerceptualTraining OriginalTonesPercentChangeMusicTraining /CRITERIA=CI(.95).
Resources	Processor Time	00:00:00.01
	Elapsed Time	00:00:00.00

Group Statistics

	Musician	N	Mean	Std. Deviation	Std. Error Mean
OriginalWordsPercentChangePerceptualTraining	Non Musician	7	27.1871	45.97602	17.37730
	Musician	7	14.7586	35.20101	13.30473
OriginalWordsPercentChangeMusicTraining	Non Musician	7	-1.5943	13.55587	5.12364
	Musician	7	-2.9729	7.26269	2.74504
OriginalTonesPercentChangePerceptualTraining	Non Musician	7	19.1514	27.04775	10.22309
	Musician	7	11.5243	30.17681	11.40576
OriginalTonesPercentChangeMusicTraining	Non Musician	7	-3.6100	18.58897	7.02597
	Musician	7	-2.8500	7.13528	2.69688

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality ..
		F	Sig.	t
OriginalWordsPercentChangePerceptualTraining	Equal variances assumed	.712	.415	.568
	Equal variances not assumed			.568
OriginalWordsPercentChangeMusicTraining	Equal variances assumed	1.590	.231	.237
	Equal variances not assumed			.237
OriginalTonesPercentChangePerceptualTraining	Equal variances assumed	.009	.927	.498
	Equal variances not assumed			.498
OriginalTonesPercentChangeMusicTraining	Equal variances assumed	3.892	.072	-.101
	Equal variances not assumed			-.101

Independent Samples Test

		t-test for Equality of Means		
		df	Sig. (2-tailed)	Mean Difference
OriginalWordsPercentChangePerceptualTraining	Equal variances assumed	12	.581	12.42857
	Equal variances not assumed	11.235	.581	12.42857
OriginalWordsPercentChangeMusicTraining	Equal variances assumed	12	.817	1.37857
	Equal variances not assumed	9.182	.818	1.37857
OriginalTonesPercentChangePerceptualTraining	Equal variances assumed	12	.628	7.62714
	Equal variances not assumed	11.859	.628	7.62714
OriginalTonesPercentChangeMusicTraining	Equal variances assumed	12	.921	-.76000
	Equal variances not assumed	7.730	.922	-.76000

Independent Samples Test

		t-test for Equality of Means		
			95% Confidence Interval of the Difference	
		Std. Error Difference	Lower	Upper
OriginalWordsPercentChangePerceptualTraining	Equal variances assumed	21.88576	-35.25641	60.11355
	Equal variances not assumed	21.88576	-35.61883	60.47597
OriginalWordsPercentChangeMusicTraining	Equal variances assumed	5.81265	-11.28610	14.04324
	Equal variances not assumed	5.81265	-11.73087	14.48801
OriginalTonesPercentChangePerceptualTraining	Equal variances assumed	15.31675	-25.74519	40.99948
	Equal variances not assumed	15.31675	-25.78925	41.04354
OriginalTonesPercentChangeMusicTraining	Equal variances assumed	7.52579	-17.15728	15.63728
	Equal variances not assumed	7.52579	-18.22038	16.70038

Appendix H

```
GLM
RevisedWordPretest1ScoreRevisedWordPosttest1ScoreRevisedWordPosttest2Sco
re
RevisedWordGeneralizationTestScoreBY Musician
/WSFACTOR=time 4 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(time*Musician) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AU
TO
/EMMEANS=TABLES(time) COMPARE ADJ(BONFERRONI)
/PRINT=DESCRIPTIVEETASQ HOMOGENEITY
/CRITERIA=ALPHA(.05)
/WSDESIGN=time
/DESIGN=Musician.
```

General Linear Model

Notes

Output Created		30-APR-2020 13:05:53
Comments		
Input	Data	C:\Users\eaelt\Documents\School\Grad School\2020 Spring\Thesis\Research Materials\Results\Statistical Tests\Research Question 3.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	16
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data for all variables in the model.

Notes

Syntax		GLM RevisedWordPretest1Scor e RevisedWordPosttest1Sco re RevisedWordPosttest2Sco re RevisedWordGeneralizatio nTestScore BY Musician /WSFACTOR=time 4 Polynomial /METHOD=SSTYPE(3) /PLOT=PROFILE (time*Musician) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO /EMMEANS=TABLES (time) COMPARE ADJ (BONFERRONI) /PRINT=DESCRIPTIVE ETASQ HOMOGENEITY /CRITERIA=ALPHA(.05) /WSDESIGN=time /DESIGN=Musician.
Resources	Processor Time	00:00:00.19
	Elapsed Time	00:00:00.14

Within-Subjects Factors

Measure: MEASURE_1

time	Dependent Variable
1	RevisedWordPretest1Score
2	RevisedWordPosttest1Score
3	RevisedWordPosttest2Score
4	RevisedWordGeneralizationTestScore

Between-Subjects Factors

Value Label			N
Musician	.00	Nonmusician	7
	1.00	Musician	7

Descriptive Statistics

Musician		Mean	Std. Deviation	N
RevisedWordPretest1Score	Nonmusician	48.2143	19.66989	7
	Musician	73.2143	23.58338	7
	Total	60.7143	24.56699	14
RevisedWordPosttest1Score	Nonmusician	69.6429	22.94371	7
	Musician	89.2857	13.36306	7
	Total	79.4643	20.71855	14
RevisedWordPosttest2Score	Nonmusician	77.6786	20.36680	7
	Musician	91.0714	13.43248	7
	Total	84.3750	17.97267	14
RevisedWordGeneralizationTestScore	Nonmusician	34.1814	4.54440	7
	Musician	49.4886	9.54426	7
	Total	41.8350	10.70782	14

Box's Test of Equality of Covariance Matrices[a]

Box's M	12.220
F	.766
df1	10
df2	688.446
Sig.	.661

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Musician
Within Subjects Design: time

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	.880	24.337[b]	3.000	10.000	.000
	Wilks' Lambda	.120	24.337[b]	3.000	10.000	.000
	Hotelling's Trace	7.301	24.337[b]	3.000	10.000	.000
	Roy's Largest Root	7.301	24.337[b]	3.000	10.000	.000
time * Musician	Pillai's Trace	.176	.710[b]	3.000	10.000	.568
	Wilks' Lambda	.824	.710[b]	3.000	10.000	.568
	Hotelling's Trace	.213	.710[b]	3.000	10.000	.568
	Roy's Largest Root	.213	.710[b]	3.000	10.000	.568

Multivariate Tests[a]

Effect		Partial Eta Squared
time	Pillai's Trace	.880
	Wilks' Lambda	.880
	Hotelling's Trace	.880
	Roy's Largest Root	.880
time * Musician	Pillai's Trace	.176
	Wilks' Lambda	.176
	Hotelling's Trace	.176
	Roy's Largest Root	.176

a. Design: Intercept + Musician Within Subjects Design: time

b. Exact statistic

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]
					Greenhouse-Geisser
time	.319	12.241	5	.032	.662

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Epsilon[b]	
	Huynh-Feldt	Lower-bound
time	.858	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Musician Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure:

MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F
time	Sphericity Assumed	15811.42	3	5270.47	33.
	Greenhouse-Geisser	15811.42	1.985	7966.91	33.
	Huynh-Feldt	15811.42	2.575	6141.47	33.
	Lower-bound	15811.42	1.000	15811.4	33.
time * Musician	Sphericity Assumed	279.039	3	93.013	.58
	Greenhouse-Geisser	279.039	1.985	140.600	.58
	Huynh-Feldt	279.039	2.575	108.384	.58
	Lower-bound	279.039	1.000	279.039	.58
Error(time)	Sphericity Assumed	5704.877	36	158.469	
	Greenhouse-Geisser	5704.877	23.816	239.544	
	Huynh-Feldt	5704.877	30.894	184.657	
	Lower-bound	5704.877	12.000	475.406	

Tests of Within-Subjects Effects

Measure:

MEASURE_1

Source		Sig.	Partial Eta Squared
time	Sphericity Assumed	.000	.735
	Greenhouse-Geisser	.000	.735
	Huynh-Feldt	.000	.735
	Lower-bound	.000	.735
time * Musician	Sphericity Assumed	.627	.047
	Greenhouse-Geisser	.563	.047
	Huynh-Feldt	.603	.047
	Lower-bound	.458	.047
Error(time)	Sphericity Assumed		
	Greenhouse-Geisser		
	Huynh-Feldt		

	Lower-bound		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	1872.988	1	1872.98	13.	.00
	Quadratic	13147.62	1	13147.6	43.	.00
	Cubic	790.810	1	790.810	24.	.00
time * Musician	Linear	218.419	1	218.419	1.5	.24
	Quadratic	46.264	1	46.264	.15	.70
	Cubic	14.356	1	14.356	.45	.51
Error(time)	Linear	1720.821	12	143.402		
	Quadratic	3602.911	12	300.243		
	Cubic	381.145	12	31.762		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	Partial Eta Squared
time	Linear	.521
	Quadratic	.785
	Cubic	.675
time * Musician	Linear	.113
	Quadratic	.013
	Cubic	.036
Error(time)	Linear	
	Quadratic	
	Cubic	

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
RevisedWordPretest1Score	Based on Mean	.308	1	12	.589
	Based on Median	.271	1	12	.612
	Based on Median and with adjusted df	.271	1	10.800	.613
	Based on trimmed mean	.312	1	12	.587
RevisedWordPosttest1Score	Based on Mean	2.848	1	12	.117
	Based on Median	2.148	1	12	.168
	Based on Median and with adjusted df	2.148	1	11.559	.169
	Based on trimmed mean	2.932	1	12	.113
RevisedWordPosttest2Score	Based on Mean	5.718	1	12	.034
	Based on Median	1.829	1	12	.201
	Based on Median and with adjusted df	1.829	1	11.261	.203
	Based on trimmed mean	5.723	1	12	.034
RevisedWordGeneralizationTestScore	Based on Mean	2.066	1	12	.176
	Based on Median	2.085	1	12	.174
	Based on Median and with adjusted df	2.085	1	10.090	.179
	Based on trimmed mean	1.877	1	12	.196

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Musician Within Subjects Design: time

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	248370.048	1	248370.048	353.744	.000	.967
Musician	4706.778	1	4706.778	6.704	.024	.358
Error	8425.411	12	702.118			

Estimated Marginal Means time

Estimates

Measure: MEASURE_1

time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	60.714	5.804	48.069	73.359
2	79.464	5.018	68.531	90.397
3	84.375	4.611	74.329	94.421
4	41.835	1.998	37.482	46.188

Pairwise Comparisons

Measure: MEASURE_1

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-18.750*	5.400	.028	-35.776	-1.724
	3	-23.661*	5.670	.008	-41.538	-5.784
	4	18.879*	4.492	.007	4.719	33.040
2	1	18.750*	5.400	.028	1.724	35.776
	3	-4.911	2.202	.274	-11.853	2.032
	4	37.629*	5.131	.000	21.454	53.805
3	1	23.661*	5.670	.008	5.784	41.538
	2	4.911	2.202	.274	-2.032	11.853
	4	42.540*	4.813	.000	27.366	57.714
4	1	-18.879*	4.492	.007	-33.040	-4.719
	2	-37.629*	5.131	.000	-53.805	-21.454
	3	-42.540*	4.813	.000	-57.714	-27.366

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

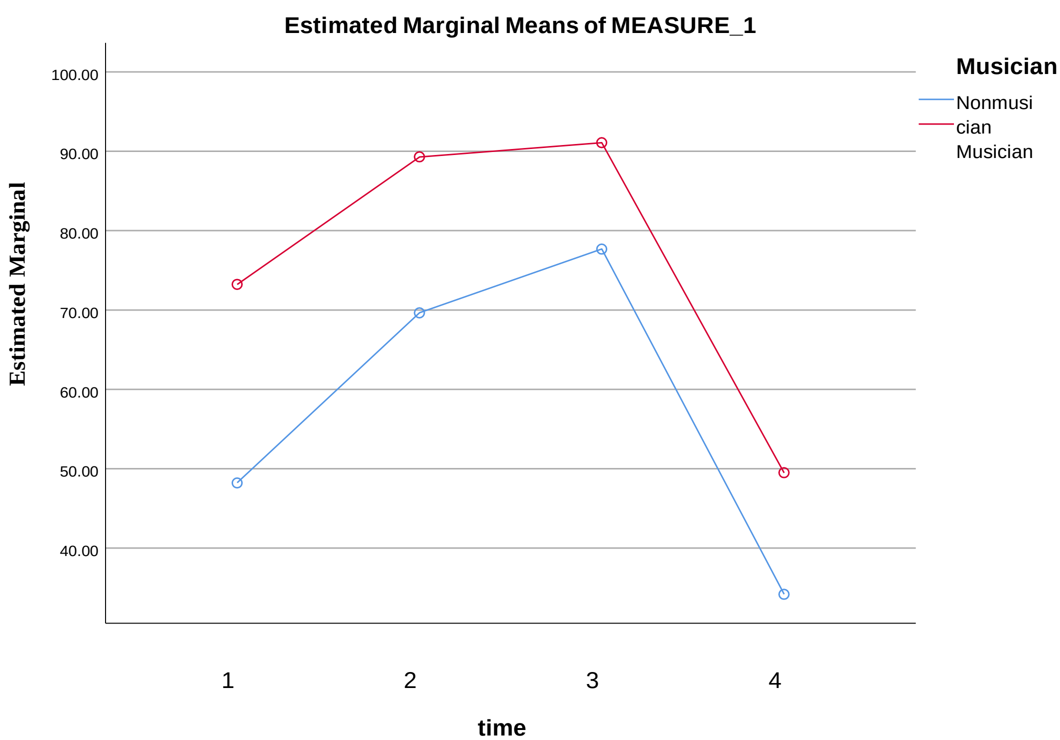
Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.880	24.337[a]	3.000	10.000	.000	.880
Wilks' lambda	.120	24.337[a]	3.000	10.000	.000	.880
Hotelling's trace	7.301	24.337[a]	3.000	10.000	.000	.880
Roy's largest root	7.301	24.337[a]	3.000	10.000	.000	.880

Each F tests the multivariate effect of time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Profile Plots




```
GLM
RevisedTonePretest1ScoreRevisedTonePosttest1ScoreRevisedTonePosttest2Sco
re
RevisedToneGeneralizationTestScoreBY Musician
```

```
/WSFACTOR=time 4 Polynomial
/METHOD=SSTYPE(3)
/PLOT=PROFILE(time*Musician) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AU
```

```
TO
/EMMEANS=TABLES(time) COMPARE ADJ(BONFERRONI)
/PRINT=DESCRIPTIVEETASQ HOMOGENEITY
/CRITERIA=ALPHA(.05)
/WSDESIGN=time
/DESIGN=Musician.
```

General Linear Model

Notes

Output Created		30-APR-2020 13:43:04
Comments		
Input	Data	C:\Users\eaelt\Documents\School\Grad School\2020 Spring\Thesis\Research Materials\Results\Statistical Tests\Research Question 3.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	16
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data for all variables in the model.

Notes

Syntax		GLM RevisedTonePretest1Scor e RevisedTonePosttest1Sco re RevisedTonePosttest2Sco re RevisedToneGeneralizatio nTestScore BY Musician /WSFACTOR=time 4 Polynomial /METHOD=SSTYPE(3) /PLOT=PROFILE (time*Musician) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO /EMMEANS=TABLES (time) COMPARE ADJ (BONFERRONI) /PRINT=DESCRIPTIVE ETASQ HOMOGENEITY /CRITERIA=ALPHA(.05) /WSDESIGN=time /DESIGN=Musician.
Resources	Processor Time	00:00:00.19
	Elapsed Time	00:00:00.16

Within-Subjects Factors

Measure: MEASURE_1

time	Dependent Variable
1	RevisedTonePretest1Score
2	RevisedTonePosttest1Score
3	RevisedTonePosttest2Score
4	RevisedToneGeneralizationTestScore

Between-Subjects Factors

	Value Label		N
Musician	.00	Nonmusician	7
	1.00	Musician	7

Descriptive Statistics

Musician		Mean	Std. Deviation	N
RevisedTonePretest1Score	Nonmusician	55.8429	17.74449	7
	Musician	81.1686	16.47945	7
	Total	68.5057	21.05580	14
RevisedTonePosttest1Score	Nonmusician	71.4257	20.78230	7
	Musician	90.9071	11.73579	7
	Total	81.1664	19.10726	14
RevisedTonePosttest2Score	Nonmusician	77.9229	19.36079	7
	Musician	92.2071	13.04468	7
	Total	85.0650	17.50641	14
RevisedToneGeneralizationTestScore	Nonmusician	45.3571	6.02574	7
	Musician	52.1429	11.22020	7
	Total	48.7500	9.34129	14

Box's Test of Equality of Covariance Matrices[a]

Box's M	9.524
F	.597
df1	10
df2	688.446
Sig.	.817

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Musician
Within Subjects Design: time

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	.842	17.736[b]	3.000	10.000	.000
	Wilks' Lambda	.158	17.736[b]	3.000	10.000	.000
	Hotelling's Trace	5.321	17.736[b]	3.000	10.000	.000
	Roy's Largest Root	5.321	17.736[b]	3.000	10.000	.000
time * Musician	Pillai's Trace	.437	2.592[b]	3.000	10.000	.111
	Wilks' Lambda	.563	2.592[b]	3.000	10.000	.111
	Hotelling's Trace	.778	2.592[b]	3.000	10.000	.111
	Roy's Largest Root	.778	2.592[b]	3.000	10.000	.111

Multivariate Tests[a]

Effect		Partial Eta Squared
time	Pillai's Trace	.842
	Wilks' Lambda	.842
	Hotelling's Trace	.842
	Roy's Largest Root	.842
time * Musician	Pillai's Trace	.437
	Wilks' Lambda	.437
	Hotelling's Trace	.437
	Roy's Largest Root	.437

a. Design: Intercept + Musician Within Subjects Design: time

b. Exact statistic

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser
time	.299	12.928	5	.025	.622

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Epsilon[b] Huynh-Feldt	Lower-bound
time	.793	.333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Musician Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F
time	Sphericity Assumed	11233.582	3	3744.527	33.636
	Greenhouse-Geisser	11233.582	1.865	6024.419	33.636
	Huynh-Feldt	11233.582	2.378	4723.248	33.636
	Lower-bound	11233.582	1.000	11233.582	33.636
time * Musician	Sphericity Assumed	651.193	3	217.064	1.950
	Greenhouse-Geisser	651.193	1.865	349.226	1.950
	Huynh-Feldt	651.193	2.378	273.799	1.950
	Lower-bound	651.193	1.000	651.193	1.950
Error(time)	Sphericity Assumed	4007.747	36	111.326	
	Greenhouse-Geisser	4007.747	22.376	179.108	
	Huynh-Feldt	4007.747	28.540	140.424	
	Lower-bound	4007.747	12.000	333.979	

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Sig.	Partial Eta Squared
time	Sphericity Assumed	.000	.737
	Greenhouse-Geisser	.000	.737
	Huynh-Feldt	.000	.737
	Lower-bound	.000	.737
time * Musician	Sphericity Assumed	.139	.140
	Greenhouse-Geisser	.168	.140
	Huynh-Feldt	.154	.140
	Lower-bound	.188	.140
Error(time)	Sphericity Assumed		
	Greenhouse-Geisser		
	Huynh-Feldt		
	Lower-bound		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	2145.975	1	2145.97	26.753	.000
	Quadratic	8395.172	1	8395.17	37.285	.000
	Cubic	692.435	1	692.435	24.207	.000
time * Musician	Linear	647.277	1	647.277	8.069	.015
	Quadratic	2.395	1	2.395	.011	.920
	Cubic	1.521	1	1.521	.053	.821
Error(time)	Linear	962.566	12	80.214		
	Quadratic	2701.924	12	225.160		
	Cubic	343.257	12	28.605		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	Partial Eta Squared
time	Linear	.690
	Quadratic	.757
	Cubic	.669
time * Musician	Linear	.402
	Quadratic	.001
	Cubic	.004
Error(time)	Linear	
	Quadratic	
	Cubic	

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
RevisedTonePretest1Score	Based on Mean	.032	1	12	.862
	Based on Median	.000	1	12	1.000
	Based on Median and with adjusted df	.000	1	10.138	1.000
	Based on trimmed mean	.029	1	12	.868
RevisedTonePosttest1Score	Based on Mean	3.340	1	12	.093
	Based on Median	3.259	1	12	.096
	Based on Median and with adjusted df	3.259	1	12.000	.096
	Based on trimmed mean	3.463	1	12	.087
RevisedTonePosttest2Score	Based on Mean	5.277	1	12	.040
	Based on Median	1.213	1	12	.292
	Based on Median and with adjusted df	1.213	1	10.329	.296
	Based on trimmed mean	5.287	1	12	.040
RevisedToneGeneralizationTestScore	Based on Mean	1.723	1	12	.214
	Based on Median	.544	1	12	.475
	Based on Median and with adjusted df	.544	1	7.427	.483
	Based on trimmed mean	1.410	1	12	.258

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Musician Within Subjects Design: time

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	281277.361	1	281277.361	470.631	.000	.975
Musician	3797.323	1	3797.323	6.354	.027	.346
Error	7171.921	12	597.660			

Estimated Marginal Means time

Estimates

Measure: MEASURE_1

time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	68.506	4.576	58.534	78.477
2	81.166	4.510	71.339	90.994
3	85.065	4.412	75.452	94.678
4	48.750	2.407	43.506	53.994

Pairwise Comparisons

Measure: MEASURE_1

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-12.661	4.216	.066	-25.953	.631
	3	-16.559*	4.187	.011	-29.760	-3.358
	4	19.756*	3.529	.001	8.630	30.882
2	1	12.661	4.216	.066	-.631	25.953
	3	-3.899	1.758	.280	-9.442	1.644
	4	32.416*	4.723	.000	17.526	47.307
3	1	16.559*	4.187	.011	3.358	29.760
	2	3.899	1.758	.280	-1.644	9.442
	4	36.315*	4.718	.000	21.440	51.190
4	1	-19.756*	3.529	.001	-30.882	-8.630
	2	-32.416*	4.723	.000	-47.307	-17.526
	3	-36.315*	4.718	.000	-51.190	-21.440

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

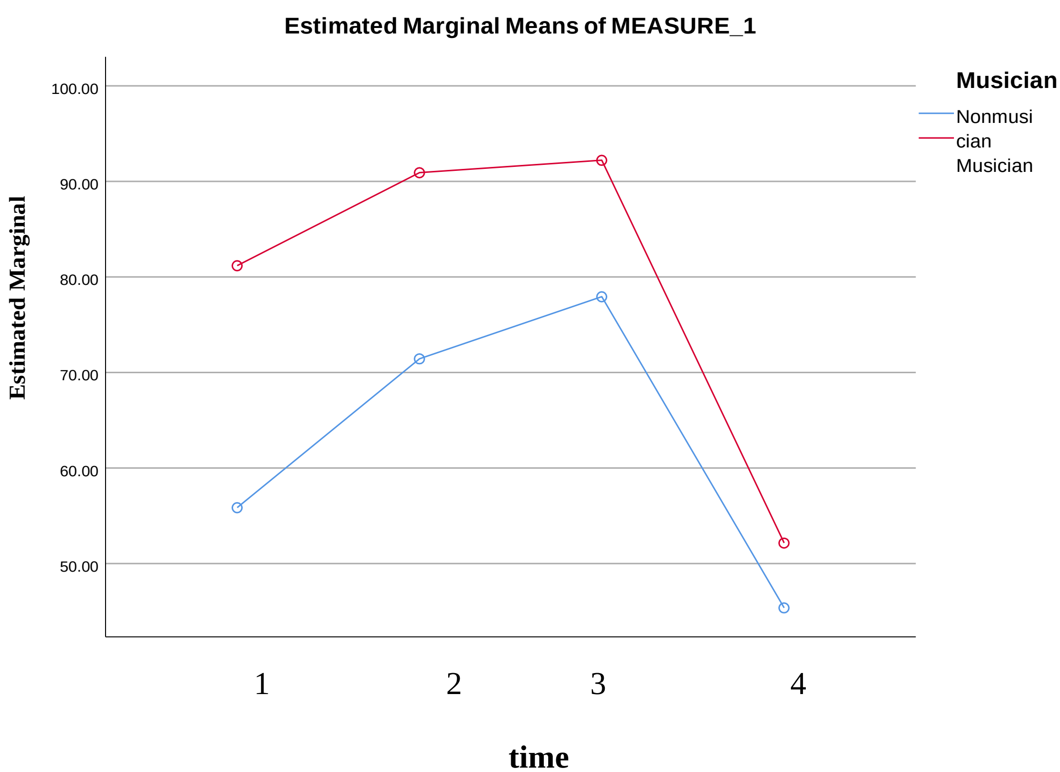
Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.842	17.736[a]	3.000	10.000	.000	.842
Wilks' lambda	.158	17.736[a]	3.000	10.000	.000	.842
Hotelling's trace	5.321	17.736[a]	3.000	10.000	.000	.842
Roy's largest root	5.321	17.736[a]	3.000	10.000	.000	.842

Each F tests the multivariate effect of time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Profile Plots



```
T-TEST GROUPS=Musician(0 1)
/MISSING=ANALYSIS
/VARIABLES=RevisedWordPretest1ScoreRevisedWordPosttest1ScoreRevisedWordP
os ttest2Score
  RevisedWordGeneralizationTestScoreRevisedTonePretest1ScoreRevisedTone
Pos ttest1Score
  RevisedTonePosttest2ScoreRevisedToneGeneralizationTestScore
/CRITERIA=CI(.95).
```

T-Test

Output Created		30-APR-2020 19:06:32
Comments		
Input	Data	C: \Users\eaelt\Documents\School\Grad School\2020 Spring\Thesis\Research Materials\Results\Statistical Tests\Research Question 3 and 1.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	16
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
Syntax		T-TEST GROUPS=Musician(0 1) /MISSING=ANALYSIS /VARIABLES=RevisedWordPretest1Score RevisedWordPosttest1Score RevisedWordPosttest2Score RevisedWordGeneralizationTestScore RevisedTonePretest1Score RevisedTonePosttest1Score RevisedTonePosttest2Score RevisedToneGeneralizationTestScore /CRITERIA=CI(.95).

Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.05

Group Statistics

	Musician	N	Mean	Std. Deviation	Std. Error Mean
RevisedWordPretest1Score	Nonmusician	7	48.2143	19.66989	7.43452
	Musician	7	73.2143	23.58338	8.91368
RevisedWordPosttest1Score	Nonmusician	7	69.6429	22.94371	8.67191
	Musician	7	89.2857	13.36306	5.05076
RevisedWordPosttest2Score	Nonmusician	7	77.6786	20.36680	7.69793
	Musician	7	91.0714	13.43248	5.07700
RevisedWordGeneralizationTestScore	Nonmusician	7	34.1814	4.54440	1.71762
	Musician	7	49.4886	9.54426	3.60739
RevisedTonePretest1Score	Nonmusician	7	55.8429	17.74449	6.70679
	Musician	7	81.1686	16.47945	6.22865
RevisedTonePosttest1Score	Nonmusician	7	71.4257	20.78230	7.85497
	Musician	7	90.9071	11.73579	4.43571
RevisedTonePosttest2Score	Nonmusician	7	77.9229	19.36079	7.31769
	Musician	7	92.2071	13.04468	4.93043
RevisedToneGeneralizationTestScore	Nonmusician	7	45.3571	6.02574	2.27752
	Musician	7	52.1429	11.22020	4.24084

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of .
		F	Sig.	t
RevisedWordPretest1Score	Equal variances assumed	.308	.589	-2.154
	Equal variances not assumed			-2.154
RevisedWordPosttest1Score	Equal variances assumed	2.848	.117	-1.957
	Equal variances not assumed			-1.957
RevisedWordPosttest2Score	Equal variances assumed	5.718	.034	-1.452
	Equal variances not assumed			-1.452
RevisedWordGeneralizationTestScore	Equal variances assumed	2.066	.176	-3.831
	Equal variances not assumed			-3.831
RevisedTonePretest1Score	Equal variances assumed	.032	.862	-2.767
	Equal variances not assumed			-2.767
RevisedTonePosttest1Score	Equal variances assumed	3.340	.093	-2.160
	Equal variances not assumed			-2.160
RevisedTonePosttest2Score	Equal variances assumed	5.277	.040	-1.619
	Equal variances not assumed			-1.619
RevisedToneGeneralizationTestScore	Equal variances assumed	1.723	.214	-1.410
	Equal variances not assumed			-1.410

		t-test for Equality of Means		
		df	Sig. (2-tailed)	Mean Difference
RevisedWordPretest1Score	Equal variances assumed	12	.052	-25.00000
	Equal variances not assumed	11.625	.053	-25.00000
RevisedWordPosttest1Score	Equal variances assumed	12	.074	-19.64286
	Equal variances not assumed	9.651	.080	-19.64286
RevisedWordPosttest2Score	Equal variances assumed	12	.172	-13.39286
	Equal variances not assumed	10.389	.176	-13.39286
RevisedWordGeneralizationTestScore	Equal variances assumed	12	.002	-15.30714
	Equal variances not assumed	8.588	.004	-15.30714
RevisedTonePretest1Score	Equal variances assumed	12	.017	-25.32571
	Equal variances not assumed	11.935	.017	-25.32571
RevisedTonePosttest1Score	Equal variances assumed	12	.052	-19.48143
	Equal variances not assumed	9.473	.058	-19.48143
RevisedTonePosttest2Score	Equal variances assumed	12	.131	-14.28429
	Equal variances not assumed	10.517	.135	-14.28429
RevisedToneGeneralizationTestScore	Equal variances assumed	12	.184	-6.78571
	Equal variances not assumed	9.195	.192	-6.78571

		t-test for Equality of Means		
		Std. Error Difference	95% Confidence Interval of the Difference	
			Lower	Upper
RevisedWordPretest1Score	Equal variances assumed	11.60714	-50.28979	.28979
	Equal variances not assumed	11.60714	-50.38045	.38045
RevisedWordPosttest1Score	Equal variances assumed	10.03554	-41.50843	2.22272
	Equal variances not assumed	10.03554	-42.11367	2.82796
RevisedWordPosttest2Score	Equal variances assumed	9.22139	-33.48454	6.69882
	Equal variances not assumed	9.22139	-33.83552	7.04981
RevisedWordGeneralizationTestScore	Equal variances assumed	3.99543	-24.01244	-6.60184
	Equal variances not assumed	3.99543	-24.41205	-6.20223
RevisedTonePretest1Score	Equal variances assumed	9.15298	-45.26834	-5.38309
	Equal variances not assumed	9.15298	-45.28040	-5.37102
RevisedTonePosttest1Score	Equal variances assumed	9.02087	-39.13621	.17336
	Equal variances not assumed	9.02087	-39.73365	.77079
RevisedTonePosttest2Score	Equal variances assumed	8.82370	-33.50948	4.94091
	Equal variances not assumed	8.82370	-33.81456	5.24599
RevisedToneGeneralizationTestScore	Equal variances assumed	4.81371	-17.27388	3.70245
	Equal variances not assumed	4.81371	-17.63994	4.06851

References

Akinlabi, A., & Liberman, M. (2000). The tonal phonology of Yoruba clitics. In B. Gerlach, & J. Grijzenhout (Eds.), *Clitics in phonology, morphology and syntax* (pp. 31 – 62). Amsterdam: John Benjamins.

Alexander, J. A., Wang, P. C. M., & Bradlow, A. R. (2005). *Lexical tone perception in musicians and non-musicians*. 397-400. Paper presented at 9th European Conference on Speech Communication and Technology, Lisbon, Portugal.

Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271–278.

Awobuluyi, A. O. (1978). *Essentials of Yoruba grammar*. Ibadan: Oxford University Press Nigeria.

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2018). Observing And Producing Pitch Gestures Facilitates The Learning Of Mandarin Chinese Tones And Words. *Studies in Second Language Acquisition*, *41*(1), 33–58.

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *The Journal of the Acoustical Society of America*, *138*(6), 3703–3716.

Chobert, J., & Besson, M. (2013). Musical Expertise and Second Language Learning. *Brain Sciences*, 3(4), 923–940.

Criado, R. (2013). A critical review of the Presentation-Practice-Production Model (PPP) in Foreign Language Teaching. In R. Monroy (Ed.), *Homenaje a Francisco Gutierrez Diez* (pp. 97-115). Murcia: Edit.um.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage Differences in Tone Perception: a Multidimensional Scaling Investigation. *Language and Speech*, *21*(1), 1–33.

Godfroid, A., Lin, C.-H., & Ryu, C. (2017). Hearing and Seeing Tone Through Color: An Efficacy Study of Web-Based, Multimodal Chinese Tone Perception Training. *Language Learning*, *67*(4), 819–857.

Good, J. (2012). How to become a “Kwa” noun. *Morphology*, *22*(2), 293–335.

Granena, G. and Long, M. (2012). ‘Age of onset, length of residence, language aptitude and ultimate L2 attainment in three linguistic domains’. *Second Language Research*, 29, 311–43.

Guan, S. (2014). Internet-Based Technology Use in Second Language Learning. *International Journal of Cyber Behavior, Psychology and Learning*, *4*(4), 69–81.

Gussenhoven, C. (2015). International Encyclopedia of Social and Behavioral Sciences Second Edition. In J. D. Wright (Ed.), *International Encyclopedia of Social and Behavioral Sciences Second Edition* (2nd ed., Vol. 23, pp. 714–721). Oxford: Elsevier.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(02), 225–257.

Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Sciences*, *31*, 213-238.

Inceoglu, S. (2015). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199.

Jiang, Y. (2017). Examining the auditory approach: Lexical effects in the perceptual judgment of Chinese L2 tone production. *Chinese as a Second Language Research*, *6*(2).

Kim, H. Y. (2014). Learning opportunities in synchronous computer-mediated communication and face-to-face interaction. *Computer Assisted Language Learning*, *27*(1), 26–43.

Kirkham, J., Lu, S., Wayland, R., & Kaan, E. (2011). Comparison of Vocalists and Instrumentalists on Lexical Tone Perception and Production Tasks. Proceedings from *ICPhS*.

Leben, W. (1971). Suprasegmental and Segmental Representation of Tone. *Studies in African Linguistics*, supplement 2, 183-200.

Levis, J. (2007). Computer Technology In Teaching And Researching Pronunciation. *Annual Review of Applied Linguistics*, 27.

Li, S. (2010). The effectiveness of corrective feedback in SLA: a meta-analysis. *Language Learning*, 60, 309–65.

Li, M., & DeKeyser, R. (2017). Perception Practice, Production Practice, And Musical Ability In L2 Mandarin Tone-Word Learning. *Studies in Second Language Acquisition*, 39(4), 593–620.

Li, M., & Dekeyser, R. (2019). Distribution of Practice Effects in the Acquisition and Retention of L2 Mandarin Tonal Word Production. *The Modern Language Journal*.

Li, B., Shao, J., & Bao, M. (2017). Effects of Phonetic Similarity in the Identification of Mandarin Tones. *Journal of Psycholinguistic Research*, *46*(1), 107–124.

Lin, W. C. (1985). Teaching Mandarin tones to adult English speakers: Analysis of difficulties with suggested remedies. *RELC Journa*l, 16(2), 31-47.

Littlewood, W. (2011). Communicative language teaching: An expanding concept for a changing world. In E. Hinkel (Ed.), *Perspectives on pedagogical grammar* (Vol. 2, pp. 541-557). New York, NY: Routledge

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

874–886.

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research*, *1624*, 28–44.

Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning*, *29*(3), 451–476.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Brocas area: an MEG study. *Nature Neuroscience*, *4*(5), 540–545.

Mennen, I., & Leeuw, E. D. (2014). Beyond Segments. *Studies in Second Language Acquisition*, *36*(2), 183–194.

Moneypenny, D. B., & Aldrich, R. S. (2018). Developing Oral Proficiency in Spanish across Class Modalities. *CALICO Journal*, *35*(3), 257–273.

Patel, A. D. (2011). Language, music, and the brain: a resourcesharing framework. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 1–25). Oxford: Oxford.

Paterson III, H. J. (2015) Phonetic transcription of tone in the IPA. In The Scottish Consortium for ICPhS 2015 (Ed.). *Proceedings of the 18th International Conference of Phonetic Sciences* Paper number 507.1-5. University of Glasgow: Glasgow, UK.

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, 127(2).

Pei, Z., Wu, Y., Xiang, X., & Qian, H. (2016). The Effects of Musical Aptitude and Musical Training on Phonological Production in Foreign Languages. *English Language Teaching*, *9*(6), 19.

Perfors, A., & Ong, J. H. (2012). Musicians are better at learning non-native sound contrasts even in non-tonal languages. Proceedings from *Building Bridges Across Cognitive Sciences Around the World: Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (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.

Pulleyblank, D. (2004). A note on tonal markedness in Yoruba. *Phonology*, 21(3), 409 – 425.

Richards, J.C. and Rodgers, T.S. (2001). *Approaches and methods in language teaching: A description and analysis* (2nd ed.). New York, NY: Cambridge University Press

Saito, K. and Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /r/ by Japanese learners of English. *Language Learning*, 62, 595–633.

Schon, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349.

Shi, H. (2018). *A method of teaching English speaking learners to produce Mandarin-Chinese tones* (dissertation). West Virginia University, Morgantown, WV.

Spada, N. and P. Lightbown. (1999). First language influence and developmental readiness in second language acquisition. *The Modern Language Journal* 83 1–21.

Swan, M. (2005). Legislation by Hypothesis: The Case of Task-Based Instruction. *Applied Linguistics*, *26*(3), 376-401.

Tsai, P.-H. (2019). Beyond self-directed computer-assisted pronunciation learning: a qualitative investigation of a collaborative approach. *Computer Assisted Language Learning*, *32*(7), 713–744.

Talamini, F., Grassi, M., Toffalini, E., Santoni, R., & Carretti, B. (2018). Learning a second language: Can music aptitude or music training have a role? *Learning and Individual Differences*, *64*, 1–7.

Ward, I. C. (1952). *An introduction to the Yoruba Language*. Cambridge: W. Heffer.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*(4), 654–662.

Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*(2), 250–267.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043.

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.

Wang, X. (2013). Perception of Mandarin Tones: The Effect of L1 Background and Training. *The Modern Language Journal*, *97*(1), 144–160.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585.

Yang, B. (2014). *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer Berlin.

Zhang, J., Wang, X., Sun, Y., Nishida, M., Zou, T., & Yamamoto, S. (2013). Improve Japanese C2L learners capability to distinguish Chinese tone 2 and tone 3 through perceptual training. Proceedings from *2013 International Conference Oriental COCOSDA Held*

Jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE).

Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone. *The Journal of the Acoustical Society of America, 137(1)*, 1452-1463.

Zheng, A., Hirata, Y., & Kelly, S. D. (2018). Exploring the Effects of Imitating Hand Gestures and Head Nods on L1 and L2 Mandarin Tone Production. *Journal of Speech, Language, and Hearing Research*, *61*(9), 2179–2195.

Vita

Elizabeth is a graduate student, as well as recent alumni, of Syracuse University. She graduated from Syracuse University in 2018 with a Bachelor of Arts in Linguistic Studies and Writing & Rhetorical Studies. Shortly after, Elizabeth matriculated to the Linguistic Studies Program's M.A, with a concentration in Teaching Languages (TESOL/TLOTE) at Syracuse University in the Fall of 2018. Her research interests include second language pedagogy, particularly of pronunciation and phonology instruction, and second language acquisition.