UNIVERSITÀ
DI PAVIA

Università degli Studi di Pavia
Dottorato di Ricerca in Scienze Linguistiche
XXXV ciclo

The Effect of Musical Training on the Perceptual Learning of Japanese Pitch Accent by Native Italian Speakers

Tutor:
Chiar.mo Prof. Cristian Pallone

Yuka Naito
Matricola 484995

ABSTRACT

The present dissertation investigates whether the effect of musical training influences perceptual learning of Japanese pitch accent. Musicians and non-musicians, who were native speakers of Italian without any experience of Japanese, engaged in Japanese pitch-accent identification and discrimination tasks before and after undergoing perceptual training. The training paradigm employed in the current study is high variability phonetic training (HVPT), for which a considerable body of literature has established the effectiveness. However, since the literature has shown contradictory findings about the role of talker variability in the effectiveness of perceptual training with HVPT, the current study also aimed to further explore whether talker variability would influence perceptual learning of Japanese pitch accent. To achieve this goal, the two categories of participants (musicians and non-musicians) were randomly assigned to a high variability (HV) training condition (stimuli produced by four talkers), or a low variability (LV) training condition (stimuli produced by one talker).

An attempt was also made to explore the effect of absolute pitch on perceptual learning of Japanese pitch accent. To date only one published study has shown its effect on the perception of Thai lexical tone by native speakers of a non-tone language (English). However, testing revealed that none of the musicians in this study possessed absolute pitch, so it was impossible to proceed with this line of investigation.

Results for pre- and post-training pitch accent identification and discrimination tasks revealed that while musicians showed pretest-posttest improvements, non-musicians showed a pretest-posttest improvement only in identification tasks and, for discrimination tasks, their pretest-posttest performance was almost the same. In addition, musicians outperformed non-musicians in all tests. These findings indicate a positive effect of musical training, although the effect barely influenced reaction time data in the discrimination tasks. Moreover, while no significant differences were found between musicians in the two training conditions, non-musicians in the HV training condition performed better than non-musicians in the LV training condition, which was actually detrimental in the

discrimination tasks. These findings suggest that whereas, for musicians, talker variability in training stimuli did not play a role in Japanese pitch-accent identification and discrimination, for non-musicians, high talker variability is more beneficial than low talker variability.

The current study contributes to the existing knowledge of Japanese pitch accent perceptual learning by native speakers of a non-tone language. It provides empirical evidence for the effects of musical training and talker variability in the process of Japanese pitch accent perceptual learning. Finally, it also offers practical insights for the L2 Japanese classroom setting: the importance of input variability for non-musicians and the possible application of the current perceptual training as an online learning tool.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to a multitude of individuals who made the completion of this dissertation possible.

First and foremost, I am deeply indebted to my supervisor, Prof. Cristian Pallone, for his consistent guidance and kind support. He has offered helpful advice and encouragement throughout, and provided me with wonderful opportunities to interact with his Italian students of Japanese. I would also like to give special thanks to Prof. Alessandro Vietti for his invaluable advice on the experimental design. It was this advice that led to the adoption of the high variability phonetic training paradigm in my dissertation. I am also grateful to him for sharing his statistical knowledge. My sincere thanks go to Prof.ssa Federica Da Milano, who has been always supportive and believed throughout that I would be able to conclude my research! She helped me not only by giving useful advice but also by recruiting musician participants from the Conservatorio di Musica "Giuseppe Verdi" in Milan. I am indebted to Prof. Lorenzo Spreafico for his indispensable advice that every experimental decision must be grounded in a well-founded reason. In addition, it was thanks to his advice that I decided to focus on talker variability in perceptual training. I would also like to thank the late Professor Gabriele Iannaccaro for his interest in my research. My study has sometimes been criticized because it was not limited to linguistics but was closely related to music. I was truly saved when he said that he wanted to be one of my advisors immediately after listening to my presentation in the second year of the doctoral program. I am very sorry that I no longer have the chance to thank him directly. I owe my deepest gratitude to our PhD coordinator, Prof.ssa Silvia Luraghi, for her tremendous, unwavering support, which enabled me to finish my dissertation. I owe special thanks to Prof.ssa Chiara Zanchi and Prof.ssa Angela Romagnoli for being a great help in recruiting non-musician and musician participants at the University of Pavia. I am also genuinely grateful to Prof. Giuseppe Pappalardo and Prof.ssa Silvia Calamai for their perceptive comments and suggestions. I welcome this opportunity to thank Prof. Paolo Di Giovine, my Master's degree thesis supervisor and the first professor who

led me to realize how intriguing linguistics was. Without him, I would not have been able to begin my PhD journey.

In addition to thanking these Italian supervisors and academics, I must express my sincere appreciation to a number of researchers outside Italy. I am enormously indebted to Prof. Irina A. Shport for her generosity in answering my numerous questions and in sharing detailed experimental information, which enabled me to base my own experiment on her methodology. I would also like to thank Dr. Gwen Brekelmans and Prof. Daiki Hashimoto not only for kindly and informatively answering my many questions but also for generously sharing their Praat scripts. My heartfelt thanks go to Prof. Mineharu Nakayama for the discussion and insightful comments and suggestions during and after the *J-SLA 2021* conference. I want also to thank Dr. Tim Joris Laméris for his useful suggestions and for being a very good *senpai*.

My online experiment would not have been possible without my Italian participants. I thank them sincerely for their time, patience, collaboration and effort.

I owe a great deal to many people for the creation of the stimuli used in the experiment. As for the stimuli used in the absolute pitch test, four people deserve to be mentioned. I am deeply grateful to Prof. Gaetano Bruno Ronsivalle for his perceptive suggestions and comments. My warmest thanks go to Miyuki Kato for her encouragement, our friendship, and for her useful comments, based on her professional experience. A very special thank you goes out to my musically talented friends, Hideko Sakamoto and her daughter Reine, for their kind help in checking the stimuli. With regard to the stimuli used in the linguistic tasks, my deepest appreciation goes to all the native Japanese speakers who kindly produced the stimuli (including those who recorded stimuli not used in the experiment for various reasons) that made this experiment possible. My heartfelt thanks go to my dear friend, Mineko Asada (Miranda-san), who was unfortunately assigned to produce the most stimuli, and to her family for their generous help in recordings. Miranda-san has also been an unfailing source of encouragement throughout, starting from my bachelor journey. I am thankful to my sister, Miki, and my long-term dear

friends, Reina Kimura and Yukari Sugawara for not only recording their voices but also checking the intelligibility of the audio files.

I heartily thank my sister-in-law, Hama, and my close friends Emanuela Fantini Perullo, Mariene Bello—colleagues at the University of Rome "La Sapienza"—and Mayumi Yamada for participating in the pilot experiment. Emanuela also corrected the Italian instructions and other texts used in the experiment; Mariene has been a much-appreciated source of encouragement and optimism throughout. I also thank them for our long-term friendship. Warm thanks go to my dearest colleague in Siena, Mayumi Yamada, in addition, for our life-long friendship, which literally saved me at many critical moments of my life.

Now I think it is essential to thank two people for their support, because without them I am convinced that I would not have been able to afford to undertake this endeavor.

I am forever grateful to Claire Archibald for always being supportive and patient throughout the writing of this dissertation, and for her help in recruiting non-musician participants. Writing the dissertation in English, my third language, has been a Herculean task for me. I am greatly indebted to her endless efforts and patience with my repeated mistakes with English prepositions and articles. I have also appreciably benefited from her constructive comments and suggestions not only on my English but also on the structure of my dissertation. I will never forget how lucky I was to have her as my English teacher.

My unbounded gratitude goes to my companion, Asham, for his unconditional support. Through thick and thin, he has always been close. I would like to specify that his support was not limited to moral, but also extended to technical expertise. My experimental work "non è tutta farina del mio sacco", as it is expressed in Italian: Asham, a professional programmer, often magically rescued me from seemingly hopeless difficult parts of the programming (in various programming languages). Without him, this dissertation would not have been possible.

Finally, I want to thank my mother and my grandmother for giving me the opportunity to learn music. My mother, Megumi, who brought me and my sister up by herself, also taught me to play the piano. As a child, I honestly did not like piano lessons, but now I realize how lucky I was to be able to learn piano at home. Indeed, learning to read music and to master the basics of piano has helped me a lot in my current research. Unfortunately, my beloved grandmother, Mieko, passed away during the writing of my dissertation, and I am deeply sorry that I can no longer thank her in person. My grandmother loved music, always cherished me, and was very supportive. It was also my grandmother who encouraged me to join the orchestra club in high school and bought me a violin. Thanks to her financial and moral support and her encouragement, I was able to pursue my interest in music. Many years later, this ultimately led me to my current doctoral research topic.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

d'	d-prime
F0	Fundamental frequency
FL	Foreign language
Gen-1	Test of generalization 1
Gen-2	Test of generalization 2
HV	High variability
HVPT	High variability phonetic training
ISI	Interstimulus interval
LV	Low variability
L2	Second language
ms	milliseconds
RT	Reaction time
SD	Standard deviation
3AFC	Three-alternative forced-choice

CHAPTER 1 INTRODUCTION

The relationship between speech and music has been of great interest in various disciplines (Patel, 2008). This is probably due to the fact that the two have common features, such as pitch, duration and intensity. Of these, the current dissertation focuses on pitch.

There are, clearly, some differences in terms of pitch between music and speech. In music, as Patel (2008, p. 13) states, pitches separated by an octave (a doubling in frequency) are heard as very similar. They are collectively referred to as a *pitch class* or *chroma* and are given the same name (e.g., all the notes called "C" on a piano keyboard). According to Patel (2008, p. 13) pitches are perceived as similar "not only by proximity in terms of pitch height but also by identity in terms of pitch chroma". On the other hand, pitch range in speech varies between speakers. It can vary from occasion to occasion within a single speaker. It seems likely that pitch in speech is relative in nature (Ladd, 2008).

In both music and speech, however, pitch is an essential attribute and time-varying pitch patterns convey information (Chandrasekaran et al., 2009). In music, melodies are created by the change in pitch, in terms of a contour code (involving changes in pitch direction between consecutive tones) and of an interval code (involving the relationship between consecutive tones on a musical scale) (Chandrasekaran et al., 2009). In speech, variations in pitch convey linguistic information, such as lexical tone, lexical pitch accent, intonation, as well as paralinguistic information (Chandrasekaran et al., 2009; Zhao & Kuhl, 2015). Given its importance in both music and speech, the present research explores pitch in the cross-domain relationship between speech and music. More specifically, it set out to investigate whether or not musically trained individuals—who were expected to be sensitive to pitch height and directions thanks to the musical training they had received—would learn to perceive Japanese pitch accent[1] better than those who were not musicians.

[1] In the present dissertation "Japanese" refers to Tokyo Japanese.

The reason why Japanese pitch accent is the target of this research is that while numerous studies have investigated the effect of musical training/experience on lexical tone perception by native speakers of non-tone languages, very little research has addressed the role of musical training/experience in Japanese pitch accent perception.

Japanese lexical pitch accent consists of a bitonal high-low accent implemented as a fundamental frequency (F0) peak near the end of the accented mora[2] followed by a steep F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006). Pitch-accent patterns, which may give rise to lexical contrast, are determined by the presence or absence of a pitch accent and (if present) its location (Kawahara, 2015; Sugiyama, 2012).

Even though, often, not much importance is attached to Japanese pitch-accent contrasts in a second language (L2)[3] Japanese classroom setting (Schaefer & Darcy, 2015; Shport, 2008, 2011, 2016), many studies (e.g., Hirano-Cook, 2011; Schaefer & Darcy, 2015; Shport, 2011, 2016; Taylor, 2012) have stressed the importance of presenting Japanese pitch-accent contrasts to L2 learners of Japanese. One reason for this is that pitch-accent information is used by native speakers to constrain the activation and selection of word candidates in the process of spoken-word recognition (Cutler & Otake, 1999; Masuda-Katsuse, 2006; Sekiguchi & Nakajima, 1999); see Section 2.1.1 for other reasons.

Numerous studies have examined the perception of Japanese pitch accents by native speakers of non-tone languages. The results of these studies suggested that acquiring Japanese pitch accent contrast is not straightforward for L2 learners, especially those who are native speakers of non-tone languages. It is also anecdotally known that, if a learner of Japanese incorrectly uses lexical pitch-accent patterns, this can interfere with the comprehension of native Japanese speakers.

[2] See Section 2.1.1 for the distinction between mora and syllable.

[3] Unless explicitly mentioned or when relevant for the discussion, the terms "L2" and "Second language" are used in a broad sense to describe any non-native language, regardless of the context of learning and regardless of the order of learning.

Note, however, that the majority of these works have been carried out on native English speakers, and very few studies have been conducted on other non-tone languages. Thus, the current research contributes to literature on cross-linguistic studies by providing empirical data for two languages (Japanese and Italian) which have not yet been well studied.

1.1. Aims of the Dissertation

The first aim of the current dissertation was to examine the effect of musical training on perceptual learning of Japanese pitch accent by Italian native speakers.

Several studies have explored the effect of musical training/experience on empirical perceptual learning of lexical tone. In these works, musicians and non-musicians underwent training on lexical tone perception. The studies then assessed musicians' and non-musicians' learning curve during training and/or their learning outcomes, comparing their performance in the pretest and the posttest. As detailed in Section 2.3, mixed findings were found for the advantage of musical expertise in lexical tone perceptual learning. Furthermore, to the best of the author's knowledge, no work has yet investigated the role of musical training in the perceptual learning of Japanese pitch accent.

The training paradigm employed in the current study is that of high variability phonetic training (HVPT), which makes use of "numerous samples, produced by multiple talkers, in varied phonetic contexts" (Thomson, 2018, p. 209). The HVPT paradigm was chosen because, as detailed in Section 2.2.2, HVPT has been extensively studied and numerous experiments have established its effectiveness, including Shport's 2011, 2016 studies, which investigated whether native English speakers naïve to Japanese could learn to identify Japanese pitch accent contrast.

Even though the benefit of HVPT has been borne out by a considerable amount of research, more recently some studies have reported contradictory findings regarding whether stimuli with multiple talkers' voices are more advantageous in perceptual training than those with a single talker's voice

(Brekelmans et al., 2022; Zhang et al., 2021). These results suggest that more research on the role of talker variability needs to be undertaken.

In addition, several investigations on the interaction between the role of talker variability and individuals' perceptual abilities have shown mixed results. According to some studies (Perrachione et al., 2011; Sadakata & McQueen, 2014), stimuli with multiple talkers' voices were beneficial for participants with strong perceptual abilities and less beneficial for participants with weaker perceptual abilities. Indeed, intuitively, it seems likely that musicians would be the type of subject to benefit more from stimuli produced by multiple talkers, since musicians generally have a good ear. But other studies (Dong et al., 2019; Qin et al., 2022) showed contrasting results for the interaction between the role of talker variability and individual perceptual abilities; see Section 2.2.3.1 for a detailed discussion. Again, conflicting findings from the studies mentioned above suggest the need for further investigations on this issue.

The second aim of the present research was thus to examine the role of talker variability and its interaction with the effect of musical training.

The third and final aim of the current study was to explore the effect of absolute pitch—the ability to name or label a note without a reference note (Burnham et al., 2015; Deutsch et al., 2006; Parncutt & Levitin, 2001); for example, naming a tone as "C", "261 Hz", or "do".

Referring to their Thai-like nonce stimuli, Chan and Leung (2020) stated that the ability of L2 learners to normalize and abstract pitch contours across various tone tokens by diverse talkers and categorize them into different tone categories is key to their success in acquiring L2 lexical tones. And indeed, although Japanese pitch accent has a lower functional load than lexical tone, it is also realized by pitch modulation.

Since absolute pitch, in other words, pitch labeling ability, involves abstracting pitch movements, categorizing and identifying them (naming or labeling notes), it is reasonable to suppose that absolute pitch would facilitate perceptual learning of Japanese pitch accent. Indeed, the results of Burnham et al.

(2015) indicated that absolute pitch is beneficial for discriminating Thai lexical tone. However, to the author' knowledge, this is the only study that has reported on the benefit of absolute pitch.

Having described the aims of the current dissertation, the next section presents its organization.

1.2. Organization of the Dissertation

As the title indicates, this section presents the structure of the present dissertation, giving a brief outline of what is to be found in chapters 2 to 7.

Chapter 2 contextualizes the research by providing background information. It falls into three parts.

The first part begins with a review of the literature on the key characteristics of Japanese pitch accent. There follow a brief overview of lexical stress and intonational pitch accent in Italian, because the participants of the present work are native Italian speakers; and a summary of cross-linguistic differences. In this part, the studies on L2 learners' perception of Japanese pitch accent are also discussed.

The second part of the chapter proceeds with a review of the literature on various training studies, especially those which have employed the high variability phonetic training (HVPT) paradigm. It pays particular attention to Shport's works (2011, 2016), since the current methodology is based on hers. Then, also discussed are works investigating the role of talker variability in perceptual training; and those exploring the interaction between talker variability and individuals' perceptual abilities.

The last part of this chapter reviews studies that have explored the effect of musical experience/training on lexical pitch perception or on perceptual learning of lexical tone (studies that have compared participants' lexical tone learning outcomes before and after perceptual training). It concludes with a discussion of three studies which have examined the effect of absolute pitch on lexical tone perception.

Chapter 3 provides a general overview of the current study's experimental design and procedure and states its main research question.

Chapters 4 to 6 cover the details of the experimental tasks conducted.

Chapter 4 presents the procedure and results for the *Identification Tasks* (identification pretest, training, posttest and test of generalization tasks), which are the core part of this experiment. In these tasks, Italian musicians and non-musicians without any experience of Japanese were asked to identify pitch-accent patterns. They were randomly assigned to one of two training conditions: a high variability training condition, with stimuli from multiple talkers, or a low variability training condition, with stimuli from a single talker. Their before and after training scores were compared in order to examine whether musical experience and/or talker variability had facilitated perceptual learning of pitch-accent patterns; and whether the two had had an interactive effect.

Chapter 5 presents the procedure and results for the *Discrimination Tasks* (discrimination pretest, posttest and generalization test tasks), in which the participants were asked to discriminate whether the first stimulus was the same as, or different from, the second stimulus in terms of target word pitch-accent pattern. Again, participants' before and after training scores were compared to examine whether musical experience and/or talker variability had facilitated perceptual learning of pitch-accent patterns; and whether the two had had an interactive effect.

Chapter 6 presents the procedure and results for the *Absolute Pitch Test* (musical note identification task). This test was administered to musicians only in order to assess whether they possessed absolute pitch.

Chapter 7 discusses the overall findings obtained in the current study, their limitations and their implications. The chapter ends with conclusions.

CHAPTER 2 LITERATURE REVIEW

This chapter presents the background information which motivates the current dissertation. The chapter starts with a review of the relevant literature on the perception and acquisition of Japanese lexical pitch accent; and on the perception of Italian lexical stress and pitch accent. Next, the literature on second language perceptual training is discussed with particular focus on high variability phonetic training, which is the training paradigm adopted in the present research. Then, previous studies which have addressed the issue of the relationship between lexical pitch perception and musical experience/training are reviewed. The chapter closes with a summary.

2.1. Pitch Accent: Some Background

To elucidate the comparison made between musicians and non-musicians in the present research, this section provides a literature review on the key characteristics of Japanese pitch accent. A concise overview of lexical stress and intonational pitch accent in Italian is also provided, because the participants in the present study's experiment are native Italian speakers. This is followed by a summary of cross-linguistic differences. Lastly, the literature on L2 learners' perception of Japanese pitch-accent patterns is discussed.

2.1.1. The Key Characteristics of Japanese Lexical Pitch Accent

In the present dissertation, and in the other studies cited in this chapter, "Japanese" refers to Tokyo Japanese, which is considered to be the standard variety (Kubozono, 2012). This variety's pronunciation is generally used in the Tokyo conurbation (Akamatsu, 1997), and it is the one primarily employed in national television broadcasts, especially NHK "Japan Broadcasting Corporation" (Akamatsu, 1997; S. J. Goss, 2015; Labrune, 2012). Finally, Tokyo Japanese is normally what is taught as Japanese in L2 classrooms (S. J. Goss, 2015; Sakamoto, 2011; Vance, 2008).

Pitch accent in Japanese is a lexical property of a given word (Kawahara, 2015; Venditti, 2005, 2006). It consists of a bitonal high-low accent implemented as a fundamental frequency (F0) peak near the end of the accented mora[4] followed by a sharp F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006). A word in Japanese is either accented or unaccented; unaccented words do not show a steep F0 fall.

Thus, pitch-accent patterns of Japanese can be described in terms of two parameters: (1) presence or absence of pitch accent, and (2) if present, location of pitch accent (Kawahara, 2015; Sugiyama, 2012); this yields n + 1 possible accentual patterns in a word consisting of *n*-syllables (Shport, 2011). To take an example, a disyllabic segmentally identical word, *hashi* (used in the present research's experiment) gives rise to three lexical contrasts based on pitch accent: *háshi* "chopsticks" (1st-syllable accented); *hashí* "bridge" (2nd-syllable accented), and *hashi* "edge" (unaccented). In the current dissertation, pitch accent is marked with an acute accent[5], following Shport (2008, 2011, 2015, 2016), and Sugiyama (2017, 2022).

So far two phonological units, syllable and mora, have been mentioned above. Here, the notion of the latter is described briefly due to the appearance of both units throughout the literature review provided in this chapter, although a full discussion of the difference is beyond the scope of the current research. The mora is a unit of syllable weight in languages which provides a distinction between light (or short) syllables and heavy (or long) syllables (Davis, 2011; Otake, 2015; Vance, 2018). A light syllable consists of a regular mora, such as (C)V, whereas a heavy syllable consists of a combination of regular and special moras (Hirano-Cook, 2011). The special moras, which can neither constitute a syllable on their own nor bear a pitch accent, fall into the following four types: (1) the second half of long vowels: /R/; (2) the second half of diphthongs (ai, oi, ui): /J/; (3) moraic nasals, or the coda nasals: /N/; and (4) moraic obstruents, or the first half of geminate consonants: /Q/ (Hirano-Cook, 2011; Kubozono, 2015, p. 11; Vance, 2018). Table 2.1 shows some examples of the differences between syllable-based division and

[4] The distinction between mora and syllable is described below.
[5] Note that there is no standard system to mark pitch accent in Japan.

mora-based division, in other words, examples of the use of special moras. As for the division of Japanese words, mora-based division is generally more familiar than syllable-based division to native Japanese speakers.

Table 2.1
Differences Between Syllable-Based Division and Mora-Based Division

Word	Meaning	Syllable count	Mora count
obaasan	grandmother, old lady	3 (o.baa.san)	5 (o.ba.a.sa.n)
gaikoku	foreign country	3 (gai.ko.ku)	4 (ga.i.ko.ku)
nippon	Japan	2 (nip.pon)	4 (ni.p.po.n)

Furthermore, the mora is also the unit of rhythm; a mora is perceived by native Japanese speakers as isochronous to other moras (Labrune, 2012; Vance, 2008, 2018). In addition, the mora plays a role as a counting unit in the meter of Japanese poetry, as in the *haiku*. Haiku poems consists of three lines; the first line is made up of five moras, the second line seven, and the third line five (5-7-5).

Since the mora plays important roles in Japanese phonetics and phonology, it is crucial to distinguish mora and syllable (Kubozono, 2018). In fact, Sugiyama (2012) argues that it is necessary to take both syllables and moras into consideration in order to describe pitch-accent patterns in Japanese, unless they overlap.

In the current study, however, the distinction is irrelevant because mora and syllable do indeed overlap in all the target words, i.e., none of the words contained any special moras and so they were simultaneously disyllabic and bimoraic. For this reason, the term “disyllabic” was consistently used to explain the stimuli employed in chapters 4 and 5 which discuss the linguistic tasks in this study’s experiment.

Coming back to the topic of Japanese pitch accent, it is worth mentioning that the pitch-accent pattern of a given word is basically arbitrary and the presence

and location of pitch are usually considered to be unpredictable except in some cases, for example, loanword accentuation (Hirano-Cook, 2011; Kawahara, 2015; Kubozono, 2018; Taylor, 2012). Since the current research is not concerned with this, the details of predictable patterning are not discussed here (but see e.g., Kawahara, 2015; Labrune, 2012).

In terms of how pitch accent is realized, F0 is the major acoustic correlate of Japanese pitch accent in production and the primary cue to pitch accent in perception (Beckman, 1986). To date, a number of studies have examined whether other acoustic properties, such as duration, amplitude, and formant information, were possible correlates of Japanese pitch accent (e.g., Beckman, 1986; Cutler & Otake, 1999; Sugiyama, 2017, 2022).

The results of Sugiyama's perceptual experiments (2017, 2022) suggested the existence of secondary cues to Japanese pitch accent, because her Japanese participants were able to identify minimal pairs of 2nd-syllable accented and unaccented disyllabic words at a rate better than chance, despite the fact that the stimulus recordings had been digitally manipulated to remove the primary cue, namely F0 (for methodological details, see Sugiyama, 2017, 2022). Nevertheless, acoustic measurements, reported in her most recent study, showed difficulty in identifying which of the acoustic properties were possible correlates of Japanese pitch accent. Specifically, consistent with other works (Beckman, 1986; Cutler & Otake, 1999; Sugiyama, 2017), pitch accent did not correlate with duration. This is not surprising, because duration is phonemic [6] (Sugiyama, 2012, 2017). Additionally, while the other works showed that amplitude might correlate somewhat with Japanese pitch accent (Beckman, 1986; Cutler & Otake, 1999; Sugiyama, 2017), a correlation analysis reported by Sugiyama (2022) suggested that the correlation was not strong between intensity difference and participants' discriminability scores (i.e., scores calculated to assess how well participants distinguished in 2nd-syllable accented and unaccented disyllabic words). Sugiyama (2022) also reported that it was difficult to observe clear differences in formant frequencies between the pitch-accent patterns of two words.

[6] Differences in length of vowels and consonants are phonemic in Japanese.

Although her perceptual experimental findings (2017, 2022) imply the existence of secondary cues, Sugiyama suggests that these cues may not have much weight for Japanese speakers, unlike what was found in equivalent studies investigating acoustic correlates of English stress accent and of Mandarin Chinese lexical tone. However, it is worth reiterating that Sugiyama's results were obtained when her participants heard digitally manipulated stimuli from which F0 information had been removed. By contrast, the stimuli used in the present research were made with natural speech. In any case, what is relevant for the current research is the general consensus among prior works which have examined acoustic correlates of Japanese pitch accent and/or perceptual cues to Japanese pitch accent (Beckman, 1986; Cutler & Otake, 1999; Sugiyama, 2017, 2022): i.e., that the dominant perceptual cue to Japanese pitch accent is F0.

Again, Japanese pitch accent mainly manifests itself in F0 (Beckman, 1986), and it is implemented as an F0 peak near the end of the accented mora, followed by a precipitous F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006). Sugiyama (2012) explored what role the elements in F0 information (namely F0 rise, F0 peak, and F0 fall) played in native speakers' production and the perception of Japanese pitch accent.

Sugiyama's production experiment (2012) found that, when her native Japanese speakers produced 20 minimal pairs (namely 2nd-syllable accented and unaccented disyllabic/bimoraic words) in isolation, the two types of words were not significantly different in either their F0 rise from the first to second mora or their F0 peak in the second mora. By contrast, when they produced the words followed by a particle in a carrier sentence, there was a greater F0 rise and a higher F0 peak for 2nd-syllable accented words than for unaccented words. But the most prominent difference to appear between two types of the words was a F0 fall from the second mora to the following particle; there was a sharp F0 fall for 2nd-syllable accented words, whereas a F0 fall did not manifested itself for unaccented words.

For the subsequent perception experiment, Sugiyama (2012) used the recordings from her production experiment described above, to create three types of stimuli: (1) 2nd-syllable accented and unaccented words produced in isolation;

(2) the two types of words extracted from a carrier sentence; (3) the two types of words and the following particle extracted from a carrier sentence. The results showed that correct word identification by Japanese participants was at chance level for words produced in isolation. Even words extracted from a carrier sentence did not yield a better correct response rate than words produced in isolation. This indicated that the F0 rise and F0 peak information contained in words extracted from a carrier sentence did not help the participants to distinguish 2nd-syllable accented and unaccented segmentally identical words. By contrast, the participants identified better the third type of stimuli, which included words and the following particle, even though accuracy was only approximately 70%. However, a correlation analysis showed that the greater the difference in the F0 *fall* between each minimal pair, the higher the participants' accuracy. This suggested the importance of the F0 fall in the perception of Japanese pitch accent: even when information about the F0 rise and the F0 peak as well as the F0 fall was available, native Japanese speakers only made use of F0 fall information for word identification.

Sugiyama's perception experiment (2012) also explored the influence of Japanese pitch accent on devoiced vowels. Vowel devoicing is a characteristic of Tokyo Japanese and it typically affects the high vowels /i/ and /u/, as in the word *kita* "north" (see e.g., Labrune, 2012; Vance, 2008; Yoshida, 2002 for a more detailed description). As she points out, prior works found that accented high vowels can be realized though the phenomenon called *ososagari* (delayed F0 fall), by which accented vowels are devoiced and the F0 peak occurs after the following syllable (e.g., Yoshida, 2002). Her results indicated that, even when there was no F0 on the devoiced vowel, the devoicing did not necessarily influence her Japanese participants' word identification performance, as long as there was the F0 contour on a particle which followed after the devoiced vowel.

Since Sugiyama's study (2012) described above showed that the presence of a particle after a target word is necessary to distinguish 2nd-syllable accented and unaccented patterns, the stimuli used in the present research's experiment consisted of target words followed by a particle embedded in carrier sentence. The target words were, as in her study, simultaneously disyllabic and bimoraic words;

however, they were not minimal pairs but triplets of segmentally identical words which contrast only in three pitch-accent patterns: 1st-syllable accented, 2nd-syllable accented, and unaccented.

Turning to pitch-accent contrast, it is worth noting that this is an important element of phrasal information in Japanese (Shport, 2016). The accentual phrase, the lowest level of phrasing, contains at most one lexical pitch-accent (e.g., Beckman & Pierrehumbert, 1986b; Pierrehumbert & Beckman, 1988; Sugiyama, 2012). The phrase may be a single word, but when words are embedded in sentences, it is quite usual for some to lose their status as separate accentual phrases (Beckman & Pierrehumbert, 1986b). In fact, some types of compounds, including noun-noun compounds and adjective-noun sequences, typically form single accentual phrases (see e.g., Beckman & Pierrehumbert, 1986b; Gussenhoven, 2004; Pierrehumbert & Beckman, 1988 for a more detailed description). The accentual phrase is defined by two delimitative tones: a phrasal high tone and a low boundary tone marking the phrase boundary (Beckman & Pierrehumbert, 1986b; Venditti, 2005). The presence and location of pitch accent alter the prosody of an accentual phrase because lexical and phrasal prosody interact (Shport, 2016). Cues to pitch-accent contrasts in multiword phrases may include, in addition to F0 peak location and the presence of a F0 fall, three other types of F0 information: (1) F0 peak height, (2) phrase-initial F0 rise to the F0 peak, and (3) the degree of F0 fall across the phrase (Shport, 2016). As for the three additional types of F0 information, (1) the F0 peak is higher for the lexical high tone of the pitch accent than the phrasal high tone; (2) the phrase initial F0 rise in accented phrases is higher than that in unaccented phrases; and (3) the F0 fall is sharper in accented phrases than that in unaccented phrases due to the lowering of all tonal targets following the lexical pitch accent in a phrase (Pierrehumbert & Beckman, 1988; Shport, 2016, p. 743; Sugiyama, 2012).

In order to take into consideration the aforementioned additional cues to pitch-accent contrasts—which may be a help for learners of Japanese to perceive Japanese pitch-accent patterns—Shport (2011, 2016) embedded her target words (triplets of segmentally identical words which contrast only in terms of pitch-accent pattern) in carrier sentences. As detailed in chapters 4 and 5, the stimuli used in the current study's experiment were identical to her stimuli.

The last point which needs to be discussed, before moving on to the next section, is the function of Japanese pitch accent. Japanese pitch accent has two main functions: the distinctive function and the culminative-delimitative function (Beckman, 1986; S. J. Goss, 2015; Kubozono, 2018; Shport, 2016). The former is the function of differentiating the meaning of homophones, such as *áme* "rain" and *ame* "candy". The latter is the function of defining the phrase by marking prominent units in a phrase (Beckman, 1986; S. J. Goss, 2015; Kubozono, 2018; Shport, 2016).

It has been argued that the primary function of Japanese pitch accent is not the distinctive function but rather the culminative-delimitative function, on account of Japanese pitch accent's relatively small distinctive load (Beckman, 1986; S. J. Goss, 2015; Kubozono, 2018; Shport, 2011, 2016). Indeed, it is often possible to distinguish homophones from their contexts. Additionally, in many minimal pairs or triplets of homophones, one of a pair/triplet is used less frequently compared to the other/s. To take an example from stimuli used in the current research's experiment, the triplet *mushi*, *múshi* "to ignore" and *mushi* "insect" are used more frequently and are more familiar than *mushí* "steaming". In fact, *mushí* "steaming" is generally employed not by itself but as a part of a compound word; for instance, *mushiatsui* (*mushi* + *atsui* "hot") "sultry, muggy", and *mushiki* (*mushi* + *ki* "container, dish, bowl") "steamer (to cook)".

This would be one of the most important reasons why, often, not much importance is attached to Japanese pitch-accent contrasts in a L2 Japanese classroom setting (Schaefer & Darcy, 2015; Shport, 2008, 2011, 2016). As possible other reasons, previous studies have pointed out time constraints in a L2 classroom, a lack of textbooks that mark pitch accents in their vocabulary sections, and a lack of knowledge of how to teach Japanese pitch-accent (Hirano, 2014; Jin, 2017; Kanamura, 2019; Oyama, 2016; Shport, 2008, 2011, 2016).

It might thus be argued that acquisition of Japanese pitch-accent contrasts is not that important for L2 learners of Japanese. Nonetheless, many works (e.g., Hirano-Cook, 2011; Schaefer & Darcy, 2015; Shport, 2011, 2016; Taylor, 2012) have underlined their importance. Based partly on arguments presented in these

works, several reasons can be advanced as to why acquiring Japanese pitch-accent contrasts is relevant for L2 learners of Japanese.

The first reason is that its distinctive function is, in fact, not negligible. Kitahara (2001) reported that approximately 13% of short words, consisting of 1-4 moras, are distinguished only by their pitch-accent patterns. He also showed that pitch-accent patterns within homophones are unevenly distributed in short words (in 2-mora words 1st-syllable accented pattern is prevalent, while in 3- and 4-mora words unaccented pattern is dominant), but pitch-accent patterns are more equally distributed in high-familiarity words (Kitahara, 2001). Matsuzaki (2000) found that the distinctive function of Japanese pitch-accent occurs in approximately 19% of the vocabulary list of about 1500 word for learners of Japanese at the elementary level[7] (300 hours of learning).

It can thus be said that there is a substantial number of homophones distinguished only by pitch-accent patterns; and it seems that these two studies did not count minimal pairs of conjugated verbs. As pointed out by Hirano-Cook (2011), however, it is also important to look at minimal pairs of conjugated forms; for example, the two verbs, *kátsu* "to win" and *kau* "to buy", whose past tense forms are a minimal pair: *kátta* "won" and *kattá* "bought". This implies that the percentage of minimal pairs distinguished only by pitch-accent patterns is not negligible; and the distinction function of Japanese pitch accent should not thus be ignored.

The second reason is that previous studies have shown that native Japanese speakers utilize pitch-accent information to constrain the activation and selection of word candidates in the process of spoken-word recognition (Cutler & Otake, 1999; Masuda-Katsuse, 2006; Sekiguchi & Nakajima, 1999). Owing to the fact that the pitch accent system is dialectally variable (see Kubozono, 2012 for a detailed account), it was claimed that pitch accent was unnecessary for native Japanese speakers to recognize Japanese utterances (Cutler & Otake, 1999). However, the results of the above-mentioned studies provided strong evidence that in the process

[7] Matsuzaki (2000) used the JLPT's (Japanese-Language Proficiency Test) vocabulary list (specifically, N3 level). Organized by the Japan Foundation and Japan Educational Exchange and Services, this is the most commonly used Japanese-language test in the world; see the JLPT website for more detailed information: https://www.jlpt.jp/e/index.html

of word recognition, native Japanese speakers exploited pitch-accent information in addition to segmental information. Otake and Cutler (1999) reported that native speakers of Japanese originating from the accentless-variety areas could also use the pitch-accent information to recognize spoken words in Tokyo Japanese, even though their response patterns differed from those of the Tokyo Japanese. This implies that speakers who had originally acquired accentless varieties learned the information as an additional component, thanks to their daily exposure to Tokyo Japanese especially via the broadcast media. Thus, it can be suggested that learning Japanese pitch-accent contrasts would be a help for L2 learners of Japanese to communicate not only with native speakers of Tokyo Japanese but also with native speakers of Japanese originating from other areas.

Moreover, as described previously, the pitch-accent patterns of words, along with phrasal and boundary tones, also form larger prosodic units at the phrase level, i.e., accentual phrases. It is anecdotally known that, if a learner of Japanese incorrectly uses lexical pitch-accent patterns, their prosody of larger units will be altered, and this can prevent communication with native Japanese speakers. Indeed, Sato (1995) showed that prosody influenced native Japanese speakers more than segments in judging nativelikeness of utterances in Japanese.

The final reason why acquiring Japanese pitch-accent contrasts is relevant for L2 learners of Japanese is based on the observations of Kanamura (2020): since pitch-accent is a basic element of the Japanese intonation system and a key part of successful communication with native speakers, it is important for learners of Japanese to acquire. The importance of acquiring pitch-accent patterns has also been underlined by the results of works based on questionnaires given to learners of Japanese. These reported that one of learners' concerns about their pronunciation and an area that needs improving, is Japanese pitch-accent (Jin, 2017; Kourakata & Nagato, 2014; Toda, 2001).

Now, having concluded this section by discussing the function of Japanese pitch accent and the importance for language learners of acquiring pitch-accent contrasts, the next section focuses on lexical stress and intonational pitch accent in Italian.

2.1.2. Brief Outline of Italian Lexical Stress and Intonational Pitch Accent

To better understand the difficulties that participants would have encountered in learning Japanese lexical pitch accent during the current research's experiment, this section provides a brief outline of lexical stress and intonational pitch accent in their mother tongue, Italian.

Lexical stress in Italian has a distinctive function (Maturi, 2007; Nespor & Bafile, 2008; Pappalardo, 2018; Schmid, 1999). For example, a trisyllabic, segmentally identical word, *capito*, gives rise to three lexical contrasts based on lexical stress: *càpito* "I arrive"; *capìto* "understood", and *capitò* "it happened".

Lexical stress position is free, but stress is generally found on the penultimate syllable (D′Imperio, 2002; Gili Fivela et al., 2015; Grice et al., 2005; Rossi, 1998). Mancini and Voghera's study (1994, p. 72) showed that 93.3% of disyllabic words, 81.1% of trisyllabic words, and 75.0% of quadrisyllabic words have lexical stress on the penultimate syllable. Additionally, stress position can be predicted in some cases by morphological information (Gili Fivela et al., 2015). To illustrate this, some examples are given now: the third-person plural active indicative present or subjunctive form, *màngiano* and *màngino* "(they) eat", or names with suffixes, such as -a/iggine, *stupidàggine* "stupidity", - igine, *vertìgine*, "vertigo", and -udine, *solitùdine* "solitude". What is more, longer and more complex words can bear secondary stress (Gili Fivela et al., 2015; Maturi, 2007; Nespor & Bafile, 2008; Schmid, 1999).

Eriksson et al. (2016) examined acoustic properties of Italian lexical stress, such as duration, F0 level, F0 variation, and spectral emphasis, which is a related parameter of intensity (see Eriksson et al., 2016 for a detailed account of spectral emphasis). They computed the values of three types of vowels based on stress level (primary, secondary, and unstressed) in target words in terms of the aforementioned acoustic parameters. Their results demonstrated that duration was the acoustic parameter most markedly affected by stress level. The effect was mainly found between primary stressed vowels and the secondary and unstressed vowels,

although the differences between all three levels were significant. Consequently, they argued that duration was the dominant acoustic correlate of Italian lexical stress.

In addition to these results, D'Imperio and Rosenthall's study (1999), having measured stressed vowels in different positions, showed that there was variation duration of stressed vowels in open, non-final syllables in Italian. In particular, a stressed open penultimate vowel is significantly longer than any other stressed vowel, including a closed penultimate vowel. Additionally, closed penultimate syllables are significantly longer than closed syllables in other positions. These findings are not surprising, because vowel duration is not contrastive in Italian (D'Imperio & Rosenthall, 1999; Schmid, 1999). It is also worth mentioning that two types of syllabic structure—CV and V—are, respectively, the most and the third most frequent syllabic types in Italian, and that these comprise slightly over 60% of the total frequency of syllabic types (Schmid, 1999).

Coming back to Eriksson and colleagues' study (2016), they reported that spectral emphasis was the second most important acoustic correlate and that F0 level was also significantly correlated with stress but not to the same degree. They concluded by ranking the acoustic correlates with Italian lexical stress: duration, spectral emphasis, and F0 level.

With regard to cues to Italian lexical stress, in a study investigating perception by native Italian speakers, Bertinetto (1980) demonstrated that duration was the most effective prominence cue, and found this hierarchy: (1) duration, (2) intensity, and (3) F0. With minor differences, subsequent works have corroborated his findings in the sense of duration being the dominant cue to Italian lexical stress in perception (e.g., Alfano, 2006; Eriksson et al., 2020). For example, Alfano (2006) found that duration played a crucial role in perception of lexical stress in Italian. Her Italian participants, however, perceived lexical stress positions better when not only duration but also F0 were present in stressed vowels, even though F0 alone did not help the participants to perceive stress positions.

The lone exception is Caccia and colleagues' study (2019). Their results suggested that F0 was the most reliable acoustic cue to Italian lexical stress in perception. However, two points are worth noting. Firstly, unlike in the series of studies conducted by Eriksson and colleagues (2016, 2020), in which the stimuli were produced by more than 30 native Italian speakers (female and male), in Caccia et al.'s study (2019) the stimuli were produced by only one native Italian speaker. Hence, it could be that that speaker's pronunciation of vowels contained some somewhat idiosyncratic acoustic properties of Italian lexical stress, peculiar to that speaker. This possibility, in addition to the relatively small sample sizes, might make the findings of this work less generalizable.

Secondly, the results of Caccia et al.'s study (2019) showed that, to a similar degree, intensity and duration also contributed to the identification of Italian lexical stress positions, even though they were much less effective compared to F0. Overall, however, what is crucial for the present research is the general agreement among prior works which have examined acoustic correlates and perceptual cues to Italian lexical stress: specifically, that duration—along with other acoustic correlates such as intensity and F0—plays a role in perception of Italian lexical stress.

Lexically stressed syllables in Italian can have a pitch accent (D'Imperio, 2002), but it is worth highlighting that Italian pitch accents are intonational, not lexical.

Intonation plays an important role in expressing the pragmatic meaning of an utterance (D'Imperio, 2002). Indeed, yes/no questions in Italian are expressed by altering only the tonal contour of the sentence without any morphological or syntactic means (D'Imperio, 2002; Krämer, 2021). For example, *Paolo ama Maria* "Paolo loves Maria" can be uttered as a statement or as a question by changing only its intonation.

As pointed out by Gili Fivela et al. (2015), numerous studies have conducted acoustic analyses of intonation in Italian in accordance with different theories and approaches. But this section limits itself to the discussion of previous works framed within the autosegmental-metrical theory of intonational phonology (see Ladd, 2008 for a review). Within this framework, tunes of Italian can be described as

sequences of high (H) and low (L) tones of three types: (1) pitch accents, (2) phrasal accents, and (3) boundary tones (Avesani, 1990; Krämer, 2021). Pitch accent has a well-defined shape, which is associated with a designated syllable in an utterance, and which marks the syllable (or the word containing the syllable) as intonationally prominent (Beckman & Pierrehumbert, 1986a, 1986b). The other two types of tone mark the edges of two different prosodic phrases: intermediate phrases and intonational phrases (Avesani, 1990).

In Italian, there are various types of intonational pitch accent. Avesani (1990) argues that there are four: two monotonal tones, H* and L*, and two bitonal tones, H+L* and L+H* (the asterisk indicates that the tone is aligned with a stressed syllable). While the first bitonal pitch accent is realized as a F0 minimum on the stressed syllable immediately preceded by a F0 peak or by a high F0 plateau, the second one is realized as a high peak on the stressed syllable immediately preceded by a local F0 minimum on the pretonic syllable (Avesani, 1990, pp. 3–4).

Since Avesani's study (1990) attempted to propose a model for the automatic synthesis of Italian intonation, it dealt with standard Italian language. However, many works have emphasized the great importance of taking into account the fact that Italian is characterized by a very strong phonetic and phonological variation across primarily geographical areas but also socio-economic background, education, means of communication, and occasion of use (Gili Fivela, 2012; Krämer, 2009, 2021; Vietti, 2019).

Italian intonation shows wide geographic variation (Gili Fivela et al., 2015; Krämer, 2021), and numerous studies have provided detailed descriptions of the characteristics of different varieties (see e.g., Gili Fivela et al., 2015 for a detailed account).

Gili Fivela et al. (2015) reviewed these works and illustrated an inventory of pitch accents in varieties of Italian. The inventory consists of nine pitch accents: other than the four types (H*, L*, H+L*, and L+H*) indicated by Avesani (1990),

another five are described: L+H*, L+<H*[8], L+¡H*[9], L*+H, L*+>H[10]. Of these pitch accents, a handful occur in all the varieties, some in only one (Gili Fivela et al., 2015). The different pitch accents cue a variety of pragmatic functions in Italian (D'Imperio, 2002); for example, the pitch accent, H+L*, is used for wh-questions in all varieties, whereas L+¡H* is used for incredulity and counter-expectational wh-questions in all varieties. Discussing each pitch accents shape associated with a specific pragmatic meaning is, however, beyond the scope of the present dissertation: see Gili Fivela et al. (2015) for a detailed discussion.

Now, having concluded this section by briefly describing intonational pitch accent, the next session focuses on differences between Japanese and Italian in terms of lexical accent.

2.1.3. Lexical Accent: Cross-Linguistic Differences Between Japanese and Italian

This section presents a brief recapitulation of two preceding sections: it summarizes several crucial differences in lexical accent between Japanese and Italian.

Firstly, in Japanese, pitch accent is a lexical property of a given word (Kawahara, 2015; Venditti, 2005, 2006). Its counterpart in Italian is lexical stress.

Secondly, the dominant acoustic correlate of Japanese pitch accent and the perceptual cue to it is F0 (Beckman, 1986; Cutler & Otake, 1999; Sugiyama, 2017, 2022). By contrast, the acoustic correlates of Italian lexical stress and cues to its perception are duration, F0, and intensity (Alfano, 2006; Bertinetto, 1980; Caccia et al., 2019; Eriksson et al., 2016, 2020), even though prior works cited here differ in which of these three acoustic correlates contributes most to the perception of Italian lexical stress.

[8] The "open" angle bracket symbolizes delayed peak (Beckman et al., 2005; Gili Fivela et al., 2015).
[9] The inverted exclamation mark indicates a superhigh peak at the end of the tone bearing unit (Gili Fivela, 2012, p. 148). The tone bearing unit is the element in the segmental structure to which tone associates (Gussenhoven, 2004, p. 29).
[10] The "close" angle bracket symbolizes early peak (Gili Fivela et al., 2015).

Thirdly, Japanese exhibits lexical contrasts in terms of pitch-accent patterns. Specifically, as stated in Section 2.1.1, these patterns can be described in two ways: (1) presence or absence of pitch accent, and (2) if present, location of pitch accent (Kawahara, 2015; Sugiyama, 2012). Italian, on the other hand, makes lexical contrasts in terms of lexical stress to which pitch accents can be associated at a post-lexical level. Thus, Italian pitch accent cannot contrast different lexical items but instead contrasts different intonational meanings.

Lastly, Japanese has only one type of pitch accent: an F0 peak near the end of the accented mora, followed by a precipitous F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006). Conversely, in Italian, as mentioned in Section 2.1.2, there are various pitch accent shapes and each shape is associated with a specific pragmatic meaning (e.g., D'Imperio, 2002; Gili Fivela et al., 2015).

It is important to bear in mind the differences illustrated so far, because they would lead to difficulties in learners' perception of Japanese pitch accent. This possibility is examined through the experiment in the following chapters.

2.1.4. L2 Learners' Perception of Japanese Pitch-Accent Patterns

This section reviews previous studies which have explored L2 learners' perception of Japanese pitch-accent patterns. To the author's knowledge, however, there is only one experimental work in which the participants were native Italian speakers. Hence, this section is divided into two main parts: the first focuses on the work of Pappalardo (2018), whose subjects were native Italian speakers; and the second focuses on studies in which participants were native speakers of other languages. As regards works exploring the effect of training on acquisition of Japanese pitch-accent patterns, these are reviewed in Section 2.2.1, with a more specific treatment of Shport's HVPT studies (2011, 2016), on which the present experiment is based, in Section 2.2.2.1. Research on the effect of musical experience on perception of Japanese pitch accent, on the other hand, is discussed in Section 2.3.1.

Pappalardo (2018) examined identification of Japanese pitch-accent patterns by 48 Italian learners of Japanese residing in Italy. The Japanese learning

experience of his participants varied in terms of time—from one to twenty years—and in terms of context of use—business settings or classroom learning settings. His target words consisted of 90 non-words (nonce words) and 10 real words. Fifty of the nonce words were presented with the particle *ga*. The other 40 nonce words included one of four special moras (see Section 2.1.1. for the special moras). The participants' task was to mark the position of a F0 fall with the symbol ¬ . In the case of unaccented words, they were instructed to write the symbol X. His study provides the percentages of correct and incorrect responses, and some information about the number of participants who had previously received a theoretical explanation of Japanese pitch accent or had previously attended a pronunciation practice lesson on Japanese pitch accent.

Although preliminary, Pappalardo's findings are valuable for the current research, whose participants were also native Italian speakers. Some relevant points from his results are listed here. Firstly, only 53.6% of responses were correct. From this value it may be reasonable to assume that perceiving correct pitch-accent patterns is difficult for native Italian speakers. Secondly, his results suggested that there was no relationship between number of years of learning experience and perceptual abilities. This seems to imply that differences in perceptual ability may stem from individual differences between participants. The present research attempts to extend the knowledge we have of such individual differences, focusing on musical training as one of the possible sources. Thirdly, as for the identification of pitch-accent pattern in the bimoraic nonce words, the unaccented pattern was most frequently identified correctly. The most difficult pitch-accent pattern was the 1st-accented syllable pattern. Lastly, even for those target words that contained a devoiced vowel, the percentage of correct responses was almost the same and only slightly lower than that for all target words. Therefore, it seems that even in the presence of a devoiced vowel, the participants were able to identify correct pitch-accent patterns in some way. Target words used in the current research's experiment were as in Shport (2011, 2016); and some of them had high vowels which could be devoiced. According to Shport (2011), the number of words in which the vowel

devoicing phenomenon could occur was balanced in the pretest/posttest (*aki*, *hashi*, *kaku*, *waki*[11]), and in the two tests of generalization (*kaki*, *mushi*, *maku*, *yoku*).

The remainder of this section discusses works in which participants were native speaker of languages other than Italian, mainly English speakers. Points of interest for the current research are the main focus of the overview.

First and foremost, it is worthwhile to mention Goss (2020)'s study, because he explored the role of individual differences in the perceptual learning of Japanese pitch accent. Goss focused on three possible sources of individual differences: (1) phonological short-term memory (PSTM) capacity; (2) auditory processing ability; and (3) learning context. His participants, who were native English speakers enrolled in a second-year Japanese class, consisted of two groups: one with 20 students studying abroad in Japan; and the other with 20 students studying at a university in the United States. The participants of the study-abroad students' group reported that they were generally aware of the presence of pitch accent but had not received perception or production training. On the other hand, the participants of the at-home students' group had received explicit, in-class instruction and correction of pitch accent and one lecture on the function of the pitch accent system. None of the participants in either group reported having had musical training for more than five years and none had studied a tone language.

While PSTM capacity was measured with a serial nonword recognition task, auditory processing ability was measured with two tasks, namely F0 discrimination and nonce word pitch discrimination. In the F0 discrimination task, participants heard two pure tones and were asked to indicate whether the second tone was of a higher or a lower pitch than the first tone, which remained constant. In the nonce word pitch discrimination task, participants were instructed to answer whether a pair of stimuli was the same or different in terms of pitch-accent patterns.

In order to measure participants' Japanese pitch accent perception ability, two tasks—with stimuli consisting of target words (real words) embedded in carrier sentences—were employed. One was a correctness judgment task, in which

[11] Vowels that could be devoiced are underlined.

participants had to answer whether a word's pitch-accent pattern was correct or incorrect. The other was a three-alternative forced-choice (3AFC) identification task in which participants were required to assign a schematized F0 contour to a word's pitch-accent pattern. The two tasks were administered twice, once before and once after the 12-week interval corresponding to the Japanese study semester, while the tasks measuring PSTM capacity and auditory processing ability were administered only before the interval.

Goss's results indicated that the participants in each learning context made only small gains on the perception tasks over the 12-week learning period. Even though the study-abroad students' group reported greater Japanese usage per week in terms of the number of classroom hours plus extracurricular time spent speaking and listening to Japanese, his statistical analysis revealed only a marginal effect of context on the posttest correctness judgement task. As he stated, his analysis was exploratory because, unlike the at-home students' group, the study-abroad students' group had not received any pedagogical intervention on pitch accent learning. But his analysis implies that neither studying aboard nor pedagogical intervention made a substantial difference in the learning of pitch-accent pattern. Furthermore, PSTM capacity predicted gains on the correctness judgment task, whereas auditory processing ability predicted gains on the identification task. Since identification tasks were also carried out in the present research's experiment, it was expected that the acoustic sensitivity of participants in the musician group would be beneficial, and that musician participants would score better on the identification tasks than participants in the non-musician group.

The second point of interest for the current research is what was suggested in Shport's 2008 study: the importance of explicit instructions and practice of Japanese pitch accent. Since she noticed that pitch-accent patterns were often not explicitly taught in L2 classroom settings, Shport examined whether intermediate learners of Japanese had acquired pitch-accent patterns only from Japanese input without formal instruction focusing on pitch accent. Her participants were 16 native speakers of American English as the experimental group and 16 native Japanese speakers as the control group. All participants of the experimental group were enrolled in the third-year Japanese language course at the University of Oregon,

and they reported that they had not received any formal instruction in Japanese pitch accent before taking part in the experiment. Her stimuli were 48 trimoraic words: half of the words were 1st-syllable accented and the other half unaccented. The stimuli were recorded by a female and a male native speaker of Japanese (half of the stimuli sets were recorded by the female speaker, the other half by the male speaker, and the order of presentation of these sets was counterbalanced). The participants carried out an AXY discrimination task in which they were instructed to listen to the stimuli set and to select the word which they perceived as having a different pitch-accent pattern from the other two words.

Shport's findings showed that the native Japanese speakers outperformed the intermediate learners of Japanese. Furthermore, the learners failed to perceive pitch-accent patterns at a higher accuracy rate than the chance level. These results suggest that it is extremely difficult for learners of Japanese to acquire Japanese pitch accent from input only, at least without explicit instructions and practice and after only two and a half years of studying Japanese. It is thus reasonable to assume that learners of Japanese need explicit instruction and practice of pitch accent to acquire it. Prior works exploring the effects of explicit instructions and practice, i.e., training, on Japanese pitch accent are reviewed in Section 2.2.1 and Section 2.2.2.1.

The third point of interest for the present research is the effects of L1. In recent years, several studies taking different approaches have shown that this is an important consideration in the study of L2 learners' perception of Japanese pitch accent.

In a series of studies using the dichotic listening paradigm, Wu et al. (2012) compared native Japanese speakers' perception of pitch accent patterns with two non-native groups' perception. Specifically, the two non-native groups were made up of native Mandarin speakers and native English speakers respectively. None of the non-native participants had any knowledge of Japanese or other pitch accent languages and the native English speakers had no previous background with tone languages. None of the participants (neither the native nor the non-native speakers) had had any formal musical training for more than six years. The task of the

participants was to identify each pair of stimuli, where each pair was presented simultaneously with one word in the left ear and the other word in the right ear.

The results showed that the native English speakers perceived Japanese pitch accent in a different way compared to the native Japanese speakers and the native Mandarin speakers, because they tended to rely more on acoustic cues in the perception of pitch accent. By contrast, native Mandarin speakers performed similarly to native Japanese speakers. According to the researchers, these findings can be accounted for by linguistic functional load. Specifically, the functional use of pitch in English is lower than in Japanese. On the other hand, lexical tone in Mandarin has a higher functional load than pitch accent in Japanese. Hence, it seems that a L1 background with linguistic pitch can play an important role in the perception of Japanese pitch accent.

Another study which showed the effects of L1 was conducted by Shport (2015). She examined whether native English speakers with no prior knowledge of Japanese or other tone languages were sensitive to the F0 peak location and the F0 fall in pitch-accent contrasts, because as discussed in Section 2.1.1, Japanese pitch-accent patterns can be described in terms of two parameters: (1) presence or absence of pitch accent, and (2) if present, location of pitch accent (Kawahara, 2015; Sugiyama, 2012). To answer this question, she carried out two tasks: an AX discrimination task and a two-alternative forced-choice (2AFC) categorization task. Her participants were native speakers of American English as the experimental group and native Japanese speakers as the control group. The findings of the discrimination task indicated that the experimental group was less sensitive to the F0 fall compared to the control group, however the experimental group was able to discriminate stimuli with or without the F0 fall at a better than chance level. Additionally, the level of sensitivity in terms of F0 peak locations was similar in both groups. The results of the categorization task, on the other hand, showed that while the control group was sensitive to the presence of the F0 fall irrespective of its magnitude, the experimental group primarily utilized the F0 peak location information and was insensitive to the F0 fall cue. Her findings imply that native language experience—of English prosody in this case (pitch accents are

intonational, as in Italian)—can influence the perception of lexical pitch-accent contrasts in Japanese.

Interestingly, the results of the two tasks in Shport's study showed the presence of large individual differences between her participants, and as one of the possible sources of such differences, she suggested musical training. This is what the current research investigated, because how extralinguistic factors (including musical experience) affect L2 prosody acquisition remains unclear (Laméris & Graham, 2020).

A relatively recent study by Wiener and Goss (2019) also reported the effects of L1 on Japanese pitch accent perception. The researchers explored learners' perception of Japanese pitch-accent patterns as a function of their participants' L1/L2/L3 experience. Specifically, the participants consisted of six groups: (1) L1-Japanese + L2-English; (2) L1-Mandarin + L2-English; (3) L1-English; (4) L1-English + L2-Japanese; (5) L1-English + L2 Mandarin; (6) L1-Mandarin + L2-English + L3-Japanese. The L2 level of groups (4) and (5) was intermediate; and the L2 English level of all non-native participants were matched for proficiency. All six groups were tested on an ABX Japanese pitch accent discrimination task in which they were asked to indicate, as quickly and accurately as possible, whether the last stimulus was the same as the first one or the second one.

Wiener and Goss's findings also indicated that L1 affects Japanese pitch accent perception. More specifically, the L1-English group discriminated pitch accent patterns less accurately than the L1-Japanese + L2-English group, while the L1-Mandarin + L2-English group discriminated pitch-accent patterns marginally more accurately than the L1-Japanese + L2-English group. However, their results did not indicate that L2-Japanese learning conferred perceptual learning effects. Wiener and Goss argued that one year of L2 classroom learning was not sufficient for the L1-English + L2-Japanese group to approach a nativelike manner of pitch accent perception. But the L1-English + L2 Mandarin group with no knowledge of Japanese reached nativelike discrimination accuracy, probably due to the Mandarin lexical tone training that they had received. Finally, the findings showed that L1-Mandarin + L2-English + L3-Japanese groups' discrimination accuracy was higher

than that of L1-Japanese + L2-English group. The authors suggested that an L2 or L3 learner's sensitivity to linguistic pitch cues is additive. This topic, of the effects of L2 learning experience, is a point of interest for the current research.

Taken together, the three studies reviewed above point to the effects of L1 on Japanese pitch accent perception. From their results, it is conceivable that the Italian participants of the current research's experiment would have difficulty because Italian, like English, does not have lexical pitch accent.

Having discussed the role of individual learner differences, the importance of explicit instruction and practice of Japanese pitch accent for L2 learners, and the effects of L1 on pitch accent perception, I now turn to the fourth point of interest for the present research: the effects of L2 learning experience. As will be discussed below, while several prior works have shown positive effects, other studies have not.

To start with the studies showing positive effects for L2 learning experience, Sakamoto (2011) examined perception of Japanese pitch accent by English learners of Japanese studying at the University of Edinburgh. Her participants consisted of three groups: (1) inexperienced learners, who had approximately two years of Japanese study and most of whom had never stayed in Japan; (2) experienced learners, who had approximately four years of Japanese study and had stayed in Japan for about one year; and (3) native Japanese speakers as the control group. Sakamoto administered to the participants three types of tasks: an ABX discrimination task, an ABX categorization task, and a three-alternative forced-choice (3AFC) identification task. In the ABX discrimination task and the ABX categorization task, participants had to answer whether the last stimulus was the same as the first stimulus or the second one. The difference between the two tasks was that while the stimuli used in the discrimination were not manipulated, those in the categorization task were manipulated in terms of the F0 peak alignment (by shifting the location of the F0 peak, seven slightly different stimuli along a F0 peak continuum were created). In the third task—the three-alternative forced-choice (3AFC) identification task—participants were required to select one of three visual stimuli with a diacritic (a square bracket: ˥) representing the audio stimulus they

had heard. The stimuli for all tasks were disyllabic nonce words accompanied by a particle. These were recorded by a female native speaker of Japanese, but in the case of the categorization task, as described earlier, the stimuli were manipulated after the recording.

Sakamoto's findings for the discrimination task indicated that the three groups did not differ significantly in their discrimination accuracy. This suggests that English-speaking learners could distinguish pitch-accent patterns as well as native Japanese speakers. Note, however, that the stimuli used in her discrimination task were recorded by only one female speaker. By contrast, the stimuli employed in the current research's discrimination tasks were recorded by different speakers. The use of different speakers was to investigate whether participants were able to distinguish pitch-accent patterns presented in stimulus pairs created combining different speakers' voices. Presenting multiple speakers made the task more like what we do in everyday settings, where we need to understand pitch-accent patterns, whoever is speaking. It is conceivable that the use of the different speakers would make the tasks more difficult than Sakamoto's discrimination task, because the participants had to handle the talker variability presented in the stimuli.

Sakamoto's results for the categorization task revealed that the English-speaking learners showed a categorical function in a similar direction to the control group, but differences arose between the two learner groups and the control group in the intermediate stimuli on the continuum. Moreover, the inexperienced learners' group had more difficulty than the experienced learners' group in categorizing the stimuli. This suggests that increased L2 learning experience can help to form the target L2 categories. Similarly, Sakamoto's results for the identification task indicated that the mean accuracy of the inexperienced learners' group was significantly lower than that of the control group, even though the inexperienced learners' group achieved better than chance level performance. By contrast, the mean accuracy of the experienced learners' group was not significantly different from that of the control group. Again, her findings suggest that as L2 learning experience increases, learners of Japanese become more nativelike in identifying pitch-accent patterns, in other words, in forming abstract linguistic categories.

A relatively recent work by Laméris and Graham (2020) showed similar results for the effects of L2 learning experience. They had two groups of participants. The first group consisted of native English speakers who were enrolled in a fourth-year Japanese class at the University of London and had spent one year in Japan in their third year of university. Laméris and Graham's English learners of Japanese was thus comparable to the experienced learners in Sakamoto's study (2011) discussed earlier. The English learners were tested with a four-alternative forced-choice (4AFC) identification task with stimuli consisting of two types: trimoraic nonce words presented in isolation or in a carrier sentence, and real Japanese filler words. In the identification task, the participants had to choose one of the four pitch-accent patterns by pressing the 1, 2, 3, or 4 keys on their keyboard. Laméris and Graham's results indicated that mean accuracy did not differ significantly between English learners and native Japanese speakers. From the findings, the researchers concluded thus that L2 learning experience seems to contribute to native-like perceptual performance in identifying pitch accent patterns.

The positive effects of L2 learning experience on pitch accent perception were also observed by Hirano-Cook (2011). She conducted a large-scale perception experiment with more than 100 English learners of Japanese who were taking or had taken Japanese language courses at a university in the United States. Her learners were made up of five groups based on their Japanese class year level. She administered to the participants a four-alternative forced-choice (4AFC) identification task in which participants were asked to indicate the location of the pitch accent using the diacritic (a square bracket: ⌝) after listening to each audio stimulus. In the case of unaccented pattern, they were told to circle "none". The stimuli for all tasks were quadrimoraic real words which were recorded by a female native speaker of Japanese. Hirano-Cook observed that mean accuracy rates increased gradually from novice to advanced learners. This result suggested that, in line with Sakamoto (2011) and Laméris and Graham(2020), learners could improve Japanese pitch accent perception as they gain L2 learning experience.

In Wu et al.'s 2017 study—a follow-up dichotic listening study subsequent to Wu et al. (2012) described previously—the authors also reported the effects of

L2 learning experience. Interestingly, their participants were not advanced L2 learners of Japanese like those in Sakamoto's study (2011) or Laméris and Graham's research (2020). Wu and colleagues' participants were composed of native Mandarin Chinese learners of Japanese and native English learners of Japanese. Both native language groups had had three months to one year of experience learning Japanese as an L2 at a university in Canada, and none had any pitch accent language experience other than Japanese. None of the participants had any formal musical training. The researchers used the same stimuli and experimental procedure as in their 2012 study, in which the participants were native Japanese speakers and non-native speakers (native Mandarin speakers and native English speakers) with no Japanese learning experience. The 2017 findings revealed that—apart from the L1 effects (L1 Mandarin experience with lexical tone), also found in Wu et al. (2012)—the English-speaking learners of Japanese showed approximation to native behavior (in contrast to L1 English naïve learners). This suggested that, even though their L1 is not a tone language and their L2-Japanese learning experience is not so long (namely, 3-12 months), learner's Japanese pitch accent perception can improve.

While the four studies discussed so far showed the effects of L2 learning experience on perception of Japanese pitch accent, Goss and Tamaoka (2019) reported on the effects of L2 vocabulary knowledge. Their participants were advanced learners of Japanese studying at a university in Japan who had lived in Japan for slightly more than two years. The participants were native Mandarin speakers and native Korean speakers. All participants were tested with tasks similar to those in Goss (2020), described above. Specifically, three tasks were used as predictor variables: (1) a serial nonword recognition task to measure phonological short-term memory (PSTM) capacity; (2) F0 discrimination task to measure auditory processing ability; and (3) lexical and grammatical knowledge tests including vocabulary and grammar items selected from the N1 and N2 level (the highest two levels) of the Japanese-Language Proficiency Test (JLPT) which is set by the Japan Foundation and Japan Educational Exchange and Services. In addition, as in Goss (2020), two linguistic tasks—with stimuli consisting of target words (real words) embedded in carrier sentences—were employed in order to measure

participants' Japanese pitch accent perception ability. One was a correctness judgment task, in which participants had to answer whether a word's pitch-accent pattern was correct or incorrect. The other was a four-alternative forced-choice (4AFC) identification task in which participants were asked to assign a schematized F0 contour to a word's pitch-accent pattern. Goss and Tamaoka found that native Mandarin speakers outperformed native Korean speakers, indicating the effects of an L1 with lexical tone. Their results also revealed that L2 lexical knowledge predicted the accuracy of both linguistic tasks regardless of the functional load of pitch in participants' L1, whereas auditory processing ability and PSTM capacity did not. The results implied that L2 lexical knowledge, rather than overall proficiency or learning experience, is closely related to pitch-accent perception ability of advanced-proficiency learners.

In contrast to the studies discussed earlier, it should be noted that several other works have shown that L2-Japanese learning experience or knowledge does not have a positive effect on Japanese pitch accent perception. Pappalardo (2018) (discussed above), whose participants were native Italian speakers as in the current research, reported results suggesting that there was no relationship between number of years of learning experience and perceptual abilities. Likewise, Shport (2008) reasoned from her findings that two and half years of Japanese learning experience without explicit instruction or practice would not be sufficient for learners of Japanese to acquire pitch accent. Wiener and Goss (2019) made a similar point, as their results indicated that English-speaking learners who had studied Japanese for one year in a L2 classroom setting did not achieve as high perceptual ability scores as native speakers of Japanese. The outcomes of Shport (2008) and Wiener and Goss (2019) were contrary to those of Wu and colleagues (2017), which indicated that 3-12 months' Japanese learning experience can improve pitch accent perception. Yet, recall that Wiener and Goss's results also showed that the L1-English + L2 Mandarin group with no knowledge of Japanese reached nativelike discrimination accuracy. According to the authors, this may be explained by the fact that the Mandarin tone training received by the L1-English + L2 Mandarin group facilitated the discrimination task. It might also be due to the high functional load of Mandarin lexical tone (Wu et al., 2012, 2017). The same reasoning might

partly be applied to explain why Wiener and Goss's L1-Mandarin + L2-English + L3-Japanese group discriminated better than the L1-Japanese + L2-English group. In addition to indicating the effects of L1-Mandarin, this result suggests that sensitivity to linguistic pitch cues carries on improving for learners in an additive manner. Finally, in a comprehensive review of works conducted by herself and her colleagues from the 1990s to 2003, Ayusawa (2003) reported that percentage of accuracy in pitch accent perception was not related to L2 learning experience and that there were large individual differences in accuracy between participants. The reviewed studies had been conducted on L2-Japanese learners with a variety of native languages and L2 proficiency levels.

In addition to the studies that show no positive effect for L2-Japanese *learning experience* or *knowledge*, Shibata and Hurtig (2007) reported findings showing that L2 *proficiency* did not affect Japanese pitch accent perception. They administered a correctness judgment task consisting of target words and a particle to participants, who had to answer whether a word's pitch-accent pattern was produced correctly or incorrectly. The participants fell into two groups. The experimental group consisted of English learners of Japanese in the United States. This group was, in turn, categorized into three sub-groups by their proficiency scores: advanced, intermediate and novice. The control group consisted of native Japanese speakers. Shibata and Hurtig's findings indicated that even advanced learners of Japanese were unable to judge lexical pitch accent at a better than chance level.

From the previous works reviewed so far, it is evident that research to date has not yet determined whether or not L2-Japanese learning experience or knowledge affect pitch accent perception. Although the effects of L2 learning experience or knowledge on Japanese pitch accent perception remain unclear, it is still important to take them into account.

Recall that the main aim of the present research was to investigate the effect of musical training on the perceptual learning of Japanese pitch accent by native Italian speakers. To this end, the performance of the two groups of participants (musicians and non-musicians) on pitch accent perception tasks was compared. If

the participants had been learners of Japanese, their L2 learning experience or knowledge could have been a confounding variable for the current research's experiment. In order to eliminate this possible confounding variable and to ensure that the two groups of participants had the same exposure to pitch-accent patterns, participants with no experience of Japanese were recruited.

Having discussed the fourth point of interest for the current research (the effects of L2 learning experience), the remainder of this section provides the last point of interest: performance as a function of different pitch-accent patterns in the case of identification tasks. Taking into account the effects of L1 shown by the prior works reviewed earlier and what Ayusawa (2003) pointed out, perceptual accuracy rate for different pitch-accent patterns depends on learners' L1. Here only studies whose participants were native Italian speakers or native speakers of other non-tone languages are discussed.

Pappalardo (2018) found that his Italian participants identified, in decreasing order of accuracy, unaccented bimoraic words, 2nd-accented bimoraic words, and 1st-accented bimoraic words. This finding, namely that the easiest pattern was the unaccented pattern, is in line with the results of Hirano-Cook (2011) and the results shown throughout the studies reviewed by Ayusawa (2003), which included native speakers of different Indo-European languages as participants.

In contrast, Laméris and Graham (2020) reported that the 1st-accented pattern was most easily identified in their identification task. Since they used trimoraic target words with a particle, they had four pitch-accent patterns. The second easiest pattern for their English-speaking participants was the 2nd-accented pattern, followed by the unaccented pattern and the 3rd-accented pattern. Similarly, Wu et al. (2012, 2017) showed that not only native English speakers but also native Mandarin speakers and native Japanese speakers perceived more accurately the 1st-accented pattern.

Taken together, these studies offer contradictory findings about which pitch-accent pattern is easiest to perceive for Italian and other non-tone language speakers. Thus, the current research also sought to examine this point.

2.2. Second Language Phonetic Training

This section is divided into four parts. The first part reviews prior works exploring the effect of training on the acquisition of Japanese pitch-accent patterns by L2 learners. The second part discusses studies using high variability phonetic training (HVPT) paradigm. These include Shport's works (2011, 2016) on which the current research was based. The third part focuses on studies investigating the role of talker variability in phonetic training. The last part provides a review of prior works which have addressed the issue of the relationship between talker variability and individuals' perceptual abilities.

2.2.1. Japanese Pitch-Accent Pattern Training

Compared to previous studies of L2 learners' perception of Japanese pitch accent reviewed in Section 2.1.4, relatively few works have been carried out on the training of Japanese pitch accent. This section reviews these works which have explored the effect of training on the acquisition of Japanese pitch-accent patterns by L2 learners. Considering the importance for the current research of Shport's HVPT studies (2011, 2016), these are reviewed separately in Section 2.2.2.1.

The training studies which are discussed here can be said to fall into two types: computer-assisted production training and instruction-oriented training. Here these two types are reviewed in order.

As regards the computer-assisted production training, Hirata (2004) assessed whether this type of training method was effective for English learners to learn pitch-accent and segmental duration contrasts. Her participants were native English speakers who were taking a second-year Japanese course at the University of Chicago. Half of them were assigned to a training group and the other half to a control group. Her production training was comprised of 10 sessions held over three and a half weeks. Each session lasted approximately 30 minutes. During the sessions, the training group worked individually in a laboratory: trainees listened to the model (target words contrasting in pitch accent and segmental duration, heard

in isolation or embedded in sentences, recorded by four native Japanese speakers) and watched the graphic representation of the model. They then produced the utterance imitating the model until the F0 contour of their production matched with that of the model.

In order to examine the effect of the training, all participants took a perception test and a production test with new target words both before and after training. Each of the perception tests contained new target words which had not been used in training. Test stimuli, consisting of target words both in isolation and in a carrier sentence, were recorded by two female native speakers of Japanese. The task used in the perception tests was 9-alternative identification task in which the participants were required to use the ability to identify not only pitch accent but also the duration of the target words (it was necessary to count the number of moras in the target words). The production tests are beyond the scope of the current research, but I briefly touch upon them here. As in the perception tests, they included new target words to examine pitch accent and durational contrast. The participants produced target words in isolation and in a carrier sentence. Then the participants' production was evaluated by two native Japanese speakers.

Hirata's findings showed that for both perception and production, the training group improved from the pretests to the posttests significantly more than the control group. In addition, for both perception and production, although the training group improved in both contexts—target words in isolation and those in a carrier sentence—they improved more when target words were embedded in a carrier sentence. The control group, on the other hand, did not improve significantly in either of these cases. The comparison of individual trainees in Hirata's study revealed also that the degree of improvement differed between participants: some of them showed greater gains in perception than in production, while others showed the opposite. Even though only eight participants took part in her experiment, this finding seems to indicate that there were individual differences between them.

Overall, Hirata's results indicated the effectiveness of her training method in both perception and production. It is conceivable that the two features—visual feedback and multiple talkers in the training stimuli—contributed to its

effectiveness. Indeed, these two features were included not only in many of the HVPT studies discussed below, but also in the current research's experiment.

Hirano-Cook (2011) employed the procedure of the second training type (instruction-oriented training) to investigate whether it was beneficial for learners to enhance their perceptual ability and production of Japanese pitch accent. Her participants were native speakers of American English who were taking the second or third-year level of Japanese language courses at a university. These students were a part of the participants in her large-scale perception experiment discussed in Section 2.1.4, and that experiment corresponded to the pretest in this training procedure. The participants were divided into two groups: the training group and the control group (which did not take part in training). Her training was aimed at establishing learners' knowledge of Japanese pitch accent and improving their self-monitoring skill. To achieve these two objectives, she gave the trainees training sessions consisting of lectures and exercises. In the lectures, she taught the trainees Japanese prosody, focusing mainly on Japanese pitch accent and its functions in comparison with English stress-accent system. In the exercises, she provided perception and production practice using various teaching techniques, such as visual representation of simplified F0 contours, hand movements, and stimuli whose speed was manipulated. The training for the experimental group consisted of six 30-minute sessions held in a laboratory over four weeks.

So as to investigate whether the training was effective, Hirano-Cook administered perception and production tests to all participants before and after the training. The content of the pretests and the posttests were the same. The task used in the perception tests was a four-alternative forced-choice (4AFC) identification task in which participants were asked to indicate the location of the pitch accent using the diacritic ¬ (a square bracket) after listening to each stimulus. In the case of the unaccented pattern, they were told to circle "none". The stimuli were quadrimoraic words recorded by a female native speaker of Japanese. Half of the stimuli were familiar to the participants, especially the training group, not only because most of these stimuli were obtained from the beginning levels of their Japanese textbook, but also because the stimuli were employed both in the training

sessions and in the two tests (pretest and posttest). The other half of the stimuli were novel in the sense that they were used only in the pretest and the posttest, and that they were composed of relatively less frequent words compared to the words of the familiar stimuli. As mentioned earlier, the production tests were beyond the scope of the current research, but I briefly describe them here. The production stimuli were the same as in the perception tasks. The participants were asked to read out the stimuli presented in two lists: one with the accent-marking diacritic ¬ , the other without. Then the participants' production was evaluated by native Japanese speakers.

Hirano-Cook's results showed that the training group showed a significant improvement in identifying Japanese pitch accent, both in familiar words and in novel words. On the other hand, the control group did not show significant improvement for either of the stimulus types. In addition, the training group's perception improved significantly from the pretest to the posttest for three pitch-accent patterns (1st, 2nd and 3rd-accented patterns), while the control group did not improve for any pattern. Interestingly, the perception pretest results revealed that learners' identification accuracy varied, suggesting the presence of individual differences between participants. In order to understand this result more clearly, Hirano-Cook used cluster analysis to subdivide the data from the training group and the control group into two levels of pretest accuracy. Analysis revealed that training was effective not only for trainees who were originally adept at identifying pitch-accent patterns but also for those who were not. As for the production tests, she found a significant pretest-posttest improvement for the training group. The control group also showed some pretest-posttest improvement, but for the stimuli presented without the accent-marking diacritic, that of the training group was significantly greater. Hirano-Cook's research on training provided evidence for the importance of explicit instructions and practice of Japanese pitch accent, as suggested by Shport (2008).

The effectiveness of explicit instructions and practice for the acquisition of Japanese pitch accent was also found by Oyama (2016). Her research differed from Hirano-Cook's (2011) in that she did not have a control group, and in that she used

a portion of her regular classroom time for pitch accent instruction; but her instructional content and duration were very similar to those of Hirano-Cook (2011). Specifically, Oyama's instruction consisted of eight 20-minute sessions twice a week over approximately one month. The sessions were devoted to lectures and exercises on Japanese prosody, focusing primarily on pitch accent features. She tested intermediate/advanced language learners in Japan who varied in terms of L1, giving them listening tests and pitch-accent knowledge test before and after the teaching sessions. She found a significant improvement in learners' performance from the pretests to the posttests. Her study provides further evidence for the importance of explicit instructions and practice of Japanese pitch accent.

Taken together, the works reviewed above indicate the effects of training on intermediate-level learners' acquisition of Japanese pitch accent in both perception and production. It is also worth mentioning that these studies employed complex training procedures in that they combined both perception and production tasks or activities. In this respect, they differ from the typical high variability phonetic training (HVPT) studies discussed below.

2.2.2. High Variability Phonetic Training

Since the training paradigm followed by the current research was that of high variability phonetic training (henceforth: HVPT), this section provides a review of HVPT studies.

Broadly speaking, "HVPT refers to perceptual training . . . in which the auditory training stimuli include numerous samples, produced by multiple talkers, in varied phonetic contexts" (Thomson, 2018, p. 209). A large number of studies applying the HVPT paradigm have been published, and although there is some variation in details such as the duration of training, the general framework of HVPT is consistent (Ingvalson et al., 2014).

The groundwork for future HVPT studies was laid out in pioneering works by Logan et al. (1991); Lively et al. (1993, 1994); and Bradlow et al. (1997, 1999).

As the first attempt, Logan et al. (1991) set out to examine whether their training would encourage Japanese learners of English to improve their identification of the English phonemes /l/ and /r/. To investigate the effects of the training, the researchers employed a pre/posttest design. In all experimental phases (pretest, training, posttest and two tests of generalization) six native Japanese speakers performed a two-alternative forced-choice (2AFC) identification task in which they were asked to identify a word presented in a minimal pair. The stimuli included minimal pairs recorded by six talkers (four female and two male): while five of them recorded the stimuli used in training, the sixth talker (a male talker) recorded those used in the second test of generalization which is described below. All minimal pairs contained the English /l/-/r/ contrast in syllable initial, medial, or final positions. No feedback was given at the pretest. The training consisted of fifteen 40-minute sessions over three weeks. During each training session, stimuli recorded by only one talker were used. Unlike at the pretest, feedback was provided during the training. The posttest conducted after the training was identical to the pretest. After the posttest, three of the participants took two tests of generalization without feedback. These tests were in order to assess whether the participants could generalize to novel stimuli. While the stimuli in the first test of generalization consisted of novel words produced by a novel talker, those in the second test of generalization consisted of novel words produced by one of the talkers who had recorded the training stimuli. It was found that the percentage of correct responses increased significantly from the pretest to the posttest, and that participants' performance depended on the phonetic context (the position of /l/ and /r/). The findings also showed that the participants who took part in the tests of generalizations were able to generalize to the novel stimuli produced by a novel talker, although their performance on novel stimuli produced by a familiar talker was better.

In a follow-up study, Lively et al. (1993) conducted two experiments to test separately two sources of variability—phonetic context variability and talker variability—presented in the stimuli used by Logan et al. (1991). Compared to Logan and colleagues' work (1991), Lively et al.'s first experiment reduced the extent of phonetic context variability of the training stimuli, but left the training and

testing procedures identical to those of Logan et al. (1991). The results indicated that even though the phonetic context variability was reduced, the training was effective for their Japanese participants to acquire English /l/ and /r/. The second experiment reduced the talker variability of the training stimuli, using only a single talker's voice. The researchers found that Japanese learners of English improved from the pretest to the posttest; however, that the training with the stimuli produced by a single talker did not lead to considerable improvements in the mean accuracy of the two tests of generalization. The findings suggest that the presence of multiple talkers in training stimuli is important for learners to achieve robust generalization to novel stimuli or novel talkers.

A number of follow-up works replicated the patterns of results found in the two studies described above, extending these seminal works. Lively et al. (1994) used a larger group of Japanese learners of English who had never lived in an English-speaking country (in the first two works the participants had been recruited in the United States). The authors demonstrated that learners' improvement in English /l/ and /r/ perception was retained three months after the conclusion of training, and that after six months without further training, participants' accuracy was still above pretest levels. In a subsequent study, Bradlow et al. (1997) expanded on previous works, increasing the number of training sessions and exploring the effects of training on English /l/ and /r/ production by native Japanese speakers resident in Japan. Their study not only replicated earlier studies findings but also showed that the training was effective in the domain of production. Bradlow et al. (1999), further extending the findings of earlier studies, showed that Japanese trainees maintained their improvement in both perception and production of English /l/ and /r/ three months after the completion of training.

In addition to the studies discussed so far, a number of works have also been conducted on perceptual learning of the English /l/-/r/ contrast by Japanese learners using the HVPT paradigm (e.g., Brekelmans et al., 2022; Hardison, 2003; Hazan et al., 2005; Iverson et al., 2005; Saito et al., 2022; Shinohara & Iverson, 2018). However, HVPT studies are not limited to exploring the perceptual learning of this consonant contrast: there have been studies examining other consonant contrasts, such as English /v/-/b/-/p/ (Hazan et al., 2005), and English /f/-/v/-/θ/ (Iino, 2019);

and many studies have been conducted on vowel contrasts (e.g., Georgiou, 2021, 2022; Iverson & Evans, 2009; Lambacher et al., 2005; Lengeris & Hazan, 2010; Nishi & Kewley-Port, 2007, 2008; J. W. S. Wong, 2015).

HVPT was also applied to the acquisition of non-native suprasegmental contrasts. Wang et al. (1999) first investigated whether HVPT can be applied to the perceptual learning of Mandarin lexical tones. Their participants were native speakers of American English who had taken one or two semesters of Mandarin Chinese language courses at Cornell University. Half of them were assigned to a training group and the other half to a control group who did not take part in the training. The training consisted of eight 40-minute sessions in a laboratory over two weeks. During the sessions, stimuli produced by four different native speakers of Mandarin were used. In all experimental phases (pretest, training, posttest, two tests of generalization, retention test phases), identification tasks were employed, and the stimuli used in the tasks consisted of monosyllabic Mandarin words having different syllabic structures. Their results showed that the identification scores of the training group improved appreciably from the pretest, and not only to the posttest (identical to the pretest), but also to the two tests of generalization (test of generalization 1 used novel stimuli produced by one of the familiar talkers in the training phase; test of generalization 2 used novel stimuli produced by a novel talker). By contrast, the control group improved only slightly from the pretest to the posttest and the two tests of generalization. In the retention test—which was identical to the pretest but administered six months after training to half of the training group and of the control group—the trainees outperformed the control participants and that trainees' improvement was retained. The researchers concluded that HVPT was also effective for the acquisition of Mandarin tones. This evidence was supported by the results of subsequent HVPT studies on Mandarin tones carried out by Li et al. (2019) and Li and Lee (2021).

Coming back to Wang et al.'s work (1999), the researchers found individual differences in trainees' initial performance and learning outcomes: some improved more than others from the pretest to the posttest. In a follow-up study, Wang et al. (2003) showed that perceptual Mandarin tone training was effective for their English-speaking learners of Mandarin to improve production of Mandarin tones,

and again, they reported individual differences in trainees' initial performance and learning outcomes.

The two seminal studies (Lively et al., 1993; Logan et al., 1991) in which the HVPT paradigm was put forth have been followed by a substantial amount of literature demonstrating the effectiveness of HVPT—including the research on the acquisition of non-native suprasegmental contrasts discussed above. See Brekelmans et al. (2022) for a detailed account of the impact of the seminal studies and of the number of phonetic training works that followed them. The large body of evidence showing the effectiveness of HVPT is one of the reasons why the HVPT paradigm was employed in the current research. However, it should be noted that mixed findings have been found by works comparing the effects of high and low talker variability input used in perceptual training. This topic is discussed in Section 2.2.3 and Section 2.2.3.1.

The second reason for using the HVPT paradigm in the present research is that HVPT can be utilized not only for L2 learners of a target language but also for naïve learners (the participants in the current experiment were native Italian speakers without any experience of Japanese). HVPT has been shown to facilitate naïve learners' acquisition of consonant contrasts (e.g., Pruitt et al., 2006), vowel contrasts (e.g., Archila-Suerte et al., 2016), lexical tones (e.g., Wang & Kuhl, 2003), and vowel length contrast (e.g., Hirata et al., 2007). Given that the current research is based on Shport's studies, which extended to perceptual training of English-speaking naïve learners on Japanese pitch-accent patterns, these studies are discussed in Section 2.2.2.1. As regards prior studies on naïve leaners which have explored the role of talker variability in perceptual training, these are reviewed in Section 2.2.3 and Section 2.2.3.1. By contrast, as regards previous perceptual training research on English naïve learners of Chinese in which examined whether musical experience has a bearing on lexical tone perceptual learning, this is discussed in Section 2.3.3.

The last reason for using the HVPT paradigm in the current research is that HVPT can be implemented online. Even though the majority of HVPT studies have been conducted in a laboratory setting (Inceoglu, 2022; Martin & Inceoglu, 2022;

Thomson, 2022), some studies (Brekelmans, 2020; Brekelmans et al., 2020, 2022; Saito et al., 2022) have successfully run their experiments entirely or partially on the Gorilla software (Anwyl-Irvine et al., 2020). In addition, several works have created web-based HVPT applications, e.g., to learn English segments (Qian et al., 2018; Thomson, 2023), and French vowels (Inceoglu, 2022). Web-based HVPT can be designed as a game (Iverson et al., 2023; Saito et al., 2022). The web-based studies mentioned above demonstrated that learners can use software tools like Gorilla online without supervision. Incidentally, the current experiment was also carried out entirely online using the Gorilla software due to the COVID-19 pandemic.

Having discussed the reasons why the present research adopted the HVPT paradigm, the next section provides a review of Shport's studies (2011, 2016).

2.2.2.1. HVPT in Shport's Studies

This entire section is dedicated to a discussion of the two studies that this research is based on—Shport (2011, 2016)—not only because these were the first HVPT studies on pitch-accent contrasts by naïve learners of Japanese, but also because the methodology of the current experiment was based largely based on them.

In these two studies, which were essentially identical, Shport investigated whether naïve learners of Japanese can learn perception of pitch-accent patterns by means of the HVPT method.

The only real difference in methodology between these works lay in the participant sample. In addition to the final number of participants, the two studies differed in whether or not learners of Japanese or a tone language were included. Specifically, in the 2011 study, out of a total of 41 native speakers of American English, six had studied Japanese and/or Mandarin Chinese. On the other hand, in the 2016 study, none of the participants (38 native speakers of American English) had any experience of Japanese or any tone language.

Recall that previous studies (reviewed in Section 2.1.4) have shown conflicting findings about the effects on Japanese pitch accent perception of L2-

Japanese learning experience or knowledge. But in the case of tone languages, Guion and Wayland (2003) suggested that the L2-tone language learning experience can play a role in Thai lexical tone perception, since they found in their experiment that English-speaking learners of Thai discriminated Thai lexical tones better than native English speakers without any experience of Thai, even though native Thai speakers outperform the English-speaking learners of Thai. It may thus be reasonable to take into account the possible effects of L2 (both Japanese and tone language) learning experience or knowledge. Indeed, as in Shport's 2016 study, the current experiment did not include Italian learners of Japanese or of any tone language.

Other than the participants, the methodology in Shport's two works was identical. To investigate the effect of the training, she employed a pre/posttest design, and randomly assigned her participants either to a training group, or to a control group who did not undergo training. In the experimental phases (pretest, training, posttest, two tests of generalization phases, all performed in a laboratory), the participants carried out a three-alternative forced-choice (3AFC) identification task in which they were asked to identify the pitch pattern of the target word. The stimuli consisted of target words embedded in carrier sentences. The disyllabic target word in each stimulus could carry any of the three pitch-accent patterns: 1st-syllable accented, 2nd-syllable accented, or unaccented.1st-syllable accented, 2nd-syllable accented, and unaccented. The use of the carrier sentences was to assist the participants in learning the pitch-accent patterns by providing additional types of F0 information (Pierrehumbert & Beckman, 1988) in a variety of sentential contexts (Lee et al., 2009; Nishinuma et al., 1996). This ensured the phonetic context variability used in the HVPT paradigm. The stimuli were recorded by six talkers (three female and three male). Specifically, four of them recorded the stimuli used in training (one of these, a female speaker, also produced the first test of generalization); the fifth talker (a male speaker) produced those used in the pretest and the posttest; and the sixth talker (a female talker) recorded those used in the second test of generalization. At the pretest, no feedback was given. The training phase consisted of three sessions of 20, 20 and 10 minutes on consecutive days over one week. Unlike at the pretest, feedback was provided during the training. The

posttest, conducted after the training, was identical to the pretest. After the posttest, the participants took the two tests of generalization without feedback. The tests of generalization were in order to assess whether the participants could generalize their knowledge to novel words and a novel talker. While the stimuli used in the first test of generalization were novel stimuli recorded by one of the talkers used in the training phase, those used in the second test of generalization were recorded by a novel talker. Details of the stimuli and the procedure, based on those in Shport's studies, are provided in Section 4.2.2 and Section 4.2.4, respectively.

The results of both of Shport's studies (2011 and 2016) were essentially the same. Overall, HVPT positively affected the accuracy of American English native speaker identification of Japanese pitch-accent pattern. More specifically, the training group improved more than the control group from the pretest to the posttest and to the first test of generalization. However, the two groups did not differ significantly in the second test of generalization. These results indicated that perceptual learning generalized to novel stimuli produced by a familiar talker (the first test of generalization) but not to those produced by a novel talker (the second test of generalization). Compared to other HVPT studies discussed previously (in Section 2.2.2), the effect of HVPT is smaller in Shport's works. It must be noted, however, that the training she gave to naïve learners of Japanese was shorter than in similar experiments—one hour—overall. And as Shport reasoned, even this had a positive learning effect on pitch-accent pattern perception. Indeed, the brevity and the efficacy of her training were the main reason why the current research was based largely on Shport's studies, considering the burden of participants. With regard to the pitch-accent patterns, Shport's results indicated that the unaccented pattern was the most difficult one. Based on other findings of Shport (2015), reviewed in Section 2.1.4, she reasoned that this difficulty may have been due to poor attention of English-speaking participants to the F0 fall cue, which is the primary acoustic cue for distinguishing between the unaccented and accented patterns.

Interestingly, large individual differences between the participants were observed in both of Shport's studies (2011 and 2016). In the 2011 work, she reported a more detailed analysis of the pretest scores than she did in the latter work. The analysis revealed that the distribution of the scores was nearly bimodal in the

training group and bimodal in the control group, and that a threshold of 60% correct responses could be set to subdivide the two groups into two levels (high-scoring level and low-scoring level), resulting in four sub-groups: Training High, Training Low, Control High, and Control Low. The difference in accuracy scores between the high-scoring participants and the low-scoring participants was observed for all tests. However, as she noted in her 2016 study, the individual differences did not influence the interpretation of the overall improvement findings due to the fact that both the training group and the control group were made up of 10 low-scoring participants and 9 high-scoring participants. Coming back to the analysis conducted in the 2011 work, individual differences between the high-scoring participants and the low-scoring participants arose in specific pitch-accent pattern confusion. While the high-scoring participants showed a tendency to confuse the unaccented pattern with the 2nd-accented pattern, the low-scoring participants tended to confuse all three pitch-accent patterns with each other.

Shport's studies attempted to further explore individual differences in perceptual learning between her participants. To this end, she administered a questionnaire to the participants to collect their information about academic and linguistic background, and musical experience. Details collected about musical experience were as follows: (1) number of years playing musical instruments; (2) the number of musical instruments played; (3) instrumental or vocal experience; (4) context of practice; and (5) self-estimated proficiency; the last two pieces of information appeared only in the 2016 study. It should be noted, however, that her investigation of individual differences was exploratory.

Based on information gathered from the questionnaire, Shport examined whether pitch-accent pattern identification accuracy correlated with individual differences. More specifically, she carried out correlation analyses for the pretest scores, score gains from the pretest to the first test of generalization, and the following variables: grade point average (GPA, academic achievement at school or college), number of languages learned, years of L2 learning experience, and number of musical instruments played, years of experience in instrument use and other variables. The analysis also included variables reported only in the 2011 study,

namely, self-assessed learning, test-taking skills, self-assessed attentiveness during the tests, and participants' major.

With regard to pretest scores, while in the 2011 work Shport found that these correlated significantly only with the participants' major and GPA, in the 2016 work she reported differently: the pretest scores of the training group and the control group correlated significantly with number of years studying L2.

What is most interesting for the current research is that both of Shport's studies (2011 and 2016) found that the training group's score gains from the pretest to the first test of generalization correlated significantly with the number of years playing musical instruments and with the number of musical instruments played, respectively. Regarding other correlations, analyses conducted in the two studies differed. Specifically, while the 2016 study did not report any other correlations, the 2011 study also reported correlations between the training group's score gains and the number of languages learned, and self-assessed attentiveness during the tests, respectively. However it should be taken into account that the participants of the 2016 work were more homogeneous than those of the 2011 work in that none of the participants had prior experience of either Japanese or any tone language. And recall that the current research adopted the 2016 study's criterion for the selection of participants. In addition, it is important to mention that the 2016 study reported no significant correlation between score gain and variables for the control group. The results for the correlation analyses discussed so far suggest that musical experience (number of years playing musical instruments; and number of musical instruments played) does not correlate with pre-existing individual differences in pitch-accent perception between naïve learners of Japanese, but that musical experience can play a role in facilitating perceptual learning of pitch-accent patterns.

The current research extended two aspects of Shport's interesting studies. Firstly, the present investigation focused primarily on the effect of musical training, comparing learning outcomes of musicians with those of non-musicians. Secondly, given Shport's results suggesting the effectiveness of HVPT, the current research sought to further explore the role of talker variability in perceptual training; and its relationship with perceptual abilities of musicians and non-musicians. These two

topics (talker variability, and its relationship with individuals' perceptual abilities) are discussed in Section 2.2.3 and Section 2.2.3.1, respectively.

2.2.3. The Role of Talker Variability in Perceptual Training

As discussed in Section 2.2.2, following the two seminal works (Lively et al., 1993; Logan et al., 1991) in which the HVPT paradigm was put forth, many studies have borne out its effectiveness. However, more recently, several studies have reported contradictory findings about the role of talker variability in perceptual training (Brekelmans et al., 2022; Zhang et al., 2021).

Brekelmans et al. (2022) carried out a large-scale replication study of Lively et al. and Logan et al.s' works in order to verify the existence of a high variability effect. To this end, they randomly assigned their participants to one of the two variability training conditions: high variability (HV, stimuli produced by five talkers), and low variability (LV, stimuli produced by a single talker). By employing a pre/posttest design based on the original studies, the researchers compared the performance of the participants in the HV training condition with that of the participants in the LV training condition.

Note that this replication study introduced some modifications to address methodological issues present in the original studies. One of the major changes was to increase the size of the participant sample. While the original works recruited Japanese-speaking learners of English who had enrolled in a university in the United States, Brekelmans et al. recruited Japanese-speaking learners of English irrespective of their present location, and conducted their experiment online. They managed to recruit 166 participants, considerably more than in the original studies (only six participants per experiment). Additionally, Brekelmans et al. carried out Bayesian analysis instead of the statistical analyses applied in the original studies. The replication study also differed from the original works in that Brekelmans et al. dropped the posttest task, which was the same as the pretest task, because their interests lay primarily in generalization. Finally, the authors added some tests of individual differences at the pretest. Specifically, these added tests were to measure

auditory processing ability, English vocabulary, attention, and familiarity with the words employed in the replication study.

Brekelmans and colleagues found that the participants in both training conditions improved from the pretest to the first test of generalization (a novel set of stimuli produced by a novel taker), indicating that the participants in both training conditions generalized successfully to a novel talker. However, their statistical analyses revealed that the evidence for greater improvement after HV training was ambiguous. Moreover, the researchers compared the performance on the two tests of generalization (test of generalization 1 and test of generalization 2, another novel set of stimuli produced by a familiar taker employed in training). Again, the researchers found the evidence for HV benefit was ambiguous. They performed further exploratory analyses. These analyses revealed that there was no evidence that either pretest accuracy or performance on various tests of individual differences (auditory processing ability, English vocabulary, attention, and familiarity with the words employed in the replication study) were linked to how training condition differences affect training outcomes. To sum up, although their findings supported the claim that the perceptual training paradigm is useful for participants to improve their perception of L2 segment contrasts, Brekelmans and colleagues' statistical analyses revealed ambiguous evidence for the benefit of the HV training condition over the LV training condition. These results suggested that if a benefit of the HV training condition exists, the effect is much smaller than that reported in the original studies.

By contrast, several studies have reported different results. In Sadakata and McQueen's study (2013), native Dutch speakers without any experience of Japanese trained to acquire a Japanese geminate-singleton fricative contrast under either of two training conditions: HV (fewer repetitions of a more varied stimuli recorded by five talkers), and LV (many repetitions of less limited stimuli recorded by a single talker). The authors found that the participants in the HV training condition generalized their learning to a novel talker and other segments better than those in the LV training condition. The benefit of the HV training condition observed in identification tests, however, was not observed in discrimination tasks: participants' discrimination performance improved irrespective of their training

conditions. Similarly, Wong (2012) reported that although participants in both training conditions (HV: six talkers and multiple phonetic contexts; LV: single talker and single phonetic context) improved after training, those in the HV training condition outperformed those in the LV training condition and generalized better in identification tasks. The effectiveness of the HV training condition over the LV training condition was also shown in Wong (2014).

The effects of talker variability on perceptual training of lexical tone have also been examined. As in training studies on segmental contrasts discussed so far, studies on lexical tone contrasts have shown mixed results.

Silpachai (2020) investigated the role of talker variability in the perceptual training of Mandarin Chinese tones by native speakers of American English who had neither any prior experience of any tone language nor any prior formal musical training. In order to examine the effects of talker variability, he randomly assigned the participants to one of two training conditions: HV training (with stimuli produced by four talkers) and LV training (with stimuli produced by a single talker). In all experimental phases (pretest, training, posttest, four tests of generalization, and retention test) his participants performed a four-alternative forced-choice (4AFC) identification task in which they were asked to identify lexical tones they had heard. All target words were monosyllabic, except for words employed in two of the four generalization tests: these were disyllabic words (with stress on the first syllable and neutral tone in the second syllable). The stimuli consisted of the target words embedded in carrier sentences. Specifically, one carrier sentence was used for the monosyllabic target words and another one was used for disyllabic target words. No feedback was given at the pretest. The training consisted of eight sessions over approximately two weeks. Each session lasted about 30-60 minutes, depending on participants' individual pace. During each training session, stimuli recorded by only one talker were used and feedback was provided, unlike at the pretest. The posttest conducted after the training was identical to the pretest. After the posttest, three of the participants took four tests of generalization without feedback. These tests were in order to assess whether the participants could generalize to novel target words and voices in monosyllabic and disyllabic target words. Six months after the posttest/tests of generalization, the retention test

(identical to the pretest) was carried out. Silpachai observed that while both training groups improved from the pretest to the posttest and the tests of generalization, the participants in the HV training condition overall outperformed those in the LV training condition immediately after training. The finding for the retention test also indicated a benefit of the HV training condition. However, both groups generalized their learning to novel monosyllabic, but not disyllabic, words recorded by a familiar and a novel talker, and neither improved their perception of Tone 1 (with a high-level pitch). Although HV training does not appear to improve perception of more tone categories or generalize lexical tonal learning to new phonetic contexts as compared to LV training, these results suggest that HV training is generally more effective than LV training.

However, the evidence on the effects of talker variability on perceptual learning of lexical tone is equivocal. Deng and colleagues (2018) used functional magnetic resonance imaging (fMRI) to explore the neural plasticity involved in HV training, comparing with that in LV training, however, it is beyond the interest of the current research to review these results, and here I only focus on their training and behavioral tests. Their participants were native speakers of American English who had neither significant musical expertise (age of onset of musical training was before seven years or continuous musical experience lasted for less than six years) nor prior experience of Mandarin Chinese or any other tone language. They were assigned to one of two training conditions: the HV training condition (using stimuli produced by four talkers) and the LV training condition (using stimuli produced by a single talker). Their training was not typical HVPT: the participants trained to acquire pitch patterns, which resembled four Mandarin tones, within a monosyllabic nonce word in order to learn the association between sounds and pictures (representing the meanings). The training consisted of nine 30-minute sessions: one session per day and no more than a 2-day gap between the sessions. While the pretest—a tone identification test—was identical to the posttest, a test of generalization (with target words produced by four novel talkers) was a tone word identification test in which the participants were asked to identify each word by selecting the corresponding picture, as in the trials during the training sessions but without feedback. Deng and colleagues observed that the participants in both

training conditions showed pretest-posttest improvement. However, they did not find a significant difference due to the training conditions. They also found no difference in outcomes between training conditions at the test of generalization. In order to explore individual differences in training outcomes, the researchers further analyzed the results for the test of generalization, dividing the participants into two groups based on the participants' pretest scores. Their analyses showed that the high-score group obtained significantly higher scores in the test of generalization than the low-score group. The authors also examined which variables (training conditions, age, gender, IQ, pretest scores, auditory working memory, phonological awareness) predicted generalization test scores. Their statistical analyses revealed that only the pretest score predicted generalization test scores.

To summarize, the studies discussed so far have provided mixed evidence for the role of talker variability in perceptual training. Several works have also investigated the relationship of talker variability with individuals' perceptual abilities. These works are reviewed in the next section.

2.2.3.1. Talker Variability and Individuals' Perceptual Abilities

This section provides a review of some studies which have examined both the role of talker variability during perceptual training and its relationship with individuals' perceptual abilities.

Perrachione et al. (2011) found that talker variability during training differently affected the learning outcomes of participants with high or low perceptual abilities. In their Experiment 1, participants were native speakers of American English speakers with no prior experience of tone languages. Participants' basic perceptual abilities for pitch were assessed by means of a pitch-contour perception test prior to training. In this test, the participants were asked to identify the pitch contours (level, rising, and falling) they had heard. Based on the results of this test, the researchers divided their participants into two groups: high-aptitude learners and low-aptitude learners. The participants in each group were randomly assigned to one of two training conditions: the HV training condition (with stimuli

produced by four talkers), and the LV training condition (with stimuli produced by one talker). In the HV training condition, there was trial-by-trial variability (talkers changed at each trial). In both training conditions, the stimuli consisted of monosyllabic nonce words superimposed with three pitch contours (level, rising, and falling), with each nonce words associated with a common object. The participants' task was to identify the corresponding objects to words that they had heard. Feedback was provided during the training sessions. After completing eight days of 1-hour training sessions, the participants took a posttest. The posttest differed from the training in that stimuli had been produced by four novel talkers. The authors reported that the high-aptitude learners outperformed the low-aptitude learners in the posttest irrespective of the training conditions. It was also observed that the high-aptitude learners received benefit from both training conditions, although the accuracy of the high-aptitude learners in the HV training condition was slightly higher than that in the LV training condition, and the HV training condition led to greater generalization to novel talkers. Interestingly, the low-aptitude learners were impaired by the HV training condition compared to the LV training condition. The accuracy of the low-aptitude learners in the LV training condition was higher than that of the low-aptitude learners in the HV training condition.

To better understand why the HV training condition was detrimental to the low-participants learners, Perrachione and colleagues conducted a follow-up experiment (Experiment 2), looking at variations across trials in the amount of talker variability in the HV training design. Specifically, the researchers introduced a HV training variant in which the stimuli from each talker were presented in separate blocks. In this experiment, new participants who met the same criteria as in Experiment 1 were presented with identical stimuli to those used in Experiment 1. The results indicated that trial-by-trial variability (mixing stimuli produced by four talkers as in the HV training condition in Experiment 1) was detrimental only to the low-aptitude learners. But they showed significant learning outcome in the talker-blocked HV training condition. On the other hand, the high-aptitude learners did not show improvement in learning outcome due to the talker-blocked condition. Given these results for the talker-blocked HV training condition, the current

research also adopted it for the HV training, in order to facilitate learning pitch-accent patterns regardless of participants' individual perceptual abilities.

Similarly to Perrachione et al. (2011), Sadakata and McQueen (2014) also examined the relationship between talker variability and participants' pitch aptitude. Their participants were native Dutch speakers who had not had substantial exposure to Mandarin Chinese. The participants trained by engaging in a two-alternative forced choice (2AFC) identification task with feedback. Specifically, participants' task during training was to identify the lexical tone of the first syllable of naturally spoken disyllabic Mandarin Chinese nonce words. Thus, the participants did not need to learn the meaning of target words, unlike in Perrachione and colleagues' work (2011). They were randomly assigned to one of the three training conditions with different levels of input variability (low/medium/high). Again, unlike in Perrachione and colleagues' study in which the two training conditions differed only in terms of the number of talkers, Sadakata and McQueen manipulated input variability in terms of number of talkers and target words. Specifically, the LV training condition used many repetitions of a limited set of words produced by a single talker; the MV training condition used fewer repetitions of a more variable set of words produced by three talkers, and the HV training condition was similar to the MV training condition but with five talkers. Irrespective of training condition, training consisted of five 15-minute sessions. In Sadakata and McQueen's study, participants' perceptual abilities were measured by means of categorization of synthesized tonal continua. In this categorization task, the researchers used stimuli from a six-step tone 2 tone 3 continuum, asking participants to identify whether the sound they had heard was more like tone 2 or tone 3. Based on this result, the participants were divided into two groups: a high perceptual aptitude group and a low perceptual aptitude group. Sadakata and McQueen reported that all participants showed pretest-posttest improvement, but that there was no significant difference between the training conditions. Neither did they find a significant difference between the training conditions in the results of a test of generalization to a novel talker and novel stimuli. However, as in Perrachione et al. (2011), Sadakata and McQueen's results showed that there was an interaction between perceptual aptitude and variability condition. Specifically, the high perceptual aptitude group

benefited from HV training, while the low perceptual aptitude group benefitted more from LV training and increased variability hindered perceptual learning in them.

Based on the findings of the two studies (Perrachione et al., 2011; Sadakata & McQueen, 2014) discussed above, it may seem reasonable to extrapolate that the musician participants in the current study would benefit more from HV training, since they are musically trained individuals and generally have a good ear. However, Dong et al. (2019) reported different findings about the interaction between perceptual aptitude and talker variability in perceptual training. They also set out to investigate whether different conditions of perceptual training would affect participants' learning of novel L2 lexical tone contrasts, and whether there was an interaction between the training conditions and individuals' perceptual aptitude. To this end, they partially replicated the previous works (Perrachione et al., 2011; Sadakata & McQueen, 2014). For example, the researchers measured perceptual aptitude by means of two tests (the pitch-contour perception test, and the categorization of synthesized tonal continua test) adapting from Perrachione et al. (2011), and Sadakata and McQueen (2014), respectively. In addition, following Perrachione et al. (2011), Dong and colleagues varied only talker variability—four talkers in the HV training condition vs. one talker in the LV training condition—but kept training stimuli identical across the training conditions. However, the authors changed the training paradigm, using a vocabulary learning task in which, after hearing a target word, the participants had to select the corresponding picture between two options displayed on the computer screen. Training consisted of six 30-minute sessions. The stimuli used in tests and training (not in the two tests for measuring perceptual aptitude) consisted of real Mandarin monosyllabic target words. The participants were native speakers of different non-tone languages (most of the participants were native English speakers). None of the participants had had prior experience of Mandarin Chinese or any other tone language. Dong et al. found that all participants improved in tone perception after training, and that there was also evidence of generalization to untrained talker and stimuli. Interestingly, although the participants in the LV training condition exhibited an advantage with the stimuli produced by a talker presented in training, Dong et al. did not find either

an overall benefit of HV training for generalization or an interaction between talker variability in training and individuals' perceptual aptitude.

A further study conducted by Qin et al. (2022) differed from the three works discussed above in that their participants were native speakers of a tone language (Mandarin Chinese), and were novice learners of Cantonese. Although the current research's participants were native speakers of Italian (non-tone language), it is worth briefly discussing Qin and colleagues' study, because they also examined whether there was an interaction between training conditions and individuals' pitch aptitude. To explore this topic, they contrasted the HV training condition (using more varied tokens produced by two talkers) and the LV training condition (using less varied tokens produced by one talker), with participants being trained on Cantonese level-tone contrasts. Participants' pitch aptitude was measured by two tests: a pitch threshold test (for assessing abilities in explicit categorization of pitch height patterns), and a pretraining discrimination test (for assessing implicit sensitivity to pitch height differences). The researchers found that the participants in the HV training condition and those in the LV training condition did not differ significantly in overnight changes between the results of the first posttest (carried out short after the training) and the results of the second posttest (after a 24-hour interval). Additionally, the findings indicated that two pitch aptitude tests respectively predicted two posttests' correct responses. As for the interaction between pitch aptitude and training conditions, only pitch aptitude measured by the pitch threshold test interacted with overnight changes in results for participants in the HV training condition. Specifically, while participants with lower aptitude show a performance decline in the second posttest, those with higher aptitude did not show this. The posttest performances in participants in the LV training condition, on the other hand, were not affected by pitch aptitude measured by the pitch threshold test. As regards the pretraining discrimination test, this might not be a predictor for the overnight difference in change between the training conditions.

In summary, the studies discussed so far (including studies examining the role of taker variability, reviewed in Section 2.2.3) have shown mixed results regarding the role of talker variability and the interaction between talker variability and individuals' perceptual abilities. To better understand these topics, the current

research sought to explore them by randomly assigning musicians and non-musicians to either an HV training condition or an LV training condition.

2.3. Effect of Musical Experience on Perception of Lexical Pitch Contrasts

This section provides a review of previous studies which have addressed the issue of the relationship between lexical pitch perception and musical experience/training. The section falls into four parts. In the first part is reviewed prior work exploring the effect of musical training on the perception of Japanese pitch-accent patterns. In the second part are discussed studies examining the effect of musical experience/training on lexical tone perception. The third part focuses on studies investigating the effect of musical experience/training on lexical tone perceptual learning. The last part provides a review of prior works which have addressed the issue of whether possessing absolute pitch affects lexical tone perception.

Note that the main aim of the current research was to examine whether musical training has any effect on the perceptual learning of Japanese pitch accent by Italian native speakers with no prior knowledge of Japanese. As can be seen in the main aim, what was of interest in the present study was the effect of formal musical training on perceptual learning of pitch accent rather than musical aptitude per se irrespective of musical training. Based on this interest, the present section discusses previous studies which have compared musicians' performance with that of non-musicians to explore whether musical experience/training, rather than musical aptitude, affects perception of lexical pitch contrast or its perceptual learning.

2.3.1. Effect of Musical Training on Japanese Pitch Accent Perception

This brief section is dedicated to a discussion of research conducted by Golob (2003). To the best of the author's knowledge, this is the only study investigating the effect of musical training on Japanese pitch accent perception.

Golob's participants were Slovenian native speakers. Her cohort was comprised of two experimental groups: a musician group and a non-musician group. The musician group, who trained for two hours twice a week in a choir group, had no prior experience of Japanese. In contrast, the non-musician group, who had no previous experience playing instruments or singing training, consisted of second- and third-year students in the Japanese language department at the University of Ljubljana. Since there were no lectures on either Japanese phonetics or Japanese phonology at the department, the researcher considered the Slovenian participants as beginners in the knowledge of Japanese phonetics. In addition to the two groups, native Japanese speakers participated in the experiment as the control group.

The task that the participants performed was an identification task. The participants were asked to indicate the location of the pitch accent after listening to each audio stimulus twice. In the case of unaccented pattern, they were told to choose "none". If they did not know the correct answer, they were asked to choose "I do not know". The answer sheet was written in the Slovenian alphabet for the musician group, because they could not read in Japanese. The stimuli were real Japanese words (consisting of 3-5 moras), some of which contained one of the special moras (see Section 2.1.1. for the special moras). The stimuli were recorded by a female native speaker of Japanese.

Golob found that the musician group outperformed the non-musician group, though the control group's performance was better than the musician group's. This result suggested that musical training plays a role in pitch-accent perception. Golob also observed that accuracy was lower on trials of stimuli containing the special mora than on trials of stimuli without the special mora, indicating that L1's syllabic system interfered with the perception of Japanese pitch-accent patterns on words with special moras.

Importantly, Golob's study showed that the musician group, without any experience of Japanese, perceived Japanese pitch accent better than the non-musician group, even though the non-musicians were learners of Japanese. This means that the effect of musical training was greater than that of L2-Japanese learning experience. However, because the two groups differed in terms of L2

learning experiences, it was difficult to determine the extent to which music training alone affected the perception of Japanese pitch accent.

Since the main purpose of the current research was to investigate the effects of music training, the present study attempted to match participants' condition. Specifically, neither the musicians nor the non-musicians recruited in the current research had any previous experience of Japanese.

2.3.2. Effect of Musical Experience on Lexical Tone Perception

Since only one work has investigated the effect of musical training on Japanese pitch accent perception, this section discusses studies examining the effect of musical experience/training on lexical tone perception by native non-tone language speakers. The reason for reviewing these studies is because lexical tone, although having a higher functional load than pitch accent in Japanese, is also realized by pitch modulation.

In Section 2.3.3 is discussed perceptual training research examining whether musical experience/training has a bearing on perceptual training of lexical tone in naïve learners of tone languages. In Section 2.3.4 are discussed studies exploring the effect of absolute pitch on lexical tone perception in non-tone language native speakers.

Alexander et al. (2005) investigated the effect of extensive experience with pitch processing in music on pitch processing in speech. To this end, the authors tested American-English speaking musicians and non-musicians with an identification task and a discrimination task of mandarin Chinese lexical tones. While the nine musicians had had eight or more years of continuous private piano or voice lessons until or beyond the year 2002, the nine non-musicians had had no more than three years of continuous private music lessons of any kind and had not studied any instrument beyond the year 1997. None of the participants had had any prior exposure to Mandarin. As mentioned above, all participants took two tasks: a two-alternative forced-choice (2AFC) identification task and an AX discrimination task. While in the identification task, the participants were required to select one of

two arrows on the computer screen representing the audio stimulus they had heard; in the discrimination task, they were asked to answer whether the two stimuli they had heard were the same or different. Stimuli used in both tasks were real monosyllabic Chinese words produced by two male and two female native speakers of Chinese. In addition, in both tasks, accuracy and reaction time were logged. Alexander and colleagues found that the musicians outperformed the non-musicians in both tasks. The reaction-time data also showed that the musicians were faster than the non-musicians. Their results suggested that musicians' experience with music pitch processing had a facilitative effect on the processing of lexical tone.

Similarly, Gottfried (2007) examined the extent to which musical training was related to better performance on perception and production of lexical tone. All participants in his experiments were native speakers of American English at Lawrence University, and none of them had studied Mandarin Chinese. Here it is worth mentioning some methodological differences from Alexander et al. (2005). Firstly, Gottfried recruited more participants: 38 participants (24 musicians and 14 non-musicians) in Experiment 1 and 42 participants (25 musicians and 17 non-musicians) in Experiment 2. In addition, Gottfried's musicians were chosen according to a more demanding set of criteria. Specifically, the musicians were students enrolled at the conservatory of music. By contrast, non-musicians were not conservatory students, although some of them had taken musical lessons for more than five years (Experiment 1) and some of them had taken part in music ensembles or taken some music lessons in high school and college. Lastly, whereas in Alexander et al., participants only carried out two tasks, in Gottfried's experiments, participants carried out a battery of tasks: a tone glide identification task and a Mandarin tone identification task (Experiment 1); and an AX discrimination task and an imitation task (Experiment 2). In Experiment 1, Gottfried found that the musicians performed significantly more accurately than the non-musicians when determining whether a sine-wave tone went up, down, or remained the same in pitch (the tone glide identification task). The musicians were also more accurate than the non-musicians in the four-alternative forced-choice (4AFC) Mandarin tone identification task, although identification accuracy for all participants was relatively low. In Experiment 2, Gottfried reported that the musicians not only

discriminated but also imitated lexical tones significantly more accurately than the non-musicians. These results indicated that their musical training facilitated their perception and production of Mandarin lexical tones, suggesting musicians' advantage in the initial stages of tone language learning.

Gottfried and Xu (2008) provided further evidence for the effect of musical experience on Mandarin lexical tone discrimination and imitation. Note that the researchers divided their English-speaking participants into the musician and non-musician categories based on participants' self-rating of musicianship on an 8-point scale. Even though the definition of musician was not as strict as that in Gottfried (2007) or in Alexander et al. (2005) discussed above, the musicians discriminated unfamiliar lexical tones significantly more accurately than non-musicians. However, Chinese native participants' performance was significantly better than that of the musicians.

The results for the effect of musical training/experience on lexical tone perception were also corroborated by the findings of Delogu et al. (2010). Delogu and colleagues' study is noteworthy because their participants were native Italian speakers, as are those in the current research. The researchers tested Italian participants with no experience of tone languages, administering an AX discrimination task on sequences of monosyllabic Chinese Mandarin words that included tonal and segmental variations. Their participants consisted of three groups: (1) university students who had not received any previous formal musical education (non-musicians); (2) learners of Chinese who had studied Chinese for at least three years in Oriental studies at university (for this group no mention of musical experience was found); (3) musicians who were either graduates or about to graduate in music at the Conservatory of Frosinone and had had at least five years of continuous practice on an instrument. The participants engaged not only in the discrimination task but also in a test for measuring musical aptitude (specifically musical ability in detecting melodic changes regardless of experience with musical training). Overall, all Italian participants showed difficulty in discriminating tonal variations as opposed to segmental variations. Delogu et al. also found that the musicians performed not only significantly better than the non-musicians on lexical tone discrimination but did so comparably to the learners of Chinese, even though

the musicians were naïve to tone languages. This suggested that musical training was effective in facilitating the discrimination of lexical tones. By contrast, no effect for musical training was observed on the discrimination of segmental variations. These results were partially replicated in a follow-up study conducted by Marie et al. (2011) with participants who were native French speakers and who had had no experience with tone languages. Both in lexical tone discrimination and in segmental discrimination, their musicians—who had played different instrument and had started music lessons around the age of seven, with 16 years of musical training on average—outperformed their non-musicians—who had not received any formal musical training.

Coming back to Delogu and colleagues' work (2010), I now briefly touch upon Delogu and colleagues' findings regarding the musical aptitude test, although musical aptitude was not what was of interest in the present study. In accordance with the results of this test, the researchers divided their participants into three groups: high, medium and low. Delogu et al. observed that the high-scoring group performed better than the other two groups, and that the medium-scoring group performed better than the low-scoring group in lexical tone discrimination. On the contrary, no difference between the three group was found in segmental discrimination. The researchers concluded that both musical aptitude (melodic abilities) and musical expertise were able to enhance the discrimination of lexical tones.

However, a recent study conducted by Götz et al. (2023) reported different findings regarding the role of musical aptitude. Götz and colleagues explored whether musicality—in terms of musical training and musical aptitude—positively influenced the perception and production of three types of Thai speech contrasts (namely, consonants, vowels, and lexical tones). Their participants were university students who were native speakers of Australian English without any previous exposure to a tone language. The participants consisted of two groups: musicians (instrumentalists/singers) who had had at least five years of continuous formal musical training; and non-musicians who had had no more than two years of musical training. The participants engaged in two tasks: (1) a three-alternative forced-choice (3AFC) identification task that fell into three parts, namely consonant

identification (three levels of voicing), vowel identification (three vowel qualities), and tone identification (three tone levels: low, mid, high); and (2) an imitation task (participants' productions were rated by two native Thai phoneticians). The participants also took a musical aptitude test to measure two aspects: tone and rhythm. Götz et al. found that the musicians performed significantly better than the non-musicians in both the perception and the production of all speech types. This result indicated a robust advantage for musicians over non-musicians in identifying and imitating Thai consonants, vowels, and lexical tones. Interestingly, the advantage conferred by musicality in Götz and colleagues' study seemed to be based primarily on formal musical training. Indeed, as regards musical aptitude, only rhythm scores (as opposed to tone scores) had a marginally significant effect on perception, and neither of the two aspects had a significant effect on production. In addition, Götz and colleagues' statistical analyses revealed that hours of weekly musical practice (reflecting formal musical training) had a significantly positive influence on both the identification task and the imitation task. This finding provided further evidence that the effect of musicality was mainly on account of formal musical training, not musical aptitude.

Turning back to studies that have examined the effect of musical experience/training only on lexical tone perception, Mok and Zuo's study (2012) also indicated that musical training enhanced lexical tone perception in native speakers of non-tone languages without any experience of tone languages. Their participants were native English and French speakers, and native Cantonese speakers. The non-tone language group (French and English); and the tone-language group (Cantonese) were both subdivided into two categories: musicians who had received more than seven years of formal musical training in any instrument or vocal singing and had played music regularly in the past two years; and non-musicians who had received no more than two years of casual musical experience and had not played music regularly in the past two years. The participants were tested with two AX discrimination tasks: one of Cantonese monosyllables and the other of pure tones resynthesized from the six Cantonese lexical tones. The researchers collected data regarding both accuracy and reaction time. The results showed that the non-tone musicians were significantly more

accurate than non-tone non-musicians in both discrimination tasks, although there was no significant difference between the two groups in reaction time. Interestingly, the non-tone musicians' accuracy was comparable to that of native Cantonese speakers (musical training had little effect on the performance of native Cantonese speakers). Hence, Mok and Zuo's findings provide further evidence for the effect of musical training on lexical tone perception by native speakers of non-tone languages.

The studies reviewed so far in this section seem to indicate a clear positive effect for musical training/experience on lexical tone perception. However, Chang and colleagues' work (2016), which examined the categorization of Mandarin Chinese lexical tones by participants differing in linguistic and musical experience, reported only a partial advantage for musical experience. Their participants consisted of three groups: native English musicians, native English non-musicians, and native Mandarin Chinese non-musicians. The native English speakers had no prior knowledge of a tone language. Their musicians had received five or more consecutive years of formal training with a western-style instrument and had played the instrument on a regular basis in the past five years, while their non-musicians had received at most three years of musical training and had not had any formal musical training within the past five years. The researchers administered to their participants two categorical perception tasks: an AX discrimination task and a two-alternative forced-choice (2AFC) identification task. Stimuli used in these tasks were Mandarin rising-level and falling-level tone continua, created by resynthesizing natural speech production of the Chinese real word syllables. Chang and colleagues found that, in the identification task, the English musicians followed the native Chinese non-musicians' pattern showing more categorical perception than the English non-musicians. However, in the discrimination task, the English musicians outperformed the Chinese non-musicians, but not the English non-musicians. This result suggested that musical experience did not have a facilitative effect on lexical tone discrimination. The authors reasoned that this was because, even though musical experience may have enhanced sensitivity to pitch, this ability may not have led to improved lexical tone discrimination, since lexical tone

discrimination required ignoring subtle differences in speech-related F0 in order to make categorical judgements.

Like Chang et al. (2016), Chen et al. (2020) investigated categorical perception of Mandarin Chinese lexical tones in participants differing in linguistic and musical experience, by testing them with an AX discrimination task and a three-alternative forced-choice (3AFC) identification. However, the results for Chen and colleagues' two tasks revealed that English musicians without exposure to tone languages showed sharper category boundaries than English non-musicians without exposure to tone languages. The results also indicated that musical experience contributed to categorical perception more consistently in native speakers of American English without exposure to tone languages than native speakers of Mandarin Chinese. In their study, both the English musicians and the Chinese musicians had received professional training.

Finally, Kirkham et al. (2011) explored the effect of different musical training backgrounds. Kirkham and colleagues divided musicians into vocalists and instrumentalists in order to examine whether one of the two backgrounds was more advantageous. Their participants were comprised of four groups: (1) native Mandarin-Chinese speaking non-musicians; (2) native English-speaking non-musicians; (3) native English-speaking instrumentalists; and (4) native English-speaking vocalists with significant vocal experience as well as instrumental experience (mainly piano). Both instrumentalists and vocalists had received at least four years of formal training and were still playing their instrument or singing. The participants engaged in an AX discrimination task and a self-paced imitation task with stimuli (real Chinese monosyllabic words). Kirkham et al. found that in both tasks, although the Chinese non-musicians outperformed all English participants, both the instrumentalists and the vocalists outperformed the English non-musicians, and there was no significant difference between the instrumentalists and the vocalists. On the strength of Kirkham et al.'s results, musicians were not divided based on musical background in the current research.

Although they differed in terms of the tasks and lexical tones used in their experiments, taken together, the studies reviewed above largely converged to

indicate the positive effect of musical training and musical experience on lexical tone perception, albeit to varying degrees.

However, it should be noted that each of the works discussed earlier defined musicians in its own way. For example, while Gottfried and Xu (2008) divided their participants into musicians and non-musicians based on participants' self-rating of musicianship on an 8-point scale, Chen et al. (2020) recruited musicians with professional training. Additionally, almost all other studies reviewed here defined musicians in terms of the number of years of musical training, but what was meant by the term "musical training" varied between studies: it could be not just simple "musical training", but also "continuous musical training", or "formal musical training", or "continuous formal musical training". Indeed, Ong et al. (2020) pointed out that there is no consensus about the definition in the literature. The fact that various studies use different terms to describe almost the same concept—such as musical training, musical experience, musical expertise, musicality, and musicianship—reflects the current lack of general agreement on the definition of musicians in the literature.

Ericsson and colleagues' study (1993) is interesting because it highlights the variability in definitions of musicianship. The researchers assessed and compared current and past amounts of practice in two age-matched musician groups: expert pianists who were students at the Music Academy of Berlin, a highly selective conservatory of music in West Berlin (Tan et al., 2018); and amateur pianists who were students in non-musical academic or vocational training programs. Over a diarized week, the expert pianists reported spending 56.75 hours on all music-related activities, with 26.71 hours dedicated to solo practice at the piano, while the amateur pianists reported spending only 7.02 hours on all music-related activities, with 1.88 hours dedicated to solo practice at the piano. The researchers calculated the accumulated amount of solo practice based on estimates of weekly practice as a function of age for each of the two groups. They found that there was a significant different between the two groups in the estimated amount of practice accumulated by age 18 years: while the expert pianists had accumulated 7,606 hours of practice, the amateur pianist had accumulated only 1,606 hours. This result was corroborated by a subsequent study (Sloboda et al., 1996) which estimated that total practice time

for students of a selective specialist music school was just under 7,000 hours. Note, however, that the amateur pianists in Ericsson and colleagues' study were accomplished amateur pianists. Indeed, they were able to successfully play Prelude No.1 in C-major by J. S. Bach, employed in the musical performance task, and they had had between five and twenty years of experience playing piano (just under 10 years, on average). Therefore, despite the considerable difference between the musicians defined as expert and those defined as amateur in Ericsson and colleagues' work, most of their "amateur" musicians would have been classified as "musicians" in many of the studies reviewed above.

Taking into account the findings of Ericsson et al. (1993) and Sloboda et al. (1996), the current research defined musicians as individuals currently engaged in formal tertiary-level musical training, including those enrolled in conservatories, musical institutes, or majoring in musicology at university. This definition is in line with Delogu et al. (2010) and Gottfried (2007). By contrast, the non-musicians were defined as people with three years or less of continuous private musical training, and who were not taking music lessons at the time of recruitment (Alexander et al., 2005; P. C. M. Wong & Perrachione, 2007). Thus, the present study excluded the intermediate category—between the expert musicians and the non-musicians—like the amateur musicians in Ericsson et al. (1993) and Sloboda et al. (1996), since it would have hindered accurate determination of the effect of musical training. As described in Section 4.2.1, since all participants were higher education students in Lombardy, the musicians were comparable to the non-musicians in terms of ages and levels of education.

2.3.3. Effect of Musical Experience on Lexical Tone Perceptual Training

This section reviews works exploring the effect of musical experience/training on perceptual training of lexical tone in native speakers of a non-tone language, via a comparison of musicians versus non-musicians. To the author's knowledge, there have been no studies exploring the effect of musical experience/training on perceptual learning of Japanese pitch accent. According to Ong et al. (2020), the training studies discussed here can be said to fall into two types: lexical tone

category training (via discrimination and/or identification), and sound-meaning mapping training (learning the associations between lexical tones and images/objects). Here these two types are reviewed in order.

As regards the former type of training, first and foremost it is worth mentioning Zhao and Kuhl's work (2015), since they employed the HVPT paradigm. They adopted a pre/posttest design in order to investigate whether musical experience influenced the perceptual learning of lexical tone categories. Their target tone categories were tone 2 (rising) and tone 3 (falling-rising) in Mandarin Chinese. Their participants were native English speakers who had neither had any formal experience with any tone languages, nor had lived in any tone-language-speaking countries for more than two months. The participants were made up of musicians and non-musicians. While the musicians had received at least eight years of private music lesson beginning before the age of 10 years, the non-musicians had received less than two years of private music lesson which had ended more than five years before. Native Mandarin-speaking non-musicians also took part in the study, but only in a pretest. In all experimental phases (pretest, training, posttest, and two tests of generalization), the researchers used stimuli produced by five female native speakers of Mandarin. While training stimuli were real monosyllabic Mandarin words, test stimuli were a nine-step tone 2 tone 3 continuum based on real monosyllabic Mandarin words. In the tests, the authors gave their participants two tasks: an AX discrimination task and an ABX identification task. In the former, the participants were instructed to judge whether the first stimulus was the same as or different from the second stimulus. These stimuli carried F0 contours that were either two steps apart on the continuum or the same. In the latter, the participants were instructed to judge whether the second stimulus was more similar to the first stimulus or the third stimulus. While the F0 contours of the first and the third stimuli remained the end points of the continuum, that of the second stimulus varied within the continuum. The training consisted of eight sessions over two weeks. Half of the English-speaking musicians and non-musicians were assigned to a training group and the other half to a control group who did not take part in the training. During the sessions, trainees engaged in a two-alternative forced choice (2AFC) identification task with feedback. The results for

the discrimination task in the pretest showed that the musicians exhibited a significantly higher overall sensitivity than both the English-speaking non-musicians and the Mandarin-speaking non-musicians, although the pattern of perception of the tone continuum showed by the Mandarin-speaking non-musicians was different from that showed by the English-speaking participants (English-speaking musicians and non-musicians exhibited a similar pattern). By contrast, the findings for the identification task in the pretest revealed no significant difference between the groups. A comparison between participants' perception in the pretest and the posttest and the two tests of generalization indicated that, in the discrimination tasks, training led to overall sensitivity improvement, but not to the formation of robust phonological categories. Identification data revealed that the participants in all conditions improved in the posttest and the first test of generalization (which used novel stimuli produced by one of the familiar talkers in the training phase). However, in the second test of generalization (which used novel stimuli produced by a novel talker), only the musicians in both the training group and the control group improved compared with their performance in the pretest. The results of Zhao and Kuhls' study suggested that there was no significant difference between the improvement of the musicians and that of the non-musicians.

Wayland and colleagues' study (2010) also used a pre/posttest design and showed similar results to those of Zhao and Kuhl's study (2015) discussed above. Wayland and colleagues' participants consisted of native speakers of different non-tone languages (most of them native English speakers). Half of the participants were musicians with at least six years of musical experience, while the other half were non-musicians with a maximum of two years of musical experience and not practicing at the moment of the experiment. All participants took part in a pretest, training, and a posttest. The training consisted of three 30-minute sessions held in a laboratory. The task used during training—a two-alternative forced choice (2AFC) identification task with feedback—differed from the task used in the pretest and posttest, an AAX categorial discrimination in which the participants were asked to answer whether the pitch contour of the last stimulus was the same or different from that of the preceding two stimuli. The researchers thus explored whether the participants were able to generalize the pitch contour identification that they had

learned during training to abstract and categorize pitch contour. The stimuli used throughout the experiment were naturally produced nonsense syllables whose F0 contours were then acoustically modified to produce linearly rising and falling F0 contours. Wayland et al. found that in the training phase, the musicians were better able to identify pitch contours than the non-musicians. However, both the musicians and the non-musicians showed comparable levels of improvement. In addition, the researchers reported that the musicians showed a significant pretest-posttest improvement, but that the musicians' performance was not significantly better than that of the non-musicians.

Taken together, the results of these two lexical tone category training studies indicated that musicians had no additional advantage: musicians and their counterparts improved to the same extent after training (Ong et al., 2020).

The second training type—sound-meaning mapping training—was employed by Wong and Perrachione (2007) to examine the learning of non-native suprasegmental patterns to achieve word identification, by native speakers of American English. None of their participants had had any exposure to a tone language. The participants fell into two categories: amateur musicians with at least six years of formal private lessons in a single instrument beginning before the age of ten; and non-musicians with no more than three years of private lessons in any instrument. The participants took a pitch pattern identification test in which they were asked to indicate the pitch pattern (level, rising or falling) of stimuli they had heard. After the pitch pattern identification test, they engaged in training. Stimuli used in the training sessions were English nonce words superimposed with three pitch patterns resembling three Mandarin lexical tones (level, rising and falling), similar to stimuli in the pitch pattern identification test. Each training session lasted approximately 30 minutes. In training sessions, participants were firstly trained to associate the image of an object with one of the nonce words: they heard each word four times with its corresponding picture. Then they were quizzed on the words: they were asked to select from among the three pictures the one corresponding to the word they had heard. Feedback was given during the quiz. At the end of each training session, a word identification test was administered. The participants were asked to identify the word they had heard by selecting the corresponding picture

from among the possible choices. Unlike during the quiz, no feedback was provided. The score of the word identification test was utilized to determine when to end training: when the participants became "successful learners", reaching higher than 95% accuracy for two consecutive sessions, or when they became "less successful learners", showing less than a 5% improvement for four consecutive sessions. Wong and Perrachione found that, although all participants showed improvement thanks to training, nine out of their 17 participants were successful learners and eight were less successful learners. Additionally, the results for the pitch pattern identification test showed that the scores of the successful learners were significantly higher than those of the less successful learners. Importantly, seven of the nine successful learners were amateur musicians. This suggested that musicians are generally better able to learn to perceive word pitch patterns.

Using a training program similar to that used by Wong and Perrachione (2007), Cooper and Wang (2012) explored how musical experience and L1 tone background influenced Cantonese word learning and lexical tone perception. Their participants consisted of native English speakers and native Thai speakers. None of the participants had any prior knowledge of Cantonese or any other tone language apart from their L1. Each language group was subdivided into musicians and non-musicians, resulting in four groups of participants: (1) English musicians, (2) English non-musicians, (3) Thai musicians, and (4) Thai non-musicians. While their musicians had received at least seven years of continuous Western instrumental musical training and were able to play an instrument at the time of the experiment, the non-musicians had received no musical training within the previous five years and less than two years of musical experience prior to that. The participants engaged in training which consisted of seven 30-minute sessions over two weeks. The stimuli used in the training sessions were monosyllabic Cantonese words produced by four native Cantonese speakers. Each of the words had a specific meaning represented by an image. During training sessions, the participants learned to identify word meaning distinguished by five Cantonese lexical tones. At the end of each training session, they took a session test on all words learned during the session. The participants also completed a Cantonese tone identification pretest and posttest, in keeping with the adopted pre/post design. Stimuli employed in the tone

identification pretest and posttest were monosyllabic Cantonese words with five Cantonese tones produced by two native Cantonese speakers. The findings for the training session tests showed that each group improved significantly from the first session to the last session. There was no significant difference between the groups on the first session test. On the other hand, on the last session test, the English musicians and the Thai non-musicians showed significantly higher accuracy rates than the English non-musicians, although the researchers did not find any significant differences between the English musicians and the Thai non-musicians, or between the Thai musicians and any of the other groups. As for the results of the pretest and the posttest, Cooper and Wang' s statistical analyses revealed overall a pretest-posttest improvement in tone identification accuracy across the groups. In addition, English musicians showed significantly higher accuracy rates than all other groups across the tests. Thai musicians' performance was also significantly better than of the Thai non-musicians across the tests. There were no significant differences between the two non-musician groups, or between the Thai musicians and the English non-musicians. These results indicated that musical experience was beneficial for non-native tone identification and tone word identification, although the combination of musical experience and L1 background with lexical tone and did not offer an added advantage for the Thai musicians.

Similarly to Cooper and Wang (2012), Maggu et al. (2018) also investigated the effects of the combination of musical and linguistic pitch experience, employing a training program similar to that used by Wong and Perrachione (2007). The participants consisted of six groups: (1) English (monolingual) musicians; (2) English (monolingual) non-musicians; (3) L1-Cantonese + L2-Mandarin musicians; (4) L1-Cantonese + L2-Mandarin non-musicians; (5) Mandarin musicians; (6) Mandarin non-musicians. While the musicians had undergone six or more years of formal musical training on any musical instrument, the non-musicians had undergone less than three years of formal musical training. Monosyllabic English nonce words were superimposed with five pitch contours of Thai lexical tones (falling, rising, high, middle, and low). The participants took part in 10 training sessions (each lasting 30-45 minutes) in which they learned to associate each nonce word with its corresponding image. At the end of each session,

the participants took a word identification test in which no feedback was given. Maggu et al. found that the English musicians outperformed the English non-musicians throughout the training sessions. They also observed that the musicians showed comparable learning curves throughout the 10 training sessions regardless of language background. By contrast, among the non-musicians, the English non-musicians had a shallower learning curve than the Mandarin and L1-Cantonese + L2-Mandarin non-musicians. The findings for the last word identification test also showed that whereas the effect of language background was not significant for the musicians, for the non-musicians it was. Indeed, both the Mandarin non-musicians and the L1-Cantonese + L2-Mandarin non-musicians outperformed the English non-musicians. In line with Cooper and Wang (2012), these findings suggested that musical experience facilitated tone word learning, although the effects of language and music pitch experience were not additive.

The studies investigating the second training type which have been reviewed so far (Cooper & Wang, 2012; Maggu et al., 2018; P. C. M. Wong & Perrachione, 2007) showed a positive effect of musical training/experience on tone word identification and lexical tone identification (in the case of Cooper & Wang, 2012).

However, Tong and Tang (2016) reported different results. They also investigated the relative contributions of L1 tone background and musical experience to non-native tone identification and tone word learning. Their participants consisted of six groups: (1) Cantonese musicians; (2) Cantonese non-musicians; (3) Mandarin musicians; (4) Mandarin non-musicians; (5) English musicians; (6) English non-musicians. The musicians had received at least seven years of continuous musical training on Western instruments, and/or had received official recognition (e.g., ABRSM: the Associated Board of the Royal Schools of Music) of level eight or above. By contrast, the non-musicians had not received any musical training within the previous five years and had no more than two years of musical experience. None of the participants had any prior knowledge of Thai or any other tone languages aside from their L1. The participants took a tone identification test before training. Stimuli used in the test were monosyllabic Thai word with five Thai lexical tones. After the test, the participants took part in four 30-minute training sessions over two days. Training procedure was similar to that

described in Cooper and Wang (2012). During the training sessions, the participants trained to identify word meaning distinguished by five Thai lexical tones. At the end of each training session, they took a tone word identification test. Here only findings relevant to the present dissertation are discussed: the effects of the difference between L1 tone-languages–Cantonese with six lexical tones versus Mandarin with four lexical tones–are beyond its scope (see Tong & Tang, 2016 for their detailed results). The findings for the tone identification test showed that the musicians outperformed the non-musicians, irrespective of their L1 background. However, the findings for the tone word identification tests conducted in each training session revealed no significant difference between the English musicians and the English non-musicians in all sessions. In addition, while the English musicians performed similarly to the Cantonese non-musicians in the first session test, English musicians' performance was significantly worse than that of the Cantonese non-musicians in the last session test. These results suggested that musical experience facilitated tone identification regardless of L1 tone background, but that the effect of musical experience decreased in later stages of perceptual training.

Dittinger et al.'s (2016) results were consistent with those of Tong and Tang (2016). The participants in Dittinger and colleagues' study were native French speakers. Half of them were professional musicians and the other half were non-musicians (no formal musical training). During perceptual training, which was similar to that in Wong and Perrachione (2007), the participants learned to associate monosyllabic Thai words with their corresponding image. Before and after training, the participants engaged in four different phonological categorization tasks in which were presented monosyllabic Thai words varying in pitch (low versus high), vowel length, aspiration, and voicing. Here only findings for pitch categorization task are discussed: the other contrasts presented in the phonological categorization were beyond the interest of the current research (see Dittinger et al., 2016 for their detailed results for the other phonological categorization tasks). Dittinger et al. found that musical training facilitated pitch categorization tasks before and after training. However, only the non-musicians showed an improvement from the pre-

training task to the post-training task, indicating that the effect of musical training decreased in the post-training task as opposed to the pre-training task.

To sum up, the studies reviewed in this section show mixed findings regarding additional musical advantage. Furthermore, the methodological differences between them—such as training methods, procedures, duration, and use of pre/posttest designs—make comparisons difficult. Note again that, like the works in Section 2.3.2, which examined the effect of musical experience/training on lexical tone perception by native non-tone language speakers, each of the studies discussed in this section also defined the musician in its own way. This reflects the fact that there is currently no consensus about the definition in the literature (Ong et al., 2020). Recall, however, that Ericsson and colleagues' study (1993) demonstrated a substantial difference between amateur musicians and expert musicians; see Section 2.3.2. for a detailed discussion. As noted in Section 2.3.2, establishing a clear difference between the two experimental groups—by contrasting non-musicians with expert musicians instead of with amateur musicians—should enable results that discern the effect of musical training.

The present study thus excluded the category intermediate between expert musicians and non-musicians in order to favor clear determination of the effect of musical training. The effect of musical training on perceptual learning of Japanese pitch accent by native Italian speakers was thus investigated via a comparison of expert musicians versus non-musicians.

2.3.4. Effect of Absolute Pitch on Lexical Tone Perception

This section provides a review of studies which have addressed whether possessing absolute pitch affects lexical tone perception, since—to the author's knowledge—no study has yet explored the effect of absolute pitch on Japanese pitch accent perception.

Absolute pitch is the ability to name a note of a particular pitch without a reference note (e.g., naming a tone as "C", "261 Hz", or "do"), or to produce a note of a particular pitch without a reference note (Burnham et al., 2015; Deutsch et al.,

2006; Parncutt & Levitin, 2001). It is different from relative pitch—the ability to identify or to produce pitches in relation to other pitches: for instance, recognizing or singing the next note of a familiar melody, or identifying musical intervals (Ashley & Timmers, 2017).

Absolute pitch is a rare ability: its incidence is often cited as one in 10,000 people in the general population (Bachem, 1955; Marvin, 2017; Takeuchi & Hulse, 1993). Deutsch et al. (2006) claimed that absolute pitch is extremely rare in the general population of the U.S. and Europe. They also reported that even among music conservatory students, the prevalence of absolute pitch was much lower for musicians who were native speakers of a non-tone language compared to Mandarin-speaking musicians, controlling for gender and age of onset of musical training; see Deutsch et al. (2006) for detailed findings. Moreover, Gregersen et al. (1999) claimed that the prevalence of absolute pitch correlated strongly with the percentages of students in schools reporting an "Asian or Pacific Islander" ethnic background. Here I will not go into further detail, but for one possible explanation of the increased incidence of absolute pitch in people with East Asian descent, see Deutsch et al. (2004); for the review for the debate on the genesis of absolute pitch, see Loui (2016), Marvin (2017) and Patel (2008).

Following Burnham et al. (2015) discussed below, the present research adopted the traditional definition of absolute pitch, reflecting pitch labeling ability (i.e., being able to name or label a note without a reference note). This was because the aim was to investigate whether the ability to identify lexical pitch (Japanese pitch accent) correlated with the ability to identify musical pitch.

To the best of the author's knowledge, only Burnham et al. (2015) have shown that musicians with absolute pitch were more accurate at lexical tone discrimination. Their participants were native speakers of Australian English without any experience of tone languages. The participants consisted of three groups: (1) musicians who possessed absolute pitch; (2) musicians who did not possess absolute pitch; and (3) non-musicians who had not received any musical training. The division of the two musician groups (recruited from two music-training institutions in Australia) was based on the results of an absolute pitch test

administered by Burnham et al. The participants were tested using an AX discrimination task in which they were asked to answer whether the two stimuli they had heard were the same or different. Stimuli used in the AX discrimination task were a monosyllable with five Thai lexical tones. Burnham and colleagues employed two interstimulus intervals (ISIs): 500 ms and 1500 ms. Length of the ISI influences which categories (acoustic, phonetic, or phonological) listeners use in order to classify stimuli: while an ISI of 500 ms encourages acoustic processing of speech stimuli, an ISI of 1500 ms encourages phonological processing (Werker & Logan, 1985). Thus, the use of the two ISIs allowed Burnham et al. to examine whether processing level interacted with musical training and absolute pitch ability. The researchers analyzed not only discrimination accuracy data but also Reaction time (RT) data. Results for the discrimination accuracy data showed that the two musician groups outperformed the non-musicians, and that the musicians with absolute pitch performed better than the musicians without absolute pitch. These indicated that, over and above musical training, absolute pitch ability positively influenced Thai lexical tone discrimination. Neither overall effect of ISI nor significant interactions with ISI were found. The RT findings revealed that the two groups of musicians were significantly faster than the non-musicians, however, there was no significant overall difference between these groups. The musicians with absolute pitch did have a RT advantage over the musicians without absolute pitch, but only at the ISI of 1500 ms. The authors reasoned that absolute pitch ability cognitively involves long-term pitch memory, which requires internal pitch standards (templates) to identify tones using labels (Parncutt & Levitin, 2001).

Inspired by Burnham and colleagues' study (2015) discussed above, the current research sought to explore the effect not only of musical training but also of absolute pitch on perceptual learning of Japanese pitch accent. As in Burnham et al. (2015), two ISIs were employed and RT data were gathered. Furthermore, the current research adopted one of the suggestions for future work provided by Burnham et al. (2015): the use of stimuli that varied in terms of the number and gender of speakers, similar to sounds in the real world.

As discussed above, only one study—Burnham et al.(2015)—has reported an advantage in perceiving lexical tones only for musicians with absolute pitch.

Two other studies—Lee and Hung (2008) and Lee et al. (2014)—that attempted to investigate the effect of absolute pitch on lexical tone perception by English-speaking musicians were unable to recruit any musicians with absolute pitch. Again, this is an extremely rare ability.

Lee and Hung (2008) examined how English-speaking musicians and non-musicians identified Mandarin lexical tone contrasts when stimuli included intact and acoustically degraded Mandarin monosyllables (with four lexical tones) produced by multiple talkers. While the musicians were students or graduate music majors of the School of Music at Ohio University, the non-musicians had not had any formal musical training or substantial musical learning experience. None of the participants had had previous experience with tone languages. The findings showed that the musicians outperformed the non-musicians, although the musical advantage reduced as the amount of F0 information in the stimuli was decreased. This indicated that musical training had a facilitative effect on Mandarin tone identification. Interestingly, Lee and Hung also gave only their musicians an absolute pitch test. This test was based on the absolute pitch task implemented by Deutsch et al. (2006). In the test, the musicians were asked to listen to synthesized musical notes of three timbres (piano, viola, and pure tone) and to identify them without a reference pitch. These notes ranged from C3 to B5. Any two consecutive notes were separated by more than an octave in order to prevent the musicians from utilizing relative pitch for the test; see Section 6.2.2. and Section 6.2.3 for a more detailed description of the absolute pitch test. Lee and Hung found that none of the musicians met the criterion for absolute pitch. Hence, it was not possible to investigate the effect of absolute pitch on Mandarin lexical tone identification in these English-speaking musicians.

A follow-up study conducted by Lee et al. (2014) examined how English-speaking musicians and non-musicians identified the pitch height (high vs. mid) of isolated Taiwanese level tones produced by multiple talkers. As in Lee and Hungs' study (2008), the participants were students at Ohio University none of whom had had previous experience of or substantial exposure to tone languages. While the musicians were music majors, the non-musicians had not had any formal musical training or substantial musical learning experience. The results showed that the

musicians were slightly better at identifying tone height than the non-musicians. Lee et al. (2014) also administered an absolute pitch test with materials identical to those used in the 2008 study. As in their previous study, the researchers did not find any musicians who met the criterion for absolute pitch. They found no correlation between musicians' performance in the Taiwanese tone identification task or in the absolute pitch test. To sum up, although there are findings for the effect of musicianship on tonal identification tasks, it has not so far been possible to investigate the effect of absolute pitch due to the scarcity of participants having this ability.

The two studies (Lee et al., 2014; Lee & Hung, 2008) discussed above revealed the difficulty of finding musicians who are both native speakers of non-tone languages and who possess absolute pitch. This difficulty is consistent with the low prevalence of absolute pitch in American conservatory-level musicians in Deutsch et al. (2006) and that in western conservatory-level (German, Polish and US) students in Miyazaki et al. (2018), although many more students of Miyazaki et al. possessed accurate relative pitch.

It was probable therefore that finding Italian musicians with absolute pitch for the current research would also be difficult. However, it was worthwhile to seek them in order to expand our understanding of the role of absolute pitch. In order to do this, the present study employed the absolute pitch test used in Lee and Hung (2008) and Lee et al. (2014).

2.4. Summary

This section summarizes the background information that motivated the present dissertation in the same order as the sections (2.1, 2.2 and 2.3).

In the first part, several crucial differences in lexical accent between Japanese and Italian emerge: e.g., whereas Japanese exhibits lexical contrasts in terms of pitch-accent patterns, Italian makes them in terms of lexical stress and has pitch accent that contrasts different intonational meanings. Given these differences and Pappalardo's (2018) findings on native Italian speakers, it was reasonable to

assume that Italian participants in the present research would also find Japanese pitch accent perception difficult. However, unlike those in Pappalardo's work, the participants in the current study were naïve learners of Japanese. This was because having participants with previous L2-Japanese learning experience or knowledge has led to unclear findings in works investigating Japanese pitch accent perception. It was also to avoid confounding factors, because this dissertation's overarching goal was to assess the effect of musical training.

The second part of this chapter has discussed various training studies, paying special attention to Shport's studies (2011, 2016), the first HVPT studies on pitch-accent contrasts by naïve learners of Japanese. Her results suggested the effectiveness of training. The current research, based largely on her methodology, introduced two categories of participants: musicians and non-musicians.

Also reviewed are studies investigating the role of talker variability in perceptual training; and those exploring the interaction between talker variability and individuals' perceptual abilities. These works have shown mixed findings. To better understand talker variability and its interaction with individual perceptual ability, the current research sought to extend Shport's works by randomly assigning musicians and non-musicians to either an HV training condition (with stimuli produced by four talkers) or an LV training condition (with stimuli produced by one talker).

The last part of this chapter has reviewed studies addressing the issue of the relationship between lexical pitch perception and musical experience/training. As for works exploring the effect of musical training on the perception of Japanese pitch-accent patterns and lexical tone, they largely converge to indicate that, to varying degrees, musical training/experience has a positive effect. However, studies investigating the effect of musical experience/training on lexical tone perceptual training have shown mixed results. As regards absolute pitch, Burnham and colleagues' study (2015) indicated it has a positive effect on Thai lexical tone discrimination.

In the light of the background information discussed in this chapter, the issues addressed in the current dissertation are as follows: (1) the effect of musical

training on perceptual learning of Japanese pitch accent by naïve native Italian speakers; (2) the effect of talker variability in training (HV training condition vs. LV training condition); and (3) the effect of absolute pitch. These issues are discussed in the chapters that follow: the next chapter provides a general overview of the current study's experimental design, and the subsequent chapters discuss details of the experiment.

CHAPTER 3 GENERAL OVERVIEW OF THIS STUDY'S EXPERIMENTAL STRUCTURE

The aim of this brief chapter is to provide a general outline of the current study's experimental design, given its complexity.

3.1. Design

Two groups of participants (musicians and non-musicians), randomly assigned to two different training conditions (high variability HV and low variability LV), performed three types of task: identification tasks, discrimination tasks and an absolute pitch test (only musicians). The schedule below shows the training and tests carried out on each experimental Day by both groups in both conditions:

Table 3.1
Overview of the Experimental Schedule

Day	Experimental steps for musicians and non-musicians
1	Pretest: identification task; discrimination task; and Training 1
2	Training 2 (+ absolute pitch test: **only for musicians**)
3	Training 3; and Posttest: identification task; discrimination task
4 – 5	Tests of generalization - identification task: two tests of generalization; - discrimination task: one test of generalization.

Note. Day: Participants could complete each day's tasks over 24 hours on days they chose, within 2 weeks of beginning.

Training: participants also performed identification tasks in training stages 1 – 3. The data were not included in analyses.

The details and data are analyzed respectively in chapters 4, 5 and 6 as follows: Chapter 4 *Identification Tasks* (identification pretest, training, posttest and test of generalization tasks); Chapter 5 *Discrimination Tasks* (discrimination pretest, posttest and generalization test tasks), and Chapter 6 *Absolute Pitch Test* (musical note identification task).

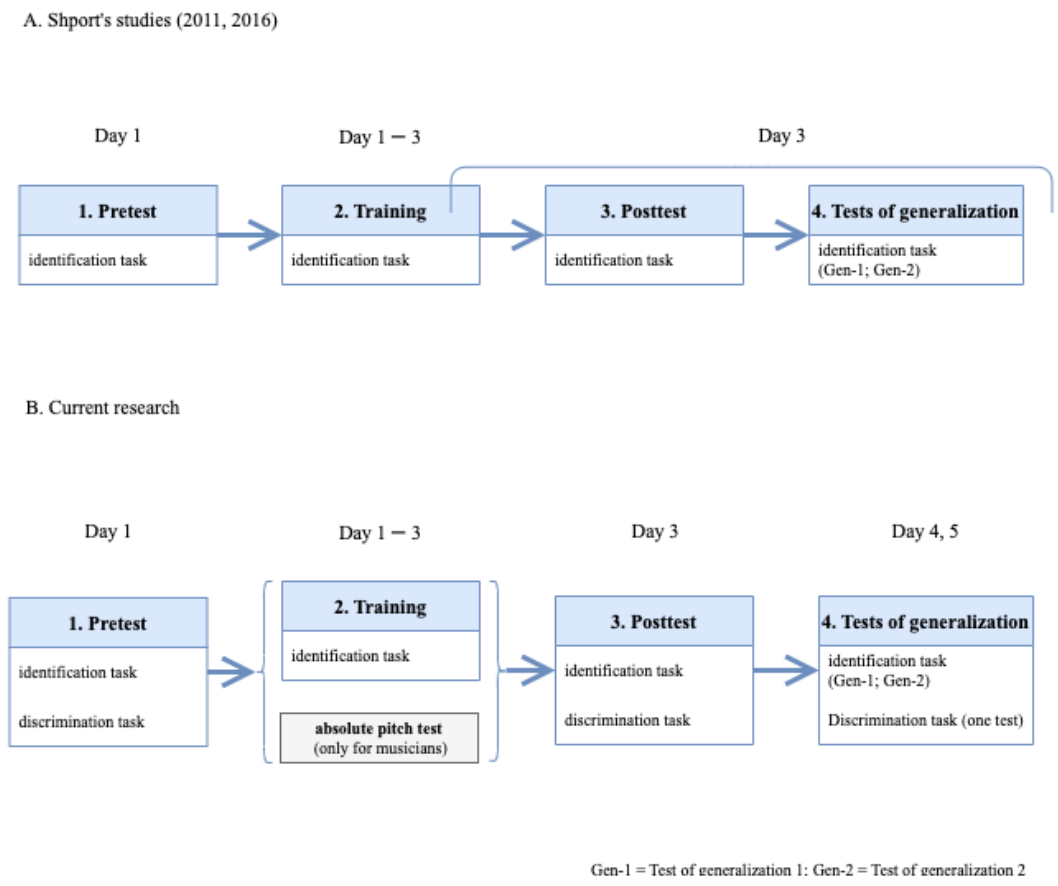
There were two reasons why the current study's experimental task results were divided into three chapters. Firstly, the two linguistic tasks (identification and discrimination) aimed to answer different research questions. Secondly, the absolute pitch test, administered only to musicians, was quite different from the linguistic tasks.

The core part of this research, presented and discussed in *Identification Tasks*, was largely based on Shport's studies (2011, 2016). The data for *Identification Tasks* were gathered in all four experimental stages: (1) pretest; (2) training; (3) posttest; and (4) tests of generalization.

A major difference between Shport's and the present study was that data were also gathered not only in *Identification Tasks*, but also in *Discrimination Tasks* and *Absolute Pitch Test* (performed at the training stage by musicians only). As the experimental phases involved more tasks than in the Shport's studies (2011, 2016), this research has a different time schedule to hers (see Figure 3.1).

Figure 3.1

Overview of the Experimental Design of Shport's Studies (2011, 2016) (Panel A) and the Current Research (Panel B)



Overall, the entire experiment took approximately four hours distributed over five experimental Days, which had to be completed within two weeks of the first Day. To lighten participants' burden, each Day was planned not to exceed one hour, which participants were required to complete. After completing a given Day, participants were instructed to wait at least eight hours before starting the next[12]. Participants were also instructed to finish one Day within 24 hours of starting it. This ensured that experimental conditions did not differ significantly between

[12] This restriction was imposed so that on one hand, participants could not perform all tasks in a single day. But it was also supposed that not all participants would be available to perform tasks every day at the same time. So, on the other hand, the interval of at least eight hours between Days was intended to facilitate participation by allowing participants to perform tasks at night and the following morning rather than at the same time every day.

participants, since all participants tested entirely online at any time they wished within the Day.

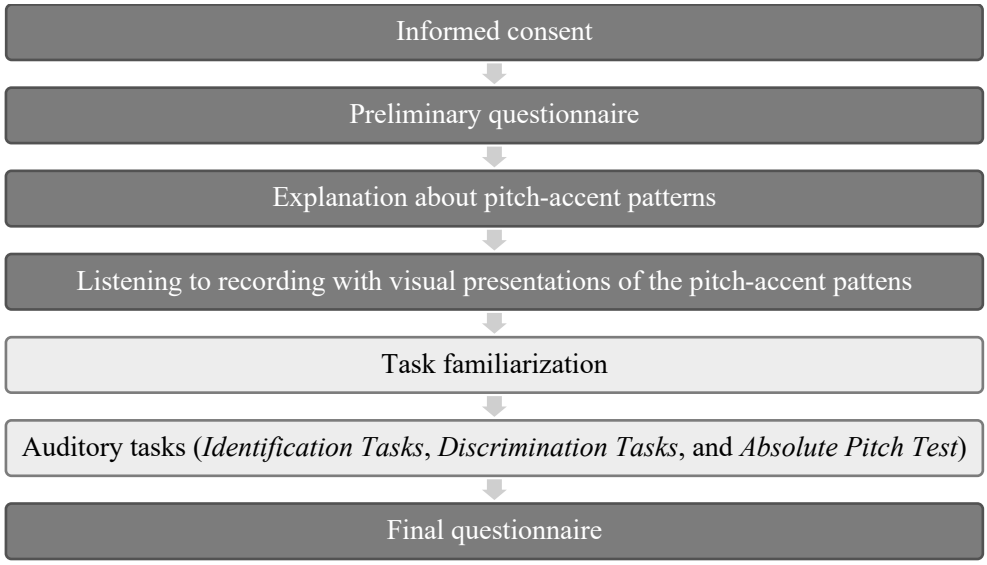
Furthermore, the order of identification tasks and discrimination tasks was counterbalanced across participants at the pretest, posttest and tests of generalization. If a participant had taken two tests of generalization based on the identification task on Day 4, he/she took a test of generalization based on the discrimination task on Day 5, and vice versa.

Lastly, it is worth mentioning another key difference between Shport's studies (2011, 2016) and this research. The perceptual training employed in this study was based on high variability phonetic training (HVPT), which was implemented by the previous studies. However, as briefly touched upon at the beginning of this section, in this study, in order to examine the effects of talker variability, the two categories of participants (musicians and non-musicians) were randomly assigned to one of the training conditions: the HV training condition and the LV training condition.

3.2. Procedure

This section describes the general procedure of the study experiment. A detailed description of the procedure for *Identification Tasks*, *Discrimination Tasks*, and *Absolute Pitch Test* are given in Chapters 4, 5 and 6. Figure 3.2 below summarizes the overall procedure of the experiment.

Figure 3.2
Summary of the Overall Procedure



Note. The two phases in light gray (Task familiarization and Auditory tasks) are detailed in Chapters 4, 5 and 6.

The study experiment was run entirely on the Gorilla software (Anwyl-Irvine et al., 2020). See Limitations section 7.2 for a discussion of the possible weaknesses of this methodological format.

With regard to privacy and the protection of participants' data, the university's data protection officer [13] provided forms to be completed with information on the research project in order to comply with privacy obligations. The completed forms, signed by myself and my supervisor, were filed with the offices of the Humanities department.

The documentation submitted included:

- A detailed breakdown of aspects regarding research data recording, sharing and storage and the undertaking to respect the university's ethical code for data treatment: *Progetto di ricerca: scheda di analisi per progetti di ricerca*; and *Dichiarazione di impegno a conformarsi*

[13] Responsabile della Protezione dei Dati dell'Università degli Studi di Pavia, phone no. +39 0382/985490, email: privacy@unipv.it

alle disposizioni del Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici e al Regolamento UE 679/2016;

- Participant informed consent to the processing of personal data for scientific research purposes: *Informativa sul trattamento dei dati personali per finalità di ricerca scientifica (ART. 13 regolamento UE 2016/679-RGDP*), see Appendix A. Note that participant consent was obtained online because the experiment was conducted remotely;
- a brief description of the research project and the processing of collected data: *Descrizione breve del progetto di ricerca e del trattamento di dati raccolti*.

The participant informed consent form (Appendix A) was provided online to each participant and acceptance was required before participants could proceed with the experimental tasks. Everyone gave their consent. A similar document, written in Japanese (for native Japanese speakers residing in Japan), was accepted by each speaker before starting the recordings (see Appendix B).

Coming back to the general procedure of this experiment, participants consented to participating in an experiment involving online auditory tasks which would require them use their own headphones or earphones. Due to the COVID-19 pandemic, all participants completed the experiment entirely online using a desktop or laptop computer[14] at home. Informed consent was obtained through the Gorilla experimental interface (Anwyl-Irvine et al., 2020). Participants could decide not to participate at any time up to the end of the final questionnaire. After completing the final questionnaire, each received a 40-euro Amazon voucher in return for their participation.

A preliminary questionnaire (see Appendix C) was administered prior to the auditory tasks. It asked participants to report age, gender, provenance, academic

[14] The use of tablets and smartphones was not allowed, because, as will be described below, participants had to press keyboard buttons to perform linguistic tasks.

background, language background and musical experience. Specifically, musical experience included the following information: (1) number of years of musical training; (2) musical instrument or vocal music practiced; (3) age of onset of musical training (only for musicians).

After completing the preliminary questionnaire, participants received an explanation about the distinctive function of pitch-accent patterns, because they had no prior experience of Japanese. In line with Shport (2011, 2016), it was explained to participants that in Japanese words have a specific pitch-accent pattern which serves to distinguish the meaning of segmentally identical words, and that these patterns have to be learned. The example of segmentally identical words offered to participants was: *hígashi* "Mr. Higashi", *higáshi* "dry Japanese sweets", and *higashi* "East" (as in Shport, 2016, p. 753).

Following the explanation, participants listened to recordings of each of these words (*higashi*) twice, looking at the corresponding visual representations of the simplified pitch-accent patterns (adopted from Shport, 2011, p. 153, 2016, p. 752). It was then explained to them that pitch-accent patterns of words were also present when words were uttered in a sentence. As examples, the following two carrier sentences were used: (1) __ *desu*. "(It) is __."; (2) *kore wa* __ *desu*. "This is __.". These two sentences were chosen to familiarize participants with the change in position of the target word in the carrier sentences, a stimulus format employed throughout *Identification Tasks* (see Table 4.2). The recordings of these carrier sentences (made by the female speaker, F6; see Section 4.2.2.2 for a more detailed description) were played twice as participants looked at visual representations of the simplified pitch-accent patterns.

The task familiarization phase followed. In this phase, which differed according to task type, participants also completed an audio check during which they were able to adjust the audio volume to comfortable listening level before the task itself started. Other details are described in the chapters on the procedures for each auditory task.

Lastly, after all the auditory tasks had been completed, a final questionnaire was administered at the end of the experiment (see Appendix D). As in Shport (2011,

2016), it asked them not only to rate, on a 4-point Likert scale, the degree of their attentiveness during the test; but also to rate, on a 5-point Likert scale, the three speakers who recorded the identification test stimuli (namely, M1, F1 and F3; see Section 4.2.2.2) and the four speakers who recorded the discrimination test stimuli (namely, M4, F4, M5 and F8; see Section 5.2.2.2), in terms of perceived difficulty of performing the task. Additionally, participants were also asked to describe any technical problems they encountered.

3.3. Main Research Question

The main research question of this dissertation, which was asked for both *Identification Tasks* and *Discrimination Tasks*, was whether musical training has any effect on the perceptual learning of Japanese pitch accent by Italian native speakers with no prior knowledge of Japanese. This was explored through two types of task (identification and discrimination) administered before and after the perceptual training.

The perceptual training employed in this study was based on high variability phonetic training (HVPT). The details of the training were described in *Identification Tasks* (Chapter 4), because the identification task was used in this training.

Other research questions, specific to each linguistic task are dealt with in Chapter 4 and 5 respectively.

CHAPTER 4 IDENTIFICATION TASKS: PERCEPTUAL TRAINING TO IDENTIFY PITCH-ACCENT PATTERNS

As the title suggests, this chapter reports on the details of the *Identification Tasks*—identification pretest, training, posttest and test of generalization—conducted in the current experiment, and on the data recorded.

4.1. Identification Tasks: Introduction

As reviewed in 2.1.4, the findings of Pappalardo's study (2018) brought out several interesting points for the present research. Firstly, the results indicated that Italian participants had difficulty in correctly perceiving pitch-accent patterns. This likely reflects several crucial differences in lexical accent between Japanese and Italian (see sections 2.1.1-2.1.3 for a detailed account of these differences). Secondly, the findings also implied that differences in perceptual ability may stem from individual differences between participants. The current research set out to extend the knowledge we have of such individual differences, focusing on musical training as one of the possible sources. Lastly, Pappalardo (2018) found that the easiest pitch-accent pattern for the participants was the unaccented pattern. This is in line with the results of Hirano-Cook (2011) and those found throughout the studies reviewed by Ayusawa (2003); but in contrast with the findings of Laméris and Graham (2020) and Wu et al. (2012, 2017) that the 1st-syllable accented pattern is the easiest, and with those of Shport (2011, 2016), who found that there was no difference between the 1st and 2nd-syllable accented patterns, but that the unaccented pattern was the most difficult. Thus, the current research also sought to examine which would be the most difficult pattern to identify.

Note that, unlike those in Pappalardo's work (2018), participants in the present study were naïve learners of Japanese. This was because participant L2-Japanese learning experience or knowledge have led to unclear findings in works investigating Japanese pitch accent perception (see Section 2.1.4). Since this dissertation's main goal was to assess the effect of musical training, it was important

to avoid confounding factors, so Japanese pitch-accent contrast was equally novel to both groups of participants.

Participants in Shport's studies (2011, 2016) were also naïve learners of Japanese (in her case, native speakers of American English). She followed the HVPT paradigm put forth by two seminal works (Lively et al., 1993; Logan et al., 1991). Its effectiveness has been demonstrated not only by these two studies but in a large body of literature. Her one-hour HVPT training also facilitated participants' perceptual learning of Japanese pitch-accent patterns, suggesting that this training paradigm is applicable to naïve learners of Japanese who were native speakers of a non-tone language.

The present research, largely based on the methodology in Shport's work (2011, 2016), explored the effect of musical training on perceptual learning of Japanese pitch accent by native Italian speakers via a comparison of expert musicians versus non-musicians.

As mentioned earlier, there is a large body of evidence showing the effectiveness of HVPT. However, mixed results have been reported by works comparing the effects of high and low talker variability input in perceptual training (e.g., Brekelmans et al., 2022; Sadakata & McQueen, 2013; J. W. S. Wong, 2012, 2014 for segmental contrasts; Silpachai, 2020; Deng et al., 2018 for lexical tone contrasts). Moreover, studies exploring the role of talker variability and the interaction between talker variability and individuals' perceptual abilities also have been shown mixed findings (Dong et al., 2019; Perrachione et al., 2011; Qin et al., 2022; Sadakata & McQueen, 2014). To develop a better understanding of these topics, the present research sought to explore them by randomly assigning musicians and non-musicians to either a HV training condition (with stimuli produced by four talkers) or a LV training condition (with stimuli produced by one talker).

With regard to the relationship between lexical pitch perception and musical experience/training, numerous works have examined the effect of musical training on the perception of Japanese pitch-accent patterns (Golob, 2003) and lexical tone (Alexander et al., 2005; Burnham et al., 2015; Chang et al., 2016; Chen et al., 2020;

Delogu et al., 2010; Gottfried, 2007; Gottfried & Xu, 2008; Götz et al., 2023; Kirkham et al., 2011; Lee et al., 2014; Lee & Hung, 2008; Marie et al., 2011; Mok & Zuo, 2012). Their results largely converge to indicate that, to varying degrees, musical training/experience had a positive effect. However, the results of experimental lexical tone perceptual training studies have provided mixed evidence for the advantage of musical expertise in lexical tone perceptual learning: some results (Cooper & Wang, 2012; Maggu et al., 2018; P. C. M. Wong & Perrachione, 2007) have indicated that musical experience/training is beneficial, whereas those of other studies (Dittinger et al., 2016; Tong & Tang, 2016; Wayland et al., 2010; Zhao & Kuhl, 2015) have indicated that the advantage of musicians over non-musicians diminished after training or that musicians showed no additional advantage. Thus, the current research set out to better understand the relationship between musical training and perceptual training on lexical pitch-accent contrast.

As regards absolute pitch, Burnham and colleagues' study (2015) indicated it had a positive effect on Thai lexical tone discrimination. In the light of Burnham et al.'s results, the present study attempted to gain a better understanding of the effect of absolute pitch.

Now, having concluded this section on the background information for *Identification Tasks*, the next section describes this study's research questions and predictions.

4.1.1. Identification Tasks: Research Questions and Predictions

The current dissertation's main aim is to provide a better understanding of the effect of musical training on perceptual learning of Japanese pitch accent by native speakers of a non-tone language (Italian), who were also naïve to Japanese. This study also examined the role of talker variability and its interaction with the effect of musical training. Moreover, it attempted to assess the effect of absolute pitch. Note that this chapter reports on these issues focusing on *Identification Tasks*; see Chapter 5 for *Discrimination Tasks*.

In the light of these aims and the background information which motivates the current dissertation discussed in Chapter 2 and Section 4.1, the present study addressed the following research questions:

RQ1: Will Italian musicians outperform Italian non-musicians in identifying Japanese pitch accent?

RQ2: After training, will the difference between musicians and non-musicians in the ability to identify Japanese pitch accent decrease or increase?

RQ3: Will the HV training condition be more beneficial for Italian musicians compared to non-musicians?

RQ4: Which pitch-accent pattern will be the most difficult to perceive for native Italian speakers?

RQ5: Will there be any difference in the ability to identify Japanese pitch accent between musicians with absolute pitch and those without absolute pitch?

Based on the literature review in Chapter 2, the following possibilities were tested in response to the research questions described above.

P1: Italian musicians would outperform Italian non-musicians in identifying Japanese pitch accent.

P2: There would be an added benefit for musicians, although a reduced or no benefit was also tested for, since evidence in the literature is mixed with regard to whether or not lexical tone perceptual training is more beneficial for musicians than non-musicians (see Section 2.3.3).

P3: Whether or not there would be an effect for talker variability and its interaction. Again, the existing literature has shown mixed findings about the role of talker variability and its interaction with individuals' perceptual abilities (see sections 2.2.3 and 2.2.3.1).

P4: Given that the current research's methodology was largely based on Shport's studies (2011, 2016), it was predicted that the unaccented pattern would be the most difficult. Again, the existing literature has shown mixed findings about

the easiest pitch-accent pattern for native speakers of non-tone languages (see sections 2.1.4, 2.2.1, and 2.2.2.1). Recall that these previous studies adopted different methodology including stimuli.

P5: It was hypothesized that musicians with absolute pitch would perform better than those without absolute pitch. However, it was also predicted that it would be very difficult to find musicians with absolute pitch.

Since Chapter 6 is dedicated to the absolute pitch test for musicians, the results for **RQ5**/**P5** are reported in Chapter 6.

Now, having outlined the experimental predictions, the next section provides details about the methodology regarding *Identification Tasks*.

4.2. Identification Tasks: Method

The methodology applied in *Identification Tasks* was mostly based on Shport (2011, 2016). Compared to these studies, however, three novel aspects were introduced. Firstly, the entire experiment, including *Identification Tasks* was carried out entirely online by means of the Gorilla software package (Anwyl-Irvine et al., 2020), rather than in a laboratory. Secondly, as described below, there were two categories of participants: musicians and non-musicians. Lastly, each category was further subdivided into two groups in the training phase: the HV (high variability) training group and the LV (low variability) training group.

4.2.1. Identification Tasks: Participants

A total of 64 adult native speakers of Italian took part. None of the participants had prior experience of Japanese or any other tone language. In addition, all of them reported having normal or corrected-to-normal vision and unimpaired hearing.

The cohort was comprised of two groups: musicians and non-musicians. Musicians were defined as individuals engaged in formal tertiary-level musical training, including those enrolled in conservatories, musical institutes, or majoring

in musicology at university, in line with Burnham et al. (2015), Delogu et al. (2010), Gottfried (2007), Lee and Hung (2008), and Lee at al. (2014). Non-musicians were defined as people with three years or less of continuous private musical training, and no music lessons at the time of recruitment (Alexander et al., 2005; P. C. M. Wong & Perrachione, 2007).

The musician group consisted of 32 students (16 male, 16 female; mean age = 24.4 years; SD = 4.9 years)[15] recruited from the Università di Pavia (majoring in Musicology), the Conservatorio di Milano and the Civica Scuola di Musica Claudio Abbado in Lombardy. The non-musician group also consisted of 32 students (16 male, 16 female; mean age = 23.1 years; SD = 2.2 years)[16] recruited from five universities in Lombardy (Università di Pavia, Università di Milano Statale, Università di Milano Bicocca, Politecnico di Milano, Università dell'Insubria). Thus, all participants were higher education students in Lombardy; they had similar ages and similar levels of education.

Each participant in the two groups was randomly assigned to one of two training conditions: HV training or LV training. Due to a slight glitch in the automated randomization process, 15 participants in each group were given the HV training condition and 17 the LV training condition.

4.2.2. Identification Tasks: Stimuli

4.2.2.1. Identification Tasks: Materials

The stimuli were Japanese carrier sentences containing target words, as in Shport (2011, 2016). Since full details are available in Shport (2011, 2016), only essential points are described in this dissertation.

Table 4.1 shows that, in total, there were 36 disyllabic target words, consisting of 12 triplets of segmentally identical words. Each word in the 12 triplets

[15] One additional musician was tested but was excluded from the sample before she had finished because she had broken the rules for the time schedule.
[16] One additional non-musician was tested but was excluded from the sample before she had finished because she had broken the rules for the time schedule.

carried one of three different pitch-accent patterns: 1st-syllable accented, 2nd-syllable accented, and unaccented. The pitch-accent patterns of the target words, already verified by Shport (2011, 2016), were further checked using the newest edition of NHK[17] accent pronunciation dictionary (NHK Hoso Bunka Kenkyujo, 2016). As described in Shport (2011, 2016), the first six triplets were used in the pretest, the training and the posttest. The second six triplets were employed in two tests of generalization (Gen-1 and Gen-2) that served to assess whether the participants could generalize to new words and a new talker.

[17] NHK, which stands for *Nippon Hoso Kyokai* "Japan Broadcasting Corporation", is a public broadcaster in Japan.

Table 4.1
Triplets of Segmentally Identical Words and Their Pitch-Accent Pattern

Experimental phase	Word triplets	Pitch-accent pattern		
		1st-syllable accented	**2nd-syllable accented**	**Unaccented**
Pretest Training Posttest	*aki*	秋 autumn	飽き weariness	空き vacancy
	hashi	箸 chopsticks	橋 bridge	端 edge
	kaku	核 core	画 stroke	格 status
	tsuru	鶴 crane	蔓 vine	釣る to fish
	umi	海 sea	膿み pus	産み giving birth
	waki	和気（藹々）harmonious	脇 side	沸き boiling
Gen-1 Gen-2	*hari*	針 needle	梁 beam	張り straining
	kaki	牡蠣 oyster	垣 fence	柿 persimmon
	maku	播く to sow	膜 membrane	巻く to wrap
	mori	守 guard	漏り leaking	森 woods
	mushi	無視 to ignore	蒸し steaming	虫 insect
	yoku	良く well	欲 greed	翌 next

Note. Adapted from Shport (2011, p. 149). Gen-1 = Test of generalization 1; Gen-2 = Test of generalization 2

As Shport (2011, 2016) pointed out, for three of the target words, the NHK dictionary provided a two-pronunciation norm in terms of lexical pitch-accent

patterns: *tsurú* or *tsúru* “vine”, *kakú* or *kaku* “stroke”, and *yóku* or *yoku* “next”. Speakers were asked to pronounce each of these three words with the pattern presented in Table 4.1, following Shport (2011, 2016). Recording details will be described below.

There were seven carrier sentences (Table 4.2). These differ in their sentential contexts, in other words, in position of the target word, sentence type, sentence length and particles. As can be seen in Table 4.2, in all sentences the target words were followed by particles that did not change their pitch-accent patterns; namely the case articles *wa*, *ga*, *o*, *ni* or present-tense copula *na* (Shport, 2016, p. 750).

Table 4.2
Variety in Sentential Context

Experimental phase		Carrier sentence	Sentence type	Word position
Pretest Training Posttest	1	__ *ga kakemasu*. "(I) can write __."	affirmative	initial
	2	*Watáshi wa __ ga kakemasu*. "I can write __."	affirmative	medial
	3	__ *ga ne*. "It's __, you know."	affirmative	initial
Only training	4	__ *ni chúuishite kudasai*. "Please pay attention to __."	imperative	initial
Gen-1 Gen-2	5	__ *o kurírkku shite.* "Click on ___."	imperative	initial
	6	__ *ga hatsuon shiyasui?* "Is ___ easy to pronounce?"	interrogative	initial
	7	*Are wa __ na no?* "Is that __?"	interrogative	medial

Note. Adapted from Shport (2011, P. 150). Gen-1 = Test of generalization 1; Gen-2 = Test of generalization 2

4.2.2.2. Identification Tasks: Speakers

The speakers, recruited in accordance with Shport (2011, 2016), were seven native speakers of Tokyo Japanese (four female[18]: F1, F3, F6, F7, and three male: M1, M2, M3; mean age = 44.7 years; age range 32 to 68 years; SD = 15.2 years). All of them were born, grew up and live in the Tokyo metropolitan area (namely, in either Tokyo, Chiba, or Saitama prefectures) and none had any speech impairment. Talker variability was thus achieved by having multiple voices, including both male and female speakers of different ages.

The seven speakers also recorded materials in accordance with Shport (2011, 2016). One female speaker (F6) recorded the materials used at the beginning of *Identification Tasks* to explain Japanese pitch-accent patterns. Her voice was also employed in the short practice provided before each task in the experiment. The other six speakers produced the stimuli described in the previous section. Specifically, as can be seen in Table 4.3, one male speaker (M1) read the pre/posttest stimuli (the first six triplets of segmentally identical words in Table 4.1 with the first three carrier sentences in Table 4.2), while four others (M2, M3, F1 and F7) uttered the training stimuli (the first six triplets in Table 4.1 with the first four carrier sentences in Table 4.2). As for the tests of generalization, one of the female speakers (F1) who recorded the stimuli for training served as a familiar talker for Test of generalization 1 (Gen-1), whereas female speaker F3 was employed as a novel talker for Test of generalization 2 (Gen-2). (They recorded the last six triplets in Table 4.1 with the last three carrier sentences in Table 4.2).

[18] Originally four female speakers (F1, F2, F3 and F6) concluded their recording. However, since F2's recordings contained too much background noise, they were substituted by F7's recordings. F4 and F5 (then substituted by F8), produced the stimuli for *Discrimination Tasks*.

Table 4.3
Speakers and Their Production of the Materials for the Experimental Phases in Identification Tasks

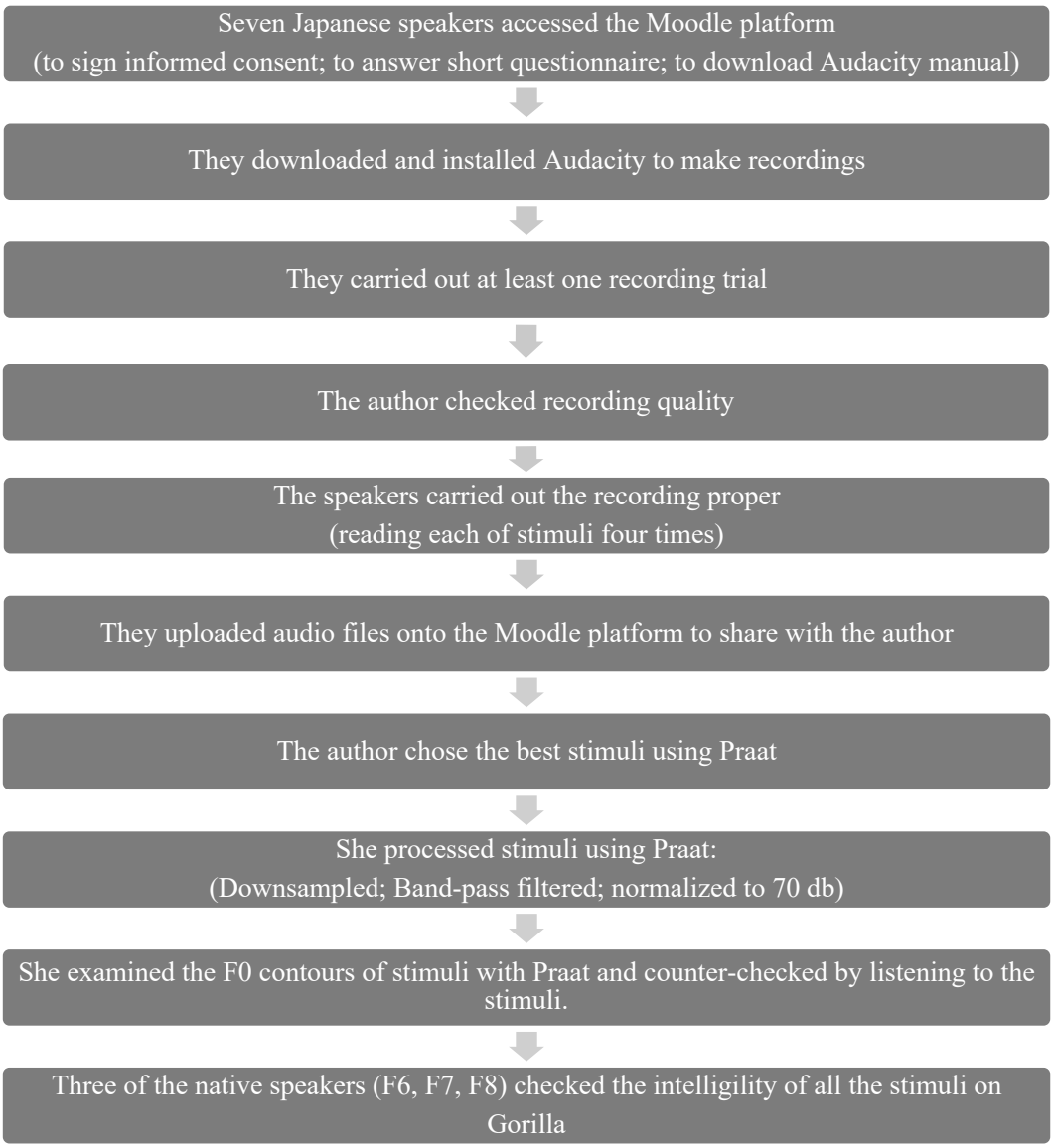
Experimental phase	Speakers	Stimuli	
		Word triplets (Table 4.1)	Carrier sentence (Table 4.2)
Explanation and task familiarization	F6	---	---
Pretest	M1	First six triplets	1 – 3
Training	M2, M3, **F1**, F7	First six triplets	1 – 4
Posttest	Same as the pretest		
Gen-1	**F1** (familiar talker: one of the speakers in the training phase)	Second six triplets	5 – 7
Gen-2	F3 (novel talker)	Second six triplets	5 – 7

Note. Gen-1 = Test of generalization 1; Gen-2 = Test of generalization 2

4.2.2.3. Identification Tasks: Recording Procedure

Due to the COVID-19 pandemic, the entire recording procedure was conducted outside the laboratory, with each of the speakers making their recording at home in Japan. Figure 4.1 presents a summary of the recording procedure.

Figure 4.1
Summary of the Recording Procedure



Firstly, the seven Japanese speakers accessed the website (https://phonetics.yukanaito.org/), built using the Moodle platform (https://moodle.org/). They filled out the informed consent (see Appendix B) and answered a short questionnaire about their age, gender and place of origin (see Appendix E). Then speakers downloaded the Audacity software manual so that they

could download and install the Audacity® recording and editing software[19] on their computer by themselves.

For the recording, the speakers were instructed to read four times at the speed of normal speech, each of the sentences with the target words[20]. In order to elicit as natural an enunciation as possible, Japanese speakers were asked to speak as if they were talking to the author (a Japanese native speaker), and they were asked to look at her photograph during their recordings.

Recordings were made with the Audacity software in PCM WAV format, at 44.1 kHz sampling rate, 32-bit sample size and mono channel. For the following reasons, speakers used a headset[21] during recordings: (1) the volume level can be kept as stable and high as possible without clipping, because the microphone remains at a uniform direction and distance from the speech sounds (Ladefoged, 2003; Podesva & Zsiga, 2014); (2) ambient noise is lower compared to using the PC's inbuilt microphone.

In order to check the difference between noise level and the volume of the spoken audio track (the signal-to-noise ratio), Japanese speakers all made at least one trial recording. The speakers were instructed on how to adjust the audio level of the microphone to keep their sound amplitude above 0.1 Pascal when they were speaking. However, some speakers' recordings continued to have a noise level of over 40 dB in the silent part of the recording: examples of background noise can be seen in Appendix F. These speakers received a Sennheiser PC 5.2 headset with a noise-cancelling microphone in return for participation (speakers without noise problems received a 3,000-yen Amazon voucher for their participation).

Once the speakers finished recording, they uploaded their audio files onto the website (https://phonetics.yukanaito.org/), to share them with the author. Then,

[19] Audacity® software is copyright © 1999-2021 Audacity Team. Web site: https://audacityteam.org/. It is free software distributed under the terms of the GNU General Public License. The name Audacity® is a registered trademark.

[20] Only F6 recorded not only the carrier sentences with the words, but also isolated words which were utilized to explain the function of Japanese pitch-accent patterns. These words will be described below.

[21] M2 used earphones and not a full headset because his recordings were made in the countryside where it was far quieter compared to the Tokyo metropolitan area, and he did not have a headset at the time of recording.

for each stimulus, the four recordings were analyzed using Praat (Boersma & Weenink, 2021). The best of the four was selected based primarily on noise level, but also on sound quality. The reason for employing one recording per stimulus and speaker, rather than multiple recordings, was to simplify the statistical analysis, since this study already had many variables. After that, as described by Brekelmans (2020) and Brekelmans et al. (2020), the files were downsampled to 22.05 kHz and band-pass filtered from 60–20,000 Hz with a smoothing factor of 10. Downsampling was used due to the fact that highest linguistically meaningful frequencies in the speech signal appear at less than 11 kHz, hence a sampling rate of 22 kHz is generally adequate (Ladefoged, 2003; Podesva & Zsiga, 2014). The band-pass filter was used to eliminate frequencies outside of those that humans can usually pick up and to avoid any spurious variation in results (G. Brekelmans, personal communication, April 20, 2021). Smoothing was applied in order to reduce the ringing effect that happens due to transition from the stop-band to the pass-band. After the use of the band-pass filter, the intensity of all recordings was normalized to 70 dB as in Brekelmans (2020) and Brekelmans et al. (2020). The resulting stimuli were then saved in PCM WAV format, at 22 kHz sampling rate, 16-bit sample size and mono channel. All processing was conducted on Praat (Boersma & Weenink, 2021).

The Gorilla software (Anwyl-Irvine et al., 2020), utilized in the present research, strongly recommends using the mp3 format for best browser compatibility. However, in order to reduce possible perceptible audio distortions, the author avoided using compressed audio formats; and since Brekelmans (2020) and Brekelmans et al. (2020), who conducted a HVPT experiment using the Gorilla software (Anwyl-Irvine et al., 2020), managed to use stimuli encoded in uncompressed WAV file format, the current study followed their treatment of recordings. The processed files were verified for browser compatibility by incorporating them into a Gorilla experiment that was then tested with updated versions of the most popular browsers (Google Chrome, Microsoft Edge, Apple Safari, and Mozilla Firefox).

Examination of the F0 contours in the audio files as visualized with Praat (Boersma & Weenink, 2021) corroborated that all speakers had produced the three

different pitch-accent pattern contrasts in each word triplet. In addition to this, all stimuli, including the materials used to explain Japanese pitch-accent patterns and those employed in the short practice provided before each task in the experiment, were checked by the author (a Japanese native speaker), who listened to them to ensure that she perceived correctly which pitch pattern was being modeled in recordings. Finally, all of the audio files were checked again for intelligibility by three of the native speakers (F6, F7, F8) who had participated in making the recordings (see Silpachai, 2020). The audio quality-checking procedure was carried out on the Gorilla software (Anwyl-Irvine et al., 2020). For the sake of time and to spread the load, the files were divided into three lots. Naturally, the speakers' own recordings were excluded from the ones they were responsible for checking. All items utilized in the experiment were identified correctly during this procedure.

4.2.3. Identification Tasks: Design

In accordance with Shport (2011, 2016), *Identification Tasks* consisted of four stages: (1) pretest; (2) training; (3) posttest; and (4) tests of generalization. Each of these stages is described in more detail below. The same task format was used throughout: a three-alternative forced-choice (3AFC) identification of pitch-accent patterns.

(See Section 3.1 for counterbalancing of identification and discrimination tasks in the overall design).

4.2.4. Identification Tasks: Procedure

Before beginning of each identification task, participants completed the task familiarization phase. They listened twice to the triplet of segmentally identical words *higashi* ("Mr. Higashi", "dry Japanese sweets", and "East") embedded in the carrier sentence *kore wa __ desu.* "This is __.", while looking at the corresponding pitch-accent patterns. Then they carried out a short practice comprised of three trials in random order, with the three-alternative forced-choice (3AFC) identification task and response keys (I. A. Shport, personal communication, April 22, 2021). The

same triplet of segmentally identical words and carrier sentence were used both in the explanation (see Section 3.2) and in the short practice. These materials, recorded by F6, were not used in any of the tests or the training. During the short practice, participants received trial-by-trial feedback on their responses. This involved seeing the number of the correct pitch-accent pattern (1, 2 or 3), together with the corresponding visual representation, in order for participants to better accustom themselves to identifying pitch-accent patterns with the corresponding number and visual representation (in accordance with a personal communication from M. Nakayama October 23, 2021). At the end of the short practice, cumulative feedback was also provided (the number of correct trials/three trials, the number of all trials).

The procedure for *Identification Tasks* was adapted from Shport's studies (2011, 2016) and comprised the same four stages: (1) pretest; (2) training; (3) posttest; and (4) tests of generalization. These are detailed below.

4.2.4.1. Identification Tasks: Pretest

The pretest followed Shport's (2011, 2016) procedure, except that it was done online.

The pretest was conducted on Day 1. It consisted of 162 trials, organized into three blocks, with one carrier sentence repeated for each of the three pitch accent patterns (see Table 4.2). The terms used and the structure are summarized in Table 4.4 below. The three blocks were ordered from easiest to most challenging in terms of sentence length and proximity from pitch accents to boundary tones, and there was an optional break between blocks. Within each block, 18 stimuli, that is 18 target words (6 triplets of segmentally identical words x 3 pitch-accent patterns) embedded in a carrier sentence, were repeated three times, resulting in 54 trials per block. In addition, stimulus presentation order was randomized, and participants could not take breaks within the block. All stimuli were recorded by M1 (see Table 4.3).

Table 4.4
Summary of Pretest Terms and Structure

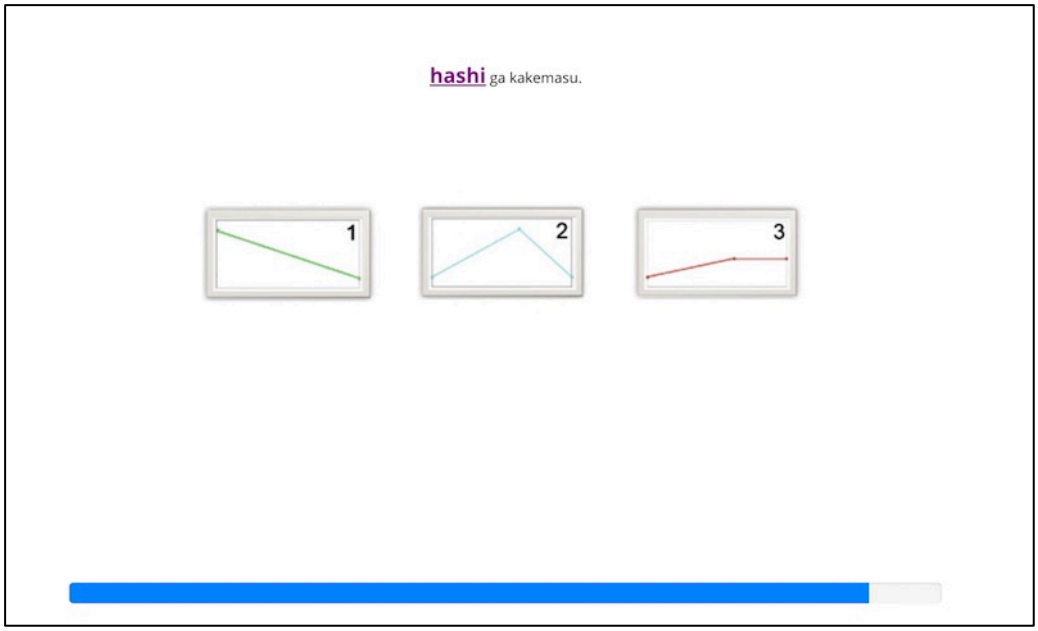
Stimuli type per block
18 stimuli = 18 target words (6 triplets of segmentally identical words x 3 pitch-accent patterns) embedded in a carrier sentence
Blocks
Participants could take an optional break between but not within blocks. Block 1 (carrier sentence 1): 18 stimuli x 3 repetitions = 54 trials; Block 2 (carrier sentence 2): 18 stimuli x 3 repetitions = 54 trials; Block 3 (carrier sentence 3): 18 stimuli x 3 repetitions = 54 trials.
Number of trials
162 trials = 54 trials x 3 blocks

Note. For each trial, the sequence of steps was in line with Shport (2011, 2016).

For each trial, the sequence of steps was in line with Shport (2011, 2016). Firstly, an underlined target word was displayed on the computer screen to attract participants' attention to the target word. The target word was then shown in a carrier sentence along with three diagrams of simplified pitch-accent patterns adopted from Shport's studies (2011, p. 153, 2016, p. 752) (Figure 4.2). As illustrated in Figure 4.1, since participants were naïve to Japanese, the sentence including the target word displayed on the screen was transcribed in Latin alphabet. After one second, a recording of the sentence was played. Participants were instructed to choose which pitch-accent pattern was correct and to answer by pressing the number 1, 2 or 3 key on their keyboard (corresponding to the numbering of the three diagrams presented on the screen, see Figure 4.2). Response time was automatically limited to five seconds. One second after they had answered, the next target word was displayed on the screen. As in Shport (2011, 2016) and

Silpachai (2020), participants were instructed to answer as quickly and accurately as they could; and if unsure, they were encouraged to make their best guess.

Figure 4.2
Presentation of Response Choices in the 3AFC Identification Task



Unlike what had happened during the short practice, no feedback was provided at any time during the pretest (as in Shport, 2011, 2016). The pretest lasted about 20 minutes, and participants were able to track their progress by looking at the progress bar (see Figure 4.2).

4.2.4.2. Identification Tasks: Training

The training took place on Day 1, 2 and 3 (as described by Shport, 2011, 2016; see Table 3.1). It took approximately one hour over the course of three days (Day 1: 20 minutes; Day 2: 20 minutes; Day 3: 10 minutes). As mentioned earlier, each Day's training and testing were conducted after a brief review of the three pitch-accent patterns (listening twice to the triplet of segmentally identical words *higashi* in a

carrier sentence while looking at the corresponding pitch-accent patterns) and a short practice.

The two categories of participants (musicians and non-musicians) were randomly assigned to one of two training conditions: HV training and LV training.

The HV group trained with stimuli recorded by four talkers (namely, F1, F7, M1 and M2). Stimuli used in the training phase were those presented at the pretest, but an additional new carrier sentence was introduced into the training on Days 2 and 3 (carrier sentence 4 in Table 4.2).

The HV training condition was practically identical to the training condition described by Sport (2011, 2016). Day 1 training (Training 1) and Day 2 training (Training 2) had 144 trials each, divided into blocks by carrier sentence and into sub-blocks by talker. Specifically, as shown in Table 4.5 and Table 4.6, 144 trials each were organized into two blocks, with one carrier sentence repeated for each of the three pitch accent patterns. Additionally, each Day's blocks occurred in a fixed order and there was an optional break between blocks. Participants could not take a break within a block. In addition, within blocks, the stimuli were spoken just once by each talker, resulting in 72 trials (6 triplets of segmentally identical words x 3 pitch-accent patterns x 4 talkers); female and male talkers were alternated, and the order of talkers was fixed (see Table 4.6). The reason for having only one talker per sub-block was to ensure positive learning outcomes for different types of learners by reducing the amount of trial-by-trial talker variability (Perrachione et al., 2011). Within each of the sub-blocks, the stimulus presentation order was randomized. Day 3 training (Training 3), consisting of 72 trials, served as a review of what participants had learned in Training 1 and Training 2, resulting in four blocks (see Table 4.5 and Table 4.6). There was an optional break between blocks, which were presented in a fixed order. Within each block, one carrier sentence was used, resulting in 18 trials per block (6 triplets of segmentally identical words x 3 pitch-accent patterns); and the trials were randomly drawn from the pool of stimuli recorded by four talkers. In addition, participants could not take a break.

Table 4.5
Summary of HV Training Structure

Stimuli type per block
18 stimuli = 18 target words (6 triplets of segmentally identical words x 3 pitch-accent patterns) embedded in a carrier sentence.
Blocks
Participants could take an optional break between but not within blocks. Day 1 and Day 2 (two blocks on each Day): Block 1 (carrier sentence 1): 18 stimuli x 4 talkers = 72 trials; Block 2 (carrier sentence 2): 18 stimuli x 4 talkers = 72 trials. *Fixed order of talkers, random order of stimuli, all stimuli used* (see Table 4.6). Day 3 (four blocks): Block 1 (carrier sentence 1 of Day 2): 18 trials; Block 2 (carrier sentence 1 of Day 1): 18 trials; Block 3 (carrier sentence 2 of Day 2): 18 trials; Block 4 (carrier sentence 2 of Day 1): 18 trials. *Random order of talkers, random order of stimuli, random selection of stimuli* (see Table 4.6).
Numbers of trials
Day 1 and Day 2: 144 trials = 72 trials x 2 blocks; Day 3: 72 trials = 18 trials x 4 blocks.

Note. For each trial, the sequence of steps was in line with Shport (2011, 2016).

Table 4.6
Overview of Block Structure Under the HV Training Condition

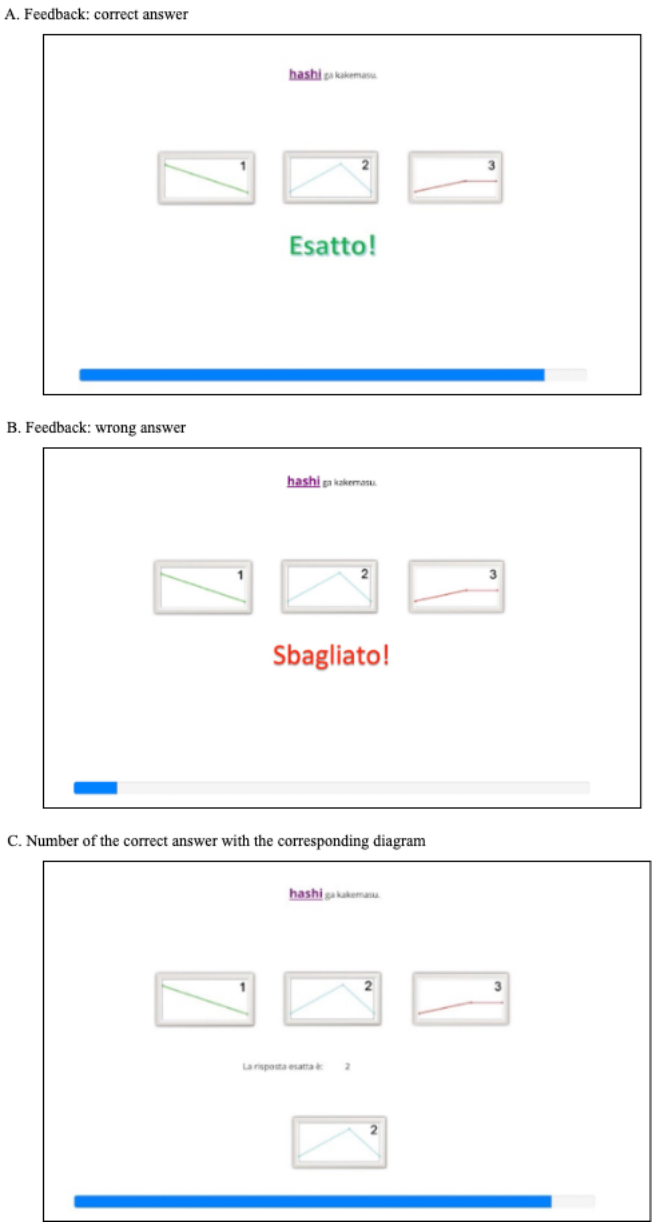
Training	Blocks and carrier sentences	Number of trials	Talkers in block
Training 1	1. __ *ga kakemas.* 2. *Watashi wa __ ga kakemas*	144	F7, M2, F1, M3 M2, F7, M3, F1
Training 2	1. __ *ga ne.* 2. __ *ni chuui shite kudasai.*	144	F1, M3, F7, M2 M3, F1, M2, F7
Training 3	1. __ *ga ne.* 2. __ *ga kakemas.* 3. __ *ni chuui shite kudasai.* 4. *Watashi wa __ ga kakemas.*	72	random order of the four talkers

The LV training group, on the other hand, trained with stimuli produced by a single talker (namely F1). In order to match tasks across training conditions, the LV training condition had the same structure as the HV training condition, but participants heard the same talker throughout all training blocks.

In both conditions, the training task was identical to the pretest task: the 3AFC identification task. However, unlike in the pretest, participants received immediate feedback on the correctness of their responses during training, as illustrated in Figure 4.3. If the correct response was chosen, the feedback in Italian was displayed: *Esatto!* "Correct!". If the wrong answer was selected, the feedback in Italian was displayed: *Sbagliato!* "Wrong!". In both cases, this feedback was followed shortly by *La risposta esatta è _.* "The right answer is _." along with a diagram of the correct answer. After a one-second delay, listeners heard the audio file of the trial one more time, irrespective of their answer. Participants were able to track their progress by looking at the progress bar (see Figure 4.2 and Figure 4.3).

Figure 4.3

Feedback in Case of Correct Answer (Panel A), That in Case of Wrong Answer (Panel B) and the Feedback Given After Both Correct and Incorrect Answers (Panel C)



4.2.4.3. Identification Tasks: Posttest and Tests of Generalization

The posttest was conducted on Day 3, after Training 3. It was identical to the pretest, except that the stimulus presentation order was randomized within each of the blocks.

Following the posttest, participants completed two tests of generalization (Gen-1 and Gen-2), on either Day 4 or Day 5, depending on the order of the identification task and the discrimination task, which was counterbalanced across participants (see Section 3.1). The order of Gen-1 and Gen-2 was also counterbalanced as described in Shport (2011, 2016).

Two tests of generalization were carried out to assess whether participants could generalize to novel stimuli. As proposed by Shport (2011, 2016), structure and procedure were identical to those of the pretest and the posttest, except that new stimuli (new target words embedded in new carrier sentences) produced by different talkers were employed (see Table 4.1–4.3).

The difference between Gen-1 and Gen-2 is that different talkers produced stimuli. The stimuli used at Gen-1 were recorded by F1, who was present in both training conditions. The stimuli employed at Gen-2, on the other hand, were produced by F3, whose voice was an unfamiliar to all participants. Thus, while Gen-1 was carried out to assess participants' ability to generalize novel stimuli produced by a familiar talker, Gen-2 was to assess participants' ability to generalize novel stimuli recorded by a novel talker.

Like what occurred during the pretest and the posttest, no feedback was provided at any time during either test of generalization (as in Shport, 2011, 2016). Both tests of generalization lasted about 20 minutes, and participants were able to track their progress by looking at the progress bar (see Figure 4.2 and Figure 4.3).

4.2.5. Identification Tasks: Analysis

A total of 41,472 trials[22] was performed by 64 participants (32 musicians and 32 non-musicians) in all tests (pretest, posttest, Gen-1, and Gen-2), i.e., 162 trials x 4 tests x 64 participants.

[22] One non-musicians performed five trials (trial number 82-86) twice in the pretest, possibly due to technical problems. Only the first attempt of these trials was included in the data analysis.

The Gorilla software package (Anwyl-Irvine et al., 2020) logged each participant's responses and reaction times (henceforth: RTs) for each trial. RTs were measured from the onset of stimulus presentation to the time when a response was given (within the 5-second response time limit) (see Section 4.2.4.1).

Participants' responses were coded as 0 for correct or 1 for incorrect. Timeouts were treated as incorrect. These participants' binary accuracy scores for each trial were used as the dependent variable, while the RT data were used only to screen for idiosyncratic data. RT data were not used as a separate dependent variable, because of the variation in both the length of the carrier sentences and the location (initial or medial position) of the disyllabic target word in the stimuli. This manipulation of target word position and carrier sentences was implemented to achieve phonetic variability in the stimuli, one of the features of the HVPT paradigm.

Recall that Japanese pitch accent consists of a bitonal high-low accent implemented as a fundamental frequency (F0) peak near the end of the accented mora followed by a sharp F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006) and that, in the unaccented pattern, there is no steep F0 fall. Therefore, in the case of target words with the 1st-syllable accented pattern, the correct pattern can be identified by listening only to the word itself, whereas in the case of target words with the 2nd-syllable accented pattern and the unaccented pattern, it is necessary to listen not only to the target words but also to the particle that follows them (see Section 2.1.1).

Thus, in order to decide the cutoff value for the minimal non-idiosyncratic response latency, the author used Praat (Boersma & Weenink, 2021) to measure some target word lengths where the target words were the first word in the stimulus (produced by three talkers). This revealed that the shortest target word length was approximately 210 ms.

As a result, responses with RTs shorter than 200 ms were removed as improbably fast reactions.

Data analyses were performed using R 4.3.2 (R Core Team, 2023). In order to address the research questions, mixed-effects binomial logistic regression models were computed using the *lme4* package (Bates et al., 2015). Models were optimized with the *bobyqa* algorithm where applicable. Model diagnosis (observation of residual qq-plots) was conducted using the *DHARMa* package (Hartig, 2022).

4.3. Identification Tasks: Results

This section falls into two parts. In response to RQ1-3, the first part reports on the overall effect of musical training; and then on that of talker variability in the stimuli used in training. The second part, in response to RQ4, focuses on participants' improvements in the three target pitch-accent patterns.

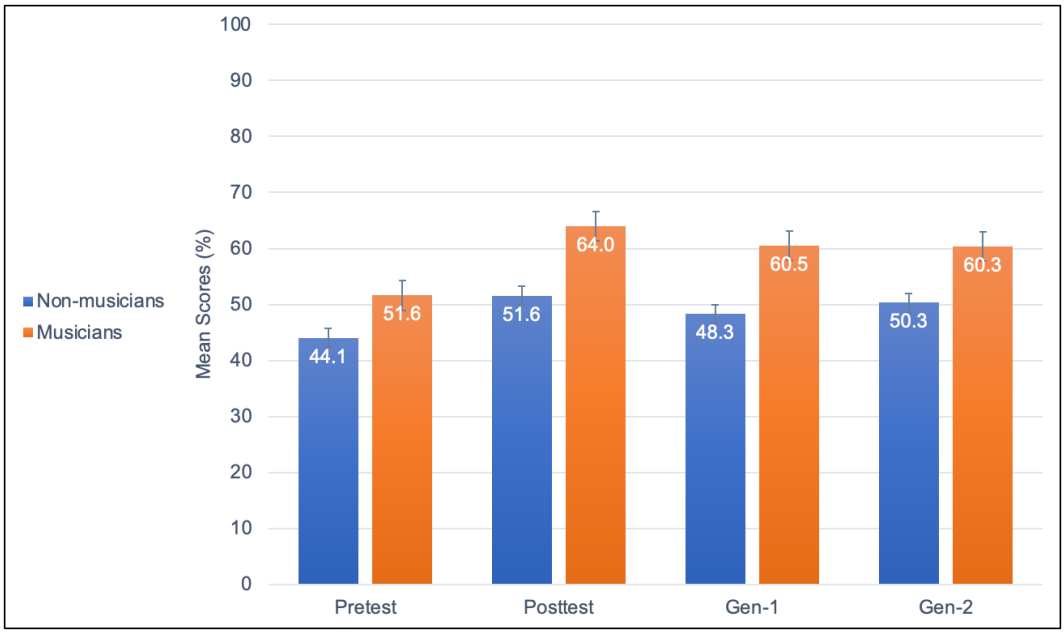
4.3.1. Identification Tasks: Effects on Overall Accuracy of Musical Training and of Talker Variability

This section first presents an overview of participants' performance in the four tests: pretest, posttest, and the two tests of generalization (Gen-1 and Gen-2). Then, it moves on to a mixed-effects model analysis to examine how the current research's predictors of interest (the effect of musical training and the role of talker variability in training) influenced participants' performance.

Figure 4.4 displays the mean scores in percentages for the four identification tests for musicians and non-musicians. Musicians are shown in orange, and non-musicians are shown in blue.

Figure 4.4

Mean Scores (%) for the Four Identification Tests: Musicians vs. Non-Musicians



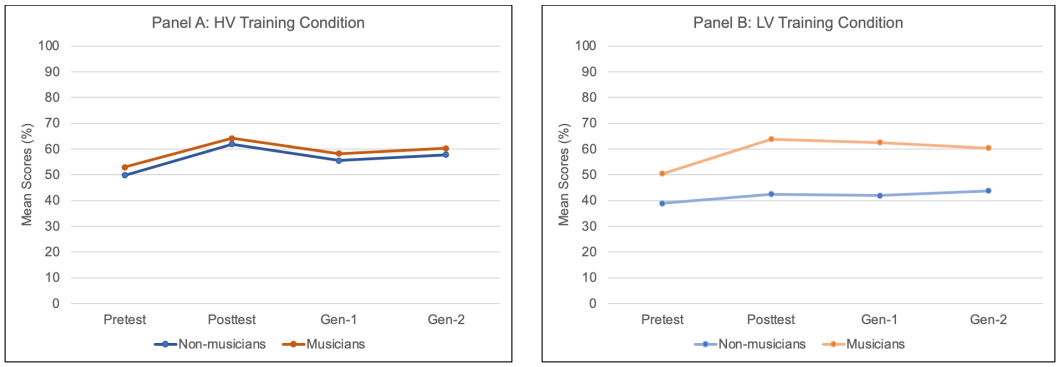
Note. Gen-1 = Test of generalization 1; Gen-2 = Test of generalization 2

Looking at Figure 4.4, musicians clearly outperformed non-musicians in all four tests. It can also be seen that, for both musicians and non-musicians, the scores for the posttest and the two tests of generalization (Gen-1 and Gen-2) are higher than the pretest scores. This indicates a positive effect of the training on both groups. However, closer inspection of Figure 4.4 reveals that musicians showed greater improvement from the pretest to the posttest and to Gen-1 and Gen-2. In fact, the difference between two groups increased noticeably after the pretest (at the pretest: 7.5%; at the posttest: 12.4%; at Gen-1: 12.2%; at Gen-2: 10%).

Panel A and Panel B in Figure 4.5 represent the mean score progress of musicians vs. non-musicians in the two training conditions (the HV training condition and the LV training condition). Again, musicians are shown in orange, and non-musicians are shown in blue; to distinguish between the two training conditions, in Panel A (the HV training condition) dark colors are used, while in Panel B (the LV training condition) light colors are used.

Figure 4.5

Mean Score Progress of Musicians vs. Non-Musicians Under the Two Training Conditions: HV (Panel A) and LV (Panel B)



Note. Gen-1 = Test of generalization 1; Gen-2 = Test of generalization 2

The most striking result to emerge from Figure 4.5 is that non-musicians and musicians achieved similar improvements in the HV training condition (Panel A), whereas in the LV training condition there was a stark difference between non-musicians and musicians, with musicians improving much more (Panel B).

Comparing the two panels in Figure 4.5, musicians improved in a similar manner in the two training conditions. By contrast, from the outset, non-musicians in the LV training condition attained lower scores than non-musicians in the HV training condition: the pretest results were 39% and 49.8%, respectively. This surprising outcome, suggesting individual differences between the two non-musician groups, is discussed in Section 7.2. It can also be seen that non-musicians under the HV training condition showed a greater pretest-posttest improvement in scores (from 49.8 to 61.9%) than those under the LV training condition, whose scores improved only slightly (from 39.0 to 42.5%). These results indicate interaction of the effects of musical training and perceptual training condition (talker variability in training stimuli).

To investigate the effects of musical training, and of LV or HV training condition on perceptual learning of Japanese pitch accent, a mixed-effects binomial logistic regression analysis was conducted. As mentioned in Section 4.2.5, participants' binary accuracy scores for each trial were used as the dependent

variable. The model contained fixed effects for: *category* (non-musician as the reference level, musician: treatment-coded); *training condition* (HV as the reference level, LV: treatment-coded) and *test* (pretest as the reference level, posttest, Gen-1, Gen-2: treatment coded). Participants and item (stimuli) were included as random effects with by-participant and by-item varying intercepts for the effect under investigation. Attempts were made to include by-participant and by-item varying slopes, but this led to convergence issues.

Since Figure 4.5 suggests interaction between category and training condition, the final model was coded in R as: glmer(Correct ~ Category + Test + Training_condition + Category : Test + Test : Training_condition + Category: Training_condition + Category : Test : Training_condition + (1|Participant_Public_ID) + (1|Item), data = data, family = 'binomial', control = glmerControl(optimizer = 'bobyqa')).

The *afex* package (Singmann et al., 2023) was used to perform likelihood ratio tests for all fixed effects. The results are summarized in Table 4.7.

Table 4.7
Likelihood Ratio Tests for All Fixed Effects: Summary of Results

Fixed effect	Result
Category	$\chi^2(1) = 7.15, p = 0.007$*
Test	$\chi^2(3) = 246.79, p < 0.001$*
Training condition	$\chi^2(1) = 2.95, p = .086$
Category × Test	$\chi^2(3) = 17.76, p < 0.001$*
Test × Training condition	$\chi^2(3) = 17.29, p < 0.001$*
Category × Training condition	$\chi^2(1) = 3.93, p = 0.047$*
Category × Test × Training condition	$\chi^2(3) = 20.57, p < 0.001$*

Note. * = significant effect

As can be seen in Table 4.7, there were significant effects of category and of test on participants' accuracy scores. In addition, there were significant interactions between: (1) category and test; (2) test and training condition; (3) category and training condition; and (4) category, test, and training condition.

In order to assess the interactions in more detail, Bonferroni-corrected multiple comparisons were calculated using the *emmeans* package (Lenth, 2023). These results are presented in Table 4.8.

Table 4.8
Interactions Between Training Condition, Musician/Non-Musician Category, and Identification Test: Results of Multiple Comparisons

Contrast	Estimate	SE	z.ratio	p
Contrasts for Test:				
HV – non-musicians:				
Pretest – Posttest	-0.56	0.06	-9.02	<0.001
– Gen-1	-0.26	0.09	-2.92	0.020
– Gen-2	-0.37	0.09	-4.06	<0.001
Posttest – Gen-1	0.30	0.09	3.25	0.007
HV – musicians:				
Pretest – Posttest	-0.54	0.06	-8.52	<0.001
– Gen-1	-0.25	0.09	-2.79	0.032
– Gen-2	-0.35	0.09	-3.85	<0.001
Posttest – Gen-1	0.28	0.09	3.08	0.012

LV – non-musicians:				
Pretest – Posttest	-0.16	0.06	-2.74	0.036
LV – musicians:				
Pretest – Posttest	-0.62	0.06	-10.62	<0.001
– Gen-1	-0.57	0.09	-6.38	<0.001
– Gen-2	-0.46	0.09	-5.21	<0.001
Contrasts for Category				
LV – musicians vs. LV – non-musicians				
Pretest	-0.53	0.24	-2.23	0.026
Posttest	-1.00	0.24	-4.19	<0.001
Gen-1	-0.96	0.24	-4.04	<0.001
Gen-2	-0.76	0.24	-3.21	0.001
Contrasts for Training Condition				
HV – non-musicians vs. LV – non-musicians				
Posttest	0.87	0.24	3.58	<0.001
Gen-1	0.60	0.24	2.46	0.014
Gen-2	0.61	0.24	2.51	0.012

Note. For brevity, only significant comparisons are presented.

As can be seen in Table 4.8, multiple comparisons revealed that the ability to identify pitch-accent patterns improved significantly from the pretest to the posttest for both non-musicians and musicians, in both training conditions, indicating the positive effect of training, irrespective of the training conditions. Non-musicians in the HV training condition and musicians in both training conditions also attained significant improvements from the pretest to both Gen-1 and Gen-2. This suggests that participants were able to generalize what they had learned during the training, since novel stimuli were used in these two tests of generalization. Although Panel A of Figure 4.5 shows a significant decrease between the posttest and Gen-1 in the performance of non-musicians and musicians in the HV training condition (the two groups' scores decreased from 64.2 to 58.3%; from 61.9 to 55.6%, respectively), the subsequent Gen-1 and Gen-2 scores remain steady at a significantly higher level than pretest scores.

Interestingly, only non-musicians in the LV training condition did not show an improvement from the pretest to either Gen-1 or Gen-2. Thus, it seems that non-musicians in the LV training condition were unable to generalize their learning to novel stimuli, even in Gen-1 (in which stimuli were produced by a familiar talker).

Multiple comparisons (see the lower half of Table 4.8) confirmed what can be seen in Panel B of Figure 4.5: in the LV training condition, musicians outperformed non-musicians in all tests. By contrast, in the HV training condition, no significant differences were found between musicians and non-musicians. This seems to imply that the HV training condition is so effective for non-musicians that they were able to improve in a similar manner to musicians.

Multiple comparisons for training condition in Table 4.8 revealed that non-musicians' accuracy results at the posttest, Gen-1 and Gen-2 were significantly better for HV training than for LV training. These findings suggests that the HV training condition was more beneficial than the LV training condition. By contrast, the difference at pretest between non-musicians in the LV and HV training conditions (see Figure 4.5) was not found to be significant. In the case of musicians, no significant differences were found between the two training conditions.

Taken together, these results indicate the effectiveness of perceptual training regardless of the training conditions. They also showed the positive effect of musical training. However statistical analyses revealed the interactions between perceptual training and musical training. It appears that non-musicians got more benefits from the HV training conditions. On the other hand, for musicians, both training conditions appeared to be effective.

4.3.2. Identification Tasks: Improvements in Pitch-Accent Pattern Accuracy

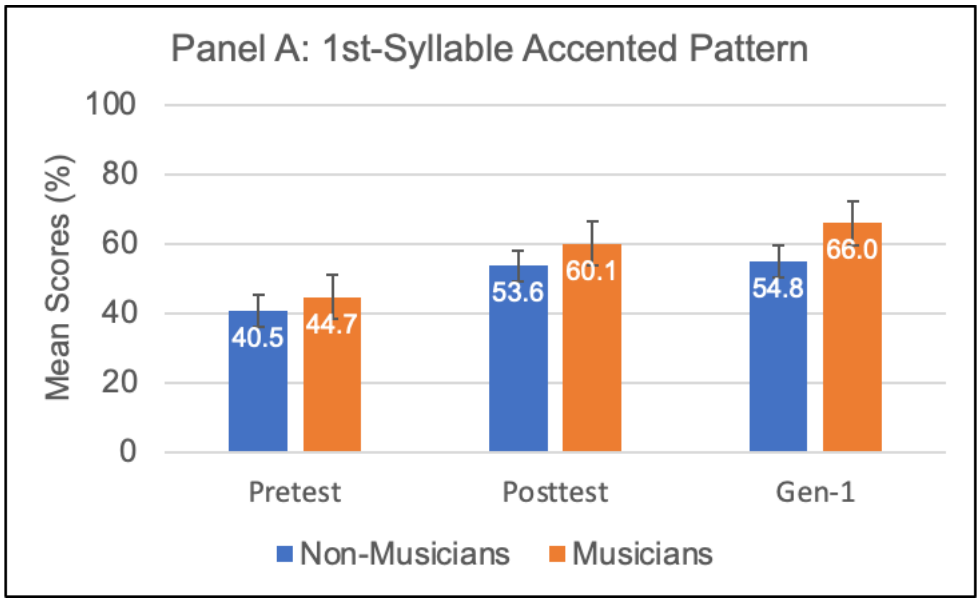
This section presents results according to pitch-accent pattern for pre- and post-training identification accuracy. Identification task results are firstly presented overall for the two categories of participants (musicians versus non-musicians); then for participant category and training condition (HV versus LV). Finally, the section presents analyses of mixed-effects models to examine whether the current research's predictors of interest (pitch-accent patterns, the effect of musical training, and the role of talker variability in training) influenced participants' accuracy.

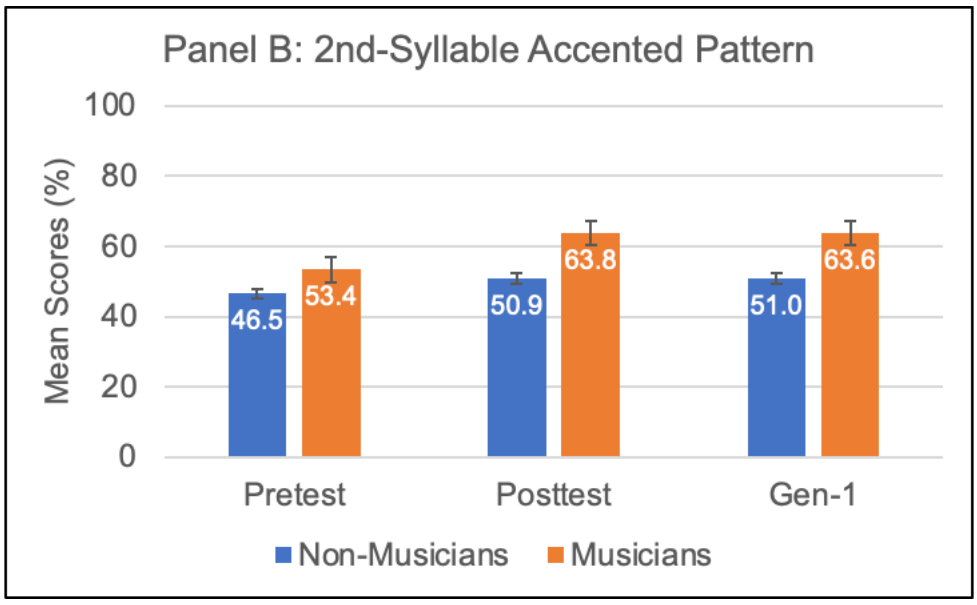
Given that similar improvements were found for all three pitch-accent patterns from the pretest to Gen-1 and Gen-2, here only pretest-posttest data and pretest-Gen-1 data are analyzed (following Shport's 2016 reasoning), to assess whether non-musicians and musicians improved their accuracy for each of the three pitch-accent patterns. This is for brevity but also because the same stimuli were used in Gen-1 and Gen-2.

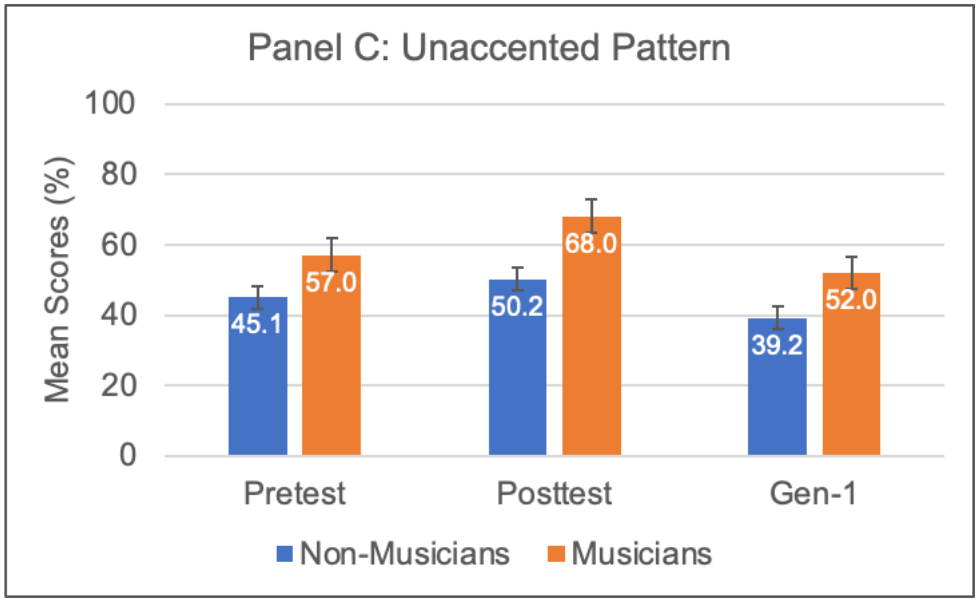
The three panels in Figure 4.6 display mean scores for pitch accent pattern identification (1st-syllable accented, 2nd-syllable accented, unaccented) in the three tests (pretest, posttest and Gen-1). The bars show Musicians (orange), and non-musicians (blue), irrespective of training condition.

Figure 4.6

Pitch-Accent Pattern, Musicians vs. Non-Musicians: Mean Scores (%) for the Three Identification Tests (Panel A: 1st-Syllable Accented Pattern; Panel B: 2nd-Syllable Accented Pattern; Panel C: Unaccented Pattern)







Note. Gen-1 = Test of generalization 1

The common aspect of the three panels in Figure 4.6 is that musicians outperformed non-musicians in all tests and for all pitch-accent patterns. In the pretest, scores for the 1st-syllable accented pattern were lower than those for the other two patterns for both non-musicians and musicians. But the figure also shows that for both musicians and non-musicians, the posttest scores were higher than the pretest scores for all pitch-accent patterns, indicating a positive effect of the training on both non-musicians and musicians.

However, closer inspection of Figure 4.6 reveals that musicians showed greater improvements than non-musicians from the pretest to the posttest and to Gen-1. In fact, for all pitch-accent patterns, the difference between non-musicians and musicians increased noticeably after the pretest (see Table 4.9).

Table 4.9
Differences (%) Between Non-Musicians and Musicians at Each of the Three Tests

	1st-syllable accented pattern	2nd-syllable accented pattern	Unaccented pattern
Pretest	4.1%	6.8%	11.9%
Posttest	6.5%	12.9%	17.8%
Gen-1	11.2%	12.7%	12.8%

Note. Gen-1 = Test of generalization 1. Percent differences are between mean scores.

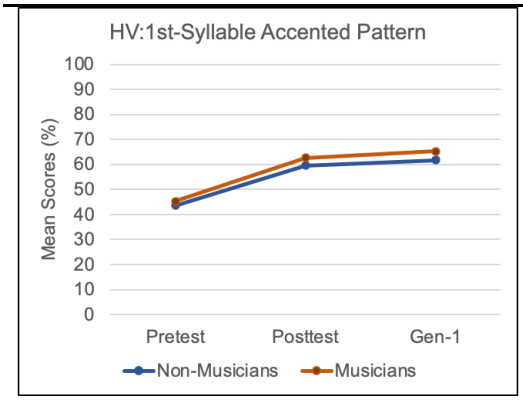
As can be observed in Table 4.9, the greatest differences between musicians and non-musicians were found for the unaccented pattern. At the same, Figure 4.6 shows that, for the unaccented pattern, both non-musicians and musicians had slightly lower scores at Gen-1 than at the pretest. By contrast, mean scores at Gen-1 for the other two accented patterns were almost the same as those at the posttest. Thus, it seems that it was difficult for participants to correctly perceive the unaccented pattern presented in novel stimuli.

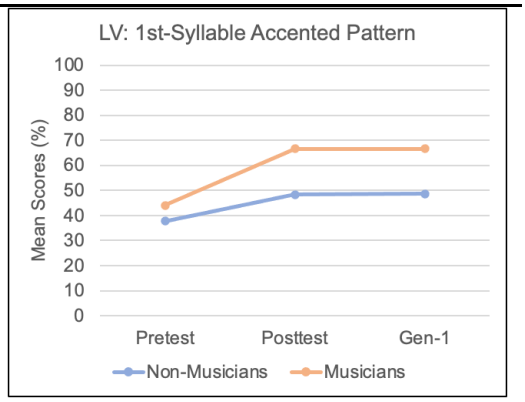
Figure 4.7 displays the progress in mean identification scores of musicians vs. non-musicians in the two training conditions (the HV training condition: Panel A; and the LV training condition: Panel B) for each pitch-accent pattern. Again, musicians are shown in orange, and non-musicians are shown in blue; to distinguish between the two training conditions, dark colors are used in Panel A, while light colors are used in Panel B.

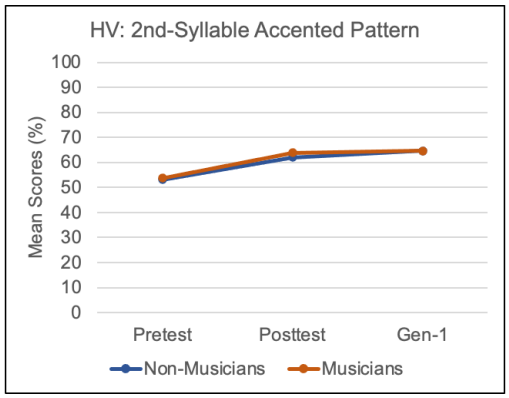
Figure 4.7

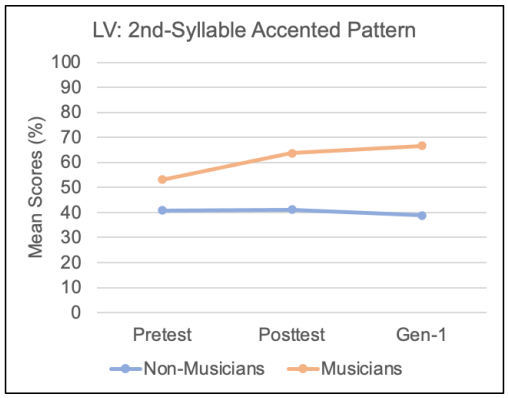
Identification Test Scores per Training Condition (HV vs. LV) for Each Pitch-Accent Pattern: Mean Scores (%) for Musicians vs. Non-Musicians

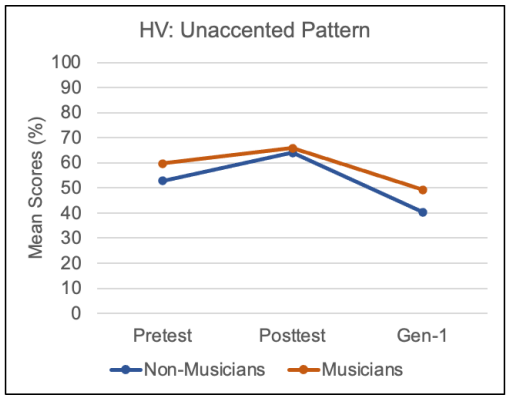
Panel A (HV Training Condition)	Panel B (LV Training Condition)

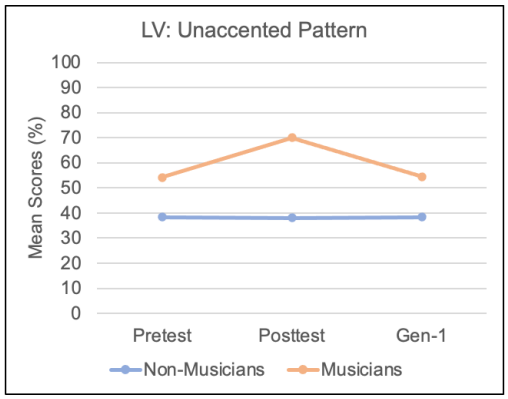












Note. Gen-1 = Test of generalization 1

What stands out in Figure 4.7 is that, as in the case of overall accuracy reported in Section 4.3.1, non-musicians and musicians attained similar improvements in the HV training condition (Panel A), whereas in the LV training condition there were stark differences for all three pitch-accent patterns between non-musicians and musicians, with musicians improving much more (Panel B).

Comparing the two sets of panels in Figure 4.7, again, as in the case of overall accent pattern accuracy, musicians improved in a similar manner in the two training conditions, while this was not the case for non-musicians. In the HV training condition, non-musicians performed similarly to musicians, but in the LV training condition, non-musicians showed only slight improvements in 1st-syllable accented pattern identification and showed hardly any improvements in the other two pitch-accent patterns.

A closer look at the line graphs for the 1st-syllable accented pattern (the top two panels in Figure 4.7) reveals that, irrespective of participants' training conditions or category (non-musicians/musicians), the 1st-syllable accented pattern had the lowest pretest scores compared to the other two pitch-accent patterns, but that pretest-posttest improvements for this pattern persisted for pretest-Gen-1.

The line graphs for the 2nd-syllable accented pattern (the middle two panels in Figure 4.7) also show pretest-posttest and pretest-Gen-1 improvements, except in the case of non-musicians in the LV training condition.

Looking at the line graphs for the unaccented pattern (the bottom two panels in Figure 4.7), it can be seen that non-musicians in the LV training condition showed no improvement at all in scores at Gen-1. Remarkably, non-musicians in the HV training condition and musicians in both training conditions attained noticeably lower scores at Gen-1 than at the posttest. Moreover, in the HV training condition, the difference in scores between non-musicians and musicians had widened at Gen-1 compared to scores at the posttest (1.8% Δ at the posttest; 8.9% Δ at Gen-1); whereas the opposite was found for the LV training condition—the difference between musicians' and non-musicians' scores decreased from the posttest to Gen-1.

These results show different improvement trends for the three pitch-accent patterns, and indicate their interaction with the effects of musical training, and perceptual training condition (talker variability in training stimuli).

To examine whether perceptual learning of Japanese pitch accent varied among three pitch-accent patterns and whether the learning was influenced by musical training, and of LV or HV training condition, mixed-effects binomial logistic regression analyses were conducted. Based on Shport (2016), pretest-posttest data and pretest-Gen-1 data were separately analyzed, since the interest of the present study was to assess improvements from the pretest to the posttest and Gen-1, not improvements from the posttest to Gen-1.

The analysis of pretest and pretest accuracy scores is provided first. As mentioned in Section 4.2.5, participants' binary accuracy scores for each trial were used as the dependent variable. The model contained fixed effects for: *pattern* (1st-syllable accented pattern as the reference level, 2nd-syllable accented pattern, unaccented pattern: treatment-coded), *category* (non-musician as the reference level, musician: treatment-coded); *training condition* (HV as the reference level, LV: treatment-coded) and *test* (pretest as the reference level, posttest: treatment coded). Participants and items (stimuli) were included as random effects with by-participant and by-item varying intercepts for the effect under investigation. Attempts were made to include by-participant and by-item varying slopes, but this led to convergence issues.

Since Figure 4.7 suggested interactions between pattern, category and training condition, the final model was coded in R as: glmer(Correct ~ Pattern + Category + Test + Training_condition + Category : Test + Test : Training_condition + Category: Training_condition + Category : Pattern + Test : Pattern + Training_condition : Pattern + Category : Test : Training_condition + Category : Test : Pattern + Test : Training_condition : Pattern + Training_condition : Pattern : CategoryRe + Category : Test : Training_condition : Pattern + (1|Participant_Public_ID) + (1|Item), data = data, family = 'binomial').

The *afex* package (Singmann et al., 2023) was used to perform likelihood ratio tests for all fixed effects. The results are summarized in Table 4.10.

Table 4.10
Likelihood Ratio Test for All Fixed Effects: Pretest-Posttest Differences

Fixed effect	Result
Pattern	$\chi^2(2) = 5.77$, $p = 0.056$
Category	$\chi^2(1) = 7.55$, $p = 0.006$*
Test	$\chi^2(1) = 239.67$, $p < 0.001$*
Training condition	$\chi^2(1) = 4.86$, $p = 0.027$*
Category × Test	$\chi^2(1) = 12.57$, $p < 0.001$*
Category × Training condition	$\chi^2(1) = 4.18$, $p = 0.041$*
Pattern × Category	$\chi^2(2) = 30.93$, $p < 0.001$*
Test × Training condition	$\chi^2(1) = 5.63$, $p = 0.018$*
Pattern × Test	$\chi^2(2) = 21.28$, $p < 0.001$*
Pattern × Training condition	$\chi^2(2) = 6.89$, $p = 0.032$*
Category × Test × Training condition	$\chi^2(1) = 17.93$, $p < 0.001$*
Pattern × Category × Test	$\chi^2(2) = 1.63$, $p = 0.443$
Pattern × Test × Training condition	$\chi^2(2) = 1.57$, $p = 0.456$
Pattern × Category × Training condition	$\chi^2(2) = 21.41$, $p < 0.001$*
Pattern × Category × Test × Training condition	$\chi^2(2) = 11.14$, $p = 0.010$*

Note. * = significant effect

Table 4.10 shows that there were significant effects of category, of test and of training condition on participants' accuracy scores. In addition, there were significant interactions between: (1) category and test; (2) category and training

condition; (3) pattern and category; (4) test and training condition; (5) pattern and test; (6) pattern and training condition; (7) category, test, and training condition; (8) pattern, category, training condition; and (9) pattern, category, test and training condition.

In order to assess the interactions in more detail, Bonferroni-corrected multiple comparisons were calculated using the *emmeans* package (Lenth, 2023). The significant results of multiple comparisons are summarized in Table 4.11 (full details are reported in Appendix G).

Table 4.11
Multiple Comparisons: Significant Interactions Between Pattern, Training Condition, Musician/Non-Musician Category, and Identification Test (Pretest and Posttest).

Contrast
Contrasts for Test:
HV – Non-Musicians:
Pretest – Posttest (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)
HV – Musicians:
Pretest – Posttest (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)
LV – Non-Musicians:
Pretest – Posttest (1st-syllable accented pattern)
LV – Musicians:
Pretest – Posttest (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)

Contrasts for Category
LV – Musicians vs. LV – Non-Musicians:
Pretest (2nd-syllable accented pattern; unaccented pattern)
Posttest (2nd-syllable accented pattern; unaccented pattern)
Contrasts for Training Condition
HV – Non-Musicians vs. LV – Non-Musicians:
Pretest (2nd-syllable accented pattern; unaccented pattern)
Posttest (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)
Contrasts for Pattern
HV – Non-Musicians:
Pretest (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)
HV –Musicians:
Pretest (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)
LV – Non-Musicians:
Posttest (1st-syllable accented pattern vs. unaccented pattern)
LV –Musicians:
Pretest (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)
Posttest (1st-syllable accented pattern vs. unaccented pattern)

Note. Gen-1 = Test of generalization 1

As can be seen in Table 4.11, multiple comparisons confirmed what is shown in Figure 4.7.

All participants, except for non-musicians in the LV training condition, showed significant pretest-posttest improvements in identifying pitch-accent patterns. Indeed, non-musicians in the LV training condition improved significantly only in 1st-syllable accented pattern identification.

In the LV training condition, there were significant differences for the 2nd-syllable accented and unaccented patterns between non-musicians and musicians at both the pretest and the posttest. By contrast, in the HV training condition, no significant differences were found between non-musicians and musicians.

Multiple comparisons for training condition (in the middle of Table 4.11) revealed that non-musicians in the HV training condition performed significantly better than those in the LV training condition in the 2nd-syllable accented and unaccented patterns in the pretest. As touched upon in Section 4.3.1, this unexpected outcome, suggesting individual differences between non-musicians, is discussed in Section 7.2. Still with regard to non-musicians, there were also significant differences in posttest scores between HV and LV training conditions for all pitch-accent patterns.

Except for non-musicians in the LV training condition, pretest scores were significantly lower for the 1st-syllable accented pattern than for the other two patterns (see the lower half of Table 4.11). This suggests that the 1st-syllable accented pattern was the most difficult pattern at the beginning.

Multiple comparisons for patterns (in the lower half of Table 4.11) also yielded that in the LV training condition, posttest scores for the 1st-syllable accented pattern differed significantly from those for the unaccented pattern, for both categories. However, as can be observed in Figure 4.7, musicians and non-musicians in the LV training condition showed different trends: while musicians got higher posttest scores for the unaccented pattern than for the 1st-syllable

accented pattern, non-musicians did not show a significant pretest-posttest improvement for the unaccented pattern.

Having concluded the analysis of pretest and posttest accuracy scores, the remainder of this section provides an analysis of pretest and Gen-1 accuracy scores.

As mentioned in Section 4.2.5, participants' binary accuracy scores for each trial were used as the dependent variable. The model contained fixed effects for: *pattern* (1st-syllable accented pattern as the reference level, 2nd-syllable accented pattern, unaccented pattern: treatment-coded), *category* (non-musician as the reference level, musician: treatment-coded); *training condition* (HV as the reference level, LV: treatment-coded) and *test* (pretest as the reference level, Gen-1: treatment coded). Participants and items (stimuli) were included as random effects with by-participant and by-item varying intercepts for the effect under investigation. Attempts were made to include by-participant and by-item varying slopes, but this led to convergence issues.

Since Figure 4.7 suggested interactions between pattern, category and training condition, the final model was coded in R as: glmer(Correct ~ Pattern + Category + Test + Training_condition + Category : Test + Test : Training_condition + Category: Training_condition + Category : Pattern + Test : Pattern + Training_condition : Pattern + Category : Test : Training_condition + Category : Test : Pattern + Test : Training_condition : Pattern + Training_condition : Pattern : CategoryRe + Category : Test : Training_condition : Pattern + (1|Participant_Public_ID) + (1|Item), data = data, family = 'binomial', control = glmerControl(optimizer = 'bobyqa')).

The *afex* package (Singmann et al., 2023) was used to perform likelihood ratio tests for all fixed effects. The results are summarized in Table 4.12.

Table 4.12
Likelihood Ratio Test for All Fixed Effects: Pretest-Gen-1 Differences

Fixed effect	Result
Pattern	$\chi^2(2) = 10.33$, $p = 0.006$*
Category	$\chi^2(1) = 7.91$, $p = 0.005$*
Test	$\chi^2(1) = 20.94$, $p < 0.001$*
Training condition	$\chi^2(1) = 2.25$, $p = 0.134$
Category × Test	$\chi^2(1) = 11.48$, $p < 0.001$*
Category × Training condition	$\chi^2(1) = 3.84$, $p = 0.050$*
Pattern × Category	$\chi^2(2) = 8.20$, $p = 0.017$*
Test × Training condition	$\chi^2(1) = 2.56$, $p = 0.109$
Pattern × Test	$\chi^2(2) = 40.60$, $p < 0.001$*
Pattern × Training condition	$\chi^2(2) = 6.44$, $p = 0.040$*
Category × Test × Training condition	$\chi^2(1) = 13.16$, $p < 0.001$*
Pattern × Category × Test	$\chi^2(2) = 3.23$, $p = 0.198$
Pattern × Test × Training condition	$\chi^2(2) = 24.63$, $p < 0.001$*
Pattern × Category × Training condition	$\chi^2(2) = 21.64$, $p < 0.001$*
Pattern × Category × Test × Training condition	$\chi^2(2) = 11.14$, $p = 0.004$*

Note. * = significant effect

In order to assess the interactions in more detail, Bonferroni-corrected multiple comparisons were conducted using the *emmeans* package (Lenth, 2023). The results of significant multiple comparisons are summarized in Table 4.13 (full details are reported in Appendix H).

Table 4.13
Multiple Comparisons: Significant Interactions Between Pattern, Training Condition, Musician/Non-Musician Category, and Identification Test (Pretest and Gem-1).

Contrast
Contrasts for Test:
HV – Non-Musicians:
Pretest – Gen-1 (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)
HV – Musicians:
Pretest – Gen-1 (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)
LV – Non-Musicians:
Pretest – Gen-1 (1st-syllable accented pattern)
LV – Musicians:
Pretest – Gen-1 (1st-syllable accented pattern; 2nd-syllable accented pattern)
Contrasts for Category
LV – Musicians vs. LV – Non-Musicians:
Pretest (2nd-syllable accented pattern; unaccented pattern)
Gen-1 (1st-syllable accented pattern; 2nd-syllable accented pattern; unaccented pattern)
Contrasts for Training Condition
HV – Non-Musicians vs. LV – Non-Musicians:
Pretest (2nd-syllable accented pattern; unaccented pattern)

Gen-1 (1st-syllable accented pattern; 2nd-syllable accented pattern)

Contrasts for Pattern

HV – Non-Musicians:

Pretest (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)

Gen-1 (1st-syllable accented pattern vs. unaccented pattern; 2nd-syllable accented pattern vs. unaccented pattern)

HV –Musicians:

Pretest (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)

Gen-1 (1st-syllable accented pattern vs. unaccented pattern; 2nd-syllable accented pattern vs. unaccented pattern)

LV – Non-Musicians:

Gen-1 (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)

LV – Musicians:

Pretest (1st-syllable accented pattern vs. 2nd-syllable accented pattern; 1st-syllable accented pattern vs. unaccented pattern)

Gen-1 (1st-syllable accented pattern vs. unaccented pattern; 2nd-syllable accented pattern vs. unaccented pattern)

Note. Gen-1 = Test of generalization 1

Again, as can be seen in Table 4.13, multiple comparisons confirmed what is shown in Figure 4.7.

Except for non-musicians in the LV training condition, all participants, showed significant pretest-Gen-1 improvements in identifying the 1st-syllable accented and 2nd-syllable accented patterns. Non-musicians in the LV training condition improved significantly only in the 1st-syllable accented pattern identification.

Recall, Figure 4.7 showed that, for the unaccented pattern, both musicians and non-musicians in the HV training condition showed pretest-Gen-1 deteriorations. Multiple comparisons revealed that this difference was significant. By contrast, in the LV training condition, no significant differences for the unaccented pattern were found for either non-musicians or musicians: indeed, non-musicians and musicians scored virtually the same at the pretest and Gen-1. Since the stimuli used in Gen-1 were novel to participants, it can be assumed that it was difficult for participants to generalize what they had learned in perceptual training to novel stimuli with the unaccented pattern.

In the LV training condition, multiple comparisons for category revealed significant pretest differences between non-musicians and musicians for the 2nd-syllable accented and the unaccented patterns. At Gen-1, there were significant differences for each of the pitch-accent patterns. By contrast, in the HV training condition no significant differences were found between non-musicians and musicians.

Multiple comparisons for training condition (in the middle of Table 4.13, and as can also be seen in Table 4.11) revealed that non-musicians in the HV training condition performed significantly better at the pretest than those in the LV training condition for the 2nd-syllable accented and unaccented patterns. There were also significant differences between LV and HV training conditions for non-musician Gen-1 scores in the 1st-syllable accented and 2nd-syllable accented patterns. Again, this reflects the difficulty in identifying the unaccented pattern at

Gen-1. Indeed, non-musicians in the two training conditions scored almost the same on Gen-1.

Multiple comparisons for pattern (in the lower half of Table 4.13, see also Table 4.11) showed that, except for non-musicians in the LV training condition, pretest scores for the 1st-syllable accented pattern were significantly lower than those for the other two pitch-accent patterns. As mentioned earlier, this suggests that the 1st-syllable accented pattern was the most difficult pattern at the beginning.

At Gen-1, however, a different picture was observed. For the unaccented pattern, all participants except for non-musicians in the LV training condition attained a lower score at Gen-1 than at the pretest. As mentioned earlier, this indicated that participants had difficulty in identifying the unaccented pattern in novel stimuli. As regards non-musicians in the LV training condition, Gen-1 scores for only the 1st-syllable accented pattern were significantly different from those for the other two pitch-accent patterns. Indeed, as can be seen in Figure 4.7, non-musicians in the LV training condition did not show any improvements for the 2nd-syllable accented and unaccented patterns. This implies that the LV training condition was of limited benefit for non-musicians compared to the HV training condition.

In summary, the results indicate that, initially, the most difficult pitch-accent pattern was the 1st-syllable accented pattern, but that, regardless of the training conditions, participants learned to identify not only the 1st-syllable accented pattern but also the other pitch-accent patterns thanks to perceptual training. However, it was observed that at Gen-1, participants found difficulty in identifying the unaccented pattern, suggesting that the unaccented pattern was the least generalizable. As regards musical training, a positive effect was found (see the overall accuracy results in Section 4.3.1). However, statistical analyses revealed interactions between perceptual training and musical training. It appears that the HV training condition was more beneficial for non-musicians, whereas for musicians, there was no stark difference in the effectiveness between the two training conditions.

4.4. Identification Tasks: Discussion

In keeping with the structure of Section 4.3, this section is also divided into two parts. In response to RQ1-3, the first part discusses the overall effect of musical training; and then that of talker variability in the stimuli used in training. The second part, in response to RQ4, focuses on participants' improvements in the three target pitch-accent patterns.

4.4.1. Identification Tasks: Discussion of Effects on Overall Accuracy of Musical Training and of Talker Variability

The current research aimed to investigate whether musical training influences perceptual learning of Japanese pitch accent by native speakers of a non-tone language (Italian), who had no experience of Japanese. To this end, Italian non-musicians and musicians engaged in Japanese pitch-accent identification before and after undergoing perceptual training conducted following a HVPT paradigm. The present study also aimed to examine the effect of talker variability in training stimuli on perceptual learning of Japanese pitch accent. To achieve this goal, musicians and non-musicians were randomly assigned to a high variability (HV) training condition (stimuli produced by four talkers), or a low variability (LV) training condition (stimuli produced by one talker).

This research also attempted to assess the effect of absolute pitch on perceptual learning of Japanese pitch accent. However, the results for this are reported in Chapter 6.

In the remainder of this section, the findings of the current research are discussed in the light of research questions (RQ1-3) and previous studies.

RQ1 addressed whether Italian musicians would outperform Italian non-musicians in identifying Japanese pitch accent. As expected, musicians clearly outperformed non-musicians in all tests. This result is consistent with that of various perception studies (Alexander et al., 2005; Burnham et al., 2015; Chang et al., 2016; Chen et al., 2020; Delogu et al., 2010; Gottfried, 2007; Gottfried & Xu, 2008; Götz et al., 2023; Kirkham et al., 2011; Lee et al., 2014; Lee & Hung, 2008; Marie et al.,

2011; Mok & Zuo, 2012), which have shown, overall, the positive effect of musical experience/training on lexical tone perception by native non-tone language speakers without any experience of the target tone language. The finding of the current research is also in line with that of Golob's study (2003), in which Slovenian musicians without any experience of Japanese perceived Japanese pitch accent better than Slovenian non-musicians who were learners of Japanese. Thus, the results of the present study further support evidence for the positive effect of musical training, providing empirical data for a pair of languages which has not yet been well studied (Japanese and Italian).

As just mentioned, previous works largely converged to indicate the facilitative effect of musical experience/training on lexical tone perception by native non-tone language speakers, albeit to varying degrees. However, as discussed in Section 2.3.3, training studies have reported mixed findings regarding additional musical advantage in the outcome of perceptual learning of lexical tone, and as far as the author knows, there has been no work exploring the effect of musical experience/training on perceptual learning of Japanese pitch accent.

RQ2 addressed whether the difference between musicians and non-musicians in the ability to identify Japanese pitch accent would decrease or increase after training. In line with the prediction, musicians showed greater improvement from the pretest to the posttest and to Gen-1 and Gen-2, in which novel stimuli were used. In fact, the difference between non-musicians and musicians widened markedly after the pretest (difference at the pretest: 7.5%; at the posttest: 12.4%; at Gen-1: 12.2%; at Gen-2: 10%). This finding indicated that there was an added benefit for musicians not only at the posttest (which was identical to the pretest) but also in the two tests of generalization, which means that musicians were better able than non-musicians to generalize to novel stimuli what they had learned. This result is dissimilar to those of some training studies (Dittinger et al., 2016; Tong & Tang, 2016; Wayland et al., 2010; Zhao & Kuhl, 2015), but in accord with those of other studies (Cooper & Wang, 2012; Maggu et al., 2018; P. C. M. Wong & Perrachione, 2007).

Of the aforementioned works, only Zhao and Kuhl (2015) employed the HVPT paradigm as the current research did. However, as just mentioned, their result differs from that of this study: they found limited evidence that musicians had an advantage in the lexical tone learning process. A possible explanation for this might be the difference in the tasks proposed. As in the present study, Zhao and Kuhl used discrimination and identification tasks before and after HVPT, but their test stimuli were presented in a nine-step tone 2 tone 3 continuum based on real monosyllabic Mandarin words. Thus, these were, more properly speaking, categorization tasks. Another possible explanation for this discrepancy in findings is the definition of musicians. Recall that the current work defined musicians as individuals currently engaged in formal tertiary-level musical training, including those enrolled in conservatories, musical institutes, or majoring in musicology at university. In Zhao and Kuhl (2015), on the other hand, musicians were defined as having received at least eight years of private music lesson beginning before the age of 10 years. Given their definition, the possibility of having included an intermediate category—between expert musicians and non-musicians, like the amateur musicians in Ericsson et al. (1993) and Sloboda et al. (1996)—cannot be ruled out. Since the current research's musicians were all expert musicians, the effect of musical training may have been more salient than in Zhao and Kuhls' study.

So far, in response to RQ1 and RQ2, the effect of musical training on perceptual learning of Japanese pitch accent has been discussed.

Before responding to RQ3 in detail it is worth remembering that both musicians and non-musicians, irrespective of the training conditions, showed pretest-posttest improvements. This indicated that the training method proposed by Shport (2011, 2016) was effective despite the brevity of the training sessions (one hour in total). Further robust evidence for its effectiveness comes from the fact that the present study's experiment was conducted entirely online and not in a laboratory as in Shport's works, and that the training was also effective for native Italian speakers (Shport's participants were native English speakers). This effectiveness offers an obvious practical implication—the use of training in L2 Japanese learning/teaching settings—which is discussed in Section 7.3.

Having mentioned the overall effectiveness of training, the focus of the remining part of this section is on RQ3, which was set to investigate the effect of talker variability in training stimuli (HV training condition vs. LV training condition) and its interaction with the effect of musical training. Specifically, RQ3 addressed whether or not the HV training condition would be more beneficial for Italian musicians compared to non-musicians.

Surprisingly, musicians and non-musicians showed different trends. On one hand, musicians in the two training conditions improved in pitch-accent pattern identification in a similar manner; indeed, no significant differences were found between musicians in the two training conditions. On the other hand, in the case of non-musicians, the difference in training condition led to differences in their results. Although, at the pretest, no significant difference was found between non-musicians in the two training conditions, non-musicians in the HV training condition performed significantly better than those in the LV training condition at the posttest, Gen-1 and Gen-2. In addition, the performance of non-musicians in the HV training condition improved from the pretest not only to the posttest but also to Gen-1 and Gen-2, although their Gen-1 scores were significantly lower than their posttest scores. By contrast, non-musicians in the LV training condition improved significantly only from the pretest to the posttest. These results indicate not only that the HV training condition was more immediately beneficial for non-musicians, but also that—unlike the LV training condition—it induced non-musicians to generalize what they had learned to the novel stimuli used in Gen-1 and Gen-2. It also bears noting that non-musicians in the HV training condition achieved similar improvements to musicians. Indeed, in the HV training condition, no significant differences in identification test scores were found between non-musicians and musicians. In stark contrast, in the LV training condition, musicians outperformed non-musicians in all tests.

Taken together, these findings suggest that whereas, for musicians, the two training conditions are comparably effective, for non-musicians, the HV training condition is more beneficial than the LV training condition.

These results are different to those of Dong et al. (2019), but similar to those of Perrachione et al. (2011), Sadakata and McQueen (2014), and Qin et al. (2022), in that the present study found an interaction between talker variability in training stimuli and musical training. However, the interaction that emerged in the current research differed from that reported by Perrachione et al. (2011), Sadakata and McQueen (2014), and Qin et al. (2022). The former two studies found that while high perceptual aptitude participants benefitted from the HV training condition, low perceptual aptitude participants received more benefit from LV training, and higher variability hindered perceptual learning in them. This detrimental effect of the HV training condition on low perceptual aptitude participants was also reported by Qin et al. (2022).

Note, however, that the abovementioned studies examined whether talker variability in training stimuli interacted with individuals' perceptual abilities, not with musical training. This difference may lead to the discrepancy in the results between the current research and these studies.

To the knowledge of the author, this study is the first of its kind to investigate the interaction between musical training and talker variability. Thus, the current research provides new evidence for this interaction.

With regard to musicians, the results suggest that talker variability did not play a role in perceptual learning of Japanese pitch accent. It may be reasonable to assume that musically trained individuals are capable of extracting abstract information about Japanese pitch accent and of applying it to novel input without the need for different samples.

By contrast, in the case of non-musicians, the findings indicate that talker variability did play a role: the HV training condition favored perceptual learning of Japanese pitch accent more than the LV training condition. This is consistent with Silpachai (2020), whose target was Mandarin Chinese lexical tone contrasts, and with Sadakata and McQueen (2013), Wong (2012), and Wong (2014), whose target was segment contrasts. The evidence implies that, unlike musicians, non-musicians need a variety of voice samples to learn to identify Japanese pitch accent, especially for generalization to new input. The observed importance of talker variability in

auditory input may have important implications for L2 Japanese learning/teaching and, in Section 7.3, these are discussed.

4.4.2. Identification Tasks: Discussion of Improvements in Pitch-Accent Pattern Accuracy

In this section, the findings of the current research are discussed in the light of research question RQ4 and previous studies.

RQ4 addressed which pitch-accent pattern would be the most difficult to perceive for native Italian speakers. As in Shport's studies (2011, 2016), there were three target pitch-accent patterns in the current research: 1st-syllable accented, 2nd-syllable accented, and unaccented. Recall that only pretest-posttest data and pretest-Gen-1 data were analyzed to examine whether non-musicians and musicians had improved their accuracy for each of the three pitch-accent patterns, as similar improvements were found for all three pitch-accent patterns from the pretest to Gen-1 and from the pretest to Gen-2 (see Section 4.3.2).

Findings for the pretest in the current research were in contrast both with what had been predicted in RQ4, and with the results of Shport (2011, 2016), Laméris and Graham (2020), and Wu et al. (2012, 2017). The present pretest results suggested that the most difficult pitch-accent pattern for all participants was the 1st-syllable accented. This is partially consistent with Pappalardo (2018) and Hirano-Cook (2011). I say partially because they reported that the easiest pattern was the unaccented one: the current research, on the other hand, found no significant differences at the pretest between the 2nd-syllable accented and unaccented patterns in participants' results irrespective of category and training condition.

With regard to the 1st-syllable accented pattern, all participants showed significant pretest-posttest improvements. However, for the other two pitch-accent patterns, only musicians in both training conditions and non-musicians in the HV training condition attained significant pretest-posttest improvements; the posttest scores of non-musicians in the LV training condition were almost the same as those

for the pretest, suggesting a limited positive effect of the LV training condition for non-musicians.

As was observed for overall accuracy in pitch accent identification (see sections 4.3.1 and 4.4.1), the results for the single pitch-accent patterns indicate that HV training condition was beneficial for non-musicians. Surprisingly, however, even at the pretest, non-musicians in the HV training condition performed significantly better than those in the LV training condition on the 2nd-syllable accented and unaccented patterns. This interesting result, suggesting individual differences between non-musicians, is discussed in Section 7.2. Non-musicians also showed significant differences in posttest scores between the HV and LV training conditions for all pitch-accent patterns. In addition, non-musicians in the HV training condition achieved similar improvements to those of musicians. Indeed, in the HV training condition, no significant differences were found between non-musicians and musicians. In clear contrast, in the LV training condition, musicians outperformed non-musicians for the 2nd-syllable accented and unaccented patterns at both the pretest and the posttest. On the other hand, as regards musicians, there were no significant differences between the two training conditions.

Interestingly, at Gen-1, a different picture was observed. Except for non-musicians in the LV training condition, all participants showed significant pretest-Gen-1 improvements in identifying the 1st-syllable accented and 2nd-syllable accented patterns. These results indicate that, at Gen-1, the most difficult pitch-accent pattern was no longer the 1st-syllable accented pattern, but the unaccented pattern; even though non-musicians in the LV training condition had difficulty in identifying both the 2nd-syllable accented and the unaccented patterns.

There is more evidence that the unaccented pattern was the most difficult pattern at Gen-1: for the unaccented pattern, both musicians and non-musicians in the HV training condition showed pretest-Gen-1 deteriorations, and in the LV training condition, non-musicians and musicians scored virtually the same at the pretest and Gen-1. Recall that the stimuli used in Gen-1 were novel to participants. Thus, it is reasonable to assume that it was the unaccented pattern which made it most difficult for participants to generalize what they had learned during training to

the novel stimuli employed at Gen-1. This partially supports the prediction for RQ4 presented in Section 4.1.1: that the most difficult pattern would be the unaccented pattern. This finding is also in line with Shport (2011, 2016).

Recall that Japanese pitch accent consists of a bitonal high-low accent implemented as a fundamental frequency (F0) peak near the end of the accented mora followed by a sharp F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006) and that, in the unaccented pattern, there is no steep F0 fall. Hence, the pitch-accent patterns of Japanese can be described in terms of two parameters: (1) presence or absence of pitch accent, and (2) if present, location of pitch accent (Kawahara, 2015; Sugiyama, 2012). In the light of the evidence that participants found the unaccented pattern difficult Gen-1, it can be speculated that, like English-speaking participants in Shport (2015), Italian participants paid more attention to F0 peak locations, which distinguish the 1st-syllable accented and 2nd-syllable accented patterns), but that they paid less attention to the presence or absence of the F0 fall cue, which serves to differentiate the unaccented pattern from the other two pitch-accent patterns.

As mentioned earlier, at the posttest (which was identical to the pretest), all participants except for the non-musicians in the LV training condition performed significantly better than at the pretest for the unaccented pattern. This indicates the positive effect of the training. However, it is conceivable that training sessions (one hour in total) were not long enough to induce participants to generalize to novel stimuli with the unaccented pattern. Future work is needed to determine if longer training sessions would be more effective.

Going back to pretest and Gen-1 accuracy scores for each pitch accent pattern, analysis revealed the same trends as for pretest and posttest accuracy scores found for the interaction between category and training condition. On one hand, as regards musicians, no significant differences between the two training conditions were found. On the other hand, as regards non-musicians, as mentioned before, those in the HV training condition identified significantly better than those in the LV training condition the 2nd-syllable accented and unaccented patterns at the pretest and the 1st-syllable accented and 2nd-syllable accented patterns at Gen-1.

Interestingly, for the unaccented pattern, non-musicians in the two training conditions scored almost the same on Gen-1, indicating the difficulty in identifying the unaccented pattern at Gen-1. At Gen-1, in the HV training condition, no significant differences were found between non-musicians and musicians, whereas, in the LV training condition, musicians outperformed non-musicians for each of the pitch-accent patterns. These findings indicate that the benefit of the HV training condition was greater than that of the LV training condition for the 1st-syllable accented and 2nd-syllable accented patterns.

To sum up, it seems that the most difficult pitch-accent pattern for participants differed before training and after training. At the pretest, the most difficult pattern was the 1st-syllable accented pattern, whereas at Gen-1, it appears that the unaccented pattern was the most difficult pattern, suggesting that the unaccented pattern was the least generalizable. Overall, perceptual training brought about improvements not only in the 1st-syllable accented pattern but also in the other pitch-accent patterns. However, the effect of the LV training condition on non-musicians' outcomes was limited to the 1st-syllable accented pattern. Indeed, as in overall accuracy (see sections 4.3.1 and 4.4.1), it is reasonable to assume that for non-musicians the HV training condition was more beneficial. By contrast, for musicians there was no stark difference in the effectiveness between the two training conditions.

CHAPTER 5 DISCRIMINATION TASKS: PITCH-ACCENT PATTERN DISCRIMINATION

This chapter focuses on the *Discrimination Tasks* conducted before and after training, namely the discrimination pretest, posttest and generalization test. The details and data are analyzed below.

5.1. Discrimination Tasks: Introduction

Since general background information for this experiment has already been presented in Chapter 4 (*Identification Tasks*), this section only provides the reasons for including *Discrimination Tasks* in this experiment. It is necessary to do so because Shport's studies (2011, 2016) on which the present research was based, only conducted identification tasks.

The first reason was that the difficulty of discrimination tasks for participants has varied from study to study. For example, Shport (2008) tested intermediate learners of Japanese, who were native speakers of American English, with an AXY discrimination task in which they listened to the stimuli set and selected the word which they perceived as having a different pitch-accent pattern from the other two words. She found that the learners failed to accurately perceive pitch-accent patterns at a higher-than-chance rate, suggesting that it is extremely difficult for learners to acquire Japanese pitch accent. On the other hand, English learners of Japanese in Sakamoto's study (2011) performed well on an ABX discrimination task—in which participants had to answer whether the last stimulus was the same as the first stimulus or the second one. Indeed, her statistical analysis revealed that there was no statistical difference between native Japanese speakers' and two groups of learners (experienced and inexperienced learners) in their discrimination accuracy. The author of the current research decided to include pitch-accent discrimination tasks in an attempt to clarify whether or not these would be difficult for the Japanese-naïve mother tongue Italian participants.

The second reason for conducting *Discrimination Tasks* in the current research is based on the findings of Sadakata and McQueen (2013). In their study,

native Dutch speakers without any experience of Japanese trained to acquire a Japanese geminate-singleton fricative contrast under either of two training conditions: HV (fewer repetitions of more varied stimuli recorded by five talkers), and LV (many repetitions of less limited stimuli recorded by a single talker). Sadakata and McQueen reported that, although the benefit of the HV training condition was observed in identification tests, it was not observed in discrimination tasks: participants' discrimination performance improved irrespective of their training condition. Thus, the current study sought to examine whether this would also be the case for perceptual learning of Japanese pitch accent.

The third reason for including *Discrimination Tasks* in the current research is based on results reported by Shport (2015). Using an AX discrimination task and a two-alternative forced-choice (2AFC) categorization task, she investigated whether native English speakers with no prior knowledge of Japanese or other tone languages would be sensitive to the F0 peak location and the F0 fall in pitch-accent contrasts. Interestingly, the findings of the two tasks showed the presence of large individual differences between the participants, and as one of the possible sources of such differences, she suggested musical training. The present study pursued this line of reasoning.

The last reason for carrying out *Discrimination Tasks* is findings shown by Burnham and colleagues' study (2015), from which the present experiment's methodology drew inspiration. They examined the effects of absolute pitch ability and musical training on Thai lexical tone discrimination. To the author's knowledge, only Burnham et al. (2015) have reported that musicians with absolute pitch were more accurate at lexical tone discrimination. Their participants—native speakers of Australian English without any experience of tone languages—consisted of three groups: (1) musicians with absolute pitch; (2) musicians without absolute pitch; and (3) non-musicians who had not had any musical training. The division of the two musician groups (1 and 2, recruited from two music-training institutions in Australia) was based on the results of an absolute pitch test administered by Burnham et al. The participants engaged in an AX discrimination task in which they were instructed to answer whether the two stimuli they had heard were the same or different. Stimuli used in the AX discrimination task were a monosyllable with five

Thai lexical tones. Burnham and colleagues used two interstimulus intervals (ISIs): 500 ms and 1500 ms. Length of the ISI influences which categories (acoustic, phonetic, or phonological) listeners use in order to classify stimuli: while an ISI of 500 ms encourages acoustic processing of speech stimuli, an ISI of 1500 ms encourages phonological processing (Werker & Logan, 1985). Hence, using the two ISIs, the researchers explored whether processing level interacted with musical training and absolute pitch ability. In Burnham and colleagues' study, two types of data were analyzed: discrimination accuracy data and reaction time (RT) data. As to the discrimination accuracy data, results showed that the two musician groups outperformed the non-musicians, and that the musicians with absolute pitch performed better than the musicians without absolute pitch. The results indicated that, over and above musical training, absolute pitch ability positively influenced Thai lexical tone discrimination. Neither overall effect of ISI nor significant interactions with ISI were found. The RT results revealed that the two groups of musicians were significantly faster than the non-musicians, however, no significant overall difference between the two musician groups was found. The musicians with absolute pitch did have a RT advantage over the musicians without absolute pitch, but only at the ISI of 1500 ms. According to the researchers, absolute pitch cognitively involves long-term pitch memory, which requires internal pitch standards (templates) in order to identify tones using labels (Parncutt & Levitin, 2001).

To develop a better understanding of the role of absolute pitch, the present research set out to explore the effect not only of musical training but also of absolute pitch on perceptual learning of Japanese pitch accent, since, to the author's knowledge, only Burnham et al. (2015) have reported that musicians with absolute pitch were more accurate at lexical tone discrimination. Like Burnham et al., the author of the current work also employed two ISIs and gathered RT data. What is more, the current study followed one of the suggestions for future work provided by Burnham et al. (2015): the use of stimuli that varied in terms of the number and gender of speakers, similar to sounds in the real world.

In the light of the reasons for including *Discrimination Tasks* in the current experiment discussed above and literature review in Chapter 2, the next section provides research questions and predictions.

5.1.1. Discrimination Tasks: Research Questions and Predictions

As mentioned earlier, the present dissertation's main aim was to investigate the effect of musical training on perceptual learning of Japanese pitch accent by native speakers of a non-tone language (Italian), who were also naïve to Japanese. This study also examined the role of talker variability and its interaction with the effect of musical training. Moreover, it attempted to assess the role of absolute pitch. Note that this chapter reports on these issues focusing on only *Discrimination Tasks*; see Chapter 4 for *Identification Tasks*.

In the light of these aims and the background information which motivates the current dissertation discussed in Chapter 2 and Section 5.1, the following research questions were addressed:

RQ1: Will Italian musicians outperform Italian non-musicians in discriminating Japanese pitch accent?

RQ2: After training, will the difference between musicians and non-musicians in the ability to discriminate Japanese pitch accent decrease or increase?

RQ3: Will the HV training condition be more beneficial for Italian musicians compared to non-musicians?

RQ4: Will there be any difference in the ability to discriminate Japanese pitch accent between musicians with absolute pitch and those without absolute pitch?

Based on the literature review in Chapter 2, the following possibilities were tested in response to the research questions described above.

P1: Italian musicians would outperform Italian non-musicians in discriminating Japanese pitch accent in terms of accuracy and reaction times.

P2: There would be an added benefit for musicians, although a reduced or no benefit was also tested for, since evidence in the literature is mixed with regard to whether or not experimental lexical tone perceptual training is more beneficial for musicians than non-musicians (see Section 2.3.3).

P3: Whether or not there would be an effect for talker variability and its interaction. Again, the existing literature has shown mixed findings about the role of talker variability and its interaction with individuals' perceptual abilities (see sections 2.2.3 and 2.2.3.1).

P4: It was hypothesized that musicians with absolute pitch would perform better than those without absolute pitch. However, it was also predicted that it would be very difficult to find musicians with absolute pitch.

Since Chapter 6 is dedicated to the absolute pitch test for musicians, the results for **RQ4**/**P4** are reported in Chapter 6.

Now, having outlined the experimental predictions, the next section provides details about the methodology employed in the discrimination tasks.

5.2. Discrimination Tasks: Method

The methodology applied in the *Discrimination Tasks* was inspired by Burnham et al. (2015). As in Burnham's, the tasks involved in the present study were AX discrimination tasks in which participants listened two stimuli (A and X), and decided whether they were the *same* or *different*. It differed, however, in four aspects. Firstly, they was carried out entirely online by means of the Gorilla software package (Anwyl-Irvine et al., 2020), rather than in a laboratory. Secondly, the stimuli were different to Burnham's, not only because the target language was not the same (in Burnham's work the target language was Thai), but also because the stimuli used in *Discrimination Tasks* consisted of the target words embedded in sentential contexts. Thirdly, the stimuli were recorded by different speakers, whereas those used in Burnham's study were recorded by one speaker. Fourthly, as described below, *Discrimination Tasks* used a pre/posttest design, whereas only a discrimination test was administered in Burnham's study (there was no training).

Lastly, each participant in the two categories of participants (musicians and non-musicians) had been randomly assigned to one of two different training conditions: the HV (high variability) training condition and the LV (low variability) training condition.

5.2.1. Discrimination Tasks: Participants

The 64 adult native speakers of Italian participants (32 musicians and 32 non-musicians) are described in *Identification Tasks* participated.

5.2.2. Discrimination Tasks: Stimuli

5.2.2.1. Discrimination Tasks: Materials

The stimuli were Japanese carrier sentences containing target words, as for *Identification Tasks*.

There were three disyllabic target words, consisting of a triplet of segmentally identical words: *hána* "girl's name" (1st-syllable accented), *haná* "flower" (2nd-syllable accented), and *hana* "nose" (unaccented). These words, presented also in Sugiyama (2012) and in Kubozono (2018), had a CVCV structure (C = consonant and V = vowel). The CV syllabic structure is common in Japanese as illustrated by the *hiragana* chart (one of two syllabaries in Japanese). This is also the most frequent syllabic type in Italian which makes up slightly over 50% of the total frequency of syllabic types (Schmid, 1999). The pitch-accent patterns of the target words were checked using the newest edition of NHK[23] accent pronunciation dictionary (NHK Hoso Bunka Kenkyujo, 2016). The two lexical words—*haná* "flower" (2nd-syllable accented), and *hana* "nose" (unaccented)—was present in the dictionary. The pronunciation of the girl's name *hána* corresponds with that accepted by the author (a native Japanese speaker), as well as by speakers M4, M5, F4 and F8 (see Section 5.2.2.2).

[23] NHK, *Nippon Hoso Kyokai* "Japan Broadcasting Corporation", is a public broadcaster in Japan.

The carrier sentence used in *Discrimination Tasks* was: __ *o kanji de koko ni kaite kudasai*. "Please write __ here in Chinese characters." The target words in the carrier sentence were followed by the accusative particle *o* which did not change their pitch-accent patterns.

5.2.2.2. Discrimination Tasks: Speakers

Five native speakers of Tokyo Japanese were recruited (three female[24]: F4, F6, F8 and three male: M4, M5; mean age = 39.8 years; age range 37 to 45 years; SD = 3.03 years). All of them were born and had grown up in the Tokyo metropolitan area (namely, in either Tokyo or Chiba prefectures), and none had any speech impairment. Talker variability was thus achieved by having multiple voices, including both male and female speakers.

As described in *Identification Tasks* (Chapter 4), participants heard an explanation of Japanese pitch-accent patterns. This had been recorded by female speaker F6, whose voice was also employed in the short practice provided before each task in the experiment. The other four speakers produced the stimuli described in the previous section. Specifically, as can be seen in Table 5.1, one male speaker (M4) and one female speaker (F4) recorded stimuli used at the pretest and the posttest, while two others (M5 and F8), as novel talkers, uttered the stimuli employed at the test of generalization.

[24] Originally three female speakers (F4, F5 and F6) concluded their recording. However, since F5's recordings contained too much background noise, they were substituted by F8's recordings.

Table 5.1
Speakers and the Materials They Produced for Discrimination Tasks

Experimental phase	Speakers
Explanation Task familiarization	F6
Pretest /Posttest	M4, F4
Test of generalization	M4, F4 and **M5, F8** (novel talkers)

5.2.2.3. Discrimination Tasks: Recording Procedure

The recording procedure was identical to that in *Identification Tasks*, except for one aspect.

Since Japanese native speakers were instructed to utter stimuli at the speed of normal speech, the length of the stimuli varied (between 1.81 seconds and slightly over 2.7 seconds). In order to have stimuli of equal duration, a Praat script was used to add an amount of silence to the last part of all stimuli except the longest, using the longest as the reference duration. This was done to account for participants' phonological working memory, a subcomponent of working memory (the temporary storage and manipulation of information) concerned with verbal and acoustic information (Baddeley, 2003; Gathercole & Baddeley, 1993). All stimuli thus had the same length (namely 2.8 seconds).

In the case of the materials used in the short practice, an amount of silence was added, so that all materials lasted 1.5 seconds.

5.2.3. Discrimination Tasks: Design

Discrimination Tasks consisted of three stages: (1) pretest; (2) posttest; and (3) test of generalization. In line with Burnham et al. (2015), the same task format was used throughout: an AX discrimination of pitch-accent patterns.

(See Section 3.1 for counterbalancing of identification and discrimination tasks in the overall design).

5.2.4. Discrimination Tasks: Procedure

Before beginning each discrimination task, participants completed the task familiarization phase. They listened twice to the triplet of segmentally identical words *higashi* ("Mr. Higashi", "dry Japanese sweets", and "East") embedded in the carrier sentence *kore wa __ desu.* "This is __.", while looking at the corresponding pitch-accent patterns. Then they carried out a short practice comprised of four trials in random order, with the AX discrimination task and response keys. The same triplet of segmentally identical words and carrier sentence were used both in the explanation (see Section 3.2) and in the short practice. These materials, recorded by F6, were not used in any of the tests. The four trials in the short practice consisted of two pairs of trials: one pair with an interstimulus interval (ISI) of 500 ms and the other with an ISI of 1500 ms. In each pair, two stimuli were presented: the first two were the same and the second two were different. During the short practice, participants received trial-by-trial visual feedback on their responses. In the case of a correct response, a green check mark flashed; in the case of an incorrect answers, a red X mark appeared. At the end of the short practice, cumulative feedback was also provided (the number of correct trials/three trials, the number of all trials).

The three stages of *Discrimination Tasks*—(1) pretest; (2) posttest; and (3) tests of generalization—are detailed below.

5.2.4.1. Discrimination Tasks: Pretest

The pretest was conducted on Day 1. It consisted of 96 trials, organized into two blocks: one with 500 ms of interstimulus interval (ISI) and the other with 1500 ms (see Table 5.2). The order of the blocks was counterbalanced across participants. Trial order within each block was randomized.

Table 5.2
Summary of Pretest Terms and Structure

Stimuli type
3 target words (1 triplet of segmentally identical words x 3 pitch-accent patterns) embedded in a carrier sentence.
Stimulus pairs per block
4 combinations of stimulus pairs in terms of speakers' voices: F4 – F4: 12 trials (six *same* trials and six *different* trials); M4 – M4: 12 trials (six *same* trials and six *different* trials); F4 – M4: 12 trials (six *same* trials and six *different* trials); M4 – F4: 12 trials (six *same* trials and six *different* trials). Total: 48 trials (24 *same* trials and 24 *different* trials)
Blocks
Participants could take an optional break between blocks and within each block, at the halfway point they took a fixed break of 10 seconds. Block 1 (500 ms ISI): 48 trials (12 trials x 4 combinations of stimulus pairs); Block 2 (1500 ms ISI): 48 trials (12 trials x 4 combinations of stimulus pairs).
Number of trials
96 trials = 48 trials x 2 blocks

There was an optional break between blocks. Half way through each block, there was a fixed break of 10 seconds, but other than that, participants could not take any breaks within the block.

Stimuli used at the pretest were recorded by two speakers, F4 and M4 (see Table 5.1). The four possible combinations of their voices were F4 – F4, M4 – M4, F4 – M4, and M4 – F4. Since each block had 48 trials, there were 12 trials for each combination of the voices. These 12 trials for each combination of the voices consisted of six *same* trials and six *different* trials, so that each block contained the same number of *same* trials and *different* trials, as in Burnham et al. (2015).

Since the target words presented three different pitch-accent patterns (1st-syllable accented, 2nd-syllable accented, and unaccented), nine possible stimulus pair combinations were generated, as can be seen in Table 5.3. Among the nine combinations, three pairs were *same* and six pairs were *different*. For each combination of the voices, therefore, in the case of *same* trials, the three *same* trials were presented twice, on the other hand, in the case of *different* trials, the six *different* trials were presented only once.

Table 5.3
Possible Stimulus Pair Combinations in Terms of the Three Pitch-Accent Patterns

	1st stimulus – 2nd stimulus	1st stimulus – 2nd stimulus	1st stimulus – 2nd stimulus
Same	1st-syllable accented – 1st-syllable accented	2nd-syllable accented – 2nd-syllable accented	unaccented – unaccented
Different	1st-syllable accented – 2nd-syllable accented	2nd-syllable accented – 1st-syllable accented	unaccented – 1st-syllable accented
	1st-syllable accented – unaccented	2nd-syllable accented – unaccented	unaccented – 2nd-syllable accented

Each trial began with a fixation point (the symbol +) displayed on the computer screen to focus participants' attention (Jiang, 2012). This lasted 1000 ms. Then, a target word was shown in a carrier sentence in which the target word and

the particle *o*[25] that followed it, were underlined to attract participants' attention to the stimulus. The particle was also underlined, because Sugiyama's results (2012) suggest that the particle was needed even for native Japanese speakers to identify that a disyllabic segmentally identical word was either the 2nd-syllable accented pattern or unaccented pattern. Since participants were naïve to Japanese, the sentence that appeared on the monitor was transcribed in Latin alphabet (see Figure 5.1).

One second after the appearance of the sentence, participants heard the first stimulus (the target word embedded in the carrier sentence) and saw an icon of a person speaking to show that audio stimulus 1 was currently being played (see Panel A in Figure 5.1). After 500 ms or 1500 ms ISI, stimulus 2 was played, accompanied by the appearance of a second icon (see Panel B in Figure 5.1).

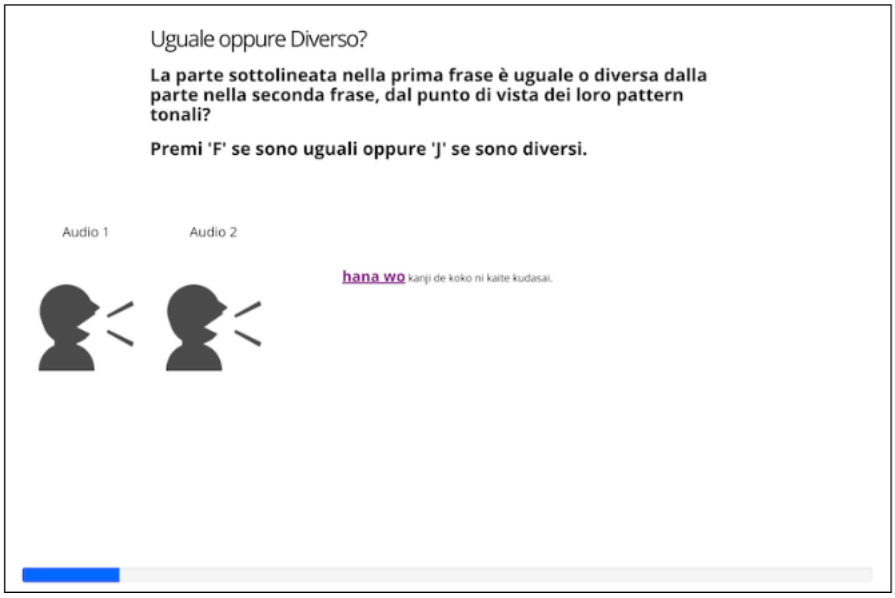
[25] As seen in Figure 5.1, only in *Discrimination Tasks*, the accusative particle を was transliterated into *wo,* instead of *o,* due to the lack of spelling control. However, note that there is no spelling standard for romanization in Japan, and that different conventions of romanization coexist. In fact, the representation of this particle by *o* is based on the Hepburn system, whereas that by *wo* is based on the *kunrei-shiki* (Rose, 2017). The representation by *wo* is also used in the *romaji* (romanization) typing method, which is the most widely used method for writing in Japanese on the computer (Rose, 2017).

Figure 5.1

Presentation of Screen Display in the AX Discrimination Task: When the First Stimulus Played (Panel A), and When the Second Stimulus Played (Panel B)[26]

A. When the first audio stimulus played:

B. When the second audio stimulus played:



[26] As can be seen in the screen displays, the expression *pattern tonali* "tonal patterns", which might not be easy for non-linguists, was used in the discrimination tasks. Recall, however, that participants had already received an explanation about the distinctive function of pitch-accent patterns prior to doing any tasks (see Section 3.2), so the expression was not new to them. In addition, before starting each discrimination task, participants completed the task familiarization phase, where this term was employed.

Participants were instructed to answer whether the underlined part in the first stimulus recording was *same* or *different* compared to that in the second stimulus recording. To respond, they had to press either the *F* key for *same* or the *J* key for *different* on their computer keyboard. These two keys were chosen, because the bumps found on both keys are designed to help users locate the correct keys without looking down. The response deadline was four seconds from the onset of the second stimulus. RTs (reaction times) were recorded from the onset of the second stimulus in an AX pair until the participant answered by pressing either *F* or *J* key within the response deadline. One second after they had answered or had missed the response deadline, the next fixation point was displayed on the screen. As in Shport (2011, 2016) and Silpachai (2020), participants were instructed to answer as quickly and accurately as they could; and if unsure, they were encouraged to make their best guess.

Unlike what had happened during the short practice, no feedback was provided at any time during the pretest (as in Burnham et al., 2015). The pretest lasted about 20 minutes, and participants were able to track their progress by looking at the progress bar (see Figure 5.1).

5.2.4.2. Discrimination Tasks: Posttest and Test of Generalization

The posttest was conducted on Day 3, after Training 3 (see chapters 3 and 4). It was identical to the pretest, except that the stimulus presentation order was randomized within each of the blocks.

Following the posttest, participants completed one test of generalization, on either Day 4 or Day 5, depending on the order of the identification task and the discrimination task, which was counterbalanced across participants (see Section 3.1).

The aim of the test of generalization was to assess whether participants could generalize to novel voices, namely those of M5 and F8. The procedure for the test of generalization was identical to that for the pretest and the posttest, except

that there were more trials (consequently, the number of blocks also increased). The terms used and the structure are summarized in Table 5.4 below.

As can be seen in Table 5.4, it consisted of 288 trials, organized into six blocks: three with 500 ms of interstimulus interval (ISI) and the other three with 1500 ms. The order of the three blocks with 500ms of ISI and the other three with 1500 ms was counterbalanced across participants. Trial order within each of three blocks based on ISI was randomized.

Table 5.4
Summary of the Test of Generalization: Terms and Structure

Stimuli type
3 target words (1 triplet of segmentally identical words x 3 pitch-accent patterns) embedded in a carrier sentence
Stimulus pairs per block
12 combinations in total: 144 trials (72 *same* trials and 72 *different* trials) (1) 4 combinations of stimulus pairs (familiar – novel voice combinations): F4 – F8: 12 trials (six *same* trials and six *different* trials); F4 – M5: 12 trials (six *same* trials and six *different* trials); M4 – F8: 12 trials (six *same* trials and six *different* trials); M4 – M5: 12 trials (six *same* trials and six *different* trials). Total: 48 trials (24 *same* trials and 24 *different* trials) (2) 4 combinations of stimulus pairs (novel – familiar voice combinations): F8 – F4: 12 trials (six *same* trials and six *different* trials); F8 – M4: 12 trials (six *same* trials and six *different* trials); M5 – F4: 12 trials (six *same* trials and six *different* trials);

M5 – M4: 12 trials (six *same* trials and six *different* trials).

Total: 48 trials (24 *same* trials and 24 *different* trials)

(3) 4 combinations of stimulus pairs (novel – novel voice combinations):

F8 – F8: 12 trials (six *same* trials and six *different* trials);

M5 – M5: 12 trials (six *same* trials and six *different* trials);

F8 – M5: 12 trials (six *same* trials and six *different* trials);

M5 – F8: 12 trials (six *same* trials and six *different* trials).

Total: 48 trials (24 *same* trials and 24 *different* trials)

Blocks

Participants could take an optional break between blocks and within each block, at the halfway point they took a fixed break of 10 seconds.

3 blocks with 500 ms ISI:

Block 1 (familiar – novel voice combinations): 48 trials;

Block 2 (novel – familiar voice combinations): 48 trials;

Block 3 (novel – novel voice combinations): 48 trials.

3 blocks with 1500 ms ISI:

Block 1 (familiar – novel voice combinations): 48 trials;

Block 2 (novel – familiar voice combinations): 48 trials;

Block 3 (novel – novel voice combinations): 48 trials.

Number of trials

288 trials = 48 trials x 6 blocks

As in the pretest described in the previous section, there was an optional break between blocks. Half way through each block, there was a fixed break of 10 seconds, but other than that, participants could not take any breaks within the block.

In addition to the stimuli used at the pretest (produced by two speakers, F4 and M4), stimuli recorded by two novel speakers, F8 and M5 (see Table 5.1), were also employed. This resulted in three possible voice combination categories: 1) familiar – novel voice combination; 2) novel – familiar voice combination; and 3) novel – novel voice combination. Each combination category had four possible speaker combinations. Specifically, in category 1) familiar – novel voice combination: F4 – F8, F4 – M5, M4 – F8, and M4 – M5; in category 2) novel – familiar voice combination: F8 – F4, F8 – M4, M5 – F4, and M5 – M4; in category 3) novel – novel voice combination: F8 – F8, M5 – M5, F8 – M5, and M5 – F8. As in the pretest, since each block had 48 trials, there were 12 trials for each speaker combination. These 12 trials for each speaker combination consisted of six *same* trials and six *different* trials, so that each block contained the same number of *same* trials and *different* trials, as in Burnham et al. (2015).

Identical to the pretest, since the target words presented three different pitch-accent patterns (1st-syllable accented, 2nd-syllable accented, and unaccented), nine possible stimulus pair combinations were generated (see Table 5.3). Among the nine combinations, three pairs were *same* and six pairs were *different*. For each speaker combination, therefore, in the case of *same* trials, the three *same* trials were presented twice, on the other hand, in the case of *different* trials, the six *different* trials were presented only once.

As during the pretest and the posttest, no feedback was provided at any time during the test of generalization (as in Burnham et al., 2015). The test of generalization lasted about 50 minutes, and participants were able to track their progress by looking at the progress bar (see Figure 5.1).

5.2.5. Discrimination Tasks: Analysis

A total of 30,720 trials[27] was performed by 64 participants (32 musicians and 32 non-musicians) in all tests (pretest, posttest, test of generalization), i.e., 162 trials x 3 tests x 64 participants.

The Gorilla software package (Anwyl-Irvine et al., 2020) logged each participant's responses, and their reaction times (henceforth: RTs) for each trial. Participants' responses were coded as 0 for correct or 1 for incorrect. RTs were measured as time of response occurring within the four seconds following the onset of the second stimulus presentation (see Section 5.2.4.1).

In order to exclude idiosyncratic responses, a cutoff value for minimal response latency was established. This was done by using Praat (Boersma & Weenink, 2021) to measure target word length (the target word was always the first word in stimuli, but these were produced by four talkers). The shortest target word length was 200 ms: this time lapse was thus used as the cutoff value and responses with impossibly fast reaction times (shorter than 200 ms) were removed.

Regarding RT data, high and low cutoff values were first set in order to identify and discard outliers. Timeouts were used as the high cutoff, whereas 350 ms was used as the low cutoff value. This low cutoff value was chosen in line with the reasoning of Abu El Adas and Levi (2022) and Laméris and Post (2023): they set 200 and 250 milliseconds as low cutoffs, respectively, because the stimuli were monosyllabic. However, the current research's target words were disyllabic (CVCV), the shortest one lasting 200 ms with its first syllable lasting about 100 ms. Recall, also, that Japanese pitch accent consists of a bitonal high-low accent implemented as a fundamental frequency (F0) peak near the end of the accented mora followed by a sharp F0 fall (Shport, 2015, 2016; Venditti, 2005, 2006). This means that even in the case of target words with the 1st-syllable accented pattern, the correct pattern only can be identified by listening to the second syllable as well

[27] In the test of generalization, three participants performed several trials more than once. Specifically, one musician performed one trial (trial number 48) twice; one non-musician performed two trials (trial number 95 and 96) twice; and one non-musician performed two trials (trial number 47 and 48) three times. This was possibly due to technical problems. Therefore, only their first attempt at the trials was included in the data analysis.

as the first syllable. Thus, in the present study the low cutoff was set by adding 100 ms—the approximate length of the first syllable of the shortest target word—to the low cutoff in the two aforementioned studies.

Subsequently, following Chan and Leung (2020) and Jiang (2012), for each participant, RT values beyond 2 standard deviations from the mean were considered as outliers and discarded. Approximately 10% of the data were affected by this treatment. Then, the RT data were log-transformed following Abu El Adas and Levi (2022), Chan and Leung and Laméris and Post (2020; 2023). As in Burnham et al. (2015), only RTs for correct responses to *different* AX pairs (AB or BA trials) were analyzed. These log RT values constitute a separate dependent variable.

The participants' responses were converted to d-prime (henceforth: d'): "a measure of sensitivity derived from proportion/percent correct data. This measure takes into account subject bias in responses by incorporating both accuracy (called 'hits') and false positives (called 'false alarms') in a single measure" (McGuire, 2010, p. 13). Following Signal Detection Theory (Macmillan & Creelman, 2005), d' = z (*Hit rate*) – z (*False alarm rate*) was calculated for each participant. *Hits* were defined as the number of correct responses when the two stimuli were *different* and the participant also responded that they were *different*. On the other hand, *false alarms* were defined as the number of incorrect responses when the two stimuli were *same* but the participant responded that they were *different*. These d' scores were used as the other separate dependent variable.

Data analyses were performed using R 4.3.2 (R Core Team, 2023). The d' scores were computed using the *psycho* package (*v0.6.1*; Makowski, 2018). In order to address the research questions, linear mixed-effects regression models were computed (dependent variable: log RT) using the *lme4* package (Bates et al., 2015). Model diagnosis (observation of residual qq-plots) was conducted using the *DHARMa* package (Hartig, 2022). For d' scores, however, a mixed ANOVA was conducted, using the *rstatix* package (Kassambara, 2023), based on the reasoning reported by Brown (2021) and Monaghan et al. (2021); for each of the tests (pretest, posttest, and the test of generalization), there was just one dependent variable measure per participant in the analysis.

5.3. Discrimination Tasks: Results

This section reports on results pertaining to RQ1-3. Since there were two dependent variables—perceptual sensitivity to differences between stimuli (when the two stimuli were *different*) and log RT—it falls into two parts. The first part presents the analysis of d' scores (a measure of participants' perceptual sensitivity to control for response biases), which were calculated using participants' raw accuracy scores (see Section 5.2.5). The second part provides the analysis of log RT data (a measure of participants' response latency).

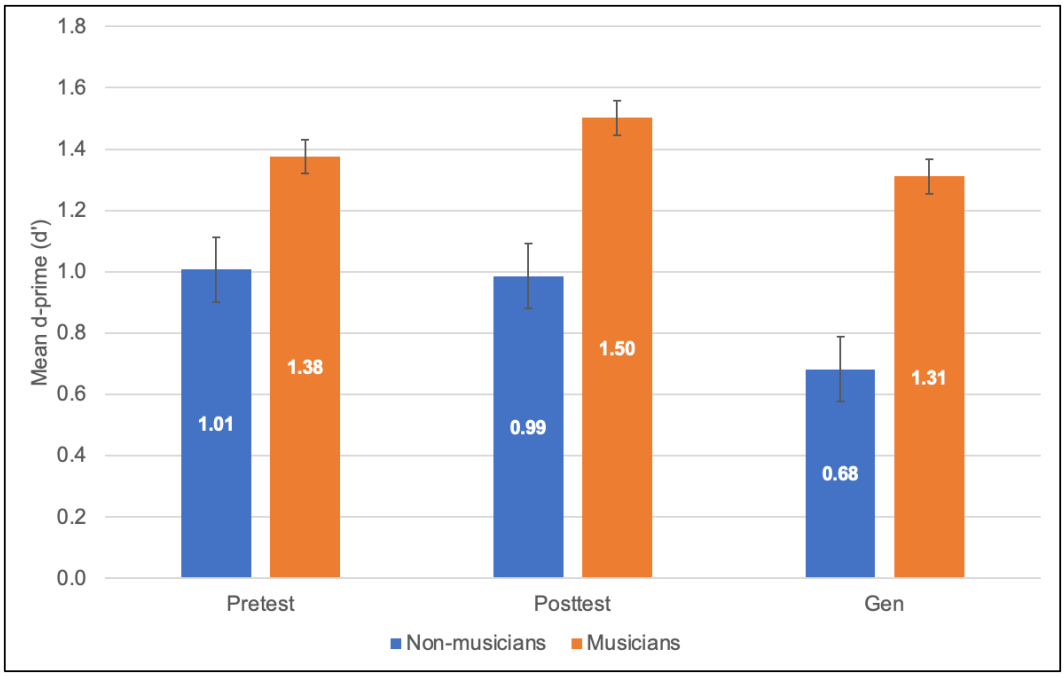
5.3.1. Discrimination Tasks: Effects on d' Scores of Musical Training and of Talker Variability

This section first provides an overview of participants' performance in the three tests: pretest, posttest, and the test of generalization (henceforth: Gen). Then, it presents the results of a mixed ANOVA to examine how participants' performance was influenced by the current research's predictors of interest (the effects of musical training, and of high variability, i.e., talker variability in training stimuli) and by ISI.

Figure 5.2 illustrates the mean d' scores for the three discrimination tests for musicians and non-musicians. Musicians are shown in orange, and non-musicians are shown in blue.

Figure 5.2

Mean d' Scores for the Three Discrimination Tests: Musicians vs. Non-Musicians



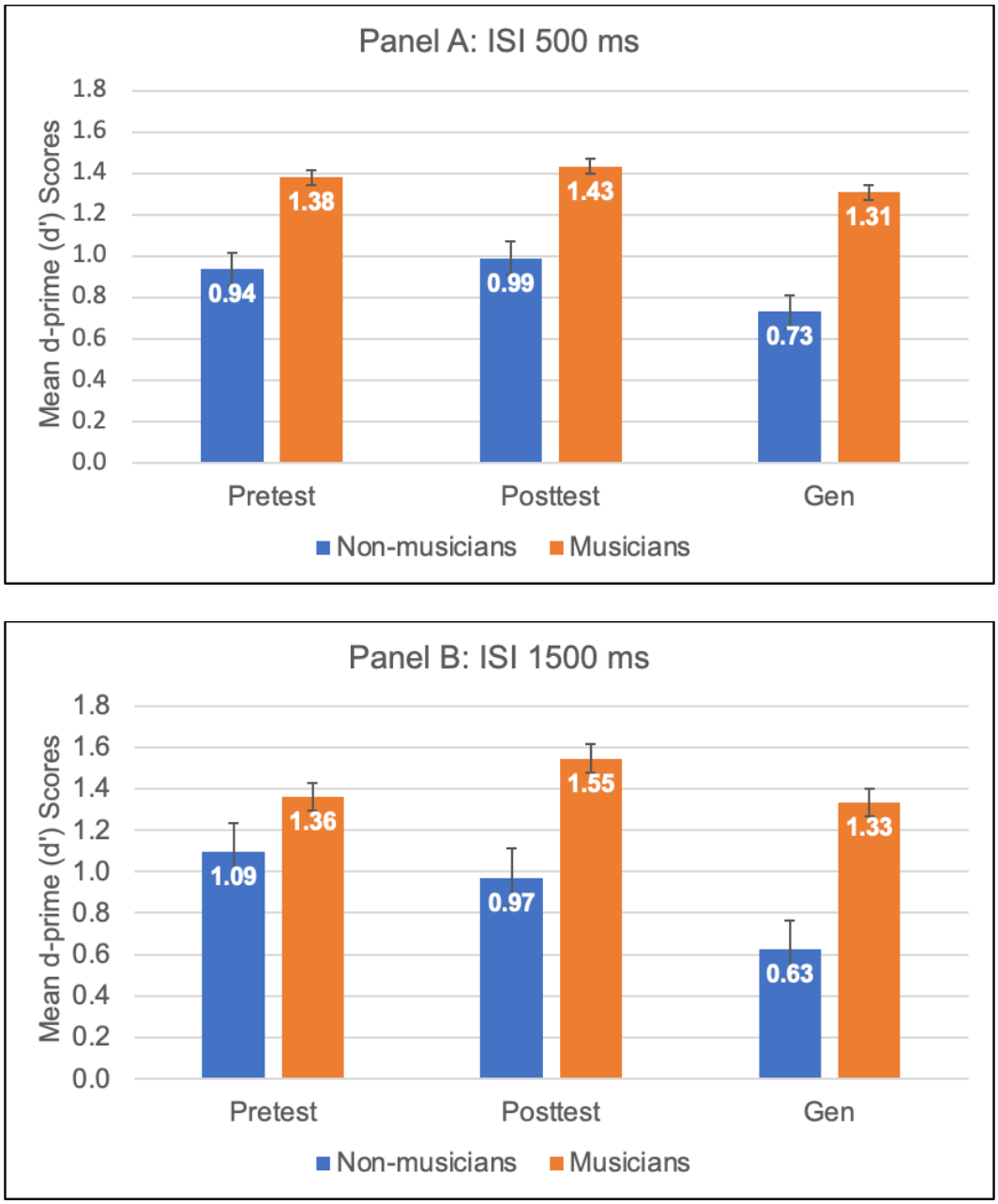
Note. Gen = Test of generalization

It can be observed clearly in Figure 5.2 that musicians outperformed non-musicians in all three tests. It is also noticeable that for musicians, posttest scores were higher than pretest scores, indicating a positive effect of the training. By contrast, for non-musicians, posttest scores were almost the same as pretest scores. At Gen, non-musicians scored considerably lower than at the pretest, while musicians' scores were only slightly lower at Gen than at the pretest. These results were not surprising, considering that novel talkers were introduced to Gen. Visual inspection of Figure 5.2 shows that the difference between musicians and non-musicians widened markedly after the pretest, especially at Gen.

Panel A and Panel B in Figure 5.3 represent the mean d' scores for the three discrimination tests for musicians and non-musicians, based on the two interstimulus intervals (ISIs: 500 ms and 1500 ms). Again, Musicians are shown in orange, and non-musicians are shown in blue.

Figure 5.3

Mean d' Scores for the Three Discrimination Tests Based on the Two ISIs (500 ms and 1500 ms): Musicians vs. Non-Musicians



Note. Gen = Test of generalization

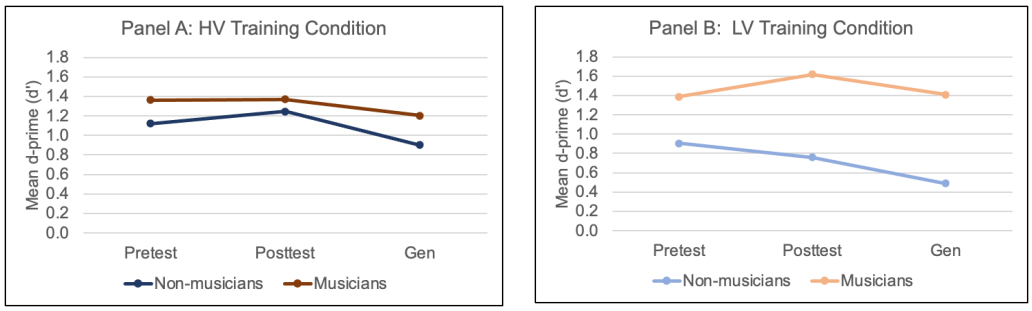
The two panels in Figure 5.3 depict almost the same trends as those shown in Figure 5.2, irrespective of the ISIs: musicians outperformed non-musicians in all three tests. While musicians showed pretest-posttest improvements in both panels,

non-musicians improved from the pretest to the posttest only when the ISI was 500 ms. A closer inspection of Figure 5.3 shows that the difference between musicians and non-musicians widened after the pretest, especially at Gen; and that it is more pronounced in Panel B than in Panel A.

So far, in response to RQ1 and RQ2, have been reported findings regarding the effect of musical training on perceptual learning of Japanese pitch accent. Now, in response to RQ3, this section focuses on the results regarding the effect of talker variability in the training stimuli (HV training condition vs. LV training condition) and its interaction with the effect of musical training.

Panel A and Panel B in Figure 5.4 display the mean before-and-after-training d' scores of musicians vs. non-musicians in the two training conditions (HV versus LV). Again, musicians are shown in orange, and non-musicians are shown in blue; to distinguish between the two training conditions, dark colors are used in Panel A (the HV training condition), while light colors are used in Panel B (the LV training condition).

Figure 5.4
Mean d' Score, Pretest-Posttest-Gen: Musicians vs. Non-Musicians Under the Two Training Conditions



Note. Gen = Test of generalization

What stands out in Figure 5.4 is that non-musicians and musicians achieved similar trends in the HV training condition (Panel A), whereas in the LV training condition there was a stark difference between non-musicians and musicians (Panel

B). It can also clearly be seen that non-musicians in the LV training condition showed deterioration in scores both from the pretest to the posttest and from the posttest to Gen. By contrast, non-musicians in the HV training condition showed similar trends to those of musicians, especially to musicians in the LV training condition; even though non-musicians in the HV training condition scored less in all tests than musicians in both training conditions.

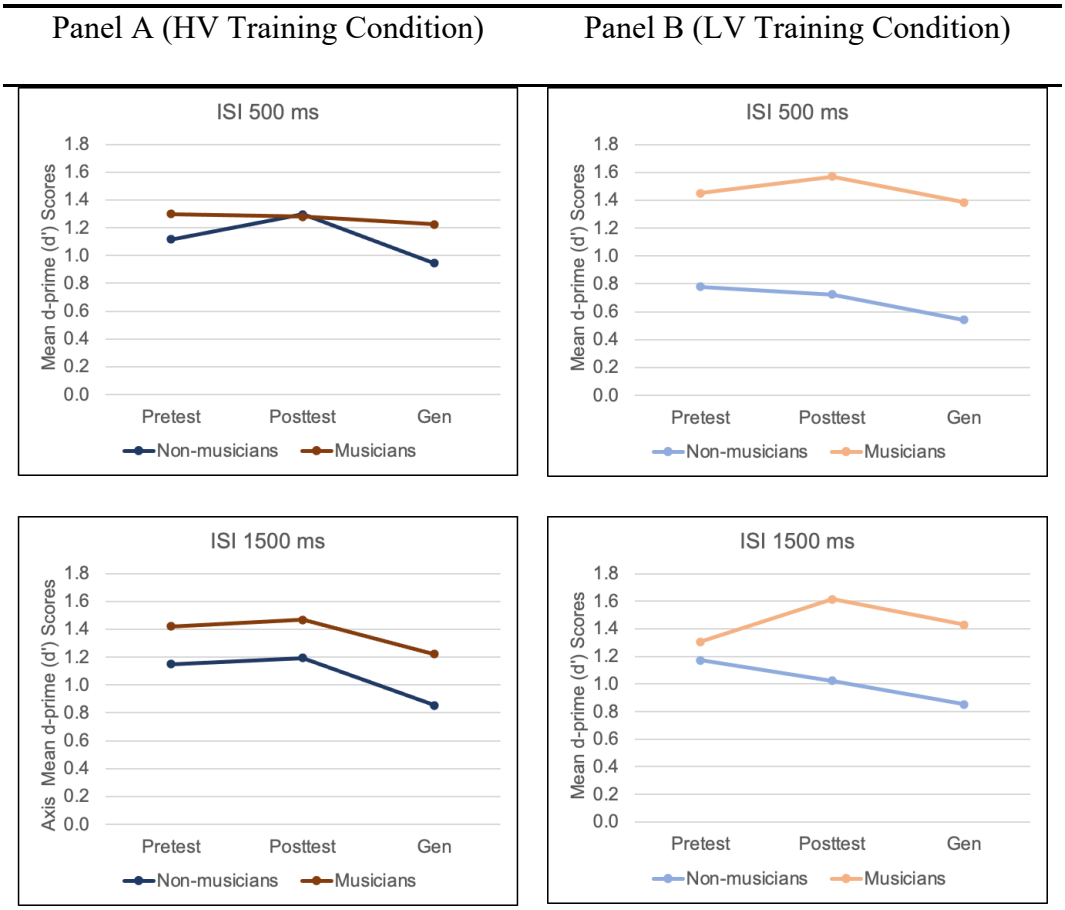
A more detailed look at Figure 5.4 reveals that from the outset, non-musicians in the LV training condition got lower scores than non-musicians in the HV training condition. This surprising result, suggesting individual differences between the two non-musician groups, is discussed in Section 7.2. Focusing on the musicians' results, those in the LV training condition (Panel B) showed a slight pretest-posttest improvement, while those in the HV training condition (Panel A) attained almost the same scores between the pretest and the posttest. As for Gen, musicians scored either nearly the same as at the pretest (for those in the LV training condition) or slightly lower than at the pretest (for those in the HV training condition), indicating that even musicians in both training conditions had difficulty in generalizing what they had learned to stimuli produced by novel talkers.

The findings indicate interaction of the effects of musical training and perceptual training condition (talker variability in training stimuli).

Figure 5.5 illustrates the pretest-posttest-Gen progression, for each ISI, in mean d' scores of musicians vs. non-musicians in the two training conditions (the HV training condition: Panel A; and the LV training condition: Panel B). Again, musicians are shown in orange, and non-musicians are shown in blue; to distinguish between the two training conditions, dark colors are used in Panel A, while light colors are used in Panel B.

Figure 5.5

Progression in Mean d' Scores per Training Condition (HV vs. LV) for Each ISI (500 ms vs. 1500 ms): Musicians vs. Non-Musicians



Note. Gen = Test of generalization

What is striking about Figure 5.5 is that, as in Figure 5.4, on one hand, non-musicians and musicians showed similar trends in the HV training condition (Panel A), especially when the ISI was set at 1500 ms. On the other hand, in the LV training condition (Panel B) there was a stark difference between non-musicians and musicians, especially when the ISI was set at 500 ms. The difference was appreciable from the outset in the case of ISI 500 ms, and in the case of ISI 1500 ms, it widened noticeably after training. In addition, only non-musicians in the LV training condition shows deteriorations from the pretest to the posttest and Gen regardless of the ISI, a trend different from those of the other three participant-training condition configurations.

An inspection of Figure 5.5 also shows that—irrespective of ISI—musicians performed better than non-musicians in all three tests, although in the HV training condition, non-musicians and musicians scored almost the same at the posttest at the ISI of 500 ms. Moreover, for all participants, Gen scores were either lower than or almost the same as pretest scores. This indicates that Gen was difficult for all participants: at Gen-1 the stimuli produced by novel talkers were introduced.

A closer look at Figure 5.5 reveals that the pretest difference between non-musicians in the two training conditions (see also Figure 5.4), is due to their results at the ISI of 500 ms. As mentioned earlier, this unexpected result is discussed in Section 7.2.

Again, the findings illustrated in Figure 5.5 suggest interaction of the effects of musical training, perceptual training condition (HV versus LV), and ISI.

To investigate these effects on perceptual learning of Japanese pitch accent, participants' d' scores were examined using a mixed ANOVA with Test (pretest, posttest, and Gen) and ISIs (500 ms and 1500 ms) as the within-subject factors and Category (non-musician and musician) and Training condition (HV and LV) as the between-subject factors. Extreme outliers, found using the *rstatix* package (Kassambara, 2023), made up about 3% of the data, and were excluded from the analysis.

The mixed ANOVA yielded a significant effect of Category, $F(1, 58) = 15.525$, $p < 0.001$, Test, $F(2, 116) = 11.28$, $p < 0.001$. There were no significant effects of either ISI or Training condition. The mixed ANOVA also revealed a significant interaction only between Category and Training condition $F(1, 58) = 4.64$, $p = 0.035$.

To further examine these significant effects, the simple two-way interaction between Category and Training condition at each level of Test was first computed, applying a Bonferroni adjustment. The simple two-way interaction yielded that there were significant simple two-way interactions between Category and Training condition at the posttest, $F(1, 120) = 9.938$, $p = 0.002$, and at Gen, $F(1, 120) = 6.808$, $p = 0.01$, but not at the pretest, $F(1, 120) = 1.478$, $p = 0.226$.

Then, two one-way models at each level of Test were run, applying a Bonferroni adjustment. The first one-way model was to investigate the effect of Training condition on d' scores at every level of Test and Category. It revealed that the simple simple main effect of Training condition on d' scores for non-musicians was significant only at the posttest, $F(1, 60) = 6.756$, $p = 0.012$, and not at the pretest, $F(1, 60) = 2.091$, $p = 0.153$, nor at Gen, $F(1, 60) = 2.091$, $p = 0.032$[28]. For musicians, no significant simple simple main effect of Training condition on d' scores was found: at the pretest, $F(1, 60) = 0.061$, $p = 0.806$, at the posttest, $F(1, 60) = 3.489$, $p = 0.067$, and at Gen, $F(1, 60) = 2.929$, $p = 0.092$. These results indicate that, even though there was no significant difference at the pretest between non-musicians, the HV training was arguably more beneficial for non-musicians than the LV training at the posttest, but that for musicians there was no significant difference between the two training conditions.

The other one-way model was to investigate the effect of Category on d' scores at every level of Test and Training condition. It yielded a significant simple simple main effect of Category on d' scores for the LV training condition at the pretest, $F(1, 66) = 14.829$, $p < 0.001$, at the posttest, $F(1, 66) = 27.8$, $p < 0.001$, and at Gen, $F(1, 66) = 29.196$, $p < 0.001$, and for the HV training condition at Gen, $F(1, 54) = 5.465$, $p = 0.023$. But no significant simple simple main effect of Category on d' scores was found for the HV training condition at the pretest, $F(1, 54) = 3.041$, $p = 0.087$, or at the posttest, $F(1, 54) = 0.205$, $p = 0.653$. Thus, this analysis confirmed the trends observable in Figure 5.4.

Grouping the data by Category and Training condition, all simple simple pairwise comparisons between Test (pretest, posttest, and Gen) were run with Bonferroni adjustment. These comparisons yielded that, while non-musicians in the LV training condition performed worse at Gen than at the pretest ($p = 0.006$), those in the HV training condition performed worse at Gen than at the posttest ($p = 0.012$). No other significant differences were found.

[28] Note that the Bonferroni adjustment applied led to statistical significance being accepted at the $p < 0.025$ level.

To sum up, statistical analyses revealed significant effects for musical training and pitch accent perceptual training, as well as for the interaction between musical training and perceptual training condition (high variability versus low variability). The findings indicate that, on one hand, for non-musicians, the HV training condition was far more effective at the posttest, because only HV training led to a pretest-posttest improvement, although this improvement was not significant. The benefit of the HV training condition for non-musicians was also confirmed by the results showing that, in the HV training condition, there were no significant differences between non-musicians and musicians at the pretest or the posttest. In the LV training condition, in contrast, non-musicians performed significantly worse than musicians in all tests, and showed after-training deteriorations. For musicians, on the other hand, there was no sharp difference in the effectiveness between the two training conditions.

5.3.2. Discrimination Tasks: Effects on Log RT of Musical Training and of Talker Variability

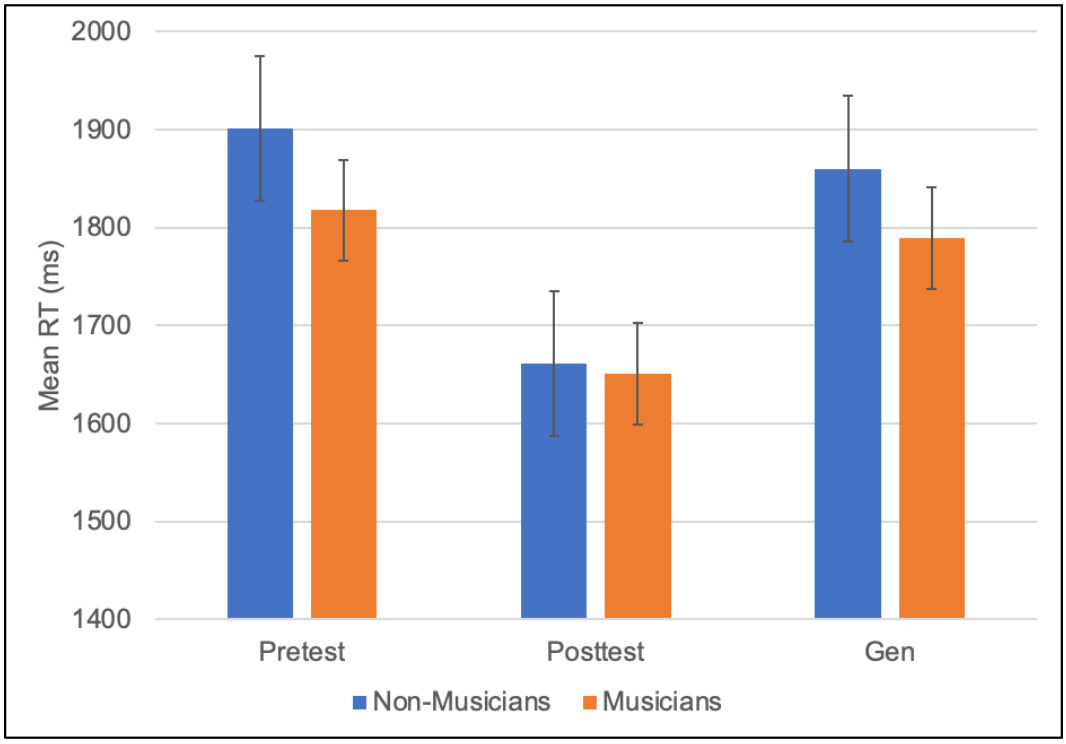
This section first presents an overview of participants' performance in the three tests: pretest, posttest, and Gen. Then, it moves on to a mixed-effects model analysis to examine how participants' performance was influenced by the current research's predictors of interest (the effects of musical training, and of high variability, i.e., talker variability in training stimuli) and by ISI.

Recall that, following Burnham et al. (2015), the log RTs analyzed and reported here were only for *correct* responses to *different* AX pairs (AB or BA trials), since the discrimination tasks in this study aimed at assessing whether participants were able to distinguish Japanese pitch-accent patterns.

To better understand and facilitate interpretation of the overall results, in the figures provided in this section, log RT values are converted to milliseconds, even though log RT was used to express the dependent variable in the mixed-effects model analysis (see Section 5.2.5).

Figure 5.6 displays mean RTs in milliseconds for the three discrimination tests for musicians and non-musicians. Musicians are shown in orange, and non-musicians are shown in blue.

Figure 5.6
Mean Reaction Times (RTs) for the Three Discrimination Tests: Musicians vs. Non-Musicians



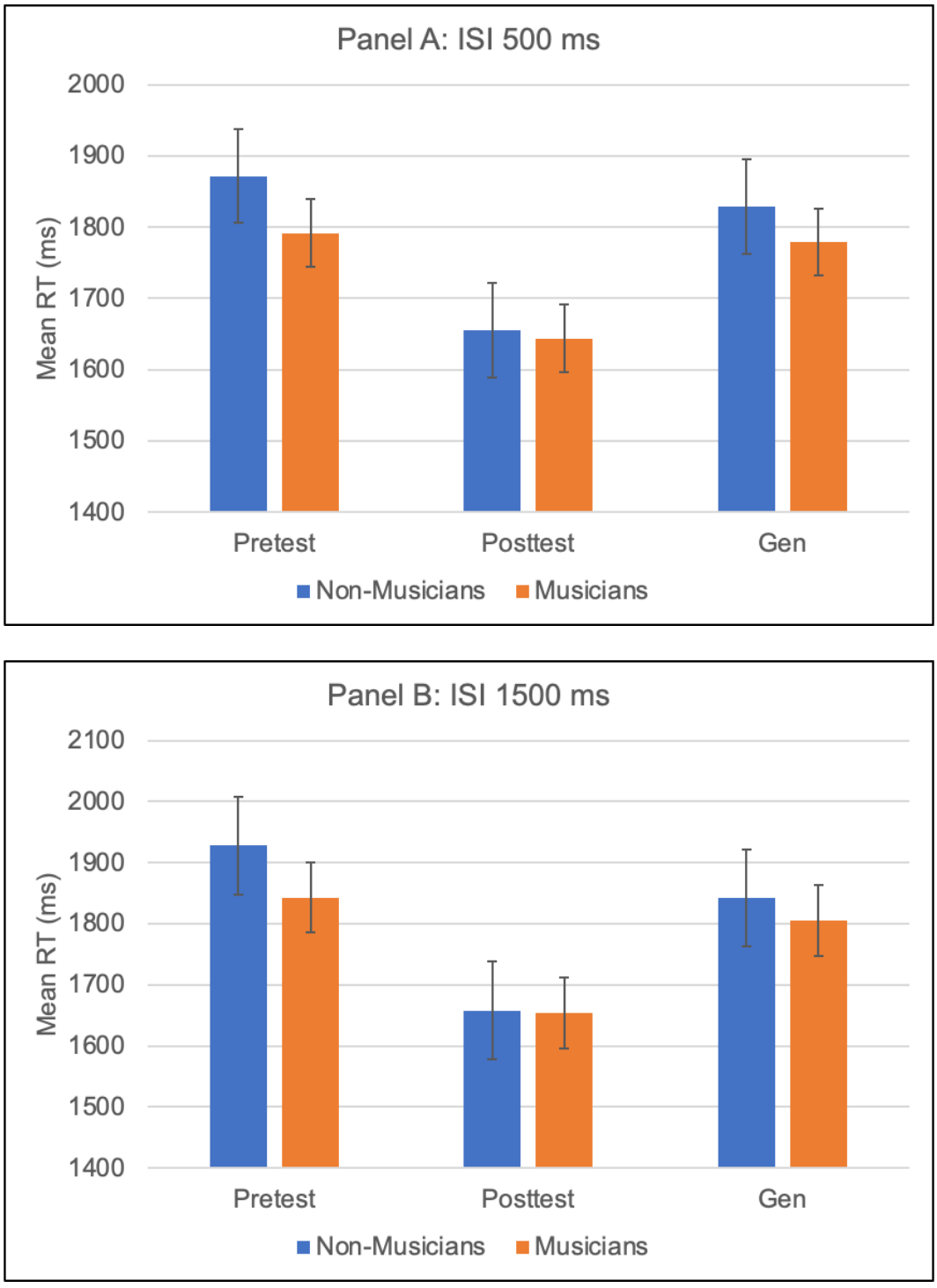
Note. Gen = Test of generalization, ms = milliseconds

What stands out in Figure 5.6 is that both non-musicians and musicians were faster at the posttest than at the pretest and at Gen. In addition, musicians responded faster than non-musicians at the pretest and at Gen, but at the posttest, there was only a slight difference between non-musicians and musicians.

Panel A and Panel B in Figure 5.7 represent the mean RTs in milliseconds for the three discrimination tests for musicians and non-musicians, for the two interstimulus intervals (Panel A, ISI:500 ms and Panel B, ISI:1500 ms). Again, musicians are shown in orange, and non-musicians are shown in blue.

Figure 5.7

Mean Reaction Times (RTs) for the Three Discrimination Tests for the Two ISIs (500 ms and 1500 ms): Musicians vs. Non-Musicians



Note. Gen = Test of generalization, ms = milliseconds

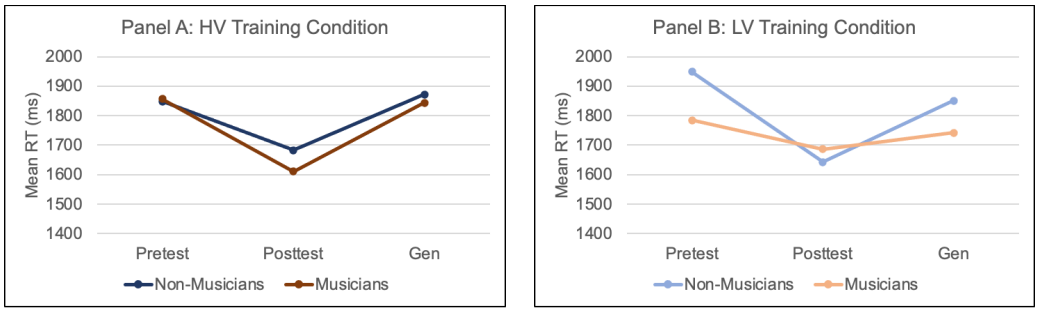
A comparison between the two panels in Figure 5.7 shows almost the same trends seen in Figure 5.6. Apparently, the ISIs did not affect participants' performance in any tests. Irrespective of ISI, both non-musicians and musicians responded faster at the posttest than at the pretest or at Gen. Musicians were faster than non-musicians at the pretest and at Gen, but at the posttest, non-musicians' response velocity was almost the same as musicians'. Closer inspection of Figure 5.6 reveals that when the ISI was set to 1500 ms, both non-musicians and musicians responded slightly more slowly, both at the pretest and at Gen, than when the ISI was set to 500 ms.

The findings reported so far have been in response to RQ1 and RQ2, which were to assess the effect of musical training on perceptual learning of Japanese pitch accent. Now, in response to RQ3, what follows focuses on the results regarding the effect of talker variability in the training stimuli (HV training condition vs. LV training condition) and its interaction with the effect of musical training.

Panel A and Panel B in Figure 5.8 depict the pretest-posttest-Gen progression in overall mean RTs of musicians vs. non-musicians in the two training conditions (the HV training condition and the LV training condition). Again, musicians are shown in orange, and non-musicians are shown in blue; to distinguish between the two training conditions, dark colors are used in Panel A (the HV training condition), while light colors are used in Panel B (the LV training condition).

Figure 5.8

Mean Reaction Times (RTs), Pretest-Posttest-Gen: Musicians vs. Non-Musicians Under the Two Training Conditions



Note. Gen = Test of generalization, ms = milliseconds

What is striking about Figure 5.8 is that musicians and non-musicians showed similar trends in the HV training condition, whereas in the LV training condition the trends differed between musicians and non-musicians. Indeed, in the LV training condition, RT progression for musicians is quite flat, although RT is slightly faster at the posttest than at the pretest and at Gen. By contrast, RT progression for non-musicians was steep, dropping dramatically from the pretest to the posttest and rising sharply from the posttest to Gen.

A more careful look at Figure 5.8 shows that in the HV training condition, while musicians and non-musicians performed with almost the same velocity at the pretest and at Gen, at the posttest musicians were faster than non-musicians. In the LV training condition, on the other hand, while musicians were faster than non-musicians at the pretest and at Gen, the situation was surprisingly reversed at the posttest: non-musicians responded slightly faster than musicians.

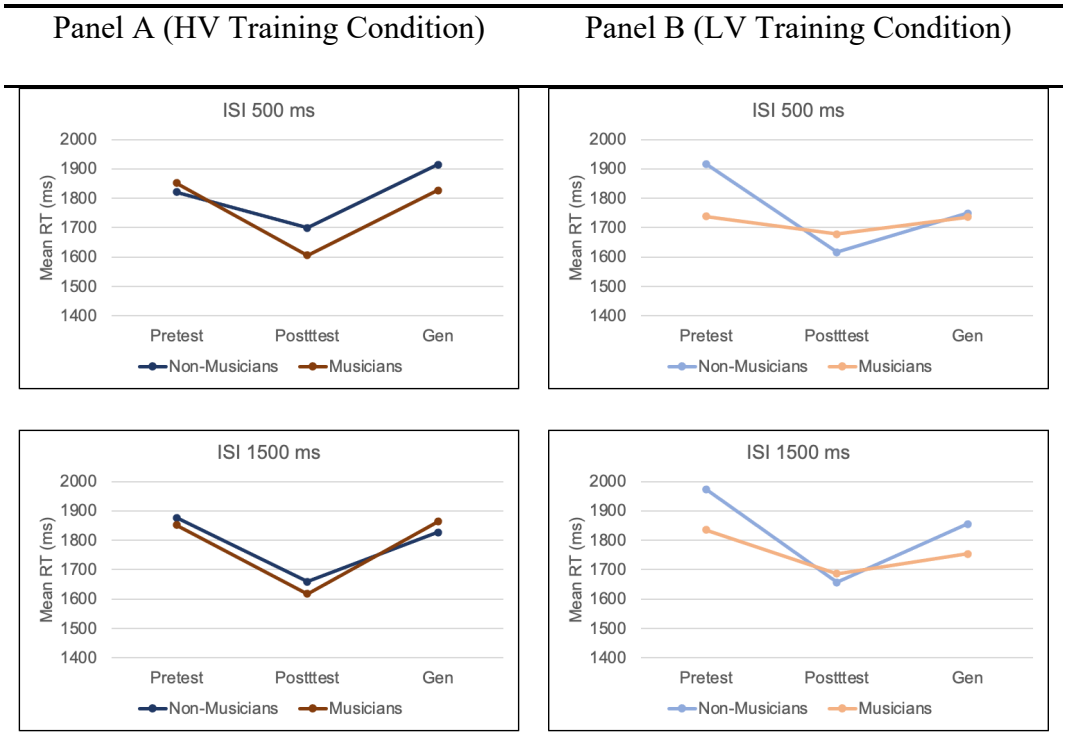
The findings indicate interaction of the effects of musical training and perceptual training condition (talker variability in training stimuli).

Figure 5.9 displays the progression, for the two ISIs (500 ms and 1500 ms), in mean RT of musicians vs. non-musicians in the two training conditions (the HV training condition: Panel A; and the LV training condition: Panel B). Again, musicians are shown in orange, and non-musicians are shown in blue; to distinguish

between the two training conditions, dark colors are used in Panel A, while light colors are used in Panel B.

Figure 5.9

Progression of Mean Reaction Times (RTs) per Training Condition (HV vs. LV) for Each ISI (500 ms vs. 1500 ms): Musicians vs. Non-Musicians



Note. Gen = Test of generalization, ms = milliseconds

A comparison between Panel A and Panel B in Figure 5.9 shows almost the same trends seen in Figure 5.8: whereas musicians and non-musicians showed similar progressions in the HV training condition, in the LV training condition the progressions differed between musicians and non-musicians.

Overall, ISI appears to have had no or little effect on participants' performance. However, Figure 5.9 shows that in the HV training condition, the difference between non-musicians and musicians was more subtle for the 1500 ms ISI than for the 500 ms ISI. By contrast, in the LV training condition, the differences between non-musicians and musicians widened at the pretest and at Gen at the 1500

ms ISI. In addition, in the LV training condition, non-musicians responded slightly faster than musicians at the posttest irrespective of ISI.

Again, the findings suggest interaction of the effects of musical training and perceptual training condition (talker variability in training stimuli), but little if any interaction of the ISIs with either of these.

To investigate the effects of musical training, of LV or HV training condition, and of ISIs on perceptual learning of Japanese pitch accent, a linear mixed-effects regression model analysis was conducted. Participants' log RTs for each trial were used as the dependent variable. The model contained fixed effects for: *ISI* (500 ms as the reference level, 1500 ms: treatment-coded); *category* (non-musician as the reference level, musician: treatment-coded); *training condition* (HV as the reference level, LV: treatment-coded) and *test* (pretest as the reference level, posttest, Gen: treatment coded). Participants and item (stimuli) were included as random effects with by-participant and by-item varying intercepts for the effects under investigation. Attempts were made to include by-participant and by-item varying slopes, but this led to convergence issues.

The final model was coded in R as: lmer(log RT ~ ISI + Category + Test + Training_condition + Category : Test + Category : Training_condition + Category : ISI + Test : Training_condition + Test : ISI + Training_condition : ISI + Category : Test: Training_condition + Category : Test : ISI + Test : Training_condition : ISI + Training_condition : ISI : Category + Category : Test : Training_condition : ISI + (1|Participant_Public_ID) + (1|Item), data = data, REML = FALSE).

The *afex* package (Singmann et al., 2023) was used to perform likelihood ratio tests for all fixed effects. The results are summarized in Table 5.5.

Table 5.5
Likelihood Ratio Tests for All Fixed Effects: Summary of Results

Fixed effect	Result
ISI	$\chi^2(1) = 0.56$, $p = 0.453$
Category	$\chi^2(1) = 0.02$, $p = 0.879$
Test	$\chi^2(2) = 114.29$, $p < 0.001$*
Training condition	$\chi^2(1) = 0.05$, $p = 0.818$
Category × Test	$\chi^2(2) = 6.92$, $p = 0.031$*
Category × Training condition	$\chi^2(1) = 0.04$, $p = 0.845$
ISI × Category	$\chi^2(1) = 0.36$, $p = 0.548$
Test × Training condition	$\chi^2(2) = 13.12$, $p = 0.001$*
ISI × Test	$\chi^2(2) = 4.14$, $p = 0.126$
ISI × Training condition	$\chi^2(1) = 0.37$, $p = 0.541$
Category × Test × Training condition	$\chi^2(2) = 11.18$, $p = 0.004$*
ISI × Category × Test	$\chi^2(2) = 1.27$, $p = 0.529$
ISI × Test × Training condition	$\chi^2(2) = 0.11$, $p = 0.944$
ISI × Category × Training condition	$\chi^2(1) = 0.81$, $p = 0.369$
ISI × Category × Test × Training condition	$\chi^2(2) = 4.58$, $p = 0.101$

Note. * = significant effect

Table 5.5 shows that there was a significant effect of test on participants' log RT values. In addition, there were significant interactions between: (1) category and test; (2) test and training condition; and (3) category, test, and training condition.

In order to assess the interactions in more detail, Bonferroni-corrected multiple comparisons were calculated using the *emmeans* package (Lenth, 2023). These results are presented in Table 5.6.

Table 5.6
Interactions Between Training Condition, Musician/Non-Musician Category, and Discrimination Test: Results of Multiple Comparisons

Contrast	Estimate	SE	z.ratio	p
Contrasts for Test:				
HV – Non-Musicians:				
Pretest – Posttest	0.05	0.01	4.82	<0.001
Posttest – Gen	-0.05	0.01	-3.96	<0.001
HV – Musicians:				
Pretest – Posttest	0.06	0.01	5.81	<0.001
Posttest – Gen	-0.05	0.01	-3.94	<0.001
LV – Non-Musicians:				
Pretest – Posttest	0.08	0.01	7.84	<0.001
Pretest – Gen	0.04	0.01	3.69	<0.001
Posttest – Gen	-0.04	0.01	-3.35	0.002
LV – Musicians:				
Pretest – Posttest	0.02	<0.01	2.63	0.026

Note. For brevity, only significant comparisons are presented. Gen = Test of generalization.

As can be seen in Table 5.6, multiple comparisons revealed that, for all participants (irrespective of the category and the training condition), there were

significant differences in the log RT values between the pretest and the posttest, which was identical to the pretest. This indicates that both HV and LV training led to faster RTs at the posttest than at the pretest. At Gen, however, except for musicians in the LV training condition, participants responded significantly slower than at the posttest. Looking back to Figure 5.8, it can be observed that only musicians in the LV training condition showed an almost flat progression: their RTs were slightly faster at the posttest than at the pretest, but RTs were slightly slower at Gen than at the posttest. Indeed, for musicians in the LV training condition, no significant differences in log RTs in were found between the pretest and Gen or between the posttest and Gen. Interestingly, only for non-musicians in the LV training condition was the pretest-Gen difference significant, with non-musicians RTs at Gen being significantly faster than at the pretest, even though at Gen novel-talker stimuli were introduced. It is worth mentioning that multiple comparisons for category and for training condition did not find any significant differences in log RT.

5.4. Discrimination Tasks: Discussion

The present research aimed to examine whether musical training influences perceptual learning of Japanese pitch accent by native speakers of a non-tone language (Italian), who had no experience of Japanese. To this end, Italian non-musicians and musicians engaged in Japanese pitch-accent discrimination and identification tasks before and after undergoing perceptual training conducted following a HVPT paradigm. As the title indicates, the discussion in this section focuses on the discrimination tasks (see Chapter 4 for the identification tasks).

A second aim was to investigate the effect of talker variability in training stimuli on perceptual learning of Japanese pitch accent. To achieve this goal, musicians and non-musicians were randomly assigned to a HV training condition (stimuli produced by four talkers), or a LV training condition (stimuli produced by one talker).

Finally, an attempt was made to assess the effect of absolute pitch on perceptual learning of Japanese pitch accent. The results for this (i.e., RQ4) are reported in Chapter 6.

Before beginning the discussion of the findings, it is worth reiterating that for the discrimination tasks, two dependent variables were analyzed: d' scores (participants' perceptual sensitivity to differences between stimuli when the two stimuli were *different*), and log RT (a measure of response latency). Based on Burnham et al. (2015), two interstimulus intervals were employed in the current work: 500 ms and 1500 ms. Their effect was explored along with the effects of musical training and perceptual training condition.

The findings of the current research are now discussed in the light of the research questions (RQ1-3) and previous studies. Note that the current research's predictors of interest (the effects of musical training and of talker variability in perceptual training stimuli) are the same as for the identification tasks (Chapter 4). To avoid repetition, this section therefore does not address those implications of the results for these predictors of interest that have already been discussed in Chapter 4, but limits itself to a discussion of the results specific to the discrimination tasks.

RQ1 addressed whether Italian musicians would outperform Italian non-musicians in discriminating between Japanese pitch-accent patterns.

For d' scores, the findings were consistent with the predictions, i.e., that overall, musicians would outperform non-musicians in distinguishing Japanese pitch-accent patterns. This accords with that of numerous perception studies (Alexander et al., 2005; Burnham et al., 2015; Chang et al., 2016; Chen et al., 2020; Delogu et al., 2010; Gottfried, 2007; Gottfried & Xu, 2008; Götz et al., 2023; Kirkham et al., 2011; Lee et al., 2014; Lee & Hung, 2008; Marie et al., 2011; Mok & Zuo, 2012), which have shown, overall, the positive effect of musical experience/training on lexical tone perception by native non-tone language speakers without any experience of the target tone language. The finding of the present research is also in line with that of Golob's study (2003), in which Slovenian musicians without any experience of Japanese perceived Japanese pitch accent better than Slovenian non-musicians who were learners of Japanese. As for the ISIs,

neither significant main effect nor significant interaction were found. This is consistent with Burnham et al. (2015).

The results for log RT data, on the other hand, indicated that musical training did not provide an overall speed advantage when perceiving pitch differences. Recall that, following Burnham et al. (2015), the log RTs analyzed and reported in the current research were only for *correct* responses to *different* AX pairs (AB or BA trials). Unlike in Burnham et al. (2015)—who used bare RT data—the Log RT data for ISIs in this study showed neither significant main effect nor significant interaction. Aside from Burnham et al. (2015), the present log RT results are in disagreement with some studies (Alexander et al., 2005; Lee & Hung, 2008), but in line with others (Lee et al., 2014; Marie et al., 2011; Mok & Zuo, 2012).

It is worth focusing on the difference in RT results between this research and Burnham and colleagues' study (2015), because the current research drew inspiration from their study. Burnham et al. found that musicians were faster than non-musicians on the Thai lexical tone discrimination task, and that all participants (musicians with/without absolute pitch and non-musicians) generally took less time in discriminating Thai lexical tones at the ISI of 1500 ms than at the ISI of 500 ms.

There are several possible reasons for the discrepancy between this study's results and those of Burnham et al. (2015). A very likely explanation is the difference in parameters for the *musician* category. In Burnham and colleagues' work, half of the musicians possessed absolute pitch. The musicians with absolute pitch not only performed significantly better, but also responded significantly faster at the ISI of 1500 ms (but not at 500 ms) than musicians without absolute pitch. Thus, their results indicated that musicians with absolute pitch had an advantage in speed and accuracy over musicians without absolute pitch. To anticipate the results of the absolute pitch test (reported in Chapter 6), none of the musicians who participated in the current research had absolute pitch.

Another possible explanation is the difference in stimuli. While the stimuli used in Burnham and colleagues were monosyllables each carrying one of five Thai lexical tones, the stimuli employed in the current research were Japanese disyllabic target words embedded in a carrier sentence.

Lastly, the experimental design regarding the ISIs was different. While the ISI was a between-participant variable in Burnham et al., in this study it was a within-participant variable, considering the small sample size and the fact that there were already two predictors of interest (the effects of musical training and talker variability).

As mentioned in Section 4.4.1, previous studies have largely converged to indicate the facilitative effect of musical experience/training on lexical tone perception by native non-tone language speakers, albeit to varying degrees. However, as discussed in Section 2.3.3, training studies have reported mixed results regarding additional musical advantage in the outcome of perceptual learning of lexical tone, and to the best of the author's knowledge, there has been no work examining the effect of musical experience/training on perceptual learning of Japanese pitch accent.

RQ2 assessed whether the difference between musicians and non-musicians in the ability to discriminate Japanese pitch-accent patterns would decrease or increase after training.

For d' scores, on one hand, in line with the prediction, the difference between musicians and non-musicians widened after training, especially at Gen. Musicians showed a pretest-posttest improvement, whereas non-musicians' performance at the posttest was almost the same as at the pretest. At Gen, while musicians performed similarly to at the pretest, non-musicians performed worse than at the pretest. Remember that at Gen, stimuli produced by novel takers were introduced, whereas the posttest had been identical to the pretest. The results regarding an advantage for musicians are dissimilar to those of some training studies (Dittinger et al., 2016; Tong & Tang, 2016; Wayland et al., 2010; Zhao & Kuhl, 2015), but in accord with those of other studies (Cooper & Wang, 2012; Maggu et al., 2018; P. C. M. Wong & Perrachione, 2007). As for the ISIs, as mentioned above, neither significant main effect nor significant interaction were found. This is consistent with Burnham et al. (2015).

For log RT data, on the other hand, musicians exhibited a speed advantage only at the pretest and at Gen. At the posttest, in contrast, musicians' response

velocity was comparable to that of non-musicians. Interestingly, non-musicians and musicians showed similar trends: while they were faster at the posttest than at the pretest, they were slower at Gen than at the posttest or at the pretest (although only slightly slower at the pretest). As for the ISIs, neither significant main effect nor significant interaction were found. This is inconsistent with the finding of Burnham et al. (2015). Note, however, that Burnham and colleagues' study did not employ a pre/posttest design. Moreover, to the best of the author's knowledge, none of the training studies mentioned in the previous paragraph, which investigated the effect of musical training/experience, either measured RTs or manipulated ISIs.

The present research's findings for log RT data imply that there was a limited benefit for musicians. However, given that no other studies have investigated this topic, future work is required to gain a better understanding of the relationship between the effect of musical training, the ISIs, and RTs.

So far, in response to RQ1 and RQ2, the effect of musical training on perceptual learning of Japanese pitch accent has been discussed. The remainder of this section focuses on RQ3, which was set to investigate the effect of talker variability in training stimuli (HV training condition versus LV training condition) and its interaction with the effect of musical training. Specifically, RQ3 addressed whether or not the HV training condition would be more beneficial for Italian musicians to discriminate Japanese pitch-accent patterns compared to non-musicians.

In terms of d' scores musicians and non-musicians showed different trends. On one hand, musicians in the two training conditions improved in pitch-accent pattern discrimination in a similar manner. Indeed, no significant differences were found between musicians in the two training conditions. On the other hand, in the case of non-musicians, the difference in training condition led to differences in their results.

Although, at the pretest, no significant difference was found between non-musicians in the two training conditions, those in the HV training condition performed significantly better than those in the LV training condition at the posttest. Additionally, only non-musicians in the HV training condition showed a pretest-

posttest improvement, although this improvement was not significant. In the HV training condition, non-musicians' d' scores were lower at Gen than at the pretest or at the posttest. Given that at Gen, novel-talker stimuli were introduced, and that even musicians in both training conditions scored less at Gen than at the posttest, it it is not surprising that non-musicians in the HV training condition performed worse at Gen than at the posttest. Non-musicians in the LV training condition also scored worst at Gen than at the other two tests. But what is peculiar to non-musicians in the LV training condition is that they showed a deterioration in scores from pretest to posttest and from posttest to Gen.

These results indicate not only that the HV training condition was more effective for non-musicians, but also that the LV training condition was detrimental to them. The benefit of the HV training condition for non-musicians was also confirmed by the results showing that, in the HV training condition, there were no significant differences between non-musicians and musicians except at Gen. In the LV training condition, in contrast, non-musicians scored significantly lower than musicians in all tests.

It bears mentioning that Shport's studies (2011, 2016)—on which the methodology for the present study was based—employed only identification tasks in all tests and in training, while her training was comparable in terms of talker variability to the HV training condition in the present study. The present results suggest that Shport's training method has a limited effect for the discrimination task, in that non-musicians in the HV training condition (who were comparable to Shport's trainees) showed a non-significant pretest-posttest improvement in discriminating Japanese pitch accent. In addition, for all participants (irrespective of the category and the training condition), no significant pretest-Gen improvements were found. Rather, except for musicians in the LV training condition, participants showed pretest-Gen deteriorations albeit to varying degrees. This may be because the type of task is different between the training session and the test. It may also be due to the brevity of the training sessions (approximately one hour in total). In order to enhance the ability of participants to generalize what they have learned to the stimuli produced by novel talkers, it might be useful to extend the length of training. It would also be interesting to include discrimination

tasks among the tasks used during training. Future work is needed to assess whether these measures would be effective.

As regards the ISIs, again, neither significant main effect nor significant interaction were found. As mentioned above (in response to RQ2), future research is required to better understand the relationship between the effect of musical training, the ISIs, training conditions, and d' scores.

Surprisingly, in the case of the two ISIs, the log RT and the d' score results were similar. Furthermore, none of the differences in log RT data between musicians and non-musicians in the two training conditions were significant, for any of the tests. These results indicate that both training conditions induced participants to respond faster at the posttest than at the pretest.

Comparing them, the log RT results and the d' score results indicate that as d' scores increased, corresponding RT values tended to drop. Conversely, as the d' score decreased, the corresponding RT value also tended to rise. However, non-musicians in the LV training condition exhibited a speed-sensitivity trade-off at the posttest: their d' scores were lower at the posttest than at the pretest, but surprisingly, at the posttest, they responded faster than musicians under the same training condition (LV). This also suggests that the LV training condition is detrimental to non-musicians.

To sum up, the d' score results and the log RT results show different trends: while musicians had a clear advantage in d' scores, their advantage in terms of log RT data was limited. In addition, as regards the effect of training condition, d' score results indicated no significant differences between the two training conditions for musicians, while for non-musicians, the HV training condition was beneficial. By contrast, the log RT results imply that the effect of HV training was comparable to that of the LV training for both groups. As regards the ISIs, neither significant main effect nor significant interaction were found for d' or for log RT results.

Before concluding this section, one limitation to the analysis of the discrimination tasks needs to be acknowledged (see Section 7.2 for other limitations to the present study).

Recall that in the current study, the stimuli used in the discrimination tasks were recorded by different speakers. This was done not only to approximate everyday settings but also to make the tasks difficult, considering that Sakamoto (2011) did not find any significant difference between English-speaking participants and native Japanese speakers in discrimination task performance.

Due to time constraints, the present research did not conduct an exploratory analysis to assess the following questions:

- whether or not participants' results differed between trials with stimulus pairs recorded by a single talker and trials with stimulus pairs that combined the voices of different speakers;
- whether or not at Gen, where new-talker stimuli were introduced, participants discriminated better between stimulus pairs recorded by familiar talkers or between those recorded by novel talkers;
- which pitch-accent pattern combination (e.g., combination between the 1st-syllable accented pattern and the 2nd-syllable accented pattern) was the most difficult for participants (G. Pappalardo, personal communication, September 15, 2023); and
- to what extent the predictors of interest in the current research (the effects of musical training and talker variability) were at play in pitch-accent discrimination with different-talker stimuli.

Future work will assess these questions.

CHAPTER 6 ABSOLUTE PITCH TEST FOR MUSICIANS

Since the absolute pitch test was quite different from the linguistic tasks reported in chapters 4 and 5, this chapter presents its details and the data analyzed.

6.1. Absolute Pitch Test: Introduction

The purpose of the absolute pitch test was to assess whether musicians possessed absolute pitch. Following Burnham et al. (2015), the current study adopted the traditional definition of absolute pitch, reflecting pitch labeling ability (i.e., being able to name or label a note without a reference note). This was because the aim was to examine whether the ability to identify lexical pitch (Japanese pitch accent) correlated with the ability to identify musical pitch.

As mentioned in Section 2.3.4, to the author's knowledge, only Burnham et al. (2015) have shown that musicians with absolute pitch have an advantage in lexical tone discrimination. Their participants—native speakers of Australian English without any experience of tone languages—consisted of three groups: (1) musicians who possessed absolute pitch; (2) musicians who did not possess absolute pitch; and (3) non-musicians who had not received any musical training. The participants were tested using an AX discrimination task with stimuli—a monosyllable with five Thai lexical tones. Burnham and colleagues employed two interstimulus intervals (ISIs): 500 ms and 1500 ms. They observed that while the two musician groups outperformed the non-musicians, the musicians with absolute pitch performed better than the musicians without absolute pitch, indicating that the positive effect of absolute pitch on Thai lexical tone discrimination. The RT data also showed an advantage for the musicians with absolute pitch over the musicians without absolute pitch, but only at the ISI of 1500 ms.

Following up on the results shown by Burnham et al. (2015), the present research sought to further investigate the effect of absolute pitch on perceptual learning of Japanese pitch accent.

However, a series of studies conducted by Lee and Hung (2008) and by Lee et al. (2014) revealed how difficult it is to find English-speaking musicians with

absolute pitch, even though their musician participants were students or graduate music majors in the School of Music at Ohio University: "expert" musicians as in Ericsson et al. (1993) and Sloboda et al. (1996); see Section 2.3.2. for definitions of expert musicians and amateur musicians in their studies. The absolute pitch test conducted by Lee and Hung (2008) and by Lee et al. (2014) showed that none of their musicians met the criterion for absolute pitch. Thus, they were unable to investigate the effect of absolute pitch on lexical tone identification in the English-speaking musicians. Additionally, they did not find a significant correlation between performance in lexical tone identification and in the absolute pitch test.

These findings are consistent with the low prevalence of absolute pitch in conservatory-level western musicians reported in Deutsch et al. (2006) and Miyazaki et al. (2018). It was probable therefore that finding Italian musicians with absolute pitch would also be difficult for the current research. However, it was worthwhile to identify the musicians with absolute pitch in order to expand our understanding of its role in Japanese pitch-accent perception.

To achieve this goal, the current research largely replicated the absolute pitch test of Lee and Hung (2008) and Lee et al. (2014) with its musician participants. One might ask why all participants were not tested for absolute pitch, since the non-musician participants might also have had it. Indeed, it would have been ideal if the non-musicians had also taken the absolute pitch test, as absolute pitch might have constituted a very important confounding variable. However, as in Lee and Hung (2008) and Lee et al. (2014), the present study tested only musicians for two reasons. Firstly, the interest of this dissertation lies in pitch identification (labeling) ability, and not in pitch production ability (e.g., the production of familiar melodies); see Levitin (1994) and Parncutt and Levitin (2001) for detailed account of pitch production ability. The current study adopted the traditional definition of absolute pitch—pitch labeling ability, which is a less common ability (Levitin, 1994). The other reason is that the labeling of musical pitch made this test almost impossible for Italian non-musicians. Prior to the experiment, some native Italian speakers who had never received musical training outside of the Italian school curriculum were asked about their knowledge of music. Their answers revealed that they had no knowledge of the musical scales from C3

to B5, which were employed for the absolute pitch test in this study. Hence, it was assumed that in practice it would be too difficult for Italian non-musicians to engage in this test. Accordingly, the test was administered only to Italian musicians. Bear in mind that, as Lee and Hung (2008) stated, the aim of this test was not to evaluate the percentage of absolute pitch possessors per se, but to measure the musicians' absolute pitch ability and to conduct correlation analyses between the absolute pitch ability and linguistic tasks (*Identification Tasks* and *Discrimination Tasks*).

This chapter is thus concerned with the research question regarding the effect of absolute pitch on the linguistic tasks: Will there be any difference in the ability to identify/discriminate Japanese pitch accent between musicians with absolute pitch and those without absolute pitch?

6.2. Absolute Pitch Test: Method

The absolute pitch test was aimed at evaluating the musician participants' ability to identify musical notes without a reference pitch. It was mostly a replication of previous studies (Lee et al., 2014; Lee & Hung, 2008), which were in turn adapted from the test used in Deutsch et al. (2006). Compared to these studies, however, one key difference was introduced; the test was conducted entirely online by means of the Gorilla software package (Anwyl-Irvine et al., 2020), rather than in a laboratory.

6.2.1. Absolute Pitch Test: Participants

Only the 32 musicians described in *Identification Tasks* and *Discrimination Tasks* participated (see Section 4.2.1 for a more detailed description).

6.2.2. Absolute Pitch Test: Stimuli

The stimuli were created based on those reported by Lee and Hung (2008), and Lee et al. (2014). Thirty-six notes, which spanned the three-octave range from C_3 (131

Hz) to B_5 (988 Hz), were generated with three timbres (pure tone, Steinway grand piano and acoustic guitar). This resulted in a total of 108 stimulus notes. Each note lasted 500 ms.

The pure tones were generated with Praat (Boersma & Weenink, 2021). Since fade-in and fade-out are necessary to prevent sudden changes of signal amplitude, which would result in audible clicks, fade-in and fade-out values of 10 ms were used (these are default values).

The other instrumental sounds, including oboe sounds which were used in the short practice, were produced with GarageBand for macOS 10.4.7. Since the sounds were initially generated in stereo, they were converted to mono channel. In addition, Praat was used to trim each note to a duration of 500 ms (from onset), because the notes generated by GarageBand lasted more than 500 ms. In order to prevent an abrupt ending in the edited sound (Moon et al., 2021; Sugiyama, 2012), each stimulus was faded out in 50 ms.

All stimuli were in WAV format, with a 44.1 kHz sampling rate, 24-bit sample size and mono channel. The intensity of all stimuli was normalized to 70 dB as for the stimuli used in *Identification Tasks* and *Discrimination Tasks*.

Since all materials (the stimuli used in the test and the oboe notes used in the short practice) were the synthesized sounds, they were checked by two Japanese female musicians; one was a professional musician with relative pitch, and the other was a musically trained individual who reported having absolute pitch. After listening to all the materials, they assured the author that they perceived all the materials as normal musical notes.

6.2.3. Absolute Pitch Test: Procedure

The absolute pitch test, i.e., the musical note identification test, followed the procedure described in Lee et al. (2014) and in Lee and Hung (2008) with some modifications, including the fact that it was carried out entirely online.

It was conducted on Day 2, before Training 2 (see chapters 3 and 4).

Before beginning the absolute pitch test, participants were told that they would be listening to a set of 36 notes, ranging from C_3 (131 Hz) to B_5 (988 Hz), with three types of timbre. It was also explained to them that the absolute pitch test was comprised of nine blocks and each block consisted of 12 notes as described below. Given this test's difficulty, foreseeable from the results of previous studies (Deutsch et al., 2006; Lee et al., 2014; Lee & Hung, 2008), participants were told that the objective of the absolute pitch test was not to evaluate their musical competence, but to analyze the correlation between this test and the linguistic tasks described in chapters 4 and 5. This was intended to avoid embarrassing participants and affecting their motivation.

After the explanation about the absolute pitch test, they completed a short practice as in Lee et al. (2014) and in Lee and Hung (2008). The purpose of the practice was to familiarize participants with the test format. In line with the previous studies cited above, the practice consisted of 12 trials of oboe tones, which were selected from the same pitch range (from C_3 to B_5) used in the test proper. Additionally, as in the test, each tone lasted 500 ms and any two consecutive notes were separated by more than an octave. These oboe tones were used only in the practice. During the practice no feedback was provided.

The terms used in the test proper and its structure are summarized in Table 6.1. As can be seen, the absolute pitch test consisted of 108 trials, organized into nine blocks: three with pure tone stimuli, another three with piano stimuli, and the other three with acoustic guitar stimuli. The order of the three blocks with pure tone stimuli, the three blocks with piano stimuli, and the three blocks with acoustic guitar stimuli was counterbalanced across participants.

Table 6.1
Summary of the Terms and Structure

Stimuli type
3 timbres (pure tone, Steinway grand piano and acoustic guitar)
Notes per timbre
36 notes, which spanned the three-octave range from C_3 (131 Hz) to B_5 (988 Hz)
Blocks
Participants took a fixed break of 10 seconds between blocks. (1) 3 blocks with pure tone notes: Block 1: 12 notes/trials; Block 2: 12 notes/trials; Block 3: 12 notes/trials. (2) 3 blocks with piano notes: Block 1: 12 notes/trials; Block 2: 12 notes/trials; Block 3: 12 notes/trials. (3) 3 blocks with acoustic guitar notes: Block 1: 12 notes/trials; Block 2: 12 notes/trials; Block 3: 12 notes/trials.
Number of trials
108 trials = 12 trials x 9 blocks

Following prior works (Deutsch et al., 2006; Lee et al., 2014; Lee & Hung, 2008), trial order within each of the three timbre blocks was not randomized, but was organized so that no two consecutive notes were separated by more than an octave. The purpose of this measure was to prevent participants from using relative pitch as a cue for the task (Deutsch et al., 2006). Additionally, trial (note) order for each timbre was varied.

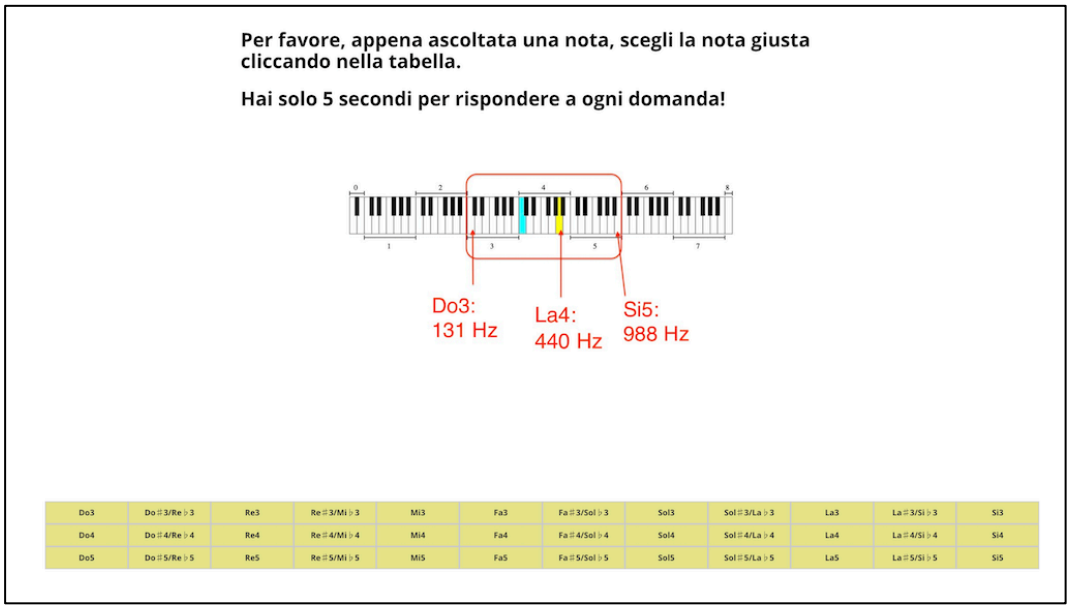
As shown in Table 6.1, the 36 notes for each timbre were divided into three blocks; consequently, each block contained 12 notes. There was a fixed break of 10 seconds between blocks. Participants could not take any breaks within the block.

For each trial, the sequence of steps was in line with Lee et al. (2014) and with Lee and Hung (2008). Firstly, participants listened to the stimulus, which lasted for 500 ms. To respond, they then had to click on the correct note name, choosing it among the 36 note names presented in the table on the computer screen (see Figure 6.1). Since participants were native Italian speakers who were receiving musical training in Italy, the Italian note names were employed (e.g., *Do*, instead of C). Two musicians who had trained in Italy checked the Italian note names presented in the table just to make sure the names were understandable for participants; one of them was a professional Japanese opera singer, and the other a native Italian speaker who had graduated from a musical institute in Italy.

The present study used a multiple-choice response format with the use of a mouse, instead of an open-ended question format with the use of pen as in prior works (Deutsch et al., 2006; Lee et al., 2014; Lee & Hung, 2008), because the absolute pitch test was conducted online. However, note that participants in the previous studies received the same explanation used in the current research, that is, that they would be listening to a set of 36 notes, ranging from C_3 (131 Hz) to B_5 (988 Hz), with three types of timbre. This meant that the participants were still selecting one answer from 36 choices, although they wrote down their responses on a sheet of paper. What is more, the use of a mouse was adopted to prevent participants' typing speed from affecting the speed or content of their responses, because they needed to meet the response deadline, which was five seconds after the end of the second stimulus.

Figure 6.1

Presentation of Screen Display in the Absolute Pitch Test



Five seconds after the response deadline, the next stimulus played.

As in the practice, no feedback was provided at any time during the test (following Lee et al., 2014; Lee & Hung, 2008). The absolute pitch test lasted about 15 minutes.

6.2.4. Absolute Pitch Test: Analysis

A total of 3,460 trials[29] was performed by 32 musicians, i.e., 108 trials x 32 musicians.

The Gorilla software package (Anwyl-Irvine et al., 2020) logged each musician's responses. Musicians' responses were coded as 0 for correct or 1 for incorrect. Timeouts were treated as incorrect. Their binary accuracy scores for each trial were used as the dependent variable. More specifically, as in Lee and Hung (2008), and Lee et al. (2014), the current research adopted four measures of accuracy scores: (1) percentage of correct responses allowing no semitone errors; (2) percentage of correct responses allowing one-semitone errors; (3) percentage of

[29] One musician performed four trials (trial number 69-72) twice, possibly due to technical problems. Therefore, only the first attempt of the trials was included in the data analysis.

correct responses allowing two-semitone errors; (4) percentage of correct responses allowing three-semitone errors. Errors up to three semitones were allowed because of the challenging nature of this test. Indeed, Deutsch et al. (2006) found that only approximately 15% of non-tone language speakers enrolled in a music conservatory were able to achieve a score of at least 85%, and this percentage was adopted as the criterion for assessing absolute pitch in Lee and Hung (2008), Lee et al. (2014) and the present research. Recall also that neither Lee and Hung (2008) nor Lee et al. (2014) found any musicians with absolute pitch.

Data analyses were performed using R 4.3.2 (R Core Team, 2023). In order to assess the effect of timbre (guitar, piano, and pure tone) on participants' performance, a mixed-effects binomial logistic regression analysis was conducted, using the *lme4* package (Bates et al., 2015). Model diagnosis (observation of residual qq-plots) was conducted using the *DHARMa* package (Hartig, 2022).

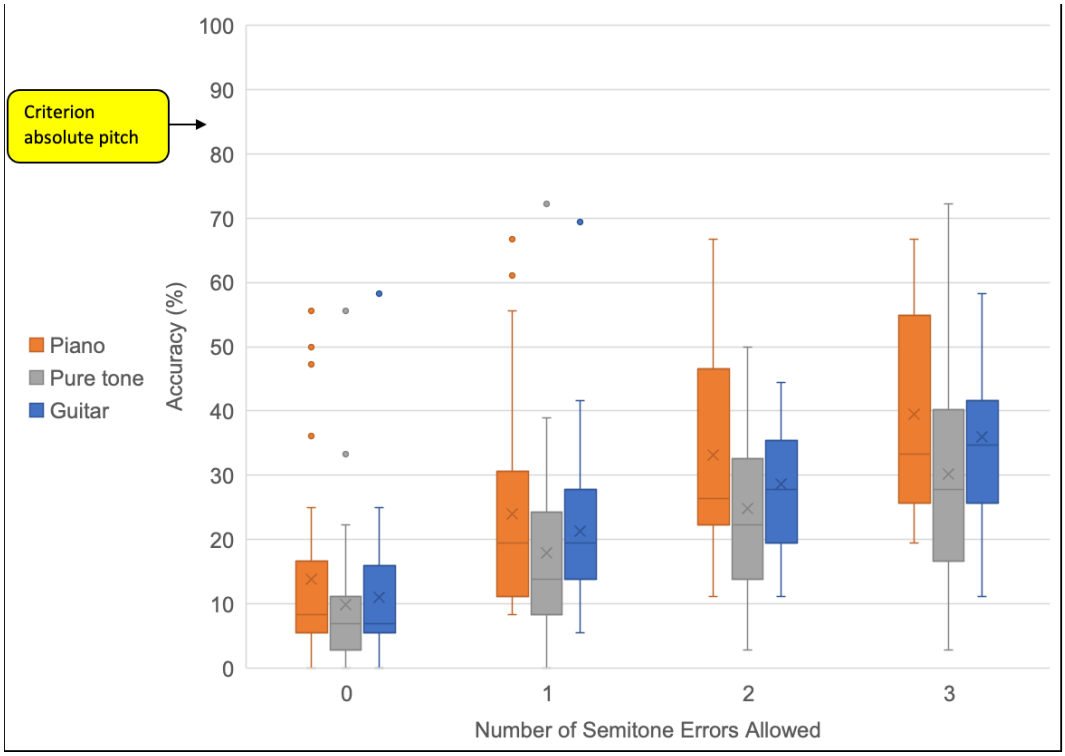
6.3. Absolute Pitch Test: Results

This section first provides an overview of performance on the absolute pitch test; then it presents mixed-effects models to examine how timbres affected the musicians' accuracy. Lastly, it provides analyses of the correlations between the absolute pitch test and each of the identification tests (pretest, posttest, and two tests of generalization), and each of the discrimination tests (pretest, posttest, and a test of generalization).

Figure 6.2 illustrates musicians' accuracy in the absolute pitch test by timbre (guitar, piano, and pure tone), and the number of semitone errors allowed (zero, one, two, and three).

Figure 6.2

Absolute Pitch Test: Musicians' Accuracy by Timbre (Piano, Pure Tone, and Guitar), as a Function of the Number of Semitone Errors Allowed



What is striking in Figure 6.2 is the overall low percentage of correct answers. When exact answers were required, percentages of accuracy were noticeably low. As the tolerance for semitone errors increases, so does the accuracy, but still the percentage of accuracy was not very high; see also Table 6.2, which shows mean accuracy on the absolute pitch test for the three timbres and the number of semitone errors allowed.

Table 6.2
Absolute Pitch Test: Mean Accuracy by Timbre (Piano, Pure Tone, and Guitar), as a Function of the Number of Semitone Errors Allowed

Timbre	Number of semitone errors allowed			
	0	1	2	3
Piano	13.9%	24.0%	33.2%	39.5%
Pure tone	9.9%	17.9%	24.8%	30.1%
Guitar	10.9%	21.3%	28.6%	35.9%

Even when errors of up to three semitones were allowed (following Lee et al., 2014; Lee & Hung, 2008), as anticipated in the preliminary results of the current research (Naito, 2023), none of the musicians met the criterion for absolute pitch of least 85% of correct responses (Deutsch et al., 2006).

Interestingly, as can be seen in Figure 6.2 and in Table 6.2, the accuracy for piano was higher than that for the other timbres. In order to evaluate the effect of timbre (guitar, piano, and pure tone) on musicians' accuracy, mixed-effects binomial logistic regression models were used for each of the four accuracy measures mentioned in Section 6.2.4: (1) percentage of correct responses allowing no semitone errors; (2) percentage of correct responses allowing one-semitone errors; (3) percentage of correct responses allowing two-semitone errors; (4) percentage of correct responses allowing three-semitone errors.

In the mixed effects models, *timbre* (piano as the reference level, guitar and pure tone; treatment coded) was treated as a fixed effects predictor. The reason why piano was used as the reference level, and tests for all pairwise comparisons were not performed, stemmed from the results of statistical analyses conducted in Lee and Hung (2008) and Lee et al. (2014). Although neither study reported a significant effect in most cases, the accuracy for piano was higher than that for viola when two-semitone errors were allowed (Lee & Hung, 2008), and higher for piano than for pure tone when three-semitone errors were allowed, (Lee et al., 2014).

Participants were included as random effects with by-participant varying intercepts and slopes for the effect under investigation. Items (stimuli) were not included because they were not repeated across participants.

Each of the full models for testing the effect of timbre on musicians' accuracy was coded in R as follows: glmer(Accuracy ~ Timbre + (1 + Timbre |Participants), data = data, family = 'binomial').

In order to test whether or not the inclusion of the participants as random effects with by-participant varying intercepts and slopes improved model fit, likelihood ratio tests were conducted. Specifically, likelihood ratio tests were carried out for the reduced models (without by-participant varying slopes) against the intercept-only models, and for the full models against the reduced model without by-participant-varying slopes. These results are provided in Table 6.3, and they show that the addition of by-participant varying intercepts and slopes led to an improvement in model fit.

Table 6.3
Mixed-Effects Model Comparisons for the Additional Parameters of Interest: Summary of Results

Additional model parameter	Number of semitone errors allowed			
	0	1	2	3
By-participant varying intercepts (reduced model)	$\chi^2(2) = 10.452$, $p = 0.0053$	$\chi^2(2) = 14.209$, $p = 0.0008$	$\chi^2(2) = 20.544$, $p < 0.0001$	$\chi^2(2) = 23.882$, $p < 0.0001$
By-participant varying intercepts and slopes (full model)	$\chi^2(5) = 23.078$, $p = 0.0003$	$\chi^2(5) = 32.878$, $p < 0.0001$	$\chi^2(5) = 22.031$, $p = 0.0005$	$\chi^2(5) = 24.092$, $p = 0.0002$

Mixed-effects modeling analyses of effect of timbre on musicians' accuracy are summarized in Table 6.4.

Table 6.4
Effect of Timbre on Musicians' Accuracy: Summary of Mixed-Effects Modeling Analyses

Number of semitone errors allowed	Level of predictor (vs. piano)	Estimate	Standard error	z-value	p
0	Guitar	-0.17	0.22	-0.76	0.446
	Pure tone	-0.32	0.21	-1.54	0.124
1	Guitar	-0.10	0.17	-0.63	0.526
	Pure tone	-0.39	0.17	-2.28	0.023
2	Guitar	-0.20	0.13	-1.54	0.122
	Pure tone	-0.46	0.15	-3.17	0.002
3	Guitar	-0.16	0.12	-1.33	0.184
	Pure tone	-0.48	0.15	-3.26	0.001

Table 6.4 showed that when exact answers were required (zero semitone errors allowed), the full model revealed no significant difference between piano and guitar (b = -0.17, SE = 0.22, p = 0.446) nor between piano and pure tone (b = -0.32, SE = 0.21, p = 0.124). By contrast, when one-semitone errors were allowed, the full model yielded a significant difference between piano and pure tone (b = -0.39, SE = 0.17, p = 0.023), but not between piano and guitar (b = -0.11, SE = 0.17, p = 0.526). When two-semitone errors were allowed, the full model again revealed a significant difference between piano and pure tone (b = -0.46, SE = 0.15, p = 0.002), but not between piano and guitar (b = -0.20, SE = 0.13, p = 0.122). Likewise, when three-semitone errors were allowed, the full model showed a significant difference

between piano and pure tone (b = -0.48, SE = 0.15, p = 0.001), but not between piano and guitar (b = -0.16, SE = 0.12, p = 0.184).

Having analyzed the effect of timbre on musicians' accuracy, the remainder of this section addresses correlation analyses. Recall that one of the research questions of *Identification Tasks* (Chapter 4) and of *Discrimination Tasks* (Chapter 5) was: "Will there be any difference in the ability to identify/discriminate Japanese pitch accent between musicians with absolute pitch and those without absolute pitch?".

As discussed earlier, the results of the absolute pitch test revealed that none of the musicians possessed absolute pitch. This means that the research question mentioned above could not be pursued. However, since it was still considered useful to investigate whether linguistic perception correlates with musical note perception, correlation analyses were carried out.

Specifically, Pearson's correlation analyses—and Spearman correlation analyses when assumptions were violated—were conducted between each of the four accuracy measures for the absolute pitch test (percentages of correct responses allowing zero-, one-, two-, and three-semitone errors) and results for each of the identification and discrimination tasks. As regards the identification tasks, percentages of musicians' correct responses excluding responses with RTs shorter than 200 ms (see Section 4.2.5) were used. As regards the discrimination tasks, musicians' d' scores were employed (see Section 5.2.5). The results are summarized in Table 6.5.

Table 6.5

Correlations Between the Four Accuracy Measures in the Absolute Pitch Test and Results in the Identification and Discrimination Tasks

		Number of semitone errors allowed			
		0	1	2	3
Identification	Pretest	$r_s = 0.41$, $p = 0.020$	$r_s = 0.36$, $p = 0.042$	$r_s = 0.38$, $p = 0.03$	$r = 0.24$, $p = 0.189$
	Posttest	$r_s = 0.26$, $p = 0.148$	$r_s = 0.27$, $p = 0.131$	$r_s = 0.31$, $p = 0.088$	$r = 0.25$, $p = 0.165$
	Gen-1	$r_s = 0.32$, $p = 0.072$	$r_s = 0.29$, $p = 0.109$	$r_s = 0.28$, $p = 0.125$	$r = 0.13$, $p = 0.484$
	Gen-2	$r_s = 0.41$, $p = 0.045$	$r_s = 0.32$, $p = 0.073$	$r_s = 0.28$, $p = 0.115$	$r = 0.20$, $p = 0.274$
Discrimination	Pretest	$r_s = 0.16$, $p = 0.393$	$r_s = 0.10$, $p = 0.567$	$r_s = 0.22$, $p = 0.223$	$r = 0.07$, $p = 0.722$
	Posttest	$r_s = 0.24$, $p = 0.188$	$r_s = 0.19$, $p = 0.295$	$r_s = 0.16$, $p = 0.37$	$r = 0.03$, $p = 0.892$
	Gen	$r_s = 0.28$, $p = 0.12$	$r_s = 0.21$, $p = 0.251$	$r_s = 0.22$, $p = 0.218$	$r_s = 0.14$, $p = 0.439$

Note. Gen-1 = Test of generalization 1; Gen-2 = Test of generalization 2; Gen = Test of generalization; r = Pearson's correlation coefficient; r_s = Spearman's correlation coefficient (rho). Shaded cells show p-values < the significance level alpha = 0.05.

As can be seen in Table 6.5, only four significant positive correlations were found. Specifically, accuracy in the identification pretest correlated with accuracy in the absolute pitch test when zero-semitone errors were allowed (r_s = 0.41, p = 0.020); when one-semitone errors were allowed (r_s = 0.36, p = 0.042); and two-semitone errors were allowed (r_s = 0.38, p = 0.03). Additionally, accuracy in identification Gen-2 correlated with that in the absolute pitch test when zero-semitone errors were allowed (r_s = 0.41, p = 0.045). By contrast, there were no significant correlations between the accuracy measures for the absolute pitch test and d' scores in the discrimination tasks.

6.4. Absolute Pitch Test: Discussion

The aim of the absolute pitch test was to assess whether or not musicians possessed absolute pitch. Since, to the author's knowledge, only Burnham et al. (2015) have found that musicians with absolute pitch were more accurate at lexical tone discrimination, the present study was designed to explore the effect of absolute pitch on perceptual learning of Japanese pitch accent in order to develop a better understanding of its effect.

The results of the absolute pitch test revealed its challenging nature, showing an overall low percentage of correct responses. Indeed, even when errors of up to three semitones were allowed (following Lee et al., 2014; Lee & Hung, 2008), none of the musicians met the criterion for absolute pitch of least 85% of correct responses (Deutsch et al., 2006). These unsurprising findings are in accord with those of Lee and Hung (2008) and Lee et al. (2014), and are also consistent with the low prevalence of absolute pitch in conservatory-level western musicians reported in Deutsch et al. (2006) and Miyazaki et al. (2018).

It is worth mentioning that absolute pitch was found by Deutsch et al. (2006) to correlate with age of onset of musical training: the younger the age at which musical training began, the more likely musicians were to meet the criteria for absolute pitch (see FIG. 1 in Deutsch et al., 2006). In their study, the probability of possessing absolute pitch had declined to zero for English-speaking musicians by

the training onset age of 8-9. In the present study, musicians' mean age of onset of musical training was 11.5 years (SD = 4.4 years), while the mean ages in Lee and Hung (2008) and in Lee et al. (2014) were 9.4 years and 8.89 years. This may explain the current research's results, which show how challenging the absolute pitch test was for the musicians participating in the experiment.

Interestingly, although overall accuracy was not very high, accuracy for the piano timbre was higher than that for the other timbres (pure tone and guitar). Statistical analyses revealed significant differences between piano and pure tone—but not between piano and guitar—when one-, two- and three-semitone errors were allowed. These results differ from the findings reported by Lee and Hung (2008) and Lee et al. (2014): in their studies accuracy for piano was higher than that for viola when two-semitone errors were allowed (Lee & Hung, 2008), and higher for piano than for pure tone when three-semitone errors were allowed (Lee et al., 2014).

A possible explanation for the relatively high accuracy for piano may be that many musicians were familiar with the sounds of the piano. Indeed, according to the preliminary questionnaire data, 19 out of the 32 musicians had practiced piano (10 had practiced piano exclusively or had practiced piano for the longest period of time compared to other instrument/s or singing). Seven out of the 32 musicians had practiced guitar (five reported having practiced guitar for the longest period of time compared to other instrument/s or singing). Conversely, it could be said that the pure tone was a sound to which the musicians were not accustomed. Hence, these results are likely to be related to the musician's familiarity with the type of timbre.

Even though musicians were more accurate at identifying piano notes without a reference note, as discussed above, none of the musicians was found to possess absolute pitch. Therefore, unfortunately, one of the research questions, "Will there be any difference in the ability to identify/discriminate Japanese pitch accent between musicians with absolute pitch and those without absolute pitch?", has to remain unanswered. Future studies are required to better understand the role of absolute pitch.

Although the present research did not find any musicians with absolute pitch, it examined whether the results for each of the identification and discrimination

tasks correlated with each of the four accuracy measures for the absolute pitch test (percentages of correct responses allowing zero-, one-, two-, and three-semitone errors), since this topic was still considered interesting to explore.

The results of the correlation analyses revealed that there were four significant positive correlations. Specifically, accuracy in the identification pretest correlated with that in the absolute pitch test when zero-, one- and two-semitone errors were allowed. Accuracy in identification Gen-2 also correlated with that in the absolute pitch test when zero-semitone errors were allowed. As regards the discrimination tasks, none of the correlations was significant.

These findings were rather unexpected, because they were inconsistent with those of Lee and Hung (2008) and Lee et al. (2014). It is also somewhat surprising that significant correlations with the accuracy measures for the absolute pitch test were found only for accuracy in the identification tasks, but not for sensitivity (d' scores) in the discrimination tasks.

However, the correlations observed in this study could be associated with the nature of absolute pitch. Since absolute pitch, in other words, pitch labeling ability, involves abstracting pitch movements, categorizing and identifying them (naming or labeling notes), it is reasonable to assume that absolute pitch would facilitate Japanese pitch-accent identification, but that discrimination tasks may not require these abilities. Indeed, almost all positive correlations were found for the identification *pretest*. This suggests that musical tone identification ability helped at the beginning of learning, but that this "helpfulness" declined sharply, as shown by the fact that there was only one correlation, between the absolute pitch test when zero-semitone errors were allowed and accuracy in identification Gen-2.

Overall, the results of the absolute pitch test in the present study confirmed the difficulty of finding musicians who are both native speakers of non-tone languages and who possess absolute pitch. However, Deutsch et al. (2006) found that the younger the age at which musical training started, the higher the prevalence of absolute pitch was: future studies on the effect of absolute pitch are thus recommended to take musicians' onset age of musical training into account, when recruiting musicians. The other issue emerging from this study's findings is the

effect of piano timbre. In future, it would be interesting to assess whether similar results could be obtained with musicians who practice instruments other than the piano, or who sing rather than playing the piano. Other ways to avoid the potentially confounding factor of timbre familiarity might be to use timbres unfamiliar to all musician participants; or to recruit musicians who practice the same instrument having a timbre different to those used in the stimuli.

CHAPTER 7 GENERAL DISCUSSION AND CONCLUSION

This last chapter starts by summarizing the findings and discussion of the present study. Next, the limitations of the experimental work are discussed. Then the implications of the results are set out. The chapter closes with conclusions.

7.1. Summary of Findings and Discussion

The aim of the current study was threefold. First and foremost, it set out to examine the effect of musical training on the perceptual learning of Japanese pitch accent by native Italian speakers with no experience of Japanese. A second aim was to investigate the effect of talker variability in training stimuli on the perceptual learning of Japanese pitch accent. For these two purposes, Italian musicians and non-musicians were randomly assigned to a high variability (HV) training condition (stimuli produced by four talkers), or a low variability (LV) training condition (stimuli produced by one talker) for training sessions (approximately one hour in total). Then, their performance on identification tasks and discrimination tasks administered before and after training was compared. As a final aim, an attempt was made to assess the effect of absolute pitch on the perceptual learning of Japanese pitch accent.

The present dissertation addressed five research questions (RQ1-5). This section summarizes and discusses the experimental results in the light of these questions and of previous studies.

To begin with, I would like to remind the reader that the dependent variables for the identification tasks and the discrimination tasks were different. For the identification tasks, participants' binary accuracy scores (0 for correct; 1 for incorrect) for each trial were employed as the dependent variable. By contrast, for the discrimination tasks, there were two dependent variables: d' scores (participants' perceptual sensitivity to differences between stimuli when the two stimuli were *different*), and log RT (a measure of response latency). In the discrimination tasks, remember also that two interstimulus intervals (500 ms and 1500 ms) were employed, following Burnham et al. (2015). Their effect was explored along with

the effects of musical training and perceptual training condition, but neither significant main effect nor significant interaction were found, either for d' score or for log RT results (see Section 5.4 for a detailed discussion on the results for ISIs).

To turn to the first research question, RQ1 addressed whether Italian musicians would outperform Italian non-musicians in identifying and in discriminating Japanese pitch accents. As expected, results for the identification tasks and d' score results for the discrimination tasks showed that overall, musicians outperformed non-musicians in identifying and in distinguishing Japanese pitch-accent patterns. These findings are in good agreement with those of various studies which have shown, overall, the positive effect of musical experience/training on lexical tone perception by native speakers of non-tone languages without any experience of the target tone language (Alexander et al., 2005; Burnham et al., 2015; Chang et al., 2016; Chen et al., 2020; Delogu et al., 2010; Gottfried, 2007; Gottfried & Xu, 2008; Götz et al., 2023; Kirkham et al., 2011; Lee et al., 2014; Lee & Hung, 2008; Marie et al., 2011; Mok & Zuo, 2012). The results are also in line with those of Golob (2003), which is, as far as the author knows, the only study investigating the effect of musical training on Japanese pitch accent perception.

Interestingly, the results for log RT data differed from the identification task scores and the d' score results for the discrimination tasks. Specifically, the findings indicated that musical training did not provide an overall speed advantage when perceiving pitch differences. Recall that, following Burnham et al. (2015)—who used bare RT data—the log RTs analyzed and reported in the current research were only for *correct* responses to *different* AX pairs (AB or BA trials). The present log RT results are in disagreement with those of some studies (Alexander et al., 2005; Burnham et al., 2015; Lee & Hung, 2008), but in line with others (Lee et al., 2014; Marie et al., 2011; Mok & Zuo, 2012).

As mentioned in Section 4.4.1, previous studies that tested perceptual ability without prior training—again, Alexander et al. (2005), Burnham et al. (2015), Chang et al. (2016), Chen et al. (2020), Delogu et al. (2010), Gottfried (2007), Gottfried and Xu (2008), Götz et al. (2023), Kirkham et al. (2011), Lee et al. (2014), Lee and Hung (2008), Marie et al. (2011), and Mok and Zuo (2012)—have largely

converged to indicate that musical experience/training facilitates lexical tone perception by native non-tone language speakers, albeit to varying degrees. However, as discussed in Section 2.3.3, studies investigating the effects of perceptual training on the learning of lexical tone have reported mixed findings regarding the advantage of musical training. To the author's knowledge, there has been no research examining the effect of musical experience/training on perceptual learning of Japanese pitch accent.

RQ2 was thus set to assess whether or not the difference between musicians and non-musicians in the ability to identify/discriminate Japanese pitch-accent patterns would decrease or increase after training. Again, the results for the identification tasks and the d' score results in the discrimination tasks are in line with the prediction, i.e., overall, the difference between non-musicians and musicians widened after the pretest. The results showing an advantage for musicians over non-musicians are dissimilar to those of some training studies (Dittinger et al., 2016; Tong & Tang, 2016; Wayland et al., 2010; Zhao & Kuhl, 2015), but in accord with those of others (Cooper & Wang, 2012; Maggu et al., 2018; P. C. M. Wong & Perrachione, 2007).

A more detailed comparison of the two sets of results (identification and d' for discrimination), however, reveals some differences. Both non-musicians and musicians showed a pretest-posttest improvement in pitch-accent pattern identification, although musicians' improvements are greater than those of non-musicians. By contrast, the d' score results for the discrimination tasks showed that non-musicians' performance at the posttest was almost the same as at the pretest, whereas musicians showed a pretest-posttest improvement. In addition, the identification task results revealed that all participants (irrespective of category) scored higher at the two tests of generalization than at the pretest, although there was an added advantage for musicians at the posttest and at the two tests of generalization. In contrast, the d' score results for the discrimination tasks showed that whereas musicians' generalization test scores were similar to those for the pretest, those of non-musicians were worse than those for the pretest: indeed, the difference between musicians and non-musicians increased considerably at Gen.

These differences may be explained by the fact that in the training sessions, the only type of task employed was the identification task. It also seems possible that the differences are due to the brevity of the training sessions (about one hour in total). Future research, involving the use of discrimination tasks in the training protocol and longer duration of training, is required to explore whether these measures would be effective to enhance the participants' performance on the discrimination tasks.

With regard to the log RT results for the discrimination tasks, again, they differed from the results for the identification tasks and the d' scores for the discrimination tasks. Specifically, musicians exhibited a speed advantage only at the pretest and at Gen. At the posttest, in contrast, musicians' response latency was comparable to that of non-musicians. Furthermore, non-musicians and musicians showed similar trends: while they responded faster at the posttest than at the pretest, they were slower at Gen than at the posttest or at the pretest (although only slightly slower at the pretest). These findings for the log RT data imply that musicians had only a limited advantage. To the best of the author's knowledge, however, none of the training studies exploring the effect of musical training/experience have measured RTs. Future studies are needed to better understand the relationship between the effect of musical training and reaction time in responding to perceptual stimuli.

So far, in response to RQ1 and RQ2, the effect of musical training on perceptual learning of Japanese pitch accent has been discussed. Now the section addresses RQ3, about the effect of talker variability in training stimuli (HV training condition versus LV training condition) and its interaction with the effect of musical training. Specifically, RQ3 was whether or not the HV training condition would be more beneficial for Italian musicians compared to non-musicians.

With respect to the two training conditions, musicians and non-musicians showed different trends in their identification task and discrimination task d' score results. On one hand, for musicians no significant differences were found between the two training conditions. On the other hand, in the case of non-musicians, after

training, non-musicians in the HV training condition performed better than those in the LV training condition.

With regard to the discrimination log RT results, unlike the results for the identification tasks and discrimination d' scores, these revealed no significant difference in reaction time between non-musicians and musicians in the two training conditions, indicating that both training conditions induced participants to respond faster at the posttest than at the pretest.

Taken together, aside from the log RT results, identification task results were similar to the discrimination d' score results. This suggests that, whereas for the musicians, the two training conditions were comparably effective, for non-musicians, the HV training condition was more beneficial.

These findings are dissimilar to those of Dong et al. (2019), but in line with those of Perrachione et al. (2011), Sadakata and McQueen (2014), and Qin et al. (2022), in that they show an interaction between talker variability in training stimuli and perceptual ability. However, the interaction found in the current research differed from that reported by the latter researchers. Perrachione et al. (2011) and Sadakata and McQueen (2014) reported that while high perceptual aptitude participants benefitted from the HV training condition, low perceptual aptitude participants received more benefit from LV training, and higher variability hindered perceptual learning. This detrimental effect of the HV training condition on low perceptual aptitude participants was also reported by Qin et al. (2022). Note, however, that these studies investigated whether talker variability in training stimuli interacted with individuals' *perceptual abilities*, not with musical training. This difference may account for the discrepancy in the results between the current research and these studies.

The current research—as far as the researcher knows, the first to do so—investigated the interaction between *musical training* and talker variability in perceptual training for pitch accent pattern identification. As regards musicians, the findings suggest that talker variability did not play a role in the learning of Japanese pitch-accent pattern identification and discrimination. It can be reasoned that musically trained individuals are already capable of extracting abstract information

about Japanese pitch accent and of applying it to novel input without the need for different samples.

In the case of non-musicians, on the other hand, the results indicate that talker variability did play a role in both tasks: the HV training condition favored perceptual learning of Japanese pitch accent more than the LV training condition. This is consistent with Silpachai (2020), whose target was Mandarin Chinese lexical tone contrasts, and with Wong (2012), and Wong (2014), whose target was segment contrasts. This finding is also partially similar to that of Sadakata and McQueen (2013). I say partially, because the researchers reported that the benefit of the HV training condition was observed only in identification tasks, but not in discrimination tasks. These differences between the present research and Sadakata and McQueens' study may be due to the difference in the training conditions. In their work, whereas HV training consisted of fewer repetitions of a more varied stimuli recorded by five talkers, LV training consisted of many repetitions of less limited stimuli recorded by a single talker. By contrast, in the current study the only difference between the two training conditions was in the number of talkers who recorded the stimuli used in the training.

A comparison between the present results for the identification tasks and the d' score results for the discrimination tasks, however, reveals some differences in the performance of non-musicians in the LV training condition. As regards the identification tasks, this group improved significantly from the pretest to the posttest and slightly (but not significantly) from the pretest to Gen-1 and Gen-2. By contrast, as regards the discrimination tasks, they showed a deterioration in d' scores from pretest to posttest and from posttest to Gen.

These d' score results indicate not only that the HV training condition was more effective for non-musicians to learn to discriminate pitch-accent patterns, but also that the LV training condition was detrimental to them. Evidence that the LV training condition was detrimental to non-musicians in the discrimination tasks is also supported by the log RT results for the discrimination tasks. Specifically, non-musicians in the LV training condition showed a negative speed-sensitivity trade-off at the discrimination posttest: while they responded faster than musicians under

the LV training condition, their perceptual sensitivity (as measured by d' scores) was lower at the posttest than at the pretest.

These results imply that, unlike musicians, non-musicians need a variety of voice samples to learn to identify and discriminate Japanese pitch- accent patterns, especially for generalization to new input. The observed importance of talker variability in auditory input may have important practical implications for L2 Japanese learning/teaching and, in Section 7.3, these are discussed.

Going on to RQ4 (which of three target pitch-accent patterns would be the most difficult to perceive for native Italian speakers), what follows focuses first on the identification tasks—without referring to absolute pitch—to then move on to the question of the relationship between absolute pitch and linguistic tasks posed in RQ5.

Recall that, because similar improvements were observed for all three pitch-accent patterns from the pretest to Gen-1 and from the pretest to Gen-2, only pretest-posttest data and pretest-Gen-1 data for the identification tasks were analyzed to examine whether non-musicians and musicians improved their accuracy for each of the three pitch-accent patterns (see Section 4.3.2).

The findings showed a before-and-after training variation in which pitch-accent pattern participants found most difficult. In the pretest results, the most difficult pattern for all participants was the 1st-syllable accented pattern. This is partially consistent with Pappalardo (2018) and Hirano-Cook (2011). I say partially because they reported that the easiest pattern was the unaccented one: the current research, in contrast, found no significant differences at the pretest between the 2nd-syllable accented and unaccented patterns in participants' results irrespective of category and training condition. All participants showed significant pretest-posttest improvements in identifying the 1st-syllable accented pattern and this pattern was no longer the most difficult one after training. With the exception of non-musicians in the LV training condition, all participants also achieved significant pretest-posttest improvements for the other two pitch-accent patterns. Non-musicians in the LV training conditions showed little change in their scores at the pretest and the

posttest. This suggests again a limited positive effect of the LV training condition for non-musicians.

Even though all participants (except for non-musicians in the LV training condition) performed significantly better at the posttest than at the pretest for the unaccented pattern, analysis of pretest-Gen-1 data showed that at Gen-1, the unaccented pattern was the most difficult pattern. This partially supports the prediction made for RQ4: that the most difficult pattern would be the unaccented pattern. This finding is also in line with Shport (2011, 2016).

It is worth mentioning that the analysis of pretest-Gen-1 data revealed the same trends as for that of pretest-posttest data. On one hand, as regards musicians, no significant differences between the two training conditions were found. On the other hand, as regards non-musicians, the effect of the LV training condition on non-musicians' outcomes was limited to the 1st-syllable accented pattern. These results support those for RQ3 about the benefits of talker variability: whereas for musicians both training conditions were effective, for non-musicians the HV training condition is more beneficial.

With regard to RQ4, it would have been ideal to conduct an exploratory analysis to assess, in the discrimination tasks, which pitch-accent pattern combination (e.g., combination between the 1st-syllable accented pattern and the 2nd-syllable accented pattern) was the most difficult for participants (G. Pappalardo, personal communication, September 15, 2023). It was not included in the present study's analyses due to time constraints. Future research will explore this question.

The remainder of this section focuses the identification tasks results with regard to RQ5, which is virtually the same as RQ4 for the discrimination tasks. Research question 5 addressed whether or not there would be any difference between musicians with absolute pitch and those without in the ability to identify and discriminate Japanese pitch accents (see Section 6.4 for a detailed discussion of the absolute pitch test results).

In the absolute pitch test, none of the musicians met the criterion for absolute pitch of least 85% of correct responses (Deutsch et al., 2006), even when errors of

up to three semitones were allowed (following Lee et al., 2014; Lee & Hung, 2008) and even for the piano timbre, for which accuracy was higher than for the other timbres (pure tone and guitar). Therefore, unfortunately, the research question has to remain unanswered.

These results are not surprising, and are in line with those of Lee and Hung (2008) and Lee et al. (2014), and also consistent with the low occurrence of absolute pitch in conservatory-level western musicians reported in Deutsch et al. (2006) and Miyazaki et al. (2018). So far, to the best of the author's knowledge, only Burnham et al. (2015) have shown that musicians with absolute pitch have an advantage in lexical tone discrimination. Thus, future work is required to better understand the role of this rare ability.

Even though the current research did not find any musicians with absolute pitch, it did investigate whether the results for any of the identification and discrimination tests correlated with any of the four accuracy measures for the absolute pitch test (percentages of correct responses allowing zero-, one-, two-, and three-semitone errors), since this topic was still considered interesting to explore.

Despite none of the musicians meeting the criterion for absolute pitch, statistical analysis revealed the following significant positive correlations: accuracy in the identification pretest correlated with that in the absolute pitch test when zero-, one- and two-semitone errors were allowed. Accuracy in the Gen-2 identification test also correlated with that in the absolute pitch test when zero-semitone errors were allowed. It is also somewhat surprising that significant correlations with the accuracy measures for the absolute pitch test were found only for accuracy in the identification tasks, but not for sensitivity (d' scores) in the discrimination tasks.

However, the correlations found in this study could be associated with the nature of absolute pitch. Since absolute pitch, in other words, pitch labeling ability, involves abstracting pitch movements, categorizing and identifying them (naming or labeling notes), it can be speculated that absolute pitch would positively influence Japanese pitch-accent identification, but that discrimination tasks may not require these abilities. Indeed, almost all significant positive correlations were found for the identification *pretest*. This suggests that musical tone identification

ability helped at the beginning of learning, but that this "helpfulness" decreased abruptly, as shown by the fact that there was only one significant positive correlation, between the absolute pitch test when zero-semitone errors were allowed and accuracy in identification Gen-2 (in which novel stimuli produced by a novel talker were used). It should be reiterated, however, that none of the musicians in the current research did actually possess absolute pitch. Future studies on the current topic are therefore recommended.

Before concluding this section, it is worth briefly discussing the question of musical aptitude. As can be seen in the main aim, what was of interest in the present study was the effect of formal musical training on perceptual learning of pitch accent rather than musical aptitude per se irrespective of musical training. However, since musical aptitude could be a confounding variable in the current research, it could be argued that it would have been ideal also to assess the musical aptitude of all participants.

Nevertheless, there were several reasons for not measuring participants' musical aptitude. The first was, as anticipated in the preliminary results of the current research (Naito, 2023), for the sake of participants' time. The whole experiment took approximately four hours (see Section 3.1). Considering participants' fatigue and work burden, musical aptitude assessment was not conducted. The second reason was that, as discussed in Section 2.3.2, the findings of Götz et al. (2023) suggested the effect of musicality was mainly on account of formal musical training, not musical aptitude. The third reason was that, as mentioned above, the present study's interest lies in musical training. Additionally, as Ong et al. (2020) pointed out, currently, there is no general agreement on the definition of musicians in the literature. Indeed, to date, various studies have used different terms to describe almost the same concept, such as musical training, musical experience, musical expertise, musicality, and musicianship. Hence, the present research defined musicians as individuals currently engaged in formal tertiary-level musical training, including those enrolled in conservatories, musical institutes, or majoring in musicology at university, in order to create as homogeneous a musicians' group as possible and to better examine the effect of musical training on perceptual learning of Japanese pitch accent. Indeed, musicians

in the current study were equivalent to expert musicians as defined in Ericsson et al. (1993) and Sloboda et al. (1996), which highlighted the difference between expert musicians and amateur musicians. The last reason, related to the previous one, was that the present study excluded the intermediate category—between the expert musicians and the non-musicians—musically trained individuals but at an amateur level, comparable to the amateur musicians in Ericsson et al. (1993) and Sloboda et al. (1996), since it would have hindered accurate determination of the effect of musical training.

Since there is a growing body of literature that explores the role of musical aptitude on L2 speech perception and production (e.g., Delogu et al., 2006; M. Li & DeKeyser, 2017; P. Li et al., 2022), a future study examining whether its effects and comparing the results with those of the present study would be very interesting.

7.2. Limitations

Although the present study contributed to a better understanding of the effects of both musical training and talker variability in perceptual training stimuli (see Section 7.3 for a more detailed discussion), several limitations need to be acknowledged. This section presents discusses these limitations.

One of the limitations is related to the fact that the current experiment was conducted entirely online due to the COVID-19 pandemic. As discussed in Section 2.2.2, some training studies (Brekelmans, 2020; Brekelmans et al., 2020, 2022; Saito et al., 2022) applying the HVPT paradigm have successfully carried out their experiments entirely or partially on the Gorilla software (Anwyl-Irvine et al., 2020), also used in the present experiment. In addition, several works have created web-based HVPT applications (e.g., Inceoglu, 2022; Qian et al., 2018; Thomson, 2023) and have demonstrated that learners can engage in HVPT without supervision. However, in a laboratory setting, for example, participants perform the experiment together at the same time, whereas in this online experiment, participants engaged in the tasks at their own convenience within a time limit (see Section 3.1). From the participant's perspective, not having to go to the laboratory was an advantage, but

the fact that there was not a uniform experimental environment and that the breaks were not the same for all participants may have been a disadvantage in terms of homogeneity of participant performance. Additionally, unlike in a laboratory experiment, it was almost impossible to monitor each participant's behavior at every step in the current online experiment. Thus, while it is possible that participants were relaxed during the tests because there was no one to monitor them, at the same time it cannot be ruled out that participants were not distracted and unable to concentrate or pay attention.

To compensate for these possible weaknesses, the measures taken by Saito and colleagues' training study (2022) may be effective. They assigned one of the authors as a personal tutor to each participant and the tutors monitored participants' daily performance so that participants could complete each task in a timely manner (see Saito et al., 2022 for a detailed account). Unfortunately, the current research was unable to take this step due to financial constraints as well as the absence of personnel to engage in this matter.

Another issue with the present study was that, from the outset, the identification task results and the d' score results for the discrimination tasks revealed differences between non-musicians in the two training conditions. More specifically, while the overall identification task accuracy results and the d' score results for the discrimination tasks showed no significant differences, the identification accuracy results for each pitch-accent pattern revealed that non-musicians in the HV training condition performed significantly better than those in the LV training condition in the 2nd-syllable accented and unaccented patterns in the pretest. This imbalance in the pretest results was clearly due to individual differences rather than to experimental factors and could have been avoided, if it had been possible to counterbalance the assignment of the two training conditions to participants based on their pretest scores, as Brekelmans et al. (2022) did. Again the problem was one of budget limitations: each participant costs researchers a pre-paid token, which the Gorilla online test system monetizes as soon as a participant's test data is downloaded by the experimenter, at any stage of the experiment. It was thus decided that experiment data would be downloaded only at the end of the

experiment, for those participants who completed all the tasks, discarding the rest. Participant's data was thus only examined when they had completed the experiment.

Considering the differences between non-musicians, it is possible that some unmeasured variables could account for these individual differences. Potential variables include learning ability, analytical skills; and spatial recognition (M. Nakayama, personal communication, October 23 and 25, 2021), which may have influenced participant's ability to interpret the diagrams of simplified pitch-accent patterns. However, in view of the anticipated participant fatigue and work burden, none of these were investigated. It would be interesting in future studies to take these variables into account.

Individual differences among non-musicians are indicative of the small sample size. Unfortunately, financial and time constraints made it difficult to have a larger sample size in the current research. A larger sample size is recommended for future studies on this topic.

The last limitation to be discussed is that this study failed to answer the research question about absolute pitch. Even though correlation analyses yielded some significant positive correlations between identification accuracy and absolute pitch test accuracy (see Section 6.3 for detailed results), the current research did not find any musicians who met its criterion for absolute pitch. The difficulty of recruiting musicians with absolute pitch will need to be overcome before the role of role of absolute pitch can be further investigated.

7.3. Implications

In spite of the limitations discussed in the previous section, the results from the current dissertation have some implications including practical implications for the learning and teaching of Japanese as a foreign (FL) or second (L2) language. This section discusses these implications.

In terms of theoretical implications, the current work's empirical findings contribute to our understanding not only of the role of talker variability in perceptual training and its interaction with individuals' perceptual abilities, but also

of the effect of musical training/experience on perceptual learning of lexical tone, because prior studies exploring these topics have reported mixed findings.

To the author's knowledge, the present research is the first of its kind to investigate the interaction between musical training (rather than individuals' perceptual abilities) and talker variability. Therefore, the current results provide new experimental evidence for this interaction.

The present study also contributes to the literature by providing empirical data for a pair of languages which has not yet been well studied (Japanese and Italian). Indeed, participants in the majority of previous works on L2 Japanese pitch accent perception in perceptual learning experiments have been English native speakers. And even though the effect of musical training/experience on lexical tone perception and perceptual learning by native English speakers have been well investigated, its effects on Japanese pitch accent perception and perceptual learning in other native languages, are almost unexplored.

The findings of the current research also have a number of practical implications.

The present results confirmed the effectiveness of the HVPT method advanced by Shport's studies (2011, 2016), on which the methodology for the present study was based. Despite having a limited effect on sensitivity in the discrimination tasks, this 1-hour training method—adapted in terms of talker variability in the current research's HV training condition—helped participants improve the perception of Japanese pitch-accent patterns. Further robust evidence for its effectiveness comes from the fact that the present study's experiment was conducted entirely online and not in a laboratory as in Shport's works, and that the training was also effective for native Italian speakers (Shport's participants were native English speakers). These results imply the potential usefulness of this training method in FL/L2 Japanese teaching settings.

Recall that, even though acquiring Japanese pitch-accent contrasts is important for FL/L2 learners of Japanese (see Section 2.1.1 for a detailed discussion), not much importance is generally attached to Japanese pitch-accent

contrasts in FL/L2 Japanese classroom settings (Schaefer & Darcy, 2015; Shport, 2008, 2011, 2016). This can be explained by shortage of time, a lack of adequate course books, and a lack of knowledge of how to teach Japanese pitch-accent (Hirano, 2014; Jin, 2017; Kanamura, 2019; Oyama, 2016; Shport, 2008, 2011, 2016). Therefore, the type of perceptual training carried out in the current experiment may be a helpful means for FL/L2 Japanese teaching to fill in the gap. Future research will develop this perceptual training into a practical online tool that can be used for FL/L2 Japanese inside or outside classroom settings as well as for FL/L2 Japanese self-study.

The results obtained in the present study clearly showed the overall positive effect of musical training (except in the case of the log RT results). If musical training can facilitate Japanese pitch accent perception, it might be possible that Japanese songs would be an aid to learn/teach Japanese pitch accent. It would be interesting to explore whether the use of Japanese songs would help to learn/teach Japanese pitch accent perception.

The present findings also suggested that the best training condition for a learner can vary based on whether or not they are musically trained. In the case of musicians, both HV and LV training conditions were effective. For non-musicians, the results suggested that a HV training condition is more beneficial, especially for the discrimination tasks: the LV training was harmful according to the d' score results for the discrimination tasks. Thus, it would be ideal to ask learners about their musical training experience in order to optimize their perceptual learning of Japanese pitch accent, and for non-musicians, it would be useful to provide varied voice samples to help them to identify and discriminate Japanese pitch accent contrasts more effectively.

7.4. Conclusions

The present study set out to investigate whether musical training influences perceptual learning of Japanese pitch accent by native Italian speakers without any experience of Japanese. As a second aim, the work examined the effect of talker

variability in perceptual pitch accent training and its interaction with the effect of musical training.

To achieve these goals, musicians and non-musicians engaged in one of the two training conditions (high variability: HV, or low variability: LV), and their performance in identifying and discriminating Japanese pitch-accent patterns before and after training was compared.

An attempt was made to explore the effect of absolute pitch, but this line of inquiry remains unanswered because none of musicians who participated in the present experiment had absolute pitch.

This study has found an overall facilitative effect of musical training (except for participants' reaction latency data): overall, Italian musicians outperformed non-musicians and the difference between non-musicians and musicians widened after training.

Investigating the role of talker variability revealed different trends for musicians and non-musicians, again, except for participants' reaction latency data. This suggested that while both training conditions are effective for musicians, for non-musicians the HV training condition is more beneficial, especially for discrimination tasks. These results imply that whereas musicians have an ability to *abstract* information about Japanese pitch accent irrespective of the number of voices they hear, non-musicians do not. It may be that non-musicians need varied voice samples to learn to identify and discriminate Japanese pitch accent contrasts.

REFERENCES

Abu El Adas, S., & Levi, S. V. (2022). Phonotactic and lexical factors in talker discrimination and identification. *Attention, Perception, & Psychophysics*, *84*(5), 1788–1804. https://doi.org/10.3758/s13414-022-02485-4

Akamatsu, T. (1997). *Japanese phonetics: Theory and practice*. LINCOM Europa.

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

Alfano, I. (2006). La percezione dell'accento lessicale: Un test sull'italiano a confronto con lo spagnolo. In R. Savy & C. Crocco (Eds.), *Analisi prosodica: Teorie, modelli e sistemi di annotazione* (pp. 632–656). EDK-Editore.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Archila-Suerte, P., Bunta, F., & Hernandez, A. E. (2016). Speech sound learning depends on individuals' ability, not just experience. *International Journal of Bilingualism*, *20*(3), 231–253. https://doi.org/10.1177/1367006914552206

Ashley, R., & Timmers, R. (Eds.). (2017). *The Routledge companion to music cognition*. Routledge.

Avesani, C. (1990). A contribution to the synthesis of Italian intonation. *First International Conference on Spoken Language Processing (ICSLP 1990)*, 833–836. https://doi.org/10.21437/ICSLP.1990-106

Ayusawa T. (2003). Acquisition of Japanese Accent and Intonation by Foreign Learners. *Journal of the Phonetic Society of Japan*, *7*(2), 47–58. https://doi.org/10.24467/onseikenkyu.7.2_47

Bachem, A. (1955). Absolute Pitch. *The Journal of the Acoustical Society of America*, *27*(6), 1180–1185. https://doi.org/10.1121/1.1908155

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckman, M. E. (1986). *Stress and Non-Stress Accent:* DE GRUYTER. https://doi.org/10.1515/9783110874020

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The Original ToBi System and the Evolution of the ToBi Framework. In S.-A. Jun (Ed.), *Prosodic Typology* (pp. 9–54). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Beckman, M. E., & Pierrehumbert, J. B. (1986a). Intonational structure in Japanese and English. *Phonology*, *3*, 255–309. https://doi.org/10.1017/S095267570000066X

Beckman, M. E., & Pierrehumbert, J. B. (1986b). Japanese prosodic phrasing and intonation synthesis. *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics*, 173–180. https://doi.org/10.3115/981131.981156

Bertinetto, P. M. (1980). The perception of stress by Italian speakers. *Journal of Phonetics*, *8*(4), 385–395. https://doi.org/10.1016/S0095-4470(19)31495-0

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer [Computer program]. Version 6.1.47, retrieved 21 May 2021 from http://www.praat.org/* [Computer software].

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English / **r** / and / **l** /: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Brekelmans, G. (2020). *Phonetic vowel training for child second language learners: The role of input variability and training task* [Doctoral dissertation, UCL (University College London)]. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10089980/

Brekelmans, G., Evans, B. G., & Wonnacott, E. (2020). *Training child learners on non-native vowel contrasts: The role of talker variability* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/63dhn

Brekelmans, G., Lavan, N., Saito, H., Clayards, M., & Wonnacott, E. (2022). Does high variability training improve the learning of non-native phoneme contrasts over low variability training? A replication. *Journal of Memory and Language*, *126*, 104352. https://doi.org/10.1016/j.jml.2022.104352

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592096035. https://doi.org/10.1177/2515245920960351

Burnham, D., Brooker, R., & Reid, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music*, *43*(6), 881–897.

Caccia, M., Presti, G., Toraldo, A., Radaelli, A., Ludovico, L. A., Ogliari, A., & Lorusso, M. L. (2019). Pitch as the Main Determiner of Italian Lexical Stress Perception Across the Lifespan: Evidence From Typical Development and Dyslexia. *Frontiers in Psychology*, *10*, 1458. https://doi.org/10.3389/fpsyg.2019.01458

Chan, R. K., & Leung, J. H. (2020). Why are Lexical Tones Difficult to Learn? Insights from the Incidental Learning of Tone-Segment Connections. *Studies in Second Language Acquisition*, *42*(1), 33–59. https://doi.org/10.1017/S0272263119000482

Chandrasekaran, B., Krishnan, A., & Gandour, J. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch

contours. *Brain and Language*, *108*(1), 1–9. https://doi.org/10.1016/j.bandl.2008.02.001

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, *139*(5), 2432–2447. https://doi.org/10.1121/1.4947497

Chen, S., Zhu, Y., Wayland, R., & Yang, Y. (2020). How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers? *PLOS ONE*, *15*(5), e0232514. https://doi.org/10.1371/journal.pone.0232514

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*, *131*(6), 4756–4769. https://doi.org/10.1121/1.4714355

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America*, *105*(3), 1877–1888. https://doi.org/10.1121/1.426724

Davis, S. (2011). Quantity. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (1st ed., pp. 103–140). Wiley. https://doi.org/10.1002/9781444343069.ch4

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*(1), 46–61. https://doi.org/10.1080/09541440802708136

Delogu, F., Lampis, G., & Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing*, *7*(3), 203–207. https://doi.org/10.1007/s10339-006-0146-7

Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. M. (2018). Training-induced brain activation and functional connectivity differentiate multi-talker and single-talker speech training. *Neurobiology of Learning and Memory*, *151*, 1–9. https://doi.org/10.1016/j.nlm.2018.03.009

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute Pitch, Speech, and Tone Language: Some Experiments and a Proposed Framework. *Music Perception*, *21*(3), 339–356. https://doi.org/10.1525/mp.2004.21.3.339

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, *119*(2), 719. https://doi.org/10.1121/1.2151799

D'Imperio, M. (2002). Italian intonation: An overview and some questions. *Probus*, *14*(1), 37–69. https://doi.org/10.1515/prbs.2002.005

D'Imperio, M., & Rosenthall, S. (1999). Phonetics and phonology of main stress in Italian. *Phonology*, *16*(1), 1–28. https://doi.org/10.1017/S0952675799003681

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations.

Journal of Cognitive Neuroscience, *28*(10), 1584–1602. https://doi.org/10.1162/jocn_a_00997

Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, *7*, e7191. https://doi.org/10.7717/peerj.7191

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Eriksson, A., Bertinetto, P. M., Heldner, M., Nodari, R., & Lenoci, G. (2016). The Acoustics of Lexical Stress in Italian as a Function of Stress Level and Speaking Style. *Interspeech 2016*, 1059–1063. https://doi.org/10.21437/Interspeech.2016-348

Eriksson, A., Šimko, J., Suni, A., Vainio, M., & Nodari, R. (2020). Lexical stress perception as a function of acoustic properties and the native language of the listener. *Speech Prosody 2020*, 449–453. https://doi.org/10.21437/SpeechProsody.2020-92

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language* (Reprinted in paperback). Psychology Press.

Georgiou, G. P. (2021). Effects of Phonetic Training on the Discrimination of Second Language Sounds by Learners with Naturalistic Access to the Second Language. *Journal of Psycholinguistic Research*, *50*(3), 707–721. https://doi.org/10.1007/s10936-021-09774-3

Georgiou, G. P. (2022). The Impact of Auditory Perceptual Training on the Perception and Production of English Vowels by Cypriot Greek Children and Adults. *Language Learning and Development*, *18*(4), 379–392. https://doi.org/10.1080/15475441.2021.1977644

Gili Fivela, B. (2012). *Testing the perception of L2 intonation*. *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, 17–30.

Gili Fivela, B., Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M., & Sorianello, P. (2015). Intonational phonology of the regional varieties of Italian. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 140–197). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199685332.003.0005

Golob, N. (2003). Perception of Japanese pitch accent by Slovene native speakers: The role of music ability. *Japanese Studies: Research and Education: Tokyo University of Foreign Studies*, *8*, 105–118.

Goss, S. (2020). Exploring variation in nonnative Japanese learners' perception of lexical pitch accent: The roles of processing resources and learning context. *Applied Psycholinguistics*, *41*(1), 25–49. https://doi.org/10.1017/S0142716419000377

Goss, S. J. (2015). *The Effects of Internal and Experience-Based Factors on the Perception of Lexical Pitch Accent by Native and Nonnative Japanese Listeners* [Doctoral dissertation, The Ohio State University]. Ohio Library and Information Network (OhioLINK). https://www.semanticscholar.org/paper/The-Effects-of-Internal-and-

Experience-Based-on-the-Goss/fda4d1467e703504db5794a855c23e3cc2112cc8

Goss, S. J., & Tamaoka, K. (2019). Lexical accent perception in highly-proficient L2 Japanese learners: The roles of language-specific experience and domain-general resources. *Second Language Research*, *35*(3), 351–376. https://doi.org/10.1177/0267658318775143

Gottfried, T. L. (2007). Music and language learning: Effect of musical training on learning L2 speech contrasts. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (Vol. 17, pp. 221–237). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.21got

Gottfried, T. L., & Xu, Y. (2008). Effect of musical experience on Mandarin tone and vowel discrimination and imitation. *The Journal of the Acoustical Society of America*, *123*(5), 3887–3887. https://doi.org/10.1121/1.2935823

Götz, A., Liu, L., Nash, B., & Burnham, D. (2023). Does Musicality Assist Foreign Language Learning? Perception and Production of Thai Vowels, Consonants and Lexical Tones by Musicians and Non-Musicians. *Brain Sciences*, *13*(5), 810. https://doi.org/10.3390/brainsci13050810

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (1999). Absolute Pitch: Prevalence, Ethnic Variation, and Estimation of the Genetic Component. *The American Journal of Human Genetics*, *65*(3), 911–913. https://doi.org/10.1086/302541

Grice, M., D'Imperio, M., Savino, M., & Avesani, C. (2005). Strategies for Intonation Labelling across Varieties of Italian. In S.-A. Jun (Ed.),

Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522. https://doi.org/10.1017/S0142716403000250

Hartig, F. (2022). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (R package version 0.4.6) [Computer software]. https://CRAN.R-project.org/package=DHARMa

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hirano H. (2014). Practice of Japanese Prosody Education for Beginners in an Integrated Japanese Course: Use of Visualized Japanese Accent and Intonation Learning Material. *NINJAL Research Papers*, *7*, 45–71. https://doi.org/10.15084/00000524

Hirano-Cook, E. (2011). *Japanese pitch accent acquisition by learners of Japanese: Effects of training on Japanese accent instruction, perception, and production* [Doctoral dissertation, University of Kansas]. KU ScholarWorks. https://kuscholarworks.ku.edu/handle/1808/8022

Hirata, Y. (2004). Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts. *Computer Assisted Language Learning*, *17*(3–4), 357–376. https://doi.org/10.1080/0958822042000319629

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *The Journal of the Acoustical Society of America*, *121*(6), 3837. https://doi.org/10.1121/1.2734401

Iino, A. (2019). Effects of HVPT on perception and production of English fricatives by Japanese learners of English. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouësny, *CALL and complexity – short papers from EUROCALL 2019* (1st ed., pp. 186–192). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.1007

Inceoglu, S. (2022). A Web-Based High Variability Phonetic Training Application for French Vowels. *Virtual PSLLT*. Virtual PSLLT. https://doi.org/10.31274/psllt.13336

Ingvalson, E. M., Ettlinger, M., & Wong, P. C. M. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, *18*(1), 35–47. https://doi.org/10.1177/1367006912456586

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. https://doi.org/10.1121/1.3148196

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278. https://doi.org/10.1121/1.2062307

Iverson, P., Herrero, B. P., & Katashima, A. (2023). Memory-card vowel training for child and adult second-language learners: A first report. *JASA Express Letters*, *3*(1), 015202. https://doi.org/10.1121/10.0016836

Jiang, N. (2012). *Conducting Reaction Time Research in Second Language Studies*. Routledge.

Jin Z. (2017). A report on using Acoustic analysis software Praat to teach Japanese speech. *Study of Classroom of the Center for Japanese Language and Culture Osaka University*, *15*, 45–62. https://doi.org/10.18910/60425

Kanamura K. (2019). Developing simple methods for teaching Japanese pronunciation: Mora rhythm and pitch accent. *The Journal of Science of Culture and Humanities*, *98*, 1–19. https://doi.org/10.15040/00000350

Kanamura K. (2020). Improvements to Phonetic Education That Would Aid in Teaching Japanese Pitch Accent: Suggestions from a Questionnaire Survey of Japanese Language Teacher. *Journal of the Phonetic Society of Japan*, *24*, 36–48. https://doi.org/10.24467/onseikenkyu.24.0_36

Kassambara, A. (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (R package version 0.7.2) [Computer software]. https://CRAN.R-project.org/package=rstatix

Kawahara, S. (2015). The phonology of Japanese accent. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 445–492). DE GRUYTER. https://doi.org/10.1515/9781614511984.445

Kirkham, J., Lu, S., Wayland, R., & Kaan, E. (2011). Comparison of Vocalists and Instrumentalists on Lexical Tone Perception and Production Tasks. *Online Proceedings of the ICPhS XVII 2011*, 1098–1101. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Kirkham,%20Joe/Kirkham,%20Joe.pdf

Kitahara, M. (2001). *Category Structure and Function of Pitch Accent in Tokyo Japanese* [Doctoral dissertation, Indiana University]. http://www.f.waseda.jp/kitahara/Paper/thesis-dist.pdf

Kourakata R., & Nagato M. (2014). A Report on Reading Aloud Activities for an Advanced Beginner's Class. *Tsukuba daigaku ryugakusei kyoiku senta nihongo kyoiku ronshu [International Student Center of the University of Tsukuba, Journal of Japanese language teaching]*, *29*, 105–117.

Krämer, M. (2009). *The phonology of Italian*. Oxford University Press.

Krämer, M. (2021). 18 Italian. In C. Gabriel, R. Gess, & T. Meisenburg (Eds.), *Manual of Romance Phonetics and Phonology* (pp. 559–596). De Gruyter. https://doi.org/10.1515/9783110550283-019

Kubozono, H. (2012). Varieties of pitch accent systems in Japanese. *Lingua*, *122*(13), 1395–1414. https://doi.org/10.1016/j.lingua.2012.08.001

Kubozono, H. (2015). Introduction to Japanese phonetics and phonology. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 1–40). DE GRUYTER. https://doi.org/10.1515/9781614511984.1

Kubozono, H. (2018). Pitch Accent. In Y. Hasegawa (Ed.), *The Cambridge Handbook of Japanese Linguistics* (pp. 154–180). Cambridge University Press.

Labrune, L. (2012). *The phonology of Japanese*. Oxford University Press.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed). Cambridge University Press.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Blackwell Pub.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247. https://doi.org/10.1017/S0142716405050150

Laméris, T. J., & Graham, C. (2020). L2 Perception and Production of Japanese Lexical Pitch: A Suprasegmental Similarity Account. *Journal of Monolingual and Bilingual Speech*, *2*(1), 106–136. https://doi.org/10.1558/jmbs.14948

Laméris, T. J., & Post, B. (2023). The combined effects of L1-specific and extralinguistic factors on individual performance in a tone categorization and word identification task by English-L1 and Mandarin-L1 speakers.

Second Language Research, *39*(3), 833–871. https://doi.org/10.1177/02676583221090068

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235–3248.

Lee, C.-Y., Lekich, A., & Zhang, Y. (2014). Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.*, *135*(3), 10.

Lee, C.-Y., Tao, L., & Bond, Z. S. (2009). Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics*, *37*(1), 1–15. https://doi.org/10.1016/j.wocn.2008.08.001

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757–3768. https://doi.org/10.1121/1.3506351

Lenth, R. V. (2023). *Estimated marginal means aka least-square means* (R package version 1.9.0) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/BF03206733

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. https://doi.org/10.1017/S0272263116000358

Li, P., Zhang, Y., Fu, X., Baills, F., & Prieto, P. (2022). *Melodic perception skills predict Catalan speakers' imitation abilities of unfamiliar languages*. 876–880. https://doi.org/10.21437/SpeechProsody.2022-178

Li, Y., & Lee, G. (2021). The Effect of Perceptual Training on Teaching Mandarin Chinese Tones. In C. Yang (Ed.), *The Acquisition of Chinese as a Second Language Pronunciation* (pp. 107–139). Springer Singapore. https://doi.org/10.1007/978-981-15-3809-4_5

Li, Y., Lee, G., & Sereno, J. A. (2019). Comparing Monosyllabic and Disyllabic Training in Perceptual Learning of Mandarin Tone. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 303–319). Department of English School of Communication & Culture Aarhus University.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087. https://doi.org/10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Loui, P. (2016). Absolute Pitch. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 81–94). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.10

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates. https://www.routledge.com/Detection-Theory-A-Users-Guide/Macmillan-Creelman/p/book/9780805842319

Maggu, A. R., Wong, P. C. M., Liu, H., & Wong, F. C. K. (2018). Experience-dependent Influence of Music and Language on Lexical Pitch Learning Is Not Additive. *Interspeech 2018*, 3791–3794. https://doi.org/10.21437/Interspeech.2018-2104

Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Mancini, F., & Voghera, M. (1994). Lunghezza, tipi di sillabe e accento in italiano. *Archivio Glottologico Italiano*, *79*, 51–77.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of Musical Expertise on Segmental and Tonal Processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715. https://doi.org/10.1162/jocn.2010.21585

Martin, I. A., & Inceoglu, S. (2022). The Laboratory, the Classroom, and Online: What Works in Each Context. In J. Levis, T. M. Derwing, & S. Sonsaat-

Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 254–272). John Wiley & Sons, Inc.

Marvin, E. W. (2017). Musical Connections: Absolute Pitch. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition* (pp. 203–212). Routledge Handbooks Online. https://doi.org/10.4324/9781315194738.ch17

Masuda-Katsuse, I. (2006). Contribution of pitch-accent information to Japanese spoken-word recognition. *Acoustical Science and Technology*, *27*(2), 97–103. https://doi.org/10.1250/ast.27.97

Matsuzaki, H. (2000). Minimal-pair of Accent for Japanese Language Learners at Elementary Level. *Bulletin of the Department of Teaching Japanese as a Second Language, Hiroshima University*, *10*, 39–46. https://doi.org/10.15027/25157

Maturi, P. (2007). *I suoni delle lingue, i suoni dell'italiano: Introduzione alla fonetica*. Il Mulino.

McGuire, G. (2010). A Brief Primer on Experimental Designs for Speech Perception Research. *Laboratory Report*, *77*(1), 2–19.

Miyazaki, K., Rakowski, A., Makomaska, S., Jiang, C., Tsuzaki, M., Oxenham, A. J., Ellis, G., & Lipscomb, S. D. (2018). Absolute Pitch and Relative Pitch in Music Students in the East and the West. *Music Perception*, *36*(2), 135–155. https://doi.org/10.1525/mp.2018.36.2.135

Mok, P. K. P., & Zuo, D. (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *The Journal of the*

Acoustical Society of America, *132*(4), 2711–2720. https://doi.org/10.1121/1.4747010

Monaghan, P., Ruiz, S., & Rebuschat, P. (2021). The role of feedback and instruction on the cross-situational learning of vocabulary and morphosyntax: Mixed effects models reveal local and global effects on acquisition. *Second Language Research*, *37*(2), 261–289. https://doi.org/10.1177/0267658320927741

Moon, C., Huang, C., & Hashimoto, D. (2021, June 11). *Hogen to ongaku—Ongakuteki chikaku ni taisuru oncho patan no eikyo—[Dialects and Music—Effect of the pitch pattern on Musical perception]* [Conference session]. Kokuritsu kokugo kenkyujo taisho gengogaku purojekuto purosodi kenkyuhan online kenkyukai [NINJAL Cross-linguistic Studies Project, Prosody Team Online Seminar], Online.

Naito, Y. (2023). Does Musical Training Influence Perceptual Learning of Japanese Pitch Accent? The Case of Native Italian Speakers. In G. P. Georgiou, A. Giannakou, & C. Savvidou (Eds.), *Advances in Second/Foreign Language Acquisition* (pp. 1–18). Springer International Publishing. https://doi.org/10.1007/978-3-031-38522-3_1

Nespor, M., & Bafile, L. (2008). *I suoni del linguaggio*. Il mulino.

NHK Hoso Bunka Kenkyujo (Ed.). (2016). *NHK Nihongo hatsuon akusento shin jiten [New dictionary of Japanese pronunciation and accent].* NHK Shuppan.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese Listeners to Perceive American English Vowels: Influence of Training Sets. *Journal of Speech,*

Language, and Hearing Research, *50*(6), 1496–1509. https://doi.org/10.1044/1092-4388(2007/103)

Nishi, K., & Kewley-Port, D. (2008). Nonnative Speech Perception Training Using Vowel Subsets: Effects of Vowels in Sets and Order of Training. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Nishinuma, Y., Arai, M., & Ayusawa, T. (1996). Perception of tonal accent by Americans learning Japanese. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, *2*, 646–649. https://doi.org/10.1109/ICSLP.1996.607444

Ong, J. H., Tan, S. H., Chan, A. H. D., & Wong, F. C. K. (2020). The Effect of Musical Experience and Congenital Amusia on Lexical Tone Perception, Production, and Learning: A Review. In H. Liu, F. Tsao, & P. Li (Eds.), *Speech Perception, Production and Acquisition: Multidisciplinary approaches in Chinese languages* (pp. 139–158). Springer.

Otake, T. (2015). Mora and mora-timing. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 493–524). DE GRUYTER. https://doi.org/10.1515/9781614511984.493

Otake, T., & Cutler, A. (1999). Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics*, *27*(3), 229–253. https://doi.org/10.1006/jpho.1999.0095

Oyama R. (2016). Acquisition of Japanese accent phrase by Japanese language learners: Effects of the Japanese teaching method by focusing on

pronunciations. *Bulletin of Center for Japanese Language and Culture*, *14*, 91–103. https://doi.org/10.14988/pa.2017.0000014443

Pappalardo, G. (2018). L'accento tonale del giapponese percepito da discenti italofoni: Un'indagine fonetico-percettiva. In P. Villani, N. Hayashi, & L. Capponcelli (Eds.), *Riflessioni sul Giappone antico e moderno* (pp. 127–150). Aracne editrice.

Parncutt, R., & Levitin, D. J. (2001). Absolute pitch. In S. Sadie (Ed.), *New Grove Dictionary of Music and Musicians* (Vol. 1, pp. 37–39). MacMillan. https://doi.org/10.1093/gmo/9781561592630.article.00070

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford University Press, USA.

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. https://doi.org/10.1121/1.3593366

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. MIT Press.

Podesva, R. J., & Zsiga, E. (2014). Sound recordings: Acoustic and articulatory data. In R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (1st ed., pp. 169–194). Cambridge University Press. https://doi.org/10.1017/CBO9781139013734.010

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and

Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684–1696. https://doi.org/10.1121/1.2161427

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *System for Adaptive High-Variability Segmental Perceptual Training: Implementation, Effectiveness, Transfer.*, *22*(1), 69–96.

Qin, Z., Jin, R., & Zhang, C. (2022). The Effects of Training Variability and Pitch Aptitude on the Overnight Consolidation of Lexical Tones. *Journal of Speech, Language, and Hearing Research*, *65*(9), 3377–3391. https://doi.org/10.1044/2022_JSLHR-22-00058

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rose, H. (2017). *The Japanese Writing System*. Multilingual Matters. https://doi.org/10.21832/ROSE8156

Rossi, M. (1998). Intonation in Italian. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 219–241). Cambridge University Press.

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, *134*(2), 1324–1335. https://doi.org/10.1121/1.4812767

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01318

Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and Multimodal High Variability Phonetic Training: Potential, Limits, and Future Directions. *Language Learning*, *72*(4), 1049–1091. https://doi.org/10.1111/lang.12503

Sakamoto, E. (2011). *Investigation of factors behind foreign accent in the L2 acquisition of Japanese lexical pitch accent by adult English speakers* [Doctoral dissertation, The University of Edinburgh]. https://era.ed.ac.uk/bitstream/handle/1842/5692/Sakamoto2011.pdf?sequence=1&isAllowed=y

Sato T. (1995). Tan'on to inritsu ga nihongo onsei no hyoka ni ataeru eikyoryoku no hikaku [The Comparison of Phone and Prosody in Evaluation of Spoken Japanese]. *Japanese language education around the globe*, *5*, 139–154. https://doi.org/10.20649/00000219

Schaefer, V., & Darcy, I. (2015). A Communicative Approach and Dialect Exposure Enhance Pitch Accent Awareness by Learners of Japanese. *Pronunciation in Second Language Learning and Teaching Proceedings*, *6*(1), Article 1. https://www.iastatedigitalpress.com/psllt/article/id/15271/

Schmid, S. (1999). *Fonetica e fonologia dell'italiano*. Paravia scriptorium.

Sekiguchi, T., & Nakajima, Y. (1999). The Use of Lexical Prosody for Lexical Access of the Japanese Language. *Journal of Psycholinguistic Research*, *28*(4), 439–454. https://doi.org/10.1023/A:1023245216726

Shibata, T., & Hurtig, R. R. (2007). Prosody Acquisition by Japanese Learners. In Z. Han (Ed.), *Understanding Second Language Process* (pp. 176–204). Multilingual Matters. https://doi.org/10.21832/9781847690159-012

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *Journal of Phonetics*, *66*, 242–251. https://doi.org/10.1016/j.wocn.2017.11.002

Shport, I. A. (2008). Acquisition of Japanese Pitch Accent by American Learners. In P. Heinrich & Y. Sugita (Eds.), *Japanese as foreign language in the age of globalization*. Iudicium-Verl.

Shport, I. A. (2011). *Cross-Linguistic Perception and Learning of Japanese Lexical Prosody by English Listeners* [Doctoral dissertation, University of Oregon]. Scholars' Bank. http://hdl.handle.net/1794/12087

Shport, I. A. (2015). Perception of acoustic cues to Tokyo Japanese pitch-accent contrasts in native Japanese and naive English listeners. *The Journal of the Acoustical Society of America*, *138*(1), 307–318. https://doi.org/10.1121/1.4922468

Shport, I. A. (2016). Training English Listeners to Identify Pitch-Accent Patterns in Tokyo Japanese. *Studies in Second Language Acquisition*, *38*(4), 739–769.

Silpachai, A. (2020). The role of talker variability in the perceptual learning of Mandarin tones by American English listeners. *Journal of Second Language Pronunciation*, 209–235.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2023). *afex: Analysis of Factorial Experiments* (R package version 1.3-0) [Computer software]. https://CRAN.R-project.org/package=afex

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology*, *87*(2), 287–309. https://doi.org/10.1111/j.2044-8295.1996.tb02591.x

Sugiyama, Y. (2012). *The production and perception of Japanese pitch accent*. Cambridge Scholars Pub.

Sugiyama, Y. (2017). Perception of Japanese Pitch Accent without F0. *Phonetica*, *74*(2), 107–123. https://doi.org/10.1159/000453069

Sugiyama, Y. (2022). Identification of Minimal Pairs of Japanese Pitch Accent in Noise-Vocoded Speech. *Frontiers in Psychology*, *13*, 887761. https://doi.org/10.3389/fpsyg.2022.887761

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*(2), 345–361. https://doi.org/10.1037/0033-2909.113.2.345

Tan, S.-L., Pfordresher, P., & Harré, R. (2018). *Psychology of music: From sound to significance* (Second edition). Routledge.

Taylor, R. L. (2012). *Eigo washa ni yoru nihongo no go akusento no shūtoku [The acquisition of Japanese lexical accent by English speakers]* [Doctoral dissertation, Nagoya University]. NAGOYA Repository. https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1429657750&disposition=inline

Thomson, R. I. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

Thomson, R. I. (2022). Perception in Pronunciation Training. In J. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching* (pp. 42–60). John Wiley & Sons, Inc.

Thomson, R. I. (2023). *English Accent Coach [Computer program]. Version 3.0. Www.englishaccentcoach.com* [Computer software].

Toda T. (2001). The effect of pronunciation practice upon the perception of Japanese accents. *Bulletin of Center for Japanese Language, Waseda University*, *14*, 67–88.

Tong, X., & Tang, Y. C. (2016). Modulation of musical experience and prosodic complexity on lexical pitch learning. *Speech Prosody 2016*, 217–221. https://doi.org/10.21437/SpeechProsody.2016-45

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge University Press.

Vance, T. J. (2018). Moras and Syllables. In Y. Hasegawa (Ed.), *The Cambridge handbook of Japanese linguistics* (pp. 135–153). Cambridge University Press.

Venditti, J. J. (2005). The J_ToBi Model of Japanese Intonation. In S.-A. Jun (Ed.), *Prosodic Typology* (pp. 172–200). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Venditti, J. J. (2006). Prosody in sentence processing. In M. Nakayama, R. Mazuka, Y. Shirai, & P. Li (Eds.), *The Handbook of East Asian Psycholinguistics* (1st ed., pp. 208–217). Cambridge University Press. https://doi.org/10.1017/CBO9780511758652.031

Vietti, A. (2019). Phonological Variation and Change in Italian. In M. Loporcaro & F. Gardani (Eds.), *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.494

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. https://doi.org/10.1121/1.1531176

Wang, Y., & Kuhl, P. K. (2003). *Evaluating the "Critical Period" Hypothesis: Perceptual Learning of Mandarin Tones in American Adults and American Children at 6, 10 and 14 Years of Age*. 15th International Congress of Phonetic Sciences (ICPhS-15), Barcelona, Spain. http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_1537.pdf

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. https://doi.org/10.1121/1.428217

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics*, *24*(1), 113–129. https://doi.org/10.1017/S0142716403000067

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*(4), 654–662. https://doi.org/10.1016/j.wocn.2010.10.001

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. https://doi.org/10.3758/BF03207136

Wiener, S., & Goss, S. (2019). Second and Third Language Learners' Sensitivity to Japanese Pitch Accent Is Additive. *Studies in Second Language Acquisition*, *41*(04), 897–910. https://doi.org/10.1017/S0272263119000068

Wong, J. W. S. (2012). Training the Perception and Production of English // and // of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, 37–40. https://assta.org/proceedings/sst/SST-12/SST2012/PDF/AUTHOR/ST120021.PDF.

Wong, J. W. S. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels /e/-/æ/ by Cantonese ESL learners with high and low L2 proficiency levels. *Interspeech 2014*, 524–528. https://doi.org/10.21437/Interspeech.2014-129

Wong, J. W. S. (2015). The Impact of L2 Proficiency on Vowel Training. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English Pronunciation* (pp. 219–239). Palgrave Macmillan UK. https://doi.org/10.1057/9781137509437_10

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. https://doi.org/10.1017/S0142716407070312

Wu, X., Kawase, S., & Wang, Y. (2017). Effects of acoustic and linguistic experience on Japanese pitch accent processing. *Bilingualism: Language and Cognition*, *20*(5), 931–946. https://doi.org/10.1017/S1366728916000559

Wu, X., Tu, J.-Y., & Wang, Y. (2012). Native and nonnative processing of Japanese pitch accent. *Applied Psycholinguistics*, *33*(3), 623–641. https://doi.org/10.1017/S0142716411000506

Yoshida, N. (2002). The Effects of Phonetic Environment on Vowel Devoicing in Japanese. *Kokugogaku*, *53*(3), 34–47.

Zhang, X., Cheng, B., & Zhang, Y. (2021). The Role of Talker Variability in Nonnative Phonetic Learning: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4802–4825. https://doi.org/10.1044/2021_JSLHR-21-00181

Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, *137*(3), 1452–1463. https://doi.org/10.1121/1.4913457

APPENDIX A

Participants: informed consent on the processing of personal data for scientific research purposes

INFORMATIVA SUL TRATTAMENTO DEI DATI PERSONALI PER FINALITA' DI RICERCA SCIENTIFICA (ART. 13 REGOLAMENTO UE 2016/679-RGDP)

Titolo del Progetto di ricerca (di seguito "Progetto"): The effects of music training/experience on the perceptual learning of Japanese Pitch accent by Italians.

Lo studio si propone di indagare l'effetto di esperienze musicali sulla percezione dell'accento melodico giapponese da parte di soggetti di madrelingua italiana. Nello studio si raccoglieranno informazioni su età, conoscenze linguistiche e esperienza musicale dei partecipanti. Inoltre verrà chiesto ai partecipanti di identificare caratteristiche e discriminare differenze tra stimoli acustici elaborati allo scopo. I dati personali e le risposte verranno raccolti ricorrendo a software dedicati. I dati e le risposte verranno trattati in forma anonima e in modalità aggregata al solo fine di ricavarne indicazioni statistiche anonime. Lo studio è condotto nel quadro del Dottorato di ricerca in scienze linguistiche in convenzione tra l'Università degli studi di Pavia e l'Università degli studi di Bergamo. Il responsabile scientifico dello studio è Cristian Pallone, tutor incaricato dal collegio del dottorato di ricerca in scienze linguistiche.

Titolare del trattamento e Responsabile protezione dati
L'Università degli Studi di Pavia, in qualità di Titolare del trattamento**,** nella persona del Magnifico Rettore (sede C.so Strada Nuova n. 65, 27100 Pavia, PEC amministrazione- centrale@certunipv.it), tratterà i Suoi dati personali soltanto nella misura in cui siano indispensabili in relazione all'obiettivo dello studio, nel rispetto di quanto previsto dalla normativa vigente in materia di protezione dei dati personali, dal Regolamento (UE) 2016/679 (RGPD) dal D. lgs. n. 196/2003 "Codice in materia di protezione dei dati personali" come modificato dal D. lgs. 101/2018 e conformemente alle "Regole deontologiche per i trattamenti a fini statistici o di ricerca scientifica" (Provvedimento del Garante n. 515 del 19/12/2018) [30] Ai sensi dell'articolo n. 37 del Regolamento (UE) 2016/679, il Titolare del Trattamento ha nominato un proprio Responsabile Protezione Dati (dati di contatto: sede C.so Strada Nuova n.65, IT-27100 Pavia, PEC amministrazione-centrale@certunipv.it; Email privacy@unipv.it). Ai sensi della normativa sopra indicata il trattamento dei Suoi dati personali da parte dei ricercatori coinvolti

[30] (This footnote was in the original document) Le Regole deontologiche si applicano all'insieme dei trattamenti effettuati per scopi statistici e scientifici -conformemente agli standard metodologici del pertinente settore disciplinare–, di cui sono titolari università, altri enti o istituti di ricerca e società scientifiche, nonché ricercatori che operano nell'ambito di dette università, enti, istituti di ricerca e soci di dette società scientifiche. Il codice **non si applica ai trattamenti** per scopi statistici e scientifici connessi con attività di tutela della salute svolte da esercenti professioni sanitarie od organismi sanitari, ovvero con attività comparabili in termini di significativa ricaduta personalizzata sull'interessato, che restano regolati dalle pertinenti disposizioni.

nell'attività di ricerca sarà improntato al rispetto dei principi di cui all'art. 5 del RGPD e, in particolare, a quelli di liceità, correttezza, trasparenza, pertinenza, non eccedenza ed in modo da garantire un'adeguata sicurezza dei dati personali.

Finalità del trattamento e base giuridica

Il trattamento dei Suoi dati personali è effettuato per la realizzazione delle finalità scientifiche del progetto: The effects of music training/experience on the perceptual learning of Japanese Pitch accent by Italians.

Il Progetto è stato redatto conformemente agli standard metodologici del settore disciplinare interessato ed è depositato presso il Dipartimento di Studi Umanistici dell'Università degli Studi di Pavia ove verrà conservato per cinque anni dalla conclusione programmata della ricerca stessa.

Il trattamento dei Suoi dati personali viene effettuato dal Titolare nell'ambito di esecuzione dei propri compiti di interesse pubblico ai sensi dell'art. 6, paragrafo 1, lett. e) del RGPD.

Il trattamento delle categorie particolari di dati personali (dati sensibili) viene effettuato per fini di ricerca scientifica ai sensi dell'art. 9, paragrafo 2, lett. g) e j) del RGPD e sulla base di un consenso esplicito da Lei prestato ai sensi dell'art. 7, comma 2, lett. a) delle Regole deontologiche per i trattamenti a fini statistici o di ricerca scientifica.

Categoria e tipologia di dati personali trattati

La realizzazione del Progetto implica il trattamento dei seguenti dati personali:

Risposte del questionario (età, conoscenze linguistiche, esperienza musicale) e dell'esperimento percettivo (test di identificazione/discriminazione di stimoli audio).

Modalità del trattamento

Il trattamento dei Suoi dati verrà effettuato mediante strumenti elettronici, es. Gorilla (https://gorilla.sc), PRAAT (https://www.fon.hum.uva.nl/praat/), Microsoft Excel e R e adottando le seguenti misure di sicurezza: anonimizzazione dei dati e loro trattamento in modalità aggregata al solo fine di ricavare indicazioni statistiche. I Suoi dati personali saranno trattati esclusivamente dal Titolare e/o da soggetti autorizzati nell'ambito della realizzazione del Progetto.

Periodo di conservazione dei dati

I Suoi dati personali saranno conservati fino al raggiungimento delle finalità del Progetto, nei limiti stabiliti dalle leggi che regolano la materia: cinque anni dalla conclusione programmata della ricerca stessa (31/12/2022).

Natura del conferimento dei dati

Il conferimento pur essendo facoltativo è necessario per le suddette finalità di ricerca è indispensabile per lo svolgimento del Progetto ed è strettamente connesso allo svolgimento di attività di interesse pubblico. Il mancato conferimento determina l'impossibilità di partecipare al Progetto.

Destinatari dei dati ed eventuale trasferimento all'estero

I Suoi dati personali potranno essere comunicati in forma anonima e/o aggregata ai seguenti soggetti:
Altre Università, istituzioni e organismi pubblici e privati aventi finalità di ricerca, esclusivamente nell'ambito di progetti congiunti, in particolare i membri del Collegio di ricerca del dottorato in scienze linguistiche in convenzione tra l'Università degli studi di Pavia e l'Università degli studi di Bergamo. Altre Università, istituzioni e organismi pubblici e privati, aventi finalità di ricerca e non partecipanti a progetti congiunti, limitatamente ad informazioni prive di dati identificativi e per scopi storici o scientifici chiaramente determinati per iscritto nella richiesta dei dati.

In tali casi, si applicano le ulteriori garanzie previste dal Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e scientifici.

I dati raccolti non saranno oggetto di trasferimento in Paesi non appartenenti all'UE.

Divulgazione dei risultati della ricerca

La divulgazione dei risultati statistici e/o scientifici (ad esempio mediante pubblicazione di articoli scientifici e/o la creazione di banche dati, anche con modalità ad accesso aperto, partecipazione a convegni, ecc.) potrà avvenire soltanto in forma anonima e/o aggregata e comunque secondo modalità che non La rendano identificabile.

Diritti dell'Interessato

In qualità di Interessato ha diritto di chiedere in ogni momento al Titolare l'esercizio di diritti di cui agli artt. 15 e ss. del RGPD e, in particolare, l'accesso ai propri dati personali, la rettifica, l'integrazione, la cancellazione, la limitazione del trattamento che la riguardi o di opporsi al loro trattamento. Ai sensi dell'art. 17, paragrafo 3, lett. d) il diritto alla cancellazione non sussiste per i dati il cui trattamento sia necessario ai fini di ricerca scientifica qualora rischi di rendere impossibile e/o pregiudicare gravemente gli obiettivi della ricerca stessa.

Per l'esercizio dei suddetti diritti può contattare il Titolare e/o il Responsabile della protezione dei dati di Ateneo ai recapiti sopraindicati.
Resta salvo il diritto di proporre reclamo al Garante per la protezione dei dati personali.

Per informazioni relative al Progetto può rivolgersi al Responsabile scientifico del progetto al seguente recapito: *omissis* [31]

Consenso

☐ Dichiaro di aver preso visione dell'INFORMATIVA SULLA PROTEZIONE DEI DATI sopra riportata ai sensi degli articoli 13 e 14 del Regolamento UE 2016/679 e, dichiarando di aver compiuto 18 anni, AUTORIZZO LA RACCOLTA E IL TRATTAMENTO DEI DATI per l'accesso alle attività del progetto "The effects of music training/experience on the perceptual learning of Japanese Pitch accent by Italians".

[31] In the original document, the email address of the author's supervisor was written. Here, it has been eliminated for reasons of privacy.

APPENDIX B

Native Japanese speakers who recorded their voice: informed consent on the processing of personal data for scientific research purposes

「音楽学習経験が日本語アクセント知覚習得に及ぼす影響：イタリア語母語話者の場合」研究の説明および同意書

パヴィア大学大学院言語学研究科
博士課程　内藤　由佳

本研究を次のように実施致します。研究の目的や実施内容等をご理解いただき、本研究にご参加いただける場合は、その旨ご同意いただければと存じます。研究に参加しない、あるいは一度参加を決めた後に途中で辞退されることになっても、不利益を被ることはありません。ご自身の意思で、研究にご参加いただけましたら幸いです。なお、本研究はベルガモ大学大学院 言語学研究科 Lorenzo Spreafico 教授の指導のもと、パヴィア大学大学院 言語学研究科 博士課程 内藤由佳が実施致します。

1. 研究の意義・目的

この研究は、日本語学習経験のないイタリア人学生を被験者として、下記の三点を主な目的として実施致します。

- 被験者がどのように日本語のアクセントを知覚するのか。
- 被験者の知覚は実験内で行うトレーニングを通じて改善するのか。
- 被験者の音楽学習経験の有無が日本語のアクセント知覚の習得に影響を及ぼすのか。及ぼす場合はどのような影響なのか。

日本語アクセントの知覚とその習得における音楽学習経験の影響が明らかになることによって、今後の外国人への日本語アクセント教育に貢献できるのではないかと思っております。

2. 研究方法、研究期間

この研究では、イタリア人学生に、日本語の短い文章を聞かせ、アクセント型の判断等をオンラインでの実験を通じてもらう予定です。オンライン実験にはソフトウェア Gorilla を用います。（https://gorilla.sc）ご協力いただく内容は、上記の実験でイタリア人に聞かせる文章となる、日本語で書かれた短い文章を複数朗読していただいたものの録音となります。具体的には以下の二点です。

1. 日本語の文章を複数朗読していただいたものの録音。
2. ご出身地等のアンケート（無記名式であり、個人を特定できる情報は伺いません）へのご回答。

研究期間は 2022 年 12 月 31 日までを予定しております。

3. 研究協力者として選定された理由

この研究にご協力いただける方は東京都またはその近郊（神奈川県・千葉県・埼玉県）ご出身かつお住まいの方（標準語が母語の方）とさせていただいております。

4. 研究への参加と撤回について

研究の趣旨をご理解いただきご参加いただければと思いますが、参加するかどうかはご自身で決定して下さい。説明を聞いてお断りいただくこともできますので、研究の辞退については、研究者に口頭もしくはメールにてお知らせ下さい。お断りになる、あるいは、一度参加を決めてから途中で辞退されることになっても、何ら不利益を被ることはありません。また、途中で参加を辞めることもできます。その際には、それまでに収集したデータを分析対象としてよいのか、廃棄を希望されるのかをお聞かせいただければ、それに従ってデータを取り扱います。

5. 研究に参加することにより期待される利益

この研究に参加することにより、直接的にあなたの利益となることはありません。音楽学習経験の有無が日本語のアクセント知覚を習得する際に及ぼす影響を明らかにすることによって、第二言語習得に影響する要因への理解が進み、今後の外国人への日本語教授法への一助となることによって社会に貢献することを期待しています。なお、ご参加時にヘッドセットをご自身でご用意いただける場合は謝礼としてAmazon ギフト券（3,000 円）を、ヘッドセットのご用意が困難な場合は、ヘッドセットを謝礼としてご送付致します。

6. 予測されるリスク、危険、心身に対する不快な状態や影響

この研究の参加には、何ら身体的な危険は伴いません。録音を中断された方は一旦研究を辞退されたこととなりますが、改めて録音に参加いただける場合は、研究者にお伝え下さい。

7. 研究成果の公表の可能性

この研究の成果は、博士論文としてまとめるとともに、できましたら複数の学会にてポスター発表を行う予定です。論文や発表では、個人が特定できない表記に致します。

また、完成後の博士論文につきましては、概要報告をご希望の場合は内容について資料及び口頭でご説明させていただきますので、ご希望なさる場合はおっしゃって下さい。

8. 守秘や個人情報、研究データの取り扱いについて

収集するデータは、定型文を朗読していただいた録音以外には、ご協力者様の性別・ご出身地・年齢です。上記以外の情報（お名前・メールアドレス等）を録音データに結びつけることはございません。また、録音データは、音声分析ソフトウェア Praat （https://www.fon.hum.uva.nl/praat/ ）を使用しての分析と、ソフトウェア Gorilla（https://gorilla.sc）を用いたイタリア人被験者を対象にしたオンライン知覚実験に使用させていただきます。なお、同意書を含む、全てのデータは内藤由佳の責

任下にて研究のため５年間保管し、５年経過後には同意書を含む、全てのデータを廃棄致します。

9.　利益相反（あるいは責務相反）について

利益相反状態あるいは責務相反状態が生じる可能性があると思われる場合は、その詳細と管理方法（本大学の研究者として社会的責任を果たすための情報開示、説明責任等）について記入して下さい。

10. 研究者、および問い合わせ先について

この研究は、パヴィア大学大学院 言語学研究科・博士課程の内藤由佳が行います。研究内容に関するご質問は、以下の連絡先までご連絡下さい。

研究実施者：　内藤由佳（パヴィア大学大学院言語学研究科・博士課程）

住所　Corso Strada Nuova n. 65, 27100 Pavia, Italy（所属キャンパスの所在住所）

連絡先：【省略】[32]

研究参加の同意書

私は、「音楽学習経験が日本語アクセント知覚習得に及ぼす影響：イタリア語母語話者の場合」研究について以上の事項について説明を受けました。研究の目的、方法等について理解し、研究に参加致します。

研究成果の報告について

・ 博士論文の概要報告　（　希望する　・　希望しない　）

[32] In the original document, the email address and mobile number of the author was written. Here, it has been eliminated for reasons of privacy.

APPENDIX C

Preliminary questionnaire for participants

(1) For musicians:

1. Qual è il tuo sesso?

☐ Maschio

☐ Femmina

☐ Preferisco non dire

2. Quanti anni hai?

3. Da dove vieni? (scegli il nome della tua regione di provenienza)

4. A quale conservatorio di musica / istituto superiore di studi musicali / università sei iscritto/a?

5. A quale anno sei attualmente iscritto/a?

6. In che cosa ti stai specializzando? (Specifica in quale strumento musicale/canto, ecc.)

7. Quanti anni avevi quando hai cominciato a frequentare lezioni di musica regolarmente?

8. Che cosa (strumenti e/o canto) hai praticato? Elenca da quello che hai praticato di più a quello che hai praticato di meno.

9. Per quanti anni in totale hai preso lezioni di musica (sia in istituto che privatamente)? Se hai preso lezioni su più di uno strumento musicale, incluso il canto, scrivi il numero di anni per ognuno. Es. Pianoforte 10 anni, violino 3 anni.

10. Oltre all'italiano, quale/i lingua/e conosci? Scrivi accanto al nome della/e lingua/e anche il tuo livello secondo il Quadro comune Europeo di riferimento (es. spagnolo B2). Se non hai idea di che livello, clicca qui[33]!

[33] The link inserted here was: https://rm.coe.int/168045bc72

(2) For non-musicians:

1. Qual è il tuo sesso?

☐ Maschio

☐ Femmina

☐ Preferisco non dire

2. Quanti anni hai?

3. Da dove vieni? (scegli il nome della tua regione di provenienza)

4. A quale Ateneo sei iscritto/a?

5. A quale anno sei attualmente iscritto/a?

6. Qual è il tuo corso di laurea/dottorato?

7. Hai mai preso lezioni private di musica (fuori dalla scuola)?

☐ Sì. Le ho prese per meno di tre anni e adesso non seguo nessuna lezione privata. (Vai alla domanda 8)

☐ No (Vai alla domanda 9).

8. Che cosa (strumenti e/o canto) hai praticato? Elenca da quello che hai praticato di più a quello che hai praticato di meno.

9. Oltre all'italiano, quale/i lingua/e conosci? Scrivi accanto al nome della/e lingua/e anche il tuo livello secondo il Quadro comune Europeo di riferimento (es. spagnolo B2). Se non hai idea di che livello, clicca qui[34]!

[34] The link inserted here was: https://rm.coe.int/168045bc72

APPENDIX D

Final questionnaire for participants

1. Eri attento/a durante i task di identificazione? I task di identificazione sono quelli dove c'è da scegliere il pattern tonale corretto fra tre proposti (l'immagine usata durante i task è questa).

[35]

Puoi dare una valutazione sul tuo livello di attenzione durante lo svolgimento dei task?

Molto attento/a, Attento/a, Ogni tanto attento/a, Non attento/a

N.B. Non ti preoccupare! Non ti giudico male in base alla tua risposta. Quindi puoi rispondere onestamente.

2. Puoi valutare i parlanti presenti nei task di identificazione in base alle tue difficoltà nello svolgimento dei task?

1) L'uomo che hai ascoltato prima e dopo l'addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone "play"!

2) La donna che hai ascoltato sia nell'addestramento che dopo l'addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone "play"!

3) La donna che hai ascoltato dopo l'addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

[35] The same image used in Figure 4.2.

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone “play”!

3. Eri attento/a durante i task di discriminazione? I task di discriminazione sono quelli dove c'è da rispondere se due audio ascoltati sono uguali oppure diversi (l'immagine usata durante i task è questa).

 [36]

Molto attento/a, Attento/a, Ogni tanto attento/a, Non attento/a

N.B. Non ti preoccupare! Non ti giudico male in base alla tua risposta. Quindi puoi rispondere onestamente.

4. Puoi valutare i parlanti presenti nei task di discriminazione in base alle tue difficoltà nello svolgimento dei task?

1) L’uomo che hai ascoltato prima e dopo l’addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone “play”!

2) La donna che hai ascoltato prima e dopo l’addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone “play”!

3) L’uomo che hai ascoltato soltanto dopo l’addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone “play”!

4) La donna che hai ascoltato soltanto dopo l’addestramento

1. Molto facile; 2. Facile; 3. Normale; 4. Difficile; 5. Molto difficile

[36] The icon was downloaded from https://icooon-mono.com/ (accessed in 03/11/2021).

Se vuoi, puoi ascoltare la sua voce cliccando sul bottone “play”!

5. Hai mai avuto problemi tecnici? (Es. blocco del browser, ecc.) Se la tua risposta è sì, puoi descrivere il problema che hai riscontrato? Lasciare vuoto se non hai avuto problemi. Premere “Avanti” per procedere.

APPENDIX E

Short questionnaire for native Japanese speakers who recorded their voice

１．ご出身地を教えていただけますか（東京都・神奈川県・埼玉県・千葉県）。
２．年齢を教えていただけますか。
３．性別を教えていただけますか。

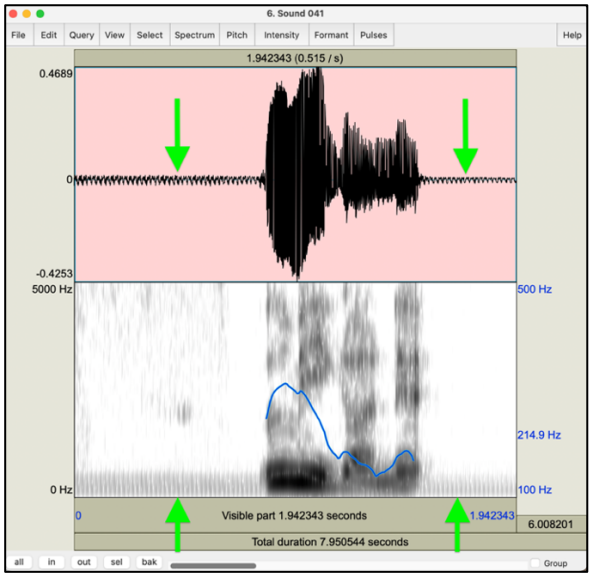
APPENDIX F

Praat software: visualization of background noise in stimulus recordings

Screenshots A. and B. of Praat sound windows show two recordings of a stimulus produced by speaker F1: *umi ga ne* "It's sea, you know". Panel A illustrates a sample which was rejected due to excessive background noise deriving from the external hard disc attached to her computer. The intensity of the silent part of the recording is approximately 50 db. Both the waveform and the spectrogram of the silent part in Panel A are different from those in Panel B, which shows a recording that was accepted. F0 contours is depicted by the blue lines.

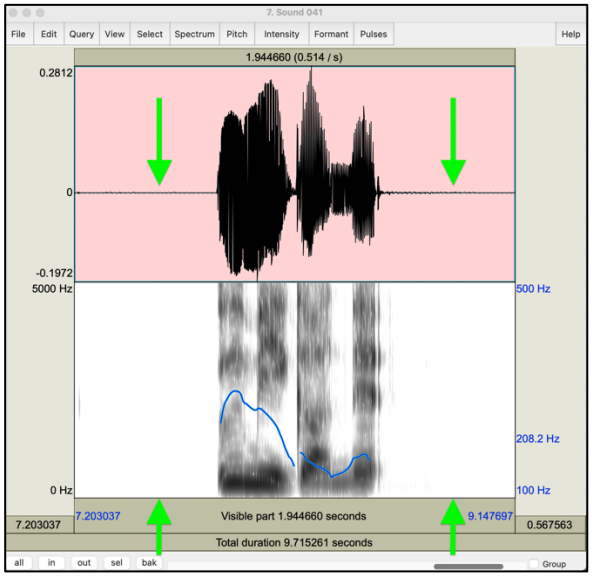
A:

recording example with background noise (the green arrow shows the wider bandwidth in the silent part of the recording).



B:

recording example with little background noise.



APPENDIX G

Multiple Comparisons: Significant Interactions Between Pattern, Training Condition, Musician/Non-Musician Category, and Identification Test (Pretest and Posttest)

Contrast	Estimate	SE	z.ratio	p
Contrasts for Test:				
HV – Non-Musicians:				
Pretest – Posttest (1st-syllable accented pattern)	-0.73	0.10	-6.80	<0.001
Pretest – Posttest (2nd-syllable accented pattern)	-0.41	0.11	-3.84	<0.001
Pretest – Posttest (unaccented pattern)	-0.52	0.11	-4.86	<0.001
HV – Musicians:				
Pretest – Posttest (1st-syllable accented pattern)	-0.78	0.11	-7.34	<0.001
Pretest – Posttest (2nd-syllable accented pattern)	-0.46	0.11	-4.34	<0.001
Pretest – Posttest (unaccented pattern)	-0.29	0.11	-2.70	0.007
LV – Non-Musicians:				
Pretest – Posttest (1st-syllable accented pattern)	-0.47	0.10	-4.81	<0.001

LV – Musicians:				
Pretest – Posttest (1st-syllable accented pattern)	-0.63	0.10	-6.23	<0.001
Pretest – Posttest (2nd-syllable accented pattern)	-0.50	0.10	-4.91	<0.001
Pretest – Posttest (unaccented pattern)	-0.75	0.10	-7.24	<0.001
Contrasts for Category				
LV – Musicians vs. LV – Non-Musicians:				
Pretest (2nd-syllable accented pattern)	-0.55	0.23	-2.44	0.015
Pretest (unaccented pattern)	-0.71	0.23	-3.13	0.002
Posttest (2nd-syllable accented pattern)	-1.04	0.23	-4.57	<0.001
Posttest (unaccented pattern)	-1.46	0.23	-6.39	<0.001
Contrasts for Training Condition				
HV – Non-Musicians vs. LV – Non-Musicians:				
Pretest (2nd-syllable accented pattern)	0.53	0.23	2.26	0.024

Pretest (unaccented pattern)	0.62	0.23	2.63	0.009
Posttest (1st-syllable accented pattern)	0.49	0.23	2.10	0.035
Posttest (2nd-syllable accented pattern)	0.93	0.23	3.96	<0.001
Posttest (unaccented pattern)	1.14	0.24	4.86	<0.001
Contrasts for Pattern				
HV – Non-Musicians:				
Pretest: 1st-syllable accented pattern vs. 2nd-syllable accented pattern	-0.43	0.15	-2.94	0.010
Pretest: 1st-syllable accented pattern vs. unaccented pattern	-0.41	0.15	-2.84	0.013
HV –Musicians:				
Pretest: 1st-syllable accented pattern vs. 2nd-syllable accented pattern	-0.37	0.14	-2.57	0.030
Pretest: 1st-syllable accented pattern vs. unaccented pattern	-0.64	0.14	-4.40	<0.001

LV – Non-Musicians:				
Posttest:				
1st-syllable accented pattern vs. unaccented pattern	0.45	0.14	3.19	0.004
LV –Musicians:				
Pretest:				
1st-syllable accented pattern vs. 2nd-syllable accented pattern	-0.41	0.14	-2.91	0.011
Pretest:				
1st-syllable accented pattern vs. unaccented pattern	-0.47	0.14	-3.30	0.003
Posttest:				
1st-syllable accented pattern vs. unaccented pattern	-0.58	0.14	-4.05	<0.001

APPENDIX H

Multiple Comparisons: Significant Interactions Between Pattern, Training Condition, Musician/Non-Musician Category, and Identification Test (Pretest and Gem-1)

Contrast	Estimate	SE	z.ratio	p
Contrasts for Test:				
HV – Non-Musicians:				
Pretest – Gen-1 (1st-syllable accented pattern)	-0.82	0.14	-5.73	<0.001
Pretest – Gen-1 (2nd-syllable accented pattern)	-0.52	0.14	-3.66	<0.001
Pretest – Gen-1 (unaccented pattern)	0.56	0.14	3.93	<0.001
HV – Musicians:				
Pretest – Gen-1 (1st-syllable accented pattern)	-0.91	0.14	-6.31	<0.001
Pretest – Gen-1 (2nd-syllable accented pattern)	-0.30	0.14	-2.12	0.034
Pretest – Gen-1 (unaccented pattern)	0.47	0.14	3.29	0.001
LV – Non-Musicians:				
Pretest – Gen-1 (1st-syllable accented pattern)	-0.50	0.14	-3.65	<0.001

LV – Musicians:				
Pretest – Gen-1 (1st-syllable accented pattern)	-1.04	0.14	-7.43	<0.001
Pretest – Gen-1 (2nd-syllable accented pattern)	-0.63	0.14	-4.48	<0.001
Contrasts for Category				
LV – Musicians vs. LV – Non-Musicians:				
Pretest (2nd-syllable accented pattern)	-0.55	0.22	-2.50	0.012
Pretest (unaccented pattern)	-0.71	0.22	-3.21	0.001
Gen-1 (1st-syllable accented pattern)	-0.82	0.22	-3.69	<0.001
Gen-1 (2nd-syllable accented pattern)	-1.24	0.22	-5.57	<0.001
Gen-1 (unaccented pattern)	-0.74	0.22	-3.31	<0.001
Contrasts for Training Condition				
HV – Non-Musicians vs. LV – Non-Musicians:				
Pretest (2nd-syllable accented pattern)	0.53	0.23	2.30	0.021

Pretest (unaccented pattern)	0.61	0.23	2.68	0.007
Gen-1 (1st-syllable accented pattern)	0.56	0.23	2.43	0.015
Gen-1 (2nd-syllable accented pattern)	1.11	0.23	4.85	<0.001
Contrasts for Pattern				
HV – Non-Musicians:				
Pretest: 1st-syllable accented pattern vs. 2nd-syllable accented pattern	-0.42	0.14	-2.99	0.008
Pretest: 1st-syllable accented pattern vs. unaccented pattern	-0.41	0.14	-2.88	0.012
Gen-1: 1st-syllable accented pattern vs. unaccented pattern	0.97	0.14	6.75	<0.001
Gen-1: 2nd-syllable accented pattern vs. unaccented pattern	1.10	0.14	7.62	<0.001
HV –Musicians:				
Pretest: 1st-syllable accented pattern vs.	-0.38	0.14	-2.65	0.024

2nd-syllable accented pattern				
Pretest:				
1st-syllable accented pattern vs. unaccented pattern	-0.65	0.14	-4.53	<0.001
Gen-1:				
1st-syllable accented pattern vs. unaccented pattern	0.73	0.14	5.08	<0.001
Gen-1:				
2nd-syllable accented pattern vs. unaccented pattern	0.50	0.14	3.51	0.001
LV – Non-Musicians:				
Gen-1:				
1st-syllable accented pattern vs. 2nd-syllable accented pattern	0.43	0.14	3.11	0.006
Gen-1:				
1st-syllable accented pattern vs. unaccented pattern	0.48	0.14	3.51	0.001
LV –Musicians:				
Pretest:				
1st-syllable accented pattern vs. 2nd-syllable accented pattern	-0.41	0.14	-2.96	0.009

Pretest:				
1st-syllable accented pattern vs. unaccented pattern	-0.46	0.14	-3.36	0.002
Gen-1:				
1st-syllable accented pattern vs. unaccented pattern	0.57	0.14	4.05	<0.001
Gen-1:				
2nd-syllable accented pattern vs. unaccented pattern	0.56	0.14	4.01	<0.001