Non-native perception, production, and lexical processing of tone

ST JOHN'S COLLEGE
UNIVERSITY OF CAMBRIDGE

Tim Joris Laméris

Faculty of Modern and Medieval Languages and Linguistics

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

St John’s College April 2022

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. Chapter 2 is based on a paper co-authored with Brechtje Post. I therefore use ‘we’ instead of ‘I’ to acknowledge the co-authorship, but the final writing is entirely my own. This dissertation contains less than 80,000 words including appendices, bibliography, footnotes, tables and equations and has less than 150 figures.

Tim Joris Laméris

April 2022

Non-native perception, production and lexical processing of tone

Tim Joris Laméris

ABSTRACT

In this dissertation, I investigate how and why adults differ in the ease with which they learn tone in a non-native language. I examine the extent to which individual variability in tone learning facility depends on factors attributable to a learner's first language, namely the function of pitch for lexical distinctions ('L1 tonal status') and the shapes of native tonal and intonational contrasts ('tone type'), as well as extralinguistic factors, namely musical experience, working memory, and pitch perception aptitude. In doing so, I aim to provide a novel and integral account of the multiplicity and diversity of factors that influence non-native tone learning facility.

The core of this dissertation consists of four empirical data chapters in the shape of journal manuscripts, which each zoom in on non-native tone learning through different lenses. Chapter 1 provides a general introduction. Chapter 2 reports a lab-based study in which 41 Mandarin and English speakers took part in a tone categorization and word identification task to investigate individual variability in pre-lexical and lexical tone perception. Chapter 3 reports two further lab-based studies to investigate pre-lexical and lexical tone processing in the spoken modality to zoom in on individual variability in production. Chapter 4 provides a comparative analysis between the perception and production tasks to discuss differences and similarities between performance in the listening and speaking modalities. Chapter 5 reports a web-based study which involved 114 speakers from typologically different languages (Dutch, Swedish, Japanese, and Thai) and which reassesses the degree to which L1-specific and extralinguistic factors determine tone perception and lexical processing. Chapter 6 provides a general discussion and conclusions.

The findings from these empirical studies show that individuals differ greatly in the ease with which they learn non-native tones, particularly at a lexical level of tone processing. Both L1-specific and extralinguistic factors explain why some individuals learn tones with more ease than others do, but these factors interact with one another in dynamic ways to determine tone learning facility. An 'L1-Modulated Domain-General Account' is proposed to formally describe the empirical findings from these studies: individual variability in tone learning facility is best captured by extralinguistic factors, but the relative effect of these factors may be modulated by a learner's language background.

Acknowledgements

I would like to acknowledge everything and everyone that has been a source of support during the almost four year-long period during which I have worked on the research that has led to this dissertation.

Acknowledgements redacted for privacy reasons.

Acknowledgements redacted for privacy reasons.

shí shì shī shì shī shì, shì shī,
shì shí shí shī.
shì shí shí shì shì shì shī, shí shí,
shì shí shī shì shì.
shì shí, shì shī shì shì shì.
shī shì shì shì shí shī, shì shī shì,
shǐ shì shí shī shì shī.
shī shì shí shì shí shī shī, shì shí shì.
shí shì shī, shī shì shǐ shì shì shí shì,
shí shì shì, shī shì shǐ shì shí shì shí shī shī.
shí shí, shǐ shí shì shí shī shī shì shí shí shī shī.
shì shì shì shì.

– Chinese story using only the syllable ‘shi’ and tone
by Yuen-Ren Chao

Translation:
‘A poet named Shī lived in a stone house and liked to eat lion flesh, and he vowed to eat ten of them. He used to go to the market in search of lions, and one day at ten o’clock, he chanced to see ten of them there. Shī killed the lions with arrows and picked up their bodies, carrying them back to his stone house. His house was dripping with water so he requested that his servants proceed to dry it. Then he began to try to eat the bodies of the ten lions. It was only then he realized that these were in fact ten lions made of stone. Try to explain the riddle.’

Contents

List of figures

List of tables

List of abbreviations

CI	Confidence interval
CrI	Credible interval
F0	Fundamental frequency
FPH	Functional Pitch Hypothesis
GUI	Graphical User Interface
Hz	Hertz
ISI	Inter-stimulus interval
L1	First language
L1er	First language speaker; native speaker
L2	Second language
MMLL	Modern and Medieval Languages and Linguistics
ms	Milliseconds
PAM	Perceptual Assimilation Model
RT	Reaction time
SD	Standard deviation
SE	Standard error
SLM	Speech Learning Model
WM	Working memory

Overview of this dissertation

The core of this dissertation consists of four empirical chapters which are articles that have been published or are in preparation for publication in international peer-reviewed journals. Each of these chapters contains its own literature review, research questions, methodology, and discussion, and can therefore be read independently. However, taken together, the research reported in these individual chapters provides a comprehensive account of non-native tone learning. The original article manuscripts have been modified in places to allow for referencing across chapters (particularly in Chapter 4, which compares data from Chapters 2 and 3). Although the empirical Chapters 2–5 each contain their own discussion, an overall evaluation and comparison of the findings will be presented in a General discussion in Chapter 6. I include appendices per chapter for ease of reading, but compile the list of references for all the chapters combined at the end of this dissertation.

Chapter 1 (General introduction) serves to contextualize the theoretical and empirical background of this dissertation and introduces key terminology.

Chapter 2 explores individual variability in pre-lexical and lexical tone processing in the listening modality by English-L1 and Mandarin-L1 speakers. It is an adaptation of Laméris & Post (2022), which has been accepted for publication in *Second Language Research.*

Chapter 3 reports on pre-lexical and lexical tone processing in the speaking modality by English-L1 and Mandarin-L1 speakers. This is an adaptation of Laméris (n.d.), which is under second-round review with *Language and Speech.*

Chapter 4 investigates the perception-production link in non-native tone learning by comparing the findings from Chapters 2 and 3. It is currently in preparation as a manuscript to be sent out for review.

Chapter 5 explores individual variability in non-native tone perception and word learning by native speakers of Dutch, Swedish, Japanese and Thai. This is an adaptation of Laméris (2022), which has been accepted for publication as a conference proceedings paper for *Speech Prosody 2022.*

The overall findings from these studies and their implications are discussed in **Chapter 6** (General discussion).

Chapter 1 General introduction

1.1 Tone

It is extremely rare to find adults who can acquire the sound system of a foreign language without leaving traces of their native tongue (Moyer, 2013). For adults, learning the sounds of a second language (L2)[1] is difficult. This difficulty of non-native speech acquisition can be broadly explained by hurdles in articulatory-motoric and phonological aspects of speech.

Articulatory-motoric learning requires the physiological reorganization of the vocal organs to produce a sound that does not occur in the first language (L1). An example is when L1 speakers of English learn Xhosa, a language with lingual ingressive sounds – also known as clicks – which require articulatory motions that are typically not employed in English (Lewis, 1994), or when English speakers acquire Hindi retroflex consonants (Hayes-Harb & Barrios, 2021). Similarly, articulatory-motoric learning is required when American English speakers learn Spanish trill /r/ consonants, which are unfamiliar articulations in their L1 (Olsen, 2012).

In addition to articulatory-motoric efforts, the L2 speech learner needs to form an awareness that certain sounds can be grouped together as abstract units, also known as phonological categories or phonemes. This acquisition of distinct categories can be referred to as phonological learning. For instance, L2 learners of Dutch from a variety of L1 backgrounds often mistakenly produce the Dutch /œy/ diphthong as /ɑu/ (Neri et al., 2006). What arguably happens in these speakers is that in the absence of a distinct /œy/ category in the L1, the L2 sound maps onto the most similar-sounding L1 category, such as /ɑu/. Therefore, to acquire the /œy/ sound, L2 learners need to create an awareness that /œy/ is a

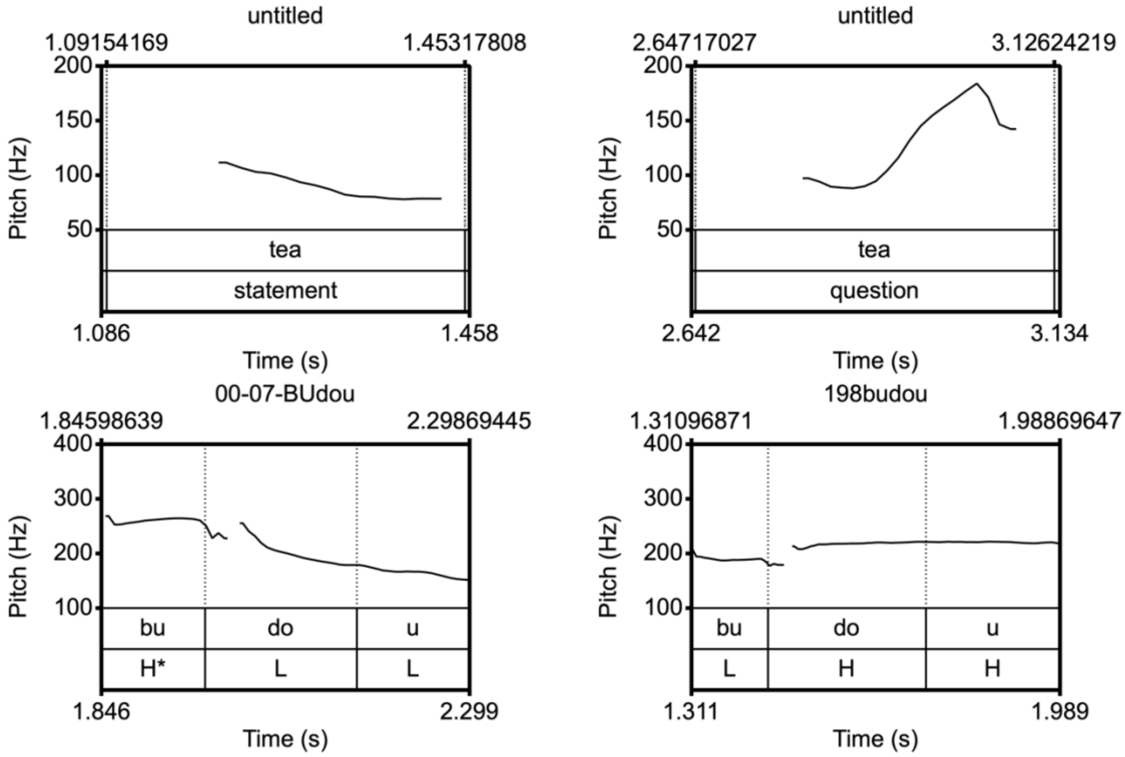
[1] Unless explicitly mentioned or when relevant for the discussion, I will use the terms “L2” and “Second language” in broad terms to describe any *non-native* language, regardless of the order in which that non-native language has been acquired after the first (native) language and any other languages.

separate phonological category, distinct from the /ɑu/ sound, cf. Best & Tyler (2007); Flege (1995).

Human speech involves more than consonants and vowels. Speech also involves pitch. Pitch is the perceptual correlate of fundamental frequency (F0; measured in Hertz), which is generated by vibrations of the vocal folds. All spoken languages employ pitch, but what pitch is used for differs across languages. In English, I can produce the sound [tʰi] with a fall or a rise in pitch, as shown in Figure 1. If I produce [tʰi] with a falling pitch pattern, I can signal a statement: ‘tea!’. If I produce it with a rising pitch pattern, I can signal a question: ‘tea?’. In this way, pitch can be used as a primary instrument for *phrasal* purposes, such as conveying utterance or discourse-level meaning. Pitch also serves to signal various paralinguistic purposes, such as different degrees of surprise or emotion (Post et al., 2015, p. 2). The use of pitch for these phrasal and paralinguistic purposes is known as ‘intonation’. Yet despite the different phrasal meanings of ‘tea!’ and ‘tea?’, in both utterances I signal the core word meaning of ‘a beverage made by infusing tea leaves in hot water’.

Figure 1

F0 traces for the English ‘tea’ and the Japanese [bɯdo:] (martial art; grape).



If a speaker of Japanese produces the sound sequence [bɯdo:] with either a falling or a rising pitch pattern (Figure 1), they can signal something different than just a change from a statement to a question. They change the core meaning of the uttered word. With a high-low pitch pattern, [bɯdo:] means ‘martial art’, whereas with a low-high pitch pattern, [bɯdo:] means ‘grape’. In this way, pitch is used as a primary instrument for *lexical* purposes. Note that in English, pitch can also play a role in distinguishing lexical meaning between otherwise similar-sounding words, but typically not on its own or as the primary instrument (Yip, 2002, pp. 3–4). For instance, <subject> can either refer to the noun [ˈsʌb.dʒɛkt] or the verb [sʌbˈdʒɛkt]. The syllables preceded by a [ˈ] are stressed syllables, which are phonetically and perceptually prominent. This prominence is achieved by a combination of parameters, including pitch, but also duration, loudness, vowel quality, and spectral tilt.

The primary use of pitch to determine a word’s core meaning, as in the case of the Japanese [bɯdo:], is known as ‘lexical pitch’, ‘lexical tone’, or simply ‘tone’. Tone languages are found across the globe, and according to some accounts, most languages in the world are in fact tonal (Yip, 2002, p. 2). In the Hmong language (Southeast Asia), for example, the sound sequence [po] can have seven different meanings, from ‘ball-like’ to ‘paternal grandmother’, depending on the tone (Esposito, 2012). Similarly, the Norwegian word [skufen] means either ‘the drawer’ or ‘the shovel’ depending on the lexical pitch pattern it is produced with (Moen & Sundet, 1996). Languages like Norwegian, as well as Japanese, are sometimes referred to as *pitch-accent* or *accentual languages* because only a small number of words are distinguished in meaning by pitch alone, and because the lexical pitch patterns are typically less complex than in fully-fledged tone languages like Hmong (Yip, 2002, p. 4). Although the distinction between tone and pitch-accent languages can be deemed to be a purely typological one, in essence they differ greatly from non-tone languages like English and Dutch because they employ pitch as a primary instrument to signal lexical distinctions between otherwise identical words. In addition, what distinguishes tone and pitch-accent languages from non-tone languages is that phonological pitch relations within a word in the former tend to be fixed, whereas in non-tone languages they can be modulated by intonation without affecting the word’s meaning. For example, both [ˈsʌb.dʒɛkt] (noun) and [sʌbˈdʒɛkt] (verb) can be realized with any permissible English intonational pitch pattern without affecting the interpretation as noun or verb. Although intonation can modulate the realization of tones and pitch accents, the overt pitch patterns in tone and pitch-accent languages on the word tend to be remarkably stable (Yip, 2002, p. 3).

1.2 Motivation, aims, and scope of this dissertation

In this dissertation I explore how adults learn tones in a second language. The motivation to research tone learning in adults stems from the observation that, while acquiring the sound system of a second language is already a challenge, acquiring the tone system in a tone language appears to be particularly difficult. Studies in advanced L2 learners of tone languages suggest that while learners can become very good at acquiring segmental features of that language (i.e., the vowels and consonants), tones present more persistent difficulty (Pelzl et al., 2019, 2020).

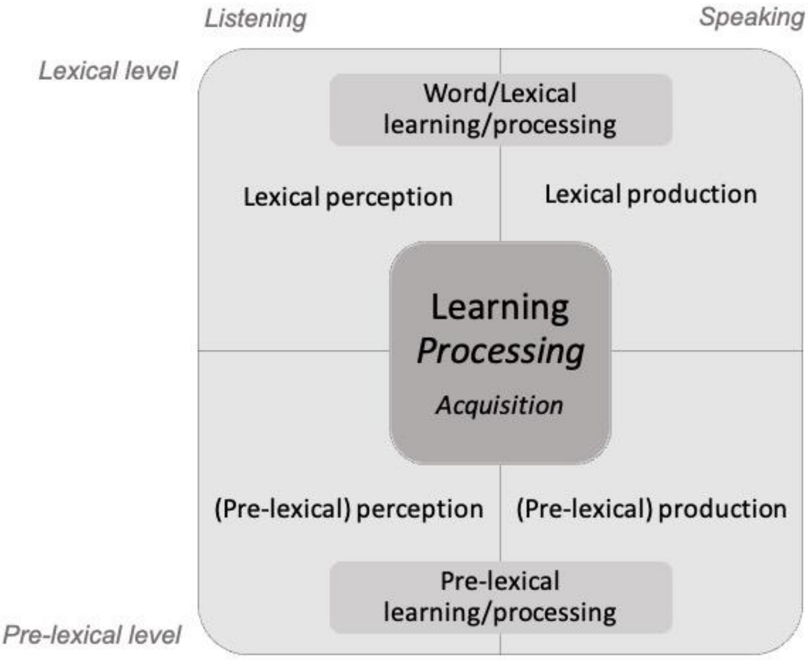
Yet, despite this apparent difficulty of learning tones, some individuals seem to learn tones more easily than others do (Perrachione et al., 2011). This is a second observation that motivates my dissertation research. Specifically, I zoom in on individual variability in tone learning *facility*, which refers to the ease with which tones are learned at early stages of learning. Note that I do not explore tone learning *capacity*, which refers to the ultimate level of attainment that an individual may reach after a long period of learning (Bowles et al., 2016, p. 775).

I explore tone learning facility by means of a series of behavioral experiments. In these experiments, non-native tone learning is assessed at different levels, namely at a *pre-lexical* level, which refers to the processing of linguistic tone that does not require an association to word meaning, and at a *lexical* level, which refers to tone processing that does. I additionally explore pre-lexical and lexical tone processing in both the listening modality, i.e., in *perception*, and in the speaking modality, i.e., in *production*.

A note on the usage of key terms: When referring to the specific level of processing in each modality (listening or speaking), I use the terms '(pre-lexical) perception' and '(pre-lexical) production', and 'lexical perception' and 'lexical production', respectively. The terms 'word learning' or 'lexical processing' are used as an encompassing term when the distinction between the listening and speaking modalities at a *lexical* level is not relevant to the discussion. Finally, tone 'learning' or 'processing', and occasionally 'acquisition' are used as overarching terms when neither the distinction between levels of processing, nor between modalities is relevant for the discussion. The use and scope of these terms are summarized in Figure 2.

Figure 2

Use and scope of terminology across modalities and levels of processing.



In all the experiments I present in this dissertation, individuals learned tonal pseudowords, following Wong & Perrachione (2007). The rationale for choosing pseudowords instead of a real-language words will be outlined in more detail in Chapters 2 and 5, the main reason being that it allowed me to create a neutral and controlled environment in which I could simulate the very start of tone learning for *ab initio* learners independent of language background. This enabled me to specifically measure individual performance in early-stage tone learning and identify individuals who performed relatively well and individuals who performed relatively poorly, thereby providing a degree of inter-learner variability in tone learning facility.

By creating this early-stage tone language learning microcosm in which the degrees of learning facility would vary between speakers, I could then explore the reasons and origins of this variability. I tested the effects of the following five factors, which I hypothesize may explain variability in learning facility. I will describe these factors, and how they were operationalized in the studies, in more detail in sections 1.3.3.1–5.

1. *L1 tonal status.* This refers to the function of pitch in a learner's L1.
2. *Tone types.* This refers to the specific shape of the tone to be learned in the target tone language, and the potential interaction with tone types (either lexical or phrasal) in the L1.
3. *Musicianship.* This refers to the degree to which an individual has long-term experience with practicing music, or the degree to which an individual has sensitivity to musical pitch or rhythm.
4. *Working memory*. This refers to an individual's working memory capacity.
5. *Pitch aptitude*. This refers to an individual's ability to perceive tones pre-lexically.

Before presenting the overall theoretical and empirical framework, at this point I would like to mention several topics within the tone learning literature that I acknowledge are of great relevance to the overall scholarship, but that are beyond the scope of this dissertation.

This dissertation concerns early-stage second language tone learning in adults, and therefore does not cover first language tone acquisition by infants or children (Antoniou & Chin, 2018; Morett, 2020; Nan et al., 2018) or in intermediate or advanced adult learners (Goss & Tamaoka, 2019; Pelzl et al., 2019). With a focus on pre-lexical and lexical processing of tones, it neither concerns tone learning at other levels, nor the interface with other aspects of language and speech such as syntax or morphology (Ajíbóyè et al., 2011; P. Tang et al., 2019) or intonation (Mennen, 2015; Ota, 2016; H. Zhang, 2018). The dissertation further zooms in on F0 (pitch) as the primary acoustic correlate of tone and makes limited reference to secondary acoustic cues that may be involved in tone learning, such as phonation and duration (S. Chen et al., 2017; Tsukada & Kondo, 2019; Y. Zhang & Kirby, 2020). The dissertation does not explicitly analyze tone or lexical pitch accent systems from theoretical perspectives such as autosegmental phonology, either (Ota, 2003).

With regard to the facilitative factors of tone learning, in this dissertation I do not probe the effects of foreign language learning aptitude (Carroll, 1981), general intelligence (Wong et al., 2020), age (Huang & Jun, 2011; Ingvalson et al., 2017), nor the effects of different learning methods such as the use of gestures (Baills et al., 2019), visual aids (Burnham et al., 2015; Reid et al., 2015) or high-variability learning paradigms (Wiener et al., 2020; Wiener & Lee, 2020; K. Zhang et al., 2018).

In the following, I present a theoretical and empirical overview of pre-lexical and lexical learning, and of each of the five factors that I hypothesize may modulate tone learning facility (L1 tonal status, tone type, musicianship, working memory, and pitch aptitude). At the end of this General introduction, I present this dissertation's research question, approach, and expected outcome.

1.3 Overview of theoretical and empirical framework

1.3.1 Pre-lexical processing and learning

Throughout this dissertation, I use the term 'pre-lexical processing' to refer to the processing of linguistic tones devoid of lexical meaning. It should be noted that pre-lexical processing in itself can refer to speech processing at two further separate levels: a phonetic and a phonological level. Phonetic processing of tones requires a listener or speaker to pay attention to fine-grained acoustic differences in the speech signal, whereas phonological processing can be described as the act of encoding pitch movements as "abstract tone categories at the syllable level" (R. K. W. Chan & Leung, 2020, p. 21). Although the distinction between phonetic and phonological processing within pre-lexical processing is relevant (as I will discuss in more detail in Chapter 4), I will mainly discuss pre-lexical tone processing in its broad sense to distinguish it from *lexical* processing, which involves linking sound to meaning.

Researchers interested in tone learning often tend to investigate pre-lexical tone processing to study tone learning at large. This is rooted in the notion that the ability to process tones pre-lexically is a steppingstone for lexical processing, in line with bottom-up accounts of speech learning (Norris et al., 2003). That is, a learner should be able to process tones accurately in a pre-lexical setting before they can use tones in words. Although top-down processes from the lexical level may exert influence on pre-lexical speech processing (McClelland & Elman, 1986), there appears to be a consensus that a learner's pre-lexical processing skill is indicative of their lexical processing. Consequently, previous studies have proposed a "phonetic-phonological-lexical continuity" in tone learning (Wong & Perrachione, 2007).

As a result, many – if not most – studies on tone learning tend to be based on experiments that uniquely examine pre-lexical processing. How can pre-lexical processing be measured? One common instrument is a *discrimination task*, which is employed widely in tone learning research (Braun & Johnson, 2011; Burnham et al., 2006; Hao, 2018; Schaefer & Darcy, 2014; Wong et al., 2020; Zhu et al., 2021). Discrimination tasks come in different formats, but a common paradigm is an AX or 'same-different' discrimination task, in which a participant listens to a sequence of identical (e.g., AA) or different tones (e.g., AB), and needs to indicate whether those tones are the same or not. Other formats are AXB tasks, in which the participant needs to indicate whether the second presented tone X was the same as tone A or B. Discrimination tasks can measure a listener's auditory sensitivity to fine-grained phonetic differences between tone stimuli that are acoustically very similar, and that for instance differ in F0 height by a few Hertz (Bent et al., 2006). Discrimination tasks can therefore be a measure of phonetic processing, although it has been suggested that the longer the inter-stimulus-interval (ISI), the more likely it is that a listener processes the tones as abstract phonological categories (X. Wang, 2013). Discrimination tasks can also be used to bridge phonetic and phonological processing by investigating whether a listener perceives tones "psychoacoustically" or "categorically" (Qin et al., 2019). One oft-employed paradigm to investigate this involves a tonal continuum. An example of a tonal continuum (Francis et al., 2003) is a set of stimuli consisting of a start stimulus representing one phonological tone category (e.g., a low-level tone of 100 Hz), and an end stimulus representing another tone category (e.g., a high-level tone of 205 Hz). In between these extremes, there is a continuum of intermediate step-stimuli that gradually approximate the end stimulus (in this case, nine step-stimuli that gradually increase in Hertz). If a listener can perceive 'within-category' differences between minimal phonetic contrasts, e.g., between the start stimulus (step 1; 100 Hz) and a minimally differing stimulus (step 3, 108 Hz), the listener is said to process tones *psychoacoustically*. Conversely, if a listener is only able to perceive 'between-category' differences (e.g., between stimuli pairs that straddle the category boundary of the low-level and high-level tone, in this case step 5), the listener is said to process tones *categorically.* Throughout this dissertation, I will regularly allude to the differences between psychoacoustic and categorical perception of tones, and it is therefore important to be aware of this distinction.

A second common instrument to measure pre-lexical tone processing is a *tone identification* or *tone categorization* task (Dong et al., 2019; C.-Y. Lee & Hung, 2008; M. Li & Dekeyser, 2017; Liu & Samuel, 2004; Wong & Perrachione, 2007). In such a task, a participant hears an auditory tone stimulus and needs to indicate which category the stimulus belongs to by selecting from multiple choices such as ‘low-level’, ‘mid-level’ or ‘high-level’ (Francis et al., 2003). Given that a participant is forced to assign the auditory stimulus to an abstract category, this type of task can be said to tap specifically into phonological processing. However, depending on the stimuli, categorization tasks can also measure some degrees of phonetic processing. For instance, a categorization task can employ a tonal continuum to measure at what point a speaker categorizes a tonal stimulus as either category, such as low-level or high-level. If a listener consistently identifies all stimuli to the left of the continuum as one category, and all stimuli to the right of the continuum as the other, this can be taken as evidence of categorical perception. Conversely, if a listener identifies stimuli in a less clear-cut way, this can be taken as evidence of more psychoacoustic perception.

In this dissertation, I used tone categorization tasks to measure pre-lexical processing of tones. I chose tone categorization tasks over discrimination tasks for two main reasons. First, tone categorization tasks have been used in several studies that addressed similar research questions as mine (Bowles et al., 2016; Cooper & Wang, 2012; Dong et al., 2019; M. Li & Dekeyser, 2017; Wong & Perrachione, 2007), and by using a same experimental method I can make more meaningful comparisons between my findings and those from earlier work. Second, I deem tone categorization tasks to be more informative than discrimination tasks, because while discrimination tasks can reveal *that* an individual may (not) be able to perceive a tonal contrast, a tone categorization task can further reveal *why* an individual may (not) be able to correctly perceive a tone. This is because tone categorization tasks can give insight into perceptual error patterns. For instance, if a participant in a tone categorization task hears an auditory stimulus X and has response options X, Y, and Z, but consistently incorrectly categorizes the stimulus as Y, it could be inferred that a confusion with the Y tone category leads to incorrect perception of the X target tone category. Indeed, throughout my studies, I found several instances of consistent error patterns in tone categorization. These provided me with additional information on how participants may have perceived specific tones. I will present and discuss these error patterns in detail in Chapters 2, 4, and 5.

Note that a tone categorization task only measures pre-lexical tone *perception* in the listening modality. To measure pre-lexical production in the speaking modality (Chapter 3), I employed an *imitation task.* An imitation task can be seen as “a production task adopting auditory instead of orthographic prompts” (Hao & de Jong, 2016, p. 152) in which upon presentation of a sound sequence, speakers are asked to repeat that sound sequence out loud and as accurately as possible. As I will discuss in detail in Chapter 3, I deem an imitation task to be an appropriate measure of pre-lexical tone production given that it does not require explicit lexical retrieval and instead relies on pre-lexical processing.

1.3.2 Lexical processing and learning

I have so far described lexical processing (word learning) in broad terms and defined it as ‘linking sound to meaning’, but what exactly does this entail?

Leach & Samuel (2007) propose that word learning consists of two aspects: *lexical configuration* and *lexical engagement.* Lexical configuration is the factual knowledge associated with a word, including its sound, meaning, and syntactic role. Lexical engagement refers to the ability of a lexical representation to activate other lexical or pre-lexical representations. For instance, in *semantic priming*, the presentation of a prime word (e.g., ‘girl’) can lead to faster responses to a semantically related target word (e.g., ‘boy’) than to an unrelated target word (e.g., ‘desk’; Finkbeiner & Nicol, 2003, p. 370). Similarly, in *phonological priming*, the presentation of a lexical entry (e.g. ‘cat’) can lead to faster activation of its pre-lexical phonemes (e.g., /k/; Leach & Samuel, 2007, p. 307).

In this dissertation, I simulated word learning of tonal pseudowords by creating experiments in which participants, after a word training session, needed to associate spoken pseudowords to images that represented the meaning of those words (‘cat’, ‘shirt’, ‘mountain’, etc.). Such sound-image paradigms are thought to represent word learning in the broad sense, entailing both lexical configuration and lexical engagement (Leach & Samuel, 2007). They have also been applied widely in previous tone word learning studies, both in the listening modality as *word identification* tasks in which upon auditory presentation of a word, a participant needs to indicate its meaning by selecting a corresponding image (Cooper & Wang, 2012; Poltrock et al., 2018; Wong & Perrachione, 2007), as well as in the speaking modality as *image-naming* tasks, in which a participant needs to produce the word that corresponds to the image that is presented. (Barcroft & Sommers, 2014; Dong et al., 2019; M. Li & Dekeyser, 2017; A. C. L. Yu et al., 2021).

I deemed it necessary to examine both pre-lexical and lexical processing of tones because – although much literature on tone learning exclusively examines the pre-lexical level – lexical processing can be argued to be more representative of real-life tone learning than pre-lexical processing alone. After all, in the real world, second language learners should not only be able to perceive differences in tones. They also need to use those tonal differences to communicate different lexical meanings. Further, I will show that, although I generally support the notion of a “continuity” (Wong & Perrachione, 2007) from pre-lexical to lexical processing, an examination of tone processing at both levels allowed me to also find evidence for a discontinuity between the two levels. Particularly, I show in Chapter 5 that whereas participants may perform quite uniformly in pre-lexical tasks, there appears to be considerably more individual variability in lexical tasks, and some individuals may experience difficulties in tone processing at a lexical level that they do not experience at a pre-lexical level. A discontinuity between pre-lexical and lexical learning has long been established in other aspects of speech, such as vowels (Díaz et al., 2012) and lexical stress (Dupoux et al., 2008), but it appears that this has only recently started to receive attention in the context of lexical tones (Ling & Grüter, 2020; Pelzl et al., 2019, 2020).

1.3.3 Facilitative factors in tone learning

A key objective of this dissertation is to identify the origins of individual variability in tone learning facility. To do so, I examined five factors that I hypothesize may affect the ease with which individuals learn tones in a second language. In this section I will describe each of these factors and explain how they may affect tone learning facility.

1.3.3.1 L1 tonal status

Throughout this dissertation, I will use the term ‘L1 tonal status’ as a typological indication that refers to the degree to which a language utilizes pitch for lexical distinctions. As discussed in detail in Chapter 5, I will describe non-tone languages like English and Dutch (in which pitch has a limited role for lexical purposes) as languages with a low tonal status, and languages like Swedish and Japanese, and Thai and Mandarin Chinese, as languages with intermediate and high tonal statuses, respectively.

In the tone learning literature, L1 tonal status is often mentioned as a factor that may explain variability in tone learning facility. This is rooted in the intuition that individuals who have no L1 experience with the lexical use of pitch may find learning tones in a second language more difficult than individuals who do have such L1 experience, as illustrated by the following citations:

> *“(..) the well-known difficulty experienced by adult speakers of non-tone languages when attempting to learn an unfamiliar tone language”*
> (Francis et al., 2008, p. 269)

> *“It is well established that the perception of non-native lexical tone contrasts is difficult for adult L2 learners (..) particularly for those whose L1 does not make use of pitch height and movement to signal changes in word meaning”*
> (Antoniou & Chin, 2018, p. 2)

> *“Lexical tones are often reported to be difficult for second language (L2) learners, especially for those whose native language (L1) is a nontone language*”
> (R. K. W. Chan & Leung, 2020, p. 2)

Several theoretical frameworks have been proposed to support this intuition. The "Feature Hypothesis" posits that "L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learners and this difficulty will be reflected in the learner's production (..)" (McAllister et al., 2002, p. 230). Although the Feature Hypothesis was based on the acquisition of duration, a "Functional Pitch Hypothesis" was later formulated to apply this notion to the acquisition of tone (Schaefer & Darcy, 2014). Similarly, a "Levels of Representation Account" has been proposed to describe the hypothesis that speakers of non-tone languages may find tones relatively difficult because "there is nothing in their native grammar that prepares them for using prosodic properties such as f0 in a lexically contrastive manner" (Francis et al., 2008, p. 269). In general, these accounts all predict that speakers of tone or pitch-accent languages may have a relative advantage compared to non-tonal speakers in L2 tone learning.

However, empirical evidence for these hypotheses is extremely mixed. This makes it difficult to assert whether there truly is an advantage based on L1 tonal status in L2 tone learning. Whereas some studies show that learners with a tonal L1 outperform their non-tonal peers in non-native tonal perception (R. K. W. Chan & Leung, 2020; Peng et al., 2010; Wayland & Guion, 2004) and word learning (Poltrock et al., 2018), other studies show that perceptual abilities are similar (Cooper & Wang, 2012; Francis et al., 2008; Gandour & Harshman, 1978; So & Best, 2010; X. Wang, 2013), and some other studies even suggest that L1 tone experience may be detrimental to L2 tone learning (Chiao, Kabak, & Braun, 2011; Francis et al., 2008; X. Wang, 2013).

In each of the empirical chapters presented in this dissertation, I examined the effect of L1 tonal status on non-native tone learning by conducting experiments with learners from various tonal backgrounds. Chapter 5 zooms in on the question of whether, all things equal, L1 tonal status truly facilitates non-native tone learning.

1.3.3.2 Tone types

Tones come in many shapes and forms, and the specific shape of a tone (henceforth: 'tone type') appears to determine the ease with which it is perceived, produced, and eventually, learned. Here, I will provide a brief overview of how tones may differ in phonetic-acoustic and phonological-categorical properties, and how this may affect tone learning facility. I will elaborate on this in detail in Chapter 2.

Some tones may be inherently easier to perceive than others. Neurological evidence suggests that humans are better at registering F0 rises than F0 falls in the brainstem (Krishnan et al., 2010). In addition, individuals have differential sensitivities to certain tonal contrasts in terms of their phonetic-acoustic properties, such as contrasts in either F0 height or contour (Francis et al., 2008; Gandour & Harshman, 1978; Qin & Jongman, 2016).

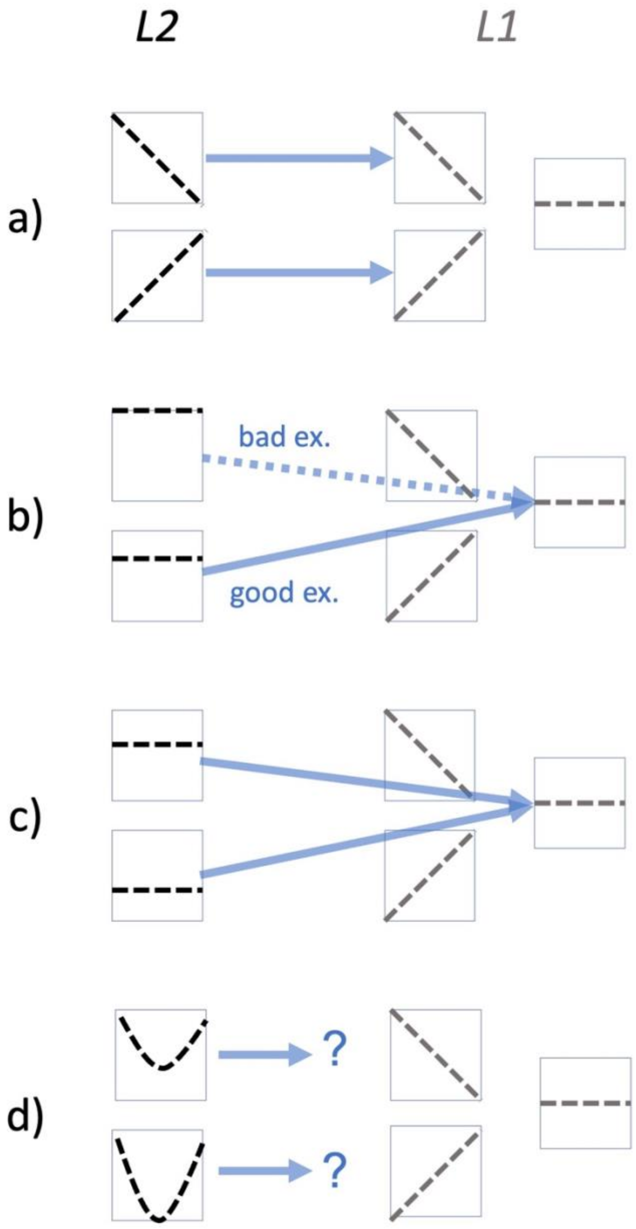
Individuals also appear to have differential sensitivities to tonal contrasts in terms of their phonological-categories properties. As described earlier, a phonological approach to speech requires the assumption that sounds with similar phonetic or articulatory properties can be grouped as distinct phonological units, also known as categories. Just as voiceless velar stops and alveolar fricatives can be described as segmental phonemes such as /k/ and /s/ to distinguish ‘cat’ from ‘sat’, so can phonetically overlapping pitch movements be described as tone categories or tonemes. For instance, Mandarin Chinese is said to have four distinct tone categories in citation form: a high-level, a mid-rising, a low-dipping, and a high-falling tone. In combination with the segments /ma/, these tonemes can respectively indicate the meanings of ‘mother’, ‘hemp’, ‘horse’ and ‘to scold’ (Antoniou & Chin, 2018; R. K. W. Chan & Leung, 2020).

An influential account that examines the effect of tone types in phonological-categorical terms is the Perceptual Assimilation Model, or PAM (Best & Tyler, 2007). Although originally designed as a model for the perception of vowels and consonants, it has in recent years been widely applied to the perception of tones (Best, 2019; J. Chen et al., 2020; So & Best, 2010).

In a nutshell, PAM assumes that listeners map – or *assimilate* – non-native tone categories to phonetically similar tone categories in the L1. That is, learners perceive non-native tones through the lens of their own tone categories. However, the difficulty of perceiving the non-native tone, and in particular the difficulty of discriminating it from other non-native tones, depends on the type of assimilation that takes place. The PAM proposes four routes of categorical assimilation, which are also visualized in Figure 3.

Figure 3

Schematic representation of tone category assimilation scenarios as predicted by PAM.



a) Two-category (easy) b) Category goodness (moderate) c) Single-category (difficult) d) No assimilation (easy/difficult)

In this scenario, the L1 has three tonal categories (a fall, a rise, and a level tone). Each of the scenarios describes an L2 tonal contrast that assimilates in different ways.

a) Two-Category Assimilation
Discrimination is easy if there are two L2 categories that map onto two different L1 categories in a one-to-one fashion.

b) Category Goodness Assimilation
Discrimination is moderately easy if there are two L2 categories that map onto one single L1 category in a many-to-one fashion and if one of those L2 categories is a better exemplar of the L1 category than the other L2 category.

c) Single-Category Assimilation:
Discrimination is difficult if there are two L2 categories that map onto one single L1 category in a many-to-one fashion and if both L2 categories are equally good or bad exemplars of the L1 category.

d) No assimilation
L2 categories remain uncategorized if they cannot clearly map onto L1 categories.

i) Discrimination is easy if each of these uncategorized L2 categories are distant from one another in terms of L1 sounds.

ii) Discrimination is difficult if each of these uncategorized L2 categories are similar to one another in terms of L1 sounds.

The PAM framework provides a compelling logic to account for tone perception facility. Its predictions have also been successfully tested in empirical studies of non-native tone perception by tonal L1 speakers (J. Chen et al., 2020; Hao, 2012; So & Best, 2010; Tsukada & Kondo, 2019; X. Wu et al., 2014). One problem, however, is that it is relatively unclear whether categorical assimilation also takes place in learners who have no tone categories in their L1 to begin with. For instance, what kind of assimilation takes place in learners whose L1 is English, which has no lexical tone categories? Some studies suggest that non-tonal L1 speakers assimilate L2 tone categories to L1 *intonational* categories, in a similar way that tonal L1 speakers assimilate L2 tones to L1 tonal categories. For instance, L. Lee & Nusbaum (1993) showed that English speakers could accurately integrate dynamic (dipping and falling) tones with segmental speech sequences, but struggled with integrating static (low and high level) tones. They suggest that English listeners may have been more attentive to dynamic tones because of their relative similarity to English intonational types. Braun & Johnson (2011) showed that Dutch listeners were more attentive to non-native disyllabic sequences in which the crucial pitch change (fall or rise) occurred on the final, but not on the first syllable. They attribute this to the fact that patterns with the pitch change on the final syllable resemble Dutch declaration and question intonational categories. Kan & Schmid (2019) found that young English-dominant heritage speakers of Cantonese performed worse than peers in Hong Kong in perception of the Cantonese high-rising and low-rising tonal contrast. They propose that the heritage speakers may have assimilated the two tones in a two-to-one fashion to the English rising question intonational type. While the findings from these studies support the hypothesis that speakers of non-tone languages assimilate non-native tonal contrasts to L1 intonational categories – which can then determine the ease with which they perceive the non-native tone type – other studies find only limited evidence. Although So & Best (2010) raise the possibility that English listeners assimilate the Mandarin rising tone to English intonational patterns, they do not find clear evidence for this. In their study, English listeners' discrimination performance per tone could not be clearly explained by any interaction with English intonational types, whereas Cantonese listeners' performance per tone could be explained by interactions with Cantonese tone types. Similarly, A. C. L. Yu et al. (2021) hypothesized that English and Urdu intonational types could affect performance in Cantonese tone discrimination and production, but this was not borne out by the data. However, they did find a strong effect of L1 tone types for speakers of Punjabi, a tone language. They suggest that an L1 intonational system "might not exert as strong an effect" (p. 21) as an L1 tonal system on L2 tone learning.

Overall, it thus appears that if L2 tone to L1 intonation assimilation does take place, the effect of such assimilation may be relatively weak compared to L2 tone to L1 tone assimilation (Best, 2019, p. 5; Reid et al., 2015; So & Best, 2010). Indeed, Best (2019, p. 5) points out that PAM was not specifically designed to address the "cross-tier perceptual relationships that are likely to come into play in non-native tone perception by listeners of non-tone L1s". This makes it relatively hard to test PAM's predictions for learners from a non-tonal L1 background.

Another limitation of PAM is that it is designed as a speech perception and not a speech production model. A model of speech perception and production that has also been applied to tone learning is the Speech Learning Model, or SLM (Flege, 1995), and I will elaborate on this model in more detail in Chapter 4.

Given its ubiquity in the literature, I will regularly refer to the tenets of PAM to discuss the effect of specific tone types on tone learning facility. I will simultaneously consider the effect of phonetic-acoustic properties of tones to explain why some learners learn specific tones more easily than other tones.

1.3.3.3 Musicianship

In addition to the effects of L1 tonal status and tone types, I will investigate the effect of musicianship on individual tone learning facility. I will use the word ‘musicianship’ as an overarching term to incorporate two measures of individual musical expertise that are frequently used in the tone learning literature: ‘musical experience’ and ‘musicality’.

Musical experience is a measure that expresses the number of years that an individual has had musical training, and in some instances the cumulative number of years of formal training per instrument (Bowles et al., 2016). It can be operationalized as either a continuous variable expressing years of practice per individual (Wong et al., 2020), or as a categorical variable to define individuals as either musicians or non-musicians. The definition of musician typically hinges on a number of criteria, the most common being 1) at least five years of continuous practice, and 2) the current ability to play an instrument (Bidelman et al., 2013; Chang et al., 2016; Choi, 2021; Wong & Perrachione, 2007).

Musicality is measured by performance in standardized tests such as the Musical Ear Test (Wallentin et al., 2010) or the Montreal Battery of Amusia (Peretz et al., 2003). These tests gauge an individual’s sensitivity to pitch, rhythm, and other features relevant to musical processing. Although by no means unrelated to one another (Cooper & Wang, 2012), musical experience and musicality are essentially mutually exclusive. An individual can have had years of musical experience but still have low sensitivity to musical pitch and rhythm, and therefore perform poorly in musicality tests. Similarly, an individual can have never practiced music in their life, but still be highly sensitive to musical pitch and rhythm and score high in musicality tests. Therefore, where adequate and relevant, I will use 'musical experience' and 'musicality' separately, but if the distinction is not crucial for purposes of the discussion, I will use the generic term 'musicianship' instead.

There are many parallels between music and speech, particularly in terms of the role that pitch plays in both domains. In music, pitch is essential for the formation of melodies. In speech, it is essential for the formation of lexical or phrasal prosody (Sadakata et al., 2020). These parallels give rise to the idea that individuals who are good at using pitch in music may also be good at using pitch in speech, and that therefore musicianship can facilitate tone learning.

The underlying rationale as to why musicianship – in particular musical experience – would facilitate tone learning has been theoretically described in the OPERA hypothesis (Patel, 2011). Although it is designed as a model for speech processing at large, I will here highlight its applications to pitch processing. The OPERA hypothesis suggests that musical experience can facilitate linguistic pitch processing because it has the potential to enhance "adaptive plasticity". Crucially, this facilitation is only expected to take place when musical experience satisfies five criteria, namely Overlap, Precision, Emotion, Repetition, and Attention: OPERA.

"Overlap" refers to the assumption that there is an overlap in the subcortical brain regions that are engaged in the processing of periodicity (the shared acoustic correlate to linguistic and musical pitch). The "Precision" criterion refers to the assumption that musical pitch processing requires more precision than linguistic pitch processing. For instance, pitch movement of just one semitone is acoustically relevant for music, but not for speech (Peretz & Hyde, 2003). Therefore musicians are only expected to show enhanced linguistic pitch processing if they have more precise pitch processing skills gained from musical practice. Third, musical experience is only assumed to enhance plasticity in linguistic pitch processing if musical practice is associated with positive "Emotion", for instance in the shape of internal satisfaction or praise from others. Finally, musical activities must be practiced frequently and across a relatively long period ("Repetition"), and musical training must enhance "Attention" to sound details at large in order to facilitate linguistic pitch perception (Patel, 2011, pp. 8–9).

Empirical evidence generally supports the OPERA hypothesis of music-to-speech transfer. In native tone processing, Mandarin-L1 musicians show better tone discrimination abilities than their non-musician counterparts (W. Tang et al., 2016), although this enhanced perceptual ability may be limited to psychoacoustic tone discrimination tasks (F. Chen & Peng, 2018; H. Wu et al., 2015). As to non-native tone processing, English-L1 musicians have been found to outperform English-L1 non-musicians, but also Mandarin-L1 non-musicians in Mandarin tone discrimination (D. Chang et al., 2016). This supports OPERA's notion that the processing of pitch in music is more precise and demanding than the processing of pitch in language, making music-derived pitch processing skills more facilitative than language-derived processing skills for tone perception (Hutka et al., 2015). However, in the same study (D. Chang et al., 2016), English-L1 musicians did not outperform Mandarin-L1 non-musicians in Mandarin tone *categorization*. This highlights that musical pitch experience may only be more beneficial than linguistic pitch experience in psychoacoustic, and not in categorical perception tasks (R. K. W. Chan & Leung, 2020; Wayland et al., 2010).

In a commentary on OPERA, Choi (2021) recently suggested that the model should incorporate an extra requirement, namely "Lack of relevant experience", in order to more precisely account for the facilitative effect of musicianship on non-native pitch processing. Choi suggests that musicianship may only facilitate non-native pitch processing if a learner has no other relevant pitch-related experience, such as prior experience with a tone language. This stems from findings from cross-linguistic studies that suggest that the facilitative effect of musicianship is particularly strong for speakers of non-tonal languages, whereas the effect for speakers of tonal languages is small or virtually non-existent, both in pre-lexical tone processing (S. Chen et al., 2020) as well as in lexical processing (Cooper & Wang, 2012; Maggu et al., 2018). In other words, individuals with enhanced musical pitch processing skills (trained musicians) only benefit from these skills compared to individuals with limited musical pitch skills (non-musicians) in linguistic pitch processing when neither of these individuals have any relevant experience with linguistic pitch processing. Therefore, throughout this dissertation, I will examine whether there is a differential in the facilitative effect of musical experience depending on an individual's linguistic pitch experience, i.e., their L1 tonal status.

1.3.3.4 Working Memory

In addition to the effect of musical experience, this dissertation examines the effect of working memory (WM) on individual differences in non-native tone learning facility. In this section, I will describe what working memory is, how it can be measured, and how working memory has been theoretically and empirically linked to non-native speech learning.

Working memory has been described as:

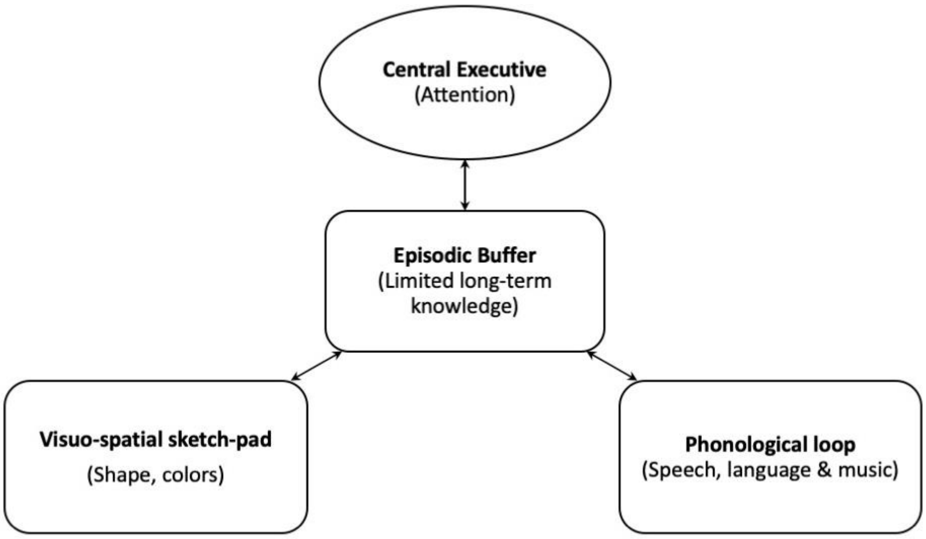
> *"The system of systems that are assumed to be necessary in order to keep things in mind while performing complex tasks such as reasoning, comprehension and learning"*
> (Baddeley, 2010, p. 136)

The canonical theoretical framework on the link between memory and speech comes from the work by Baddeley & Hitch, who present a model of working memory and the "phonological loop" (Baddeley, 2003; Baddeley & Hitch, 1974). This model was recently reviewed in Baddeley & Hitch (2019). In this review, they provide a breakdown of the four components of working memory, which I will summarize hereunder.

A key tenet of working memory is that it not only acts as an information store, but also as a processing system for that information. The central component of working memory is an attentional control system called the "central executive". This central executive is connected to a storage system for visual information called the "visuo-spatial sketch pad" and a store for verbal and acoustic information called the "phonological loop". The fourth component of working memory is called the "episodic buffer" (Baddeley & Hitch, 2019). This buffer stores a limited amount of long-term semantic and episodic information, which it combines with information from the visuo-spatial sketchpad and the phonological loop. Unlike the central executive, the episodic buffer is primarily concerned with information storage rather than with attentional control. A schematic overview of the most recent model of working memory is provided in Figure 4.

Figure 4

The model of working memory by Baddeley & Hitch' (2019).



Of all these components, the phonological loop is believed to be the most relevant to language. This is because it stores and processes verbal and acoustic information. The visuo-spatial sketchpad integrates visual information and is deemed less relevant to language, although it may be involved in certain reading tasks that require the reader to pay attention to the layout and composition of a text (Baddeley, 2003, p. 200). As a consequence, most literature on the link between working memory and language focuses on the phonological loop.

The phonological loop itself can be broken down into two components: a temporary store and a subvocal rehearsal system. The temporary store holds memory traces for a few seconds, and these memory traces decay unless they are stored and reactivated by the subvocal rehearsal system (Baddeley & Hitch, 2019).

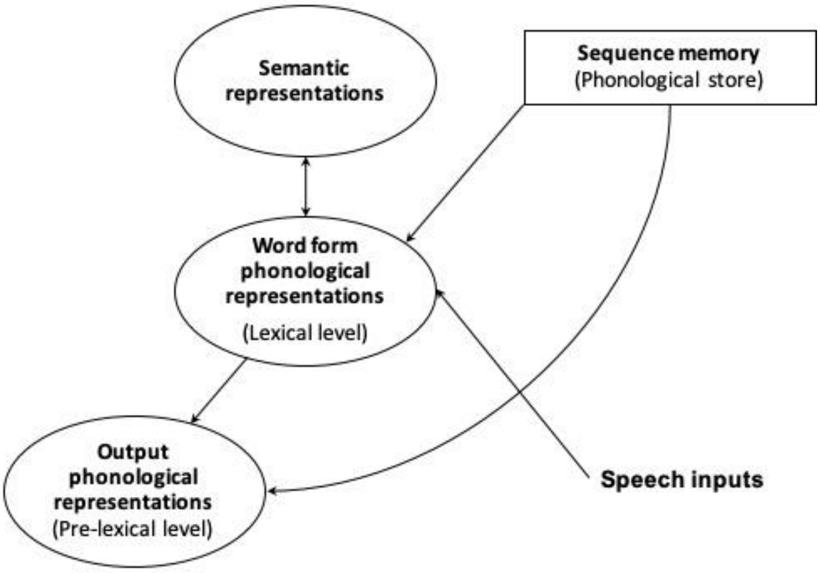
This theoretical formulation of the phonological loop allows for the design of tasks that gauge an individual's capacity of temporary storage and rehearsal, and such tasks have been used widely to measure working memory capacity. In practice, working memory capacity in the phonological loop is assumed to be measurable by *string recall tasks* in which participants must recall a string of items (digits, letters, or words) upon visual or auditory presentation. It is assumed that, when asked to recall a string, participants will subvocalize the string that is presented to them (i.e., they would repeat this in their head), and this will allow them to recall that string (Mattys & Baddeley, 2019, p. 1121).

Decades of empirical research provide a strong link between the phonological loop and language processing, particularly in a second language. Baddeley & Hitch (2019, p. 101) describe the phonological loop as “a confluence point for language-related material”, and refer to a meta-analysis of over 3,700 learners from 79 samples by Linck et al. (2014) that shows strong links between phonological loop capacity and second language proficiency. In general, working memory capacity has been found to be positively correlated with performance in second language proficiency, particularly in word learning (Atkins & Baddeley, 1998; Kormos & Sáfár, 2008), and to some extent in speech perception (Goss, 2020), although the strength of such correlations may depend heavily on the nature of the working memory task and the aspect of word learning or speech perception that is measured (Bidelman et al., 2013; Hutka et al., 2015).

The question arises at this point why it would be theoretically intuitive that individuals who are good at recalling sequences would also be good at processing non-native sounds and learning non-native words. A consideration as to why this may be the case can be found in Gupta (2003), who presents a theoretical account of the link between working memory (which he refers to as ‘sequence memory’) and language processing at a pre-lexical and at a lexical level. A schematic overview of the sequence memory model is provided in Figure 5.

Figure 5

Gupta's model of sequence memory and word learning (2003).


Semantic representations
Sequence memory (Phonological store)
Word form phonological representations (Lexical level)
Output phonological representations (Pre-lexical level)
Speech inputs

Gupta's model, which is primarily designed as a model for L1 word learning, relates "sequence memory" to processes involved in word learning. Sequence memory is described as a short-term sequencing mechanism that takes snapshots of linguistic representations. It is assumed to be quantifiable by string recall tasks. The sequence memory is similar to the temporary store in the phonological loop, but differs in that it is not a store in which items are entered, and rather a "serial ordering device that sets up associations to a sequence of activations in the lexical system" (Gupta, 2003, p. 1215). At a pre-lexical level, sequence memory can support the retention and recall of sequences of individual sub-lexical items that together form a word, i.e., individual phonemes. Thus, individuals with good sequence memory capacity will show high accuracy in pre-lexical tasks such as imitation tasks, which require an individual to immediately repeat an auditorily presented word without necessarily having to process that word at a lexical level (Gupta, 2003, p. 1230). Although the notion of 'accuracy' is not defined by Gupta, it is plausible that accuracy here refers to accurate oral production of the individual phonemes. At a lexical level, sequence memory can support the retention and recall of sequences of words. Overall, these facilitations of sequence memory on both short-term pre-lexical and lexical processing are assumed to provide a basis for long-term word learning. Gupta underpinned this theory with a number of experiments in which there was evidence for a correlation between English-L1 participants' ability to recall digits in a string recall task and their ability to learn novel words in a word learning task, operationalized by image-naming (Gupta, 2003).

Individuals vary in their capacity to recall strings and sequences: some will be able to recall relatively long strings (e.g., sequences of up to 8 digits), whereas others will only be able recall relatively short strings (e.g., sequences of up to 3 digits). The maximum length of a sequence recalled is typically referred to as an individual's *span*. In string recall tasks involving digits, i.e., a *digit span task*, normally developing adults typically have a span of up to 7 digits, with a margin of +/- 2 (Miller, 1956).

In this dissertation, I measured individuals' working memory (or, under Gupta's definition, their sequence memory) by means of a backwards digit span task. In a backwards digit span task, a participant is presented with a string of digits (e.g., 1-2-3) and is required to recall that sequence in reverse order (i.e., 3-2-1). A backwards digit span task is generally viewed as a reliable measure of working memory given that backwards recall not only requires storage and rehearsal of string information in the phonological loop, but also manipulation of that string in the central executive (Kormos & Sáfár, 2008, p. 263; Oberauer et al., 2000). However, it has also been argued that at least for some adults, backwards digit recall may only require short-term memory storage which does not involve processing and attentional capacities from the central executive (St Clair-Thompson, 2010). At the same time, some studies suggest that specific measures of working memory or short-term memory may in fact tap into the same information storage and processing capacities (Colom et al., 2006).

Acknowledging that each measure of working memory capacity may have its advantages and disadvantages, the choice for a memory span task with digits instead of words was motivated by the fact that I conducted experiments with participants from six different language backgrounds. A word span task would have required control for phonotactic (dis)similarity with words in each L1 (Baddeley, 2003, p. 191; Gathercole, 1995). I deemed digits more language-independent, although it has been suggested that the syllable count of digits across languages may have a small effect on individual digit recall (Schmidt et al., 2020). The choice for backwards digit span over forward span was primarily motivated by the fact that the former is more demanding, and therefore more likely to tap into both short-term memory and more complex working memory capacities related to speech learning.

As will be discussed in more detail in the individual chapters, the role of WM on non-native tone processing has been only sparsely studied, and previous studies present mixed findings: some studies suggest that WM facilitates non-native tone processing (Bowles et al., 2016; Goss, 2020; Ingvalson et al., 2017) whereas others fail to find a clear link (Goss & Tamaoka, 2019; Perrachione et al., 2011). Given this relatively unclear link between WM and non-native tone processing, the studies reported in the individual chapters will assess whether and how WM facilitates tone perception, production, and word learning, whilst simultaneously accounting for L1-specific factors and individual musical experience.

1.3.3.5 Pitch aptitude

The fifth and final factor of which I will investigate its facilitative effect on tone learning is pitch aptitude, which refers to the ability to perceive tones at the pre-lexical level. It is measured by accuracy in a pre-lexical tone categorization task on meaningless syllables (Wong & Perrachione, 2007), and is believed to facilitate lexical tone processing.

Previous studies have investigated the effect of pitch aptitude on lexical tone processing, and all refer to the same concept and use the same instrument to measure it (accuracy in a tone categorization task), but use different terminologies in doing so, namely "pitch identification" (Wong & Perrachione, 2007), "basic perceptual abilities for pitch" (Perrachione et al., 2011), "pitch ability" or "pitch processing" (Bowles et al., 2016), "individual aptitude" (Dong et al., 2019), or "phonological processing" (Ling & Grüter, 2020). For coherence, I will henceforth use the term 'pitch aptitude' to refer to this concept.

In the abovementioned studies, pitch aptitude was generally found to facilitate lexical tone processing. I therefore deemed it necessary to also investigate the effect of pitch aptitude on performance in my word identification and image-naming tasks.

Recall that pitch aptitude is measured by a tone categorization task, which was also the instrument I used to measure pre-lexical tone perception. Pitch aptitude in this dissertation is therefore only used as an additional factor to investigate *lexical* tone perception and tone production facility. I did not investigate the effect of another factor on pre-lexical tone perception facility in addition to L1 tonal status, tone type, musical experience, and working memory. I acknowledge that it is possible to investigate whether in turn, pre-lexical tone perception is facilitated by additional pitch-related abilities, such as the individual auditory ability to perceive just noticeable differences (JNDs) in F0 on non-speech stimuli as measured by an “adaptive pitch test” (Y. S. Chang et al., 2017; Goss & Tamaoka, 2019; Mandel, 2009; Wiener & Goss, 2019). However, assessing the effect of such individual auditory abilities fell outside the scope of my studies.

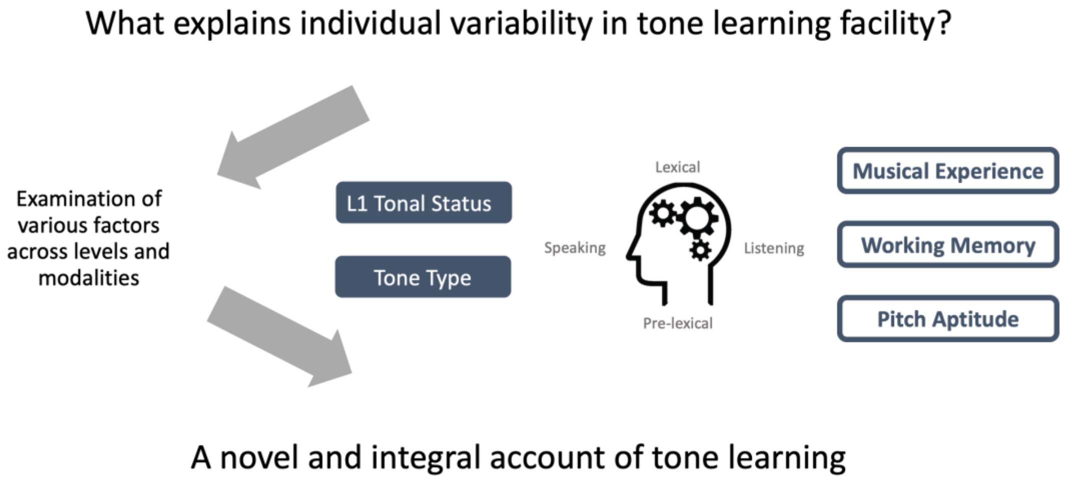
1.4 Research question, approach, and expected outcome

Having introduced the key principles of pre-lexical and lexical tone learning and the factors that may facilitate the ease with which individuals learn tones in a non-native language, I finally formulate this dissertation's research question, approach, and expected outcome.

In this dissertation, I ask what explains individual variability in tone learning facility. I approach this question by assessing to what extent L1 tonal status, tone type, musical experience, working memory, and pitch aptitude affect the ease with which individuals process tone, at a pre-lexical and at a lexical level, and in the listening and speaking modalities. As I will outline in more detail in each empirical chapter, whereas previous studies have separately investigated the effects of these individual factors on tone learning at a specific level of processing or in a specific modality, there appears to be no comprehensive study that has examined an array of factors on tone learning at large. I therefore expect that the outcome of this dissertation will be a novel and integral empirical and theoretical account of tone learning. The research question, approach, and expected outcome is schematically summarized in Figure 6.

Figure 6

Research question, approach, and expected outcome.


What explains individual variability in tone learning facility?
Examination of various factors across levels and modalities
L1 Tonal Status
Tone Type
Lexical
Speaking
Listening
Pre-lexical
Musical Experience
Working Memory
Pitch Aptitude
A novel and integral account of tone learning

Chapter 2 Tone categorization and word identification[2]

Adult second language learners often show considerable individual variability in the ease with which they learn lexical tones. It is known that factors pertaining to a learner's L1 (such as L1 tonal status or L1 tone type) as well as extralinguistic factors (such as musical experience and working memory) modulate tone learning facility. However, how such L1-specific and extralinguistic factors affect performance together in dynamic ways is less well understood. Therefore, to unpack the potential interactions between these factors for individual learners, we assessed the combined effects of L1 tonal status, L1 tone type, and musical experience and working memory on L2 tone perception and word learning of tonal pseudowords by English-L1 and Mandarin-L1 adult learners, by using a pre-lexical tone categorization task and a lexical word identification task. We found that L2 tone perception and word learning were primarily facilitated by extralinguistic factors, but that the degree to which learners rely on these factors is modulated by their L1 tonal status, as for instance musical experience facilitated perception and word learning for English, but not for Mandarin participants. We also found clear effects of L1 tone type, as Mandarin participants tended to struggle with categorizing and lexically processing level tone contrasts, which do not occur in Mandarin.

[2] Adapted from: **Laméris, T. J.**, & Post, B. (2022). 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*, 1–39. https://doi.org/10.1177/02676583221090068

2.1 Introduction

In tone languages, fundamental frequency (F0) acts as a primary acoustic cue to change a word's core lexical meaning (Yip, 2002). For adult L2 learners, lexical tones are thought to be relatively difficult to master. In particular, while they may overcome difficulties in processing tones devoid of lexical meaning in tone *perception* (X. Wang, 2013), it appears that linking tones to a lexical item in *word learning* presents considerably more persistent difficulty (Pelzl et al., 2019, 2020). Yet, as with all aspects of speech, some learners appear to perceive tones and learn tone words more easily than others do, reflecting the large degree of individual variability in L2 learners' speech learning *facility*, i.e., the ease with which non-native sounds are learned in the early stages (Bowles et al., 2016, pp. 774–775; Kachlicka et al., 2019). To better understand what accounts for this individual variability, this Chapter examines how factors pertaining to a learner's L1, as well as extralinguistic factors, jointly affect L2 tone perception and word learning facility.

We will use the term 'L1-specific factors' to refer to linguistic factors pertaining to a learner's L1, and zoom in on *L1 tonal status* (i.e., does the L1 use tones for lexical purposes?) and *L1 tone type,* (i.e., what types of F0-based units, either tonal or intonational, exist in the L1?). In addition, we will use the term 'extralinguistic factors' to refer to individual factors not related to the L1, and focus in this Chapter on musical experience and working memory. As we will review in section 2.2, all these factors are known to modulate L2 tone perception and word learning facility. However with a few notable exceptions[3] (R. K. W. Chan & Leung, 2020; D. Chang et al., 2016; S. Chen et al., 2020; Cooper & Wang, 2012), most previous studies either only assess the effects of L1-specific factors, controlling for or not measuring the effect of extralinguistic factors (Braun et al., 2014; J. Chen et al., 2020; So & Best, 2010), or they assess extralinguistic factors, but in participants of the same L1 (Bowles

[3] -Chan & Leung (2020) investigated the effect of tonal status (L1 Cantonese and L1 English) and musical experience on 'phonological learning' (in between pre-lexical and lexical learning) of Thai tones.
-Chang et al. (2016) investigated the effect of tonal status (L1 Mandarin and L1 English) and musical experience on Mandarin and musical tone perception.
-Chen et al. (2020) investigated the effect of tonal status (L1 Mandarin and L1 English) and musical experience on tone perception of meaningless syllables.
-Cooper & Wang (2012) investigated the effect of tonal status (L1 Thai and L1 English) and musical experience on Cantonese tone perception and word learning.

et al., 2016; Wong et al., 2020). Therefore, instead of looking at these factors separately, we examine the combined effects of L1-specific and extralinguistic factors to try to provide a more complete and accurate account of individual variability in L2 tone learning. More specifically, this Chapter investigates how L1 tonal status, L1 tone type, musical experience and working memory – factors that have not been investigated simultaneously in previous studies – work together to modulate performance in a tone categorization task (representing *tone perception*) and in a pseudoword word identification task (representing *tone word learning*) by a group of tonal (Mandarin-L1) and non-tonal (English-L1) learners.

We will first review the existing literature on the effects of L1 tonal status, L1 tone type, and musical experience and working memory on tone perception and tone word learning.

2.2 Background

2.2.1 L1-specific factors in non-native tone perception

There is ample evidence that L1 tonal status modulates individual performance in tone perception. In comparison to non-tonal peers, L1 speakers of a tonal language (henceforth: 'tonal L1ers') tend to process tones predominantly in the left brain hemisphere (Klein et al., 2001; Y. Wang et al., 2004), perceive L2 tones in a categorical rather than in a psychoacoustic way (Hallé et al., 2004), and tend to be better at identifying tones spoken by multiple speakers (Y. S. Chang et al., 2017). Some studies further show that the stronger the lexical role of pitch in the L1, the better the sensitivity to pitch in an L2 (Schaefer & Darcy, 2014), and that not only L1 but also L2 knowledge of a tonal language can facilitate non-native pitch perception (Wiener & Goss, 2019). While this suggests that tonal L1ers perceive tones *differently* than their non-tonal peers, by no means do they always perform *better*, as evidenced by findings of tone identification and discrimination tasks in which tonal L1ers do not outperform their non-tonal peers (Cooper & Wang, 2012; Francis et al., 2008; Gandour & Harshman, 1978; So & Best, 2010; X. Wang, 2013). Note however, that there are findings that do suggest a comparative advantage in tone perception for tonal L1ers (R. K. W. Chan & Leung, 2020; Peng et al., 2010; Wayland & Guion, 2004).

One reason why L1 tonal status alone may not explain individual differences in L2 tone perception is because the factor of L1 tone type needs to be considered. Simply put, rather than L2 tones overall, it is often specific L2 tones that may be easy or difficult to perceive, depending on the tone types in a learner's L1. Note that we will use the term 'L1 tone type' as an overarching expression to describe specific F0-based units (which can be either lexical or intonational tones) occurring in the L1 in terms of 1) phonological-categorical and 2) phonetic-acoustic properties, following the distinction proposed by K. Yu et al. (2017).

Previous studies have suggested that L1 tone type (in phonological-categorical terms) affects L2 tone perception because listeners may assimilate L2 tones to F0-based categories in the L1 (S. Chen et al., 2020; Hao, 2012; So & Best, 2010). This notion of categorical assimilation is rooted in models of L2 speech perception such as the Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007) that propose that the ease with which non-native sounds are perceived depends on the relative similarity between L1 and L2 sounds. For example, L1 speakers of Mandarin, which only has one high-level tone, appear to struggle with discriminating Cantonese mid-level and low-level tones (Qin & Jongman, 2016; Zhu et al., 2021). It has been suggested that this is because Mandarin listeners tend to assimilate Cantonese level tones to the single Mandarin level tone, making them therefore relatively difficult to perceive accurately (Qin & Jongman, 2016, p. 334; Zhu et al., 2021, p. 4224).

Crucially, non-tonal listeners may be less affected by categorical assimilation because they simply do not have competing lexical tone categories in their L1. Although they may assimilate L2 tones to intonational categories, effects of such assimilation on L2 tone perception may be relatively weak (Best, 2019, p. 5; Reid et al., 2015; So & Best, 2010, 2014), arguably because intonational categories have a "weaker (less categorical) mental representation" than lexical tone categories (Francis et al., 2008, p. 269). As a result, even though they may fail to form abstract L2 tone categories (R. K. W. Chan & Leung, 2020, p. 10), non-tonal listeners may in some instances perceive L2 tones more accurately than tonal listeners by processing them in a psychoacoustic manner (A. Chen et al., 2018; Peng et al., 2010; X. Wang, 2013; K. Yu et al., 2019).

An alternative account describing the effect of L1 tone type on L2 tone perception focuses on phonetic-acoustic rather than phonological-categorical properties. For instance, speakers of Mandarin appear to pay relatively more attention to differences in F0 contour and direction, whereas English speakers may pay relatively more attention to F0 height when processing pitch, which could potentially explain the difficulty for Mandarin speakers to perceive level tone contrasts in an L2 (Francis et al., 2008; Gandour & Harshman, 1978; Qin & Jongman, 2016).

Finally, we note that attentional differences between listeners of different L1s to secondary cues of lexical tones may also modulate L2 tone perception (S. Chen et al., 2017). For instance, laryngeal phonation (creaky voice) facilitates perception of low-register tones in Cantonese-L1 listeners (K. M. Yu & Lam, 2014) and of low-dipping tones in Mandarin-L1 listeners (R. Yang, 2015). In this study, we will zoom in on F0 as the primary acoustic cue to lexical tone and only manipulated F0 between the stimuli, but we will consider the possible effect of the absence of other acoustic cues on participants' tone perception in the discussion (section 2.6).

2.2.2 L1-specific factors in non-native tone word learning

Whereas accounting for individual differences in L2 tone *perception* based on L1 tonal status alone remains relatively complex, particularly because of the effect of L1 tone type on the perception of specific L2 tones, it appears that individual differences in L2 tone *word learning* can be more easily accounted for by L1 tonal status.

For instance, Pelzl et al. (2019) report that English-L1 advanced L2 learners of Mandarin can accurately perceive pitch in a pre-lexical tone categorization task, but may not all be able to "repurpose it as a lexical cue" (p. 80) in lexical tasks. In an eye-tracking study, Ling & Grüter (2020) similarly found that English-L1 intermediate learners of Mandarin had "considerably more difficulty in using tone alone to distinguish between words" (p. 19).

It is crucial to note that these studies involved Mandarin participants listening to their own L1, thereby perhaps naturally yielding an advantage of L1 tonal status in comparison to non-tonal participants. However, evidence for a facilitative effect of L1 tonal status in L2 tone word learning is also found in studies in which tonal L1ers were exposed to a different tone language. For instance, Poltrock et al. (2018) showed that Mandarin participants outperformed French listeners in recalling Cantonese pseudowords that contrasted in tone.

Chan & Leung (2020) investigated the effects of L1 tonal status on the incidental "phonological learning", which was defined as an intermediate step between tone perception and tone word learning (p. 4). They show that Cantonese participants outperformed English participants in the phonological learning of Thai tones, and suggest that Cantonese L1 tonal status facilitated the formation of syllable-level tone categories required for utilizing tones at the word level.

It thus appears that L1 tonal status on its own may facilitate L2 tone word learning, given tonal L1ers' familiarity with using pitch to indicate lexical meaning (Cooper & Wang, 2012, p. 4765). However, to the best of our knowledge, there are no studies that examine whether in addition to L1 tonal status, L1 tone type also modulates L2 tone word learning in a similar way that it is known to modulate L2 tone perception. To address this gap in the literature, we first ask:

RQ1: *How do Mandarin participants' L1 tonal status and L1 tone types affect individual performance in a tone categorization task and a word identification task of tonal pseudowords with a rising, a falling, a mid-level, and a low-level tone, and how does this compare to performance by English participants?*

2.2.3 Extralinguistic factors: musical experience and working memory

There has been an increasing interest in recent years to explain individual variability in L2 tone learning by not only looking at learners' L1-specific, but also extralinguistic factors. Here, we focus on two of these factors, musical experience and working memory, and review previous studies that have investigated their role in L2 tone perception and word learning.

Musical experience is one of the most investigated extralinguistic factors in the L2 tone perception and word learning literature, possibly due to the shared cognitive processing of pitch in music and language (Perrachione et al., 2013; Sadakata et al., 2020). For tone perception, studies with Mandarin speakers have revealed improved pitch sensitivity and tone discrimination abilities in trained musicians compared to non-musicians (W. Tang et al., 2016; H. Wu et al., 2015). In a large-scale study involving over 400 Cantonese native speakers, years of musical training was found to be the strongest predictor of performance in a tone discrimination task (Wong et al., 2020). However, some studies show no clear effect of musical experience on tone perception (R. K. W. Chan & Leung, 2020), and it has been suggested that a facilitative effect of musical experience on L2 tone perception may be task-dependent (D. Chang et al., 2016).

Studies on L2 tone word learning generally find a facilitative effect of musical experience. In one of the earliest studies on the subject, Wong & Perrachione (2007) report that English learners with musical experience performed better than non-musicians, both in pre-lexical perception of tones on meaningless syllables and in the learning of tonal pseudowords. Bowles et al. (2016) found similar facilitative effects of musical experience in a large study of L2 Mandarin word learning by 160 English-L1 participants.

As this Chapter focuses on the combined effects of L1-specific and extralinguistic factors, a key question is whether L1-specific factors such as L1 tonal status interact with extralinguistic factors like musical experience. Studies that have investigated this suggest that this is indeed the case. For instance, S. Chen et al. (2020) showed that English-L1 musicians had a stronger categorical perception of tones than non-musicians, whereas no such difference was found between Mandarin-L1 musicians and non-musicians. This suggests that the facilitative effect of musical experience on L2 tone perception may be weaker for tonal L1ers. Such an interaction between L1 tonal status and musical experience was also found in L2 tone word learning by Cooper & Wang (2012), who showed that musical experience only benefited English, but not Thai participants in Cantonese tone word learning. The authors suggest that English participants may have drawn on their pitch acuity gained through musical practice “to enhance their ability to utilize linguistic pitch in a higher-level linguistic context” (p. 4765). By contrast, the Thai participants may not have needed to additionally draw on skills gained through musical experience because they already benefited from their L1 tonal status in tone word learning, making musical experience less relevant. This suggests that there is a dynamic interplay between L1-specific and extralinguistic factors in tone word learning, and highlights the importance of accounting for both of these types of factors in investigating L2 tone learning facility.

As a second extralinguistic factor, we assessed the effect of individual learners’ working memory (WM) on performance in our tone categorization and tone word identification tasks. We deemed it necessary to include a measure of WM because our word identification task replicates word learning, for which WM has been found to be facilitative

(Baddeley, 2003; Kormos & Sáfár, 2008). In addition, we want to further investigate the role of WM in facilitating pre-lexical and lexical processing of pitch following conflicting findings in the literature. Findings from previous studies suggest that WM may not facilitate pre-lexical pitch processing, either in language or in music, although this may depend on how cognitively demanding the task is (Bidelman et al., 2013; Hutka et al., 2015). As for lexical pitch processing, studies in English-L1 participants suggest that WM facilitates word-level processing of Japanese pitch (Goss, 2020), and moderately facilitates Mandarin tone word learning (Bowles et al., 2016). However, findings from Chinese-L1 and Korean-L1 advanced learners of Japanese lexical pitch (Goss & Tamaoka, 2019) and English-L1 beginners learning tonal pseudowords (Perrachione et al., 2011) revealed no such facilitative effect. Given this relatively unclear link between WM and pre-lexical and lexical pitch processing, we therefore re-assess whether WM facilitates performance in tone categorization and word identification in English and Mandarin participants.

Finally, since our study measured both tone perception (in a tone categorization task) and tone word learning performance (in a word identification task), we will also investigate whether performance in one task predicts performance in the other. Indeed, studies that investigated the link between pre-lexical and lexical pitch processing suggest that L2 tone perception ability may in fact be one of the strongest facilitators of L2 tone word learning in English speakers (Bowles et al., 2016; Ling & Grüter, 2020; Perrachione et al., 2011; Wong & Perrachione, 2007, p. 565). However, evidence from the cross-linguistic study by Cooper & Wang (2012) suggests that L1 tonal status may attenuate the facilitative effect of tone categorization ability on tone word learning, as English-L1 participants did but Thai-L1 participants did not benefit from tone categorization accuracy (i.e., *pitch aptitude*) in Cantonese tone word learning. This leaves it relatively unclear what extralinguistic factors do facilitate tone word learning in tonal L1 participants given that, based on Cooper & Wang (2012), neither musical experience nor pitch aptitude appear to strongly do so.

In sum, the literature to date has mainly investigated how individual variability in L2 tone perception and word learning is modulated by learners' L1-specific or extralinguistic factors, but only a handful of studies have examined the combined effect of such factors. Yet, findings that suggest that musical experience facilitates L2 tone word learning in English but not in Thai listeners (Cooper & Wang, 2012) highlight that simultaneously accounting for an array of L1-specific and extralinguistic factors may provide a more refined view of how individual factors modulate L2 tone learning facility. Therefore, our study combines L1-specific and extralinguistic factors, which were only partially addressed in previous studies, to better understand the relative weighting of and interactions between these factors on performance in L2 tone perception and in word learning. We therefore ask as our second research question:

RQ2: *How do Mandarin participants' L1 tonal status and L1 tone types interact with musical experience and working memory to determine performance in our tone categorization and word identification tasks, and how does this compare to English participants?*

2.3 Methods

We assessed the combined effects of L1 tonal status, L1 tone type, musical experience and working memory (WM) in tone perception and word learning by means of two behavioral tasks: A tone categorization task and a tone word identification task.

2.3.1 Participants

The study was approved by the Research Ethics Committee of the Faculty of MMLL at the University of Cambridge. 21 native speakers of English (11 female; mean age: 20.98) and 20 native speakers of Mandarin Chinese (10 female; mean age: 22.63) participated in this study. Participants were all recruited at the University of Cambridge, participated voluntarily and were paid for their participation. Within each group, half of the participants were musicians, which we defined as participants who were actively practicing music and who had more than 6 years of formal musical training (Cooper & Wang, 2012; Wong & Perrachione, 2007). An overview of the participants is given in Table 1 and a detailed description is provided in Appendix 2.1–2.

None of the participants claimed to be simultaneous bilinguals (i.e., being fully proficient in two languages acquired since birth), but many had knowledge of a second language and some had some exposure to a heritage language. Some speakers in the

Mandarin group reported to have some knowledge of another Chinese language or dialect (including Wu and Cantonese)[4]. None of the English participants had knowledge of a pitch-accent or tone language.

Participants' working memory was estimated by a backwards digit span task, as outlined later in this section. The measure of musical experience was computed as the number of years of playing a musical instrument including formal instruction. Equivalence tests with Cohen's d set at 0.5 (Lakens et al., 2018) revealed no significant differences in the age, WM, or musical experience between the two groups.

Table 1

Participant demographics.

	English		Mandarin	
	(*n* = 21)		(*n* = 20)	
Age (years)	20.98 (1.56)		22.63 (3.32)	
WM score (%)	57.90 (23.19)		72.09 (23.54)	
Pitch perception aptitude (%)	89.79 (13.91)		95.22 (4.92)	
Musical experience (years)	MU (n=11)	NM (n=10)	MU (n=10)	NM (n=10)
MU= musicians, NM= non-musicians	13.32 (2.84)	0.90 (1.66)	14.84 (5.09)	0.98 (0.97)

Values are means with standard deviations in brackets.

2.3.2 Stimuli

Two sets of audio stimuli were used: a set of vowels ([i] [a] and [ɛ]) for the tone categorization task and a set of pseudowords (/nɔn/, /lɔn/, /jɑɹ/ and /juɹ/; see Table 2, page 44) for the word identification task. These stimuli carried either a rising, a falling, a mid-level, or a low-level tone, resulting in 3 x 4 = 12 tone stimuli and 4 x 4 = 16 word stimuli. The four tones were chosen explicitly to assess the effect of L1 tone type on Mandarin participants, with the rising and falling being exemplars of the rising and falling tones in Mandarin, but mid-level and low-level tones both being similar to the single Mandarin high-level tone in terms of pitch contour. Following the predictions of the Perceptual Assimilation Model (Best,

[4] We note that some of the Chinese L2s reported by our Mandarin speakers have level tone contrasts unlike Mandarin, which may have affected performance on our mid- and low-level tones. However, a visual inspection of performance by participants who reported a L2 with level tone contrasts versus participants who did not, did not reveal notable differences (see Appendix 2.12–13). In addition to the fact that all participants reported that Mandarin was their L1 and the language they used the most, we therefore deemed it fit to group these participants together.

2019; Best & Tyler, 2007), this may make the rising and falling tones relatively easy, and the mid-level and low-level tones relatively difficult to process for Mandarin speakers. As to the similarity with English intonational types, the rising and falling tones resemble canonical rising LH and falling HL intonational types in Southern British English. The mid-level and low-level tones do not clearly map onto any intonational type, as there are no clear intonational categories in English that are distinguished exclusively by pitch height (Grabe et al., 2003, 2004; Ladd, 2012, p. 91). If we assume that English listeners assimilate non-native tones to intonational types, this too may make the rising and falling tones relatively easy and the level tones relatively difficult. However, given that it remains relatively unclear if tone-to-intonation assimilation occurs in the first place – and even if it does, whether it exerts a strong effect on non-native tone processing (Best, 2019, p. 5; So & Best, 2010; A. C. L. Yu et al., 2021, p. 21) – we refrain from making strong predictions about the relative difficulty of the tone types for the English listeners based on their similarity to English intonational types.

To avoid either of the groups being favored by listening to stimuli produced by a speaker of their own native language (Braun & Johnson, 2011), stimuli were recorded by two native speakers of Italian, who were trained singers. To ensure that participants would not be influenced by voice familiarity across tasks and to help abstract away from the F0 traces to tone categories, the female voice was used in the tone categorization task and the male voice in the word identification task.

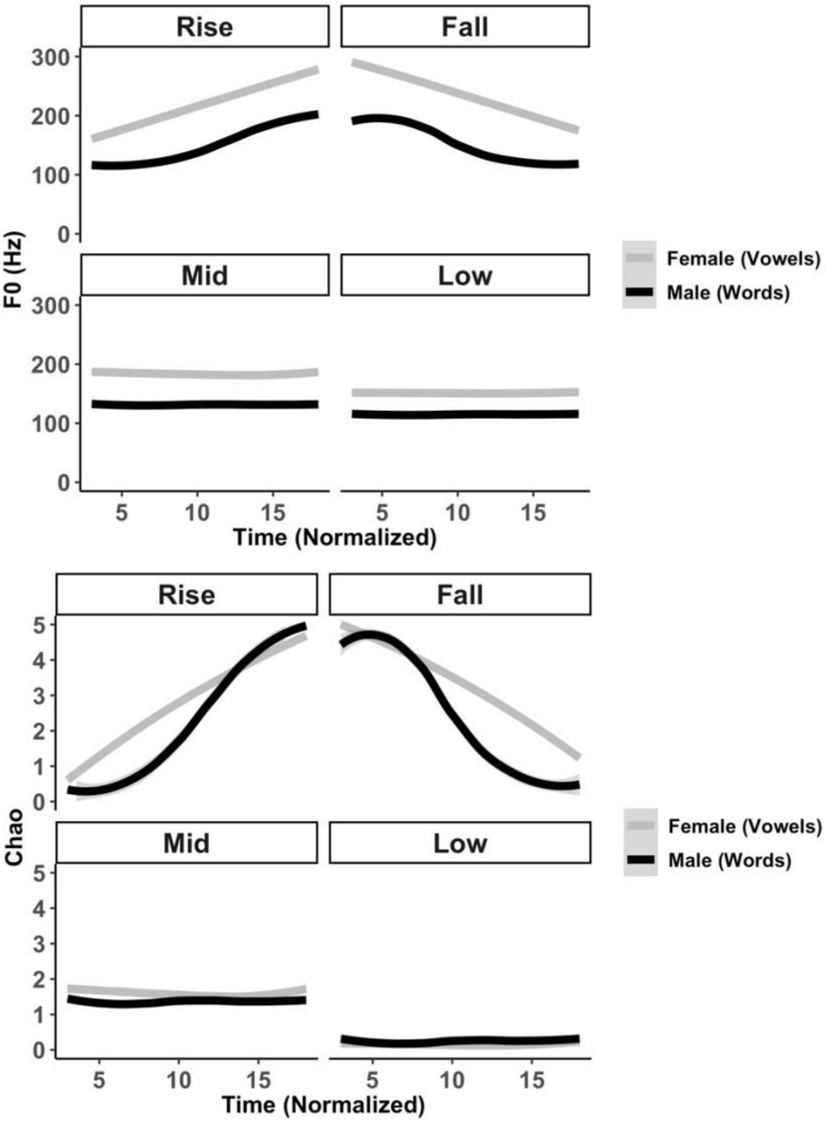
Stimuli were recorded in a sound-attenuated booth at a sampling frequency of 48 KHz. The speakers were instructed to produce stimuli with a flat tone at a comfortable pitch level. The F0 contour of this naturally produced flat tone was taken as a baseline tone (the mid-level tone). The speakers were also instructed to naturally produce stimuli with a rising, falling, and low-level tone. Based on the F0 onset and end values of these natural productions, the mid-level tone stimuli were then resynthesized using Pitch-Synchronous Overlap and Add (PSOLA) in *Praat* (Boersma & Weenink, 2019) to create stimuli for the other tones. This ensured that tone minimal quadruplets only differed in F0 and not in other acoustic cues. Both the male and the female tones had the same relative tone values in terms of Chao numerals (Chao, 1968) and the stimuli in the tone categorization and word identification tasks were therefore deemed to belong to the same four tone categories: namely 15 (rise); 51 (fall); 22 (mid-level); and 11 (low-level). For visualization, the F0 and Chao-normalized tone traces are shown in Figure 7.

After resynthesis, the average intensity of stimuli was set to 70 dB (using the ‘scale intensity’ command in *Praat*). Five trained phoneticians deemed the synthesized stimuli to sound as natural as the original mid-level stimuli.

In the tone categorization task, each tone was represented by an arrow (Figure 8). In the word identification task, each pseudowords was linked to an image to establish a sound-meaning connection (Figure 9). The images were gathered from a database by Rossion & Pourtois (2004) and represent 16 high-frequency nouns (Battig & Montague, 1969; van Overschelde et al., 2004). Care was taken to select words that were semantically unrelated to each other to facilitate word learning (Nation, 2000).

Figure 7

Smoothed F0 and Chao numeral traces for the four tones.



Shading ribbons, where present, indicate a 95% Confidence Interval.

Figure 8

Visual stimuli in tone categorization task.

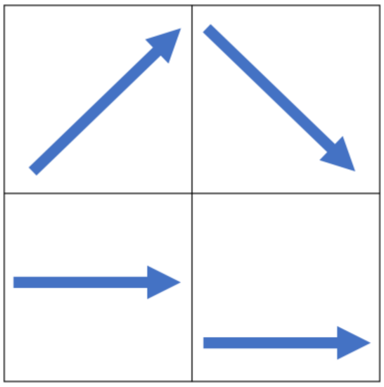


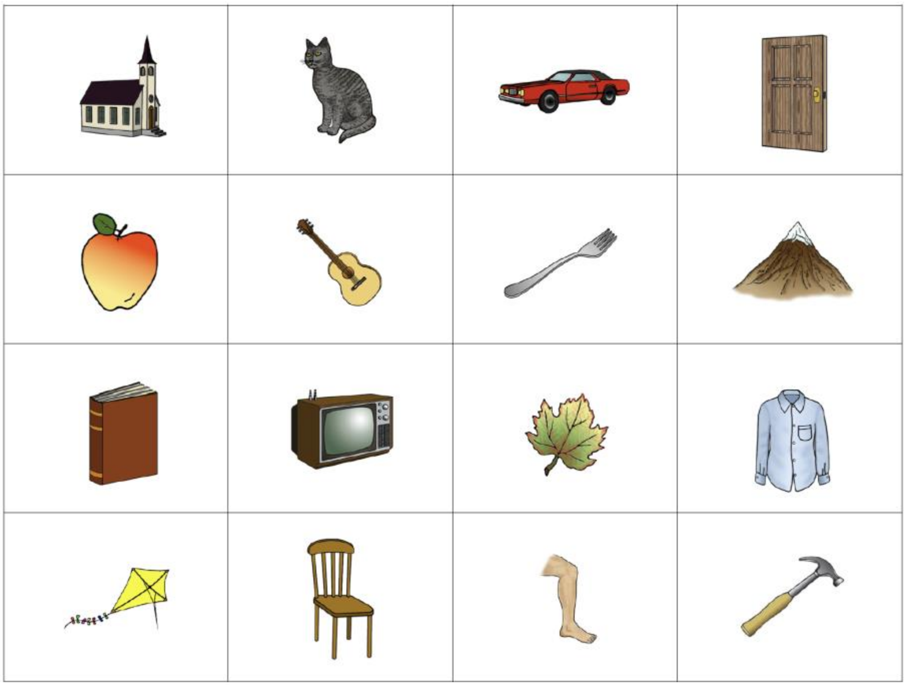
Table 2

Pseudowords.

	Tone 1	Tone 2	Tone 3	Tone 4
	(Rising 15)	(Falling 51)	(Mid-level 22)	(Low-level 11)
Segment 1	/nɔn15/	/nɔn51/	/nɔn22/	/nɔn11/
Meaning	*television*	*book*	*cat*	*fork*
Segment 2	/lɔn15/	/lɔn51/	/lɔn22/	/lɔn11/
meaning	*chair*	*leg*	*apple*	*church*
Segment 3	/jɑɹ15/	/jɑɹ51/	/jɑɹ22/	/jɑɹ11/
meaning	*mountain*	*kite*	*leaf*	*shirt*
Segment 4	/juɹ15/	/juɹ51/	/juɹ22/	/juɹ11/
meaning	*door*	*guitar*	*car*	*hammer*

Figure 9

Visual stimuli for pseudowords.



2.3.3 Procedure

A battery of eight tasks (including training sessions) was conducted over two consecutive days (Table 3). Note that in addition to the tone categorization and word identification tasks, participants also completed an image-naming task, which is reported in Chapter 3.

Participants were told that they were taking part in a study that investigated the effects of audiovisual presentation on L2 word learning. After signing a consent form, participants completed the tasks individually. The researcher (myself) only intervened at the start of new tasks to provide instructions. Written instructions for each task were in English or Mandarin. The experiment was carried out over two days to limit the total time spent in one session and to facilitate word recall after a night of sleep (Dumay & Gaskell, 2007).

All tasks were administered in a sound-attenuated booth and run on a touchscreen tablet laptop (*DELL Inspiron 13 5000 Series*) through the *OpenSesame* software (Mathôt et al., 2012). Participants listened to audio stimuli over *Beyerdynamic DT 990* headphones at a comfortable listening level.

Table 3

Overview of tasks.

DAY 1	
Description	Duration (minutes)
Tone categorization	5
Word training (imitation)	10
*Word production**	*5*
Word identification	15
DAY 2	
Description	Duration (minutes)
Working memory	5–10
Word training (imitation)	10
*Word production**	*5*
Word identification	15

**Reported in Chapter 3.*

2.3.3.1 Tone categorization task

In the tone categorization task, participants listened to a vowel carrying one of the four tones and were asked to identify the tone by touching the corresponding arrow on the touchscreen. They were encouraged to make their choice as quickly as possible and to guess if unsure. Time-out was 5000 ms after presentation of the audio stimulus.

One practice session with 16 trials (4 presentations per tone) including feedback was held at the beginning. The feedback consisted of a green circle if the response was correct or a red cross if the response was incorrect, followed by the correct sound-arrow combination. In the practice session, the vowel [o] was used, which was not used in the main session. The practice session was followed by a main session in which there were 72 trials (6 presentations per stimulus) without feedback in a randomized order.

2.3.3.2 Word training

The word training session consisted of imitation (listen-and-repeat), which was expected to be a relatively effective way to quickly memorize novel L2 words (Baills et al., 2019; M. Li

& Dekeyser, 2017)[5]. Participants were presented with the individual pseudowords (the audio stimuli) and their meaning (the images). They were asked to repeat the words out loud and pronounce them as accurately as possible, whilst simultaneously trying to memorize the words. No feedback was given regarding their pronunciation.

After a familiarization with the images and their meanings in participants' native language to ensure that participants considered the images to be analogous to a word in their L1, each of the 16 pseudowords was audiovisually presented 4 times, resulting in 64 trials in total. Participants had 5000 ms to repeat the word before the next audiovisual stimulus was presented. The first two trials were in a pseudorandomized order for all participants: each audiovisual stimulus was presented twice in a row (e.g., the word for 'cat', followed by the participant's imitation, followed by one more trial (presentation + imitation) for 'cat'), and the order was such that no segmental or tonal minimal pair followed one another. The last two presentations were fully randomized for each participant individually.

The same word training was conducted on day 2. The only difference was that the image familiarization was not conducted, and that the pseudorandomized presentation order was the reverse of that of day 1. The imitations were voice-recorded, and these production results are discussed in Chapter 3.

2.3.3.3 Word identification task

The word identification task involved image-matching to replicate L2-to-L1 word recall and tone word learning, following Barcroft & Sommers (2014); Cooper & Wang (2012). Participants would hear a pseudoword and were then prompted to identify the meaning of that

[5] We take note of empirical evidence that suggests that production during training may disrupt perceptual learning of the non-native sound to be learned, at least in certain pre-lexical tasks and when production and perception are required within the same trial (Baese-Berk & Samuel, 2016). Although our study did not investigate the effect of different training paradigms, it is worth noting that our participants reached relatively high word identification scores after only two training sessions (involving both imitation and word identification with feedback) in comparison to similar tone word learning studies that only involved feedbacked word identification trials: Participants in Cooper & Wang (2012) completed seven 30-minute training sessions spread out over two weeks to learn 15 Cantonese tone words (3 syllables x 5 tones), and mean word identification of accuracy was 67%. In addition, the imitation task was included in our study because participants were also tested on their word production (image-naming), which was expected to benefit from training in the same modality (Baese-Berk, 2019; M. Li & Dekeyser, 2017). I report on the production results from the imitation and image-naming tasks in Chapter 3.

word by making a 16-way choice on the touchscreen. The options were displayed on a 4x4 answer board, similar to Figure 9 on page 45. Participants were encouraged to make their choice as quickly as possible and to guess if unsure. Time-out per trial was set to 10 s.

Participants started with a practice block in which they received feedback to familiarize themselves with the task format, but also to further help them memorize the words through perceptual training (M. Li & Dekeyser, 2017). The feedback consisted of a green circle if the response was correct or a red cross if the response was incorrect, followed by the correct sound-image combination. Each stimulus was presented twice, totaling 32 trials, in a randomized order. This practice block lasted about 5 minutes.

The practice block was followed by a main block without feedback. To avoid that participants would associate the audio stimulus with the physical position of the image on the answer board rather than with the actual image, the images' positions were shuffled in the main block. In the main block, each stimulus was presented 6 times, totaling 96 trials, in a randomized order. There was a small break after the participants had completed two-thirds of the task. The exact same task was repeated on day 2, with the only difference being that the images' positions on the answer boards were again shuffled in the practice and main blocks.

2.3.3.4 Working memory task

Working memory was operationalized through a backwards digit span task, as one of the proxies of WM associated with retention of phonological and lexical information required for L2 perception and word learning (Baddeley, 2003; Goss, 2020, p. 28; Kormos & Sáfár, 2008).

Participants were instructed to repeat out loud in their native language and in backward order a sequence of digits presented to them on the screen. Each of the digits was presented one by one for 750 ms with an inter-stimulus-interval (ISI) of 250 ms. After a practice session, participants were presented with a block of five 2-digit sequences (e.g., 1-7; 6-3; 2-5; 8-4; 9-5). Participants would move onto a next block of five n+1-digit sequences (e.g., 5-8-2; 6-9-4; etc.) and continue to do so if they correctly repeated at least three sequences per block. If participants did not reach this threshold, the task was aborted at the end of a block. The maximum attainable block consisted of five 8-digit sequences.

A percentage working memory score was calculated by dividing the total number of digits from fully correctly recalled sequences by the maximum attainable score (175). Mean working memory scores per group are reported above in Table 1.

2.3.4 Statistical procedures

All analyses were performed in *R 4.1.1* (R Core Team, 2021). Figures were generated with the *ggplot2* package (Wickham, 2016). We present descriptive statistics and results from mixed-effects models to assess the effects of L1-specific and extralinguistic factors on performance in the tone categorization and word identification tasks. Null responses and responses with unnaturally fast reaction times (< 250 ms) were removed, excluding 0.84% and 1.42% of data points from each task, respectively. Because accuracy scores in the tone categorization task revealed a ceiling effect, we analyzed reaction times (RTs) as a main proxy of performance. For RT data, only data for correctly categorized items were analyzed. RT data were log-transformed and outliers (2.5 SDs from the mean) were removed, following Chan & Leung (2020). For the word identification task, in which there was considerably more variability in accuracy (% correctly recalled words), accuracy scores rather than RT were analyzed as a proxy of performance.

(Generalized) linear mixed-effects models were computed in the *lme4* package (Bates et al., 2015) and fitted with the *bobyqa* optimizer where applicable. Model diagnosis (observation of residual QQ plots) was carried out with the *DHARMa* package (Hartig, 2020). We adhered to a maximum Variance Inflation Factor (VIF) threshold of 5 (O'Brien, 2007) in all final models. None of the models showed multicollinearity. Post-hoc power simulations were carried out using the *simr* package (Green & MacLeod, 2016)[6].

We built models with fixed effects and interactions of interest to this Chapter's research questions. The model for tone categorization (dependent variable: log RT) contained fixed effects for *L1* (English, Mandarin; contrast-coded), *Tone* (Rise, Fall, Mid-level, Low-

[6] The observed power in our models, using the *simr* package (Green & MacLeod, 2016), following Wiener et al. (2020) for 100 simulations was 92.00% (CI: 84.84, 96.48) for *Musical Experience* and 77.00% (67.51, 84.83) for the *L1:Musical Experience* interaction in the tone categorization model. In the word identification model, it was 100.0% (96.38, 100.00) for *Musical Experience* and 95.00% (88.72, 98.36) for the *L1:Musical Experience* interaction. We acknowledge the limitations of post-hoc power analyses (Hoenig & Heisey, 2001).

level; contrast-coded), *Musical Experience* (a continuous variable expressing years of playing a musical instrument; scaled and centered), and *Working Memory* (a continuous variable expressing working memory score; scaled and centered), and the three-way interactions *L1:Tone:Musical Experience* and *L1:Tone:Working Memory.*

The final model for word identification (dependent variable: correct/incorrect) contained the same fixed effects and interactions as the tone categorization model, but in addition contained a fixed effect of *Tone Categorization* (a continuous variable expressing log RTs in the tone categorization task; centered and scaled), and an *L1:Tone:Tone Categorization* interaction to see to what extent tone perception predicts performance in tone word learning. All final models contained *Subject* (individual participant) and *Item* (stimulus) as random intercepts. Attempts were made to include random slopes but this led to convergence issues. To assess the interactions in more detail, Bonferroni-corrected multiple comparisons were generated using the *emmeans* package (Lenth, 2020).

2.4 Predictions

Based on the literature reviewed in section 2.2, we make the following predictions for our tasks in response to our research questions:

P1: Mandarin participants are expected to have slower reaction times for mid-level and low-level tones. English participants may be better at quickly categorizing level tones as opposed to contour tones. We therefore expect an *L1:Tone* interaction in the tone categorization task.

Although we are not aware of any previous literature that has investigated the effect of tone type in tone word learning, we expect the general familiarity with associating F0 to lexical meaning (i.e., *L1 tonal status)*, rather than the familiarity with specific pitch contours (i.e., *tone type*), to be a stronger predictor of performance in the word identification task. Mandarin participants are thus expected to overall outperform English participants in accurately recalling tonal pseudowords.

P2: It is expected that musical experience will not necessarily facilitate tone categorization in Mandarin speakers, but it may do so for ‘difficult’ mid-level and low-level tones, which are expected to be relatively challenging and may be identified faster by musicians as opposed to non-musicians. Musical experience is not expected to strongly predict word identification performance in Mandarin speakers. For English speakers however, musical experience is expected to be a strong predictor of performance in both tone categorization and word identification. We therefore expect an *L1:Musical Experience* interaction. In both groups, working memory is only expected to facilitate word identification performance.

2.5 Results

We first present an overview of performance in the tone categorization and word identification tasks in section 2.5.1, after which we present model results in section 2.5.2 to investigate how our predictors of interest (*L1, Tone, Musical Experience* and *Working Memory*) affected variability in performance.

2.5.1 Overview of performance and individual variability

2.5.1.1 Tone categorization

Figure 10 shows accuracy scores and log-transformed reaction times (RTs) for the tone categorization task. Descriptive statistics are reported in Table 4. A visual inspection reveals no stark difference between the English and Mandarin group, either in terms of accuracy or reaction times. As mentioned earlier, because of a ceiling effect observed for the accuracy scores, we will focus on log RTs in subsequent analyses as a measure of tone categorization performance.

Figure 10

Tone categorization: Accuracy (% correct) and log RT per L1.

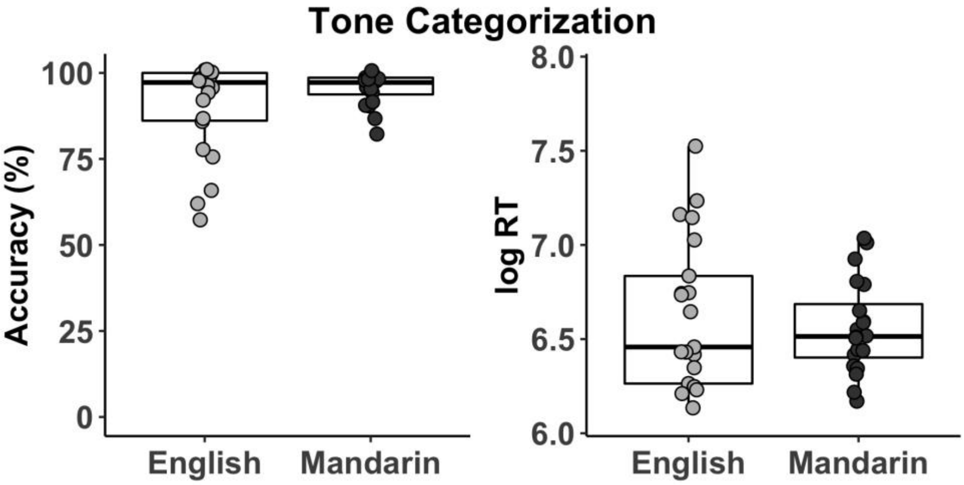


Table 4

Tone categorization: Descriptive statistics.

	English	Mandarin
Accuracy (%)	89.8 (13.9)	95.2 (4.9)
Reaction time (log RT)	6.6 (0.4)	6.5 (0.5)

Values are means with standard deviations in brackets.

2.5.1.2 Word identification

Figure 11 shows accuracy and log RT for the word identification task on days 1 and 2. Descriptive statistics are reported in Table 5. A visual inspection suggests that participants improved their accuracy scores from day 1 to day 2, but that large individual differences exist both in the English and the Mandarin group, with some participants attaining high word identification accuracy and some performing worse. RTs were not the focus of our analysis for the word identification task, but a visual inspection suggests that log RTs did not differ greatly between groups or across days.

Figure 11

Word identification: Accuracy (% correct) and log RT per day and L1.

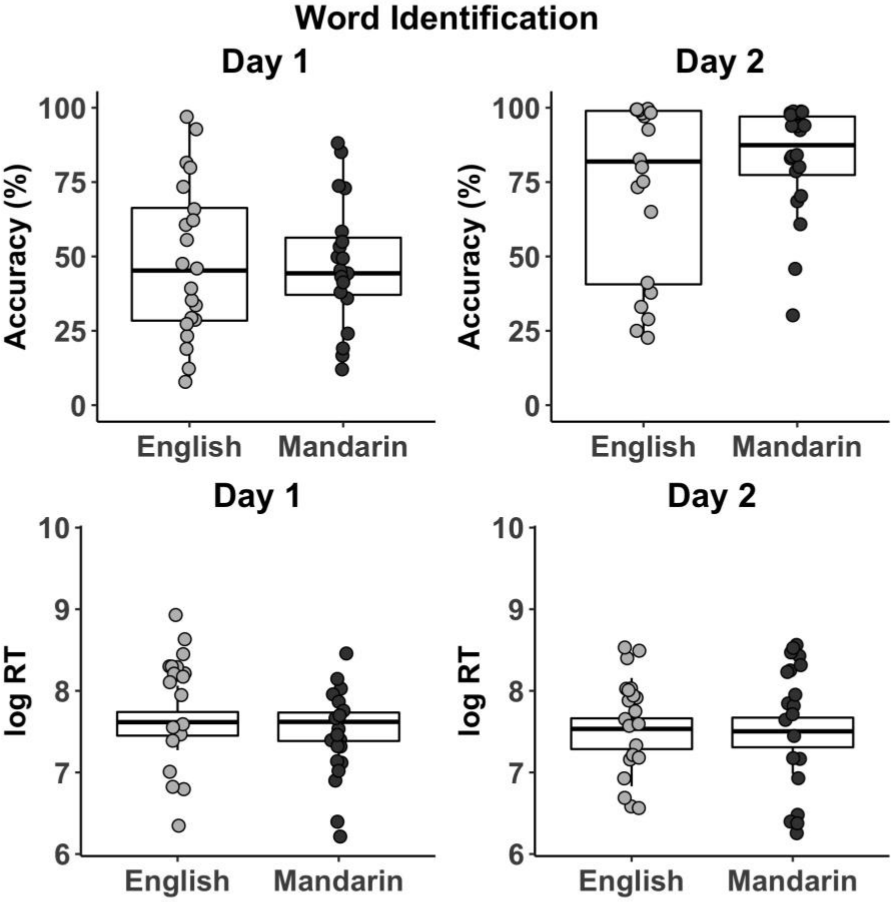


Table 5

Word identification: Descriptive statistics.

	English	Mandarin
Day 1 accuracy (%)	48.4 (26.5)	47.5 (20.9)
Day 1 reaction time (log RT)	7.5 (0.3)	7.5 (0.2)
Day 1 % of tone-only errors*	52.0 (29.8)	60.9 (23.4)
Day 2 accuracy (%)	73.8 (29.5)	82.4 (4.9)
Day 2 reaction time (log RT)	7.5 (0.3)	7.5 (0.2)
Day 2 % of tone-only errors	73.2 (34.9)	65.5 (31.2)

**Tone-only errors are discussed in section 2.5.2.3. Values are means with standard deviation in brackets.*

2.5.2 Model results

To account for the observed individual variability across the tone categorization and word identification tasks, this section highlights significant effects and interactions found in our models. Note that we only present data from the main block on day 2 of the word identification task. This is for brevity but also because we consider data from day 1 to be intermediate, as the word training had not been fully completed then.

A summary of all significant (*p* < 0.05) effects and interactions is provided in Table 6 (full details are in Appendix 2.3–4) .

Following our research questions, we will first address the effects and interactions of *L1* and *Tone* in sections 2.5.2.1–3, after which we will highlight the effects of *Musical Experience* and *Working Memory* in sections 2.5.2.4–5.

Table 6

Summary of significant effects and interactions.

Tone Categorization Task (log RT)*	Word Identification Task (accuracy)**
Musical Experience *L1:Tone* *L1:Musical Experience* *L1:Tone:Musical Experience*	*Tone* *Musical Experience* *Working Memory* *L1:Tone* *L1:Musical Experience* *L1:Working Memory* *Tone:Tone Categorization*

* *lmer(logRT ~ L1*Tone*Musical Experience + L1*Tone*Working Memory + (1|Subject) + (1|Item))*
** *glmer(correct ~ L1*Tone*Musical Experience + L1*Tone*Working Memory + L1*Tone*Tone Categorization + (1|Subject) + (1|Item))*

2.5.2.1 L1:Tone interaction

As shown by the log RTs and accuracy scores in the tone categorization and word identification tasks in Figure 10–11, overall performance between both groups was comparable, and there was no significant effect of *L1* alone in either of the tasks. However, in both tasks, there were significant *L1:Tone* interactions.

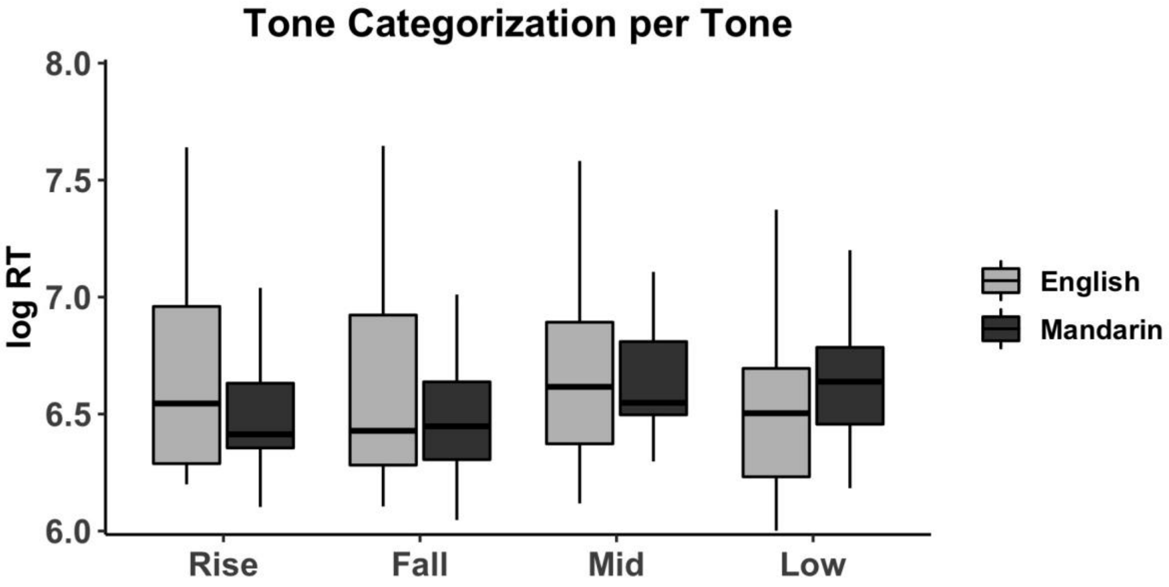
To investigate these interactions in more detail, we first focus on significant multiple comparisons (fully reported in Appendix 2.5–6)[7]. For tone categorization, there were no

[7] Note that in the tables, multiple pairwise comparisons are made with reference to the latter element in a pair, as obtained by the list(pairwise~) command in *emmeans*. For instance, in Appendix 2.5, the "Fall-Mid" comparison with a negative b-estimate of -0.20 indicates that, compared to mid-level tones, falling tones were identified with smaller (faster) reaction times. Changing the reference to falling tones by using the list(revpairwise~) command yields the exact same output, but reverses the sign of the b-estimate and z-score or t-score. For ease of reading, we report the estimate with the sign as relevant to the comparison mentioned in the main text, which may in some cases differ from the sign mentioned in the output table.

significant comparisons between groups, nor between tones within the English group. Within the Mandarin group, mid-level (*b* = 0.20, *SE* = 0.06, *p* = 0.027) and low-level tones (*b* = 0.20, *SE* = 0.06, *p* = 0.047) were categorized significantly slower in comparison to falling tones. A visualization of log RT per tone between groups in Figure 12 shows that indeed, log RTs are similar between groups, and similar between tones within the English group, but that within the Mandarin group, mid-level and low-level tones were categorized more slowly.

Figure 12

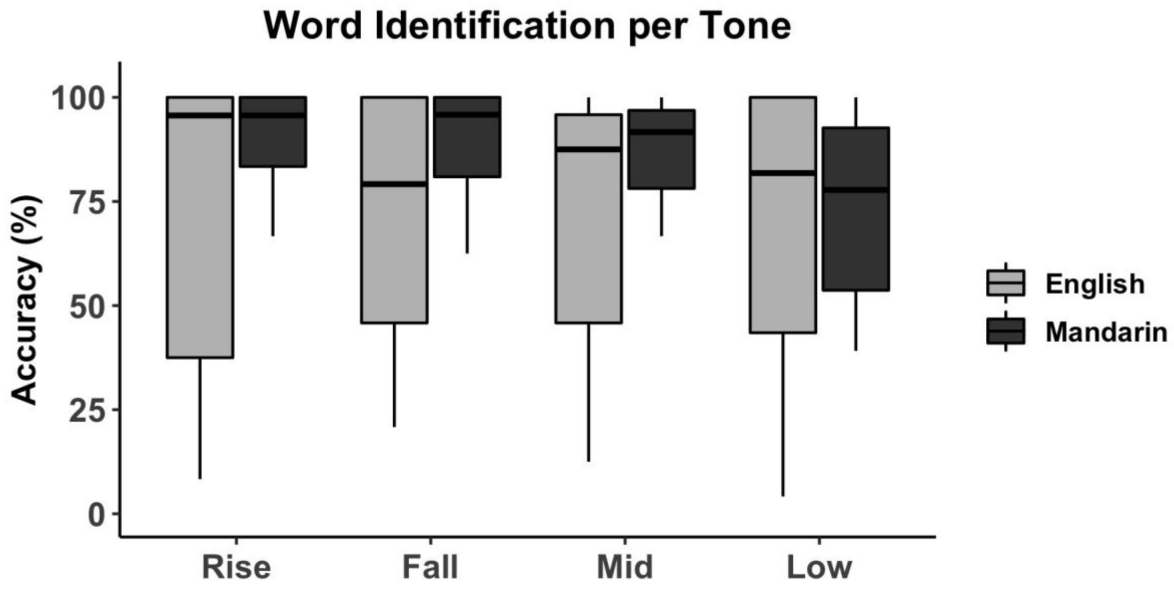
Tone categorization: log RT per tone and L1.



For word identification, multiple comparisons revealed that Mandarin participants were significantly less likely than English participants to identify words carrying a low-level tone (*b* = -1.11, *SE* = 0.47, *p* = 0.018). There were no significant comparisons between tones within the English group. Within the Mandarin group, words carrying a low-level tone were significantly less likely to be identified than words with a rising (*b* = -1.15, *SE* = 0.35, *p* = 0.005) and a falling tone (*b* = -1.23, *SE* = 0.34, *p* = 0.002). A visualization of word identification accuracy per tone between groups in Figure 13 reflects the finding that whereas English participants' word identification accuracy did not vary much between tones, Mandarin participants' accuracy was lower for words carrying a low-level tone.

Figure 13

Word identification (day 2): Accuracy per tone and L1.

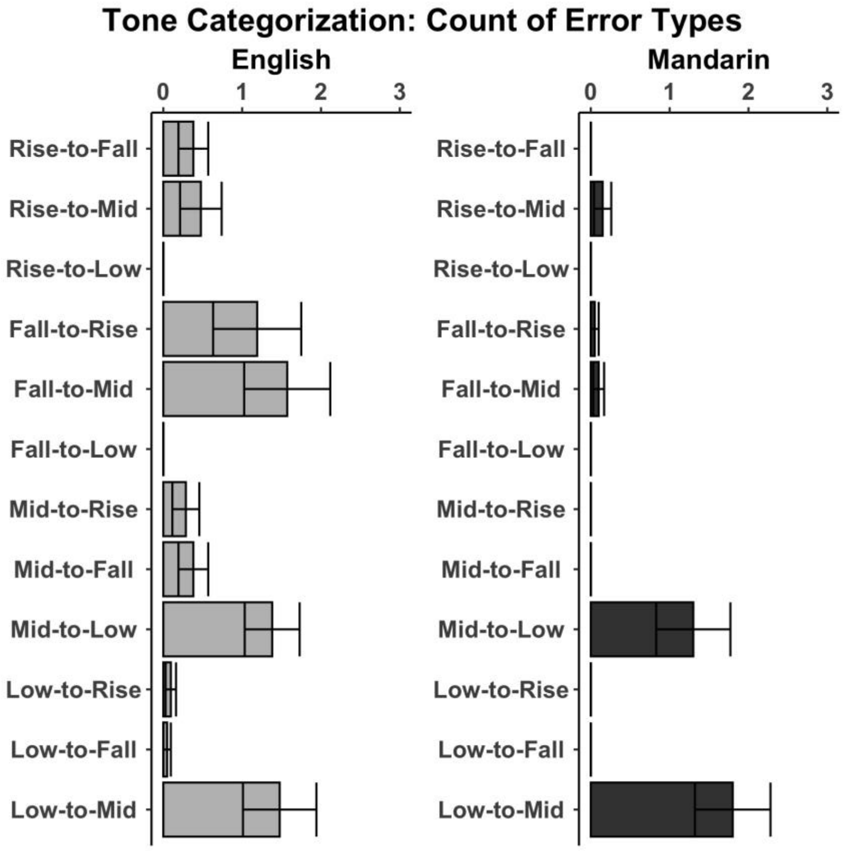


2.5.2.2 Error types in tone categorization

To further investigate how tone type affected tone categorization performance, this section presents error types. Figure 14 displays the count of error types in tone categorization averaged over each participant. For instance, a ‘rise-to-fall’ error indicates that upon hearing a vowel with a rising tone, a participant miscategorized that as a falling tone. A visual inspection of the distribution of all possible 12 error types suggests that English participants miscategorized tones relatively across the board, whereas Mandarin participants predominantly miscategorized mid-level tones as low-level tones and vice versa. Mixed-effects models and multiple comparisons (Appendix 2.7) revealed that in the English group, some error types occurred significantly more often than others: fall-to-mid and low-to-mid errors were likelier to occur in comparison to 5 and 3 other error types, respectively. In the Mandarin group, only the mid-to-low and low-to-mid errors were likelier to occur in comparison to 1 and 3 other error types, respectively.

Figure 14

Tone categorization: Count of error types per L1 and error type.



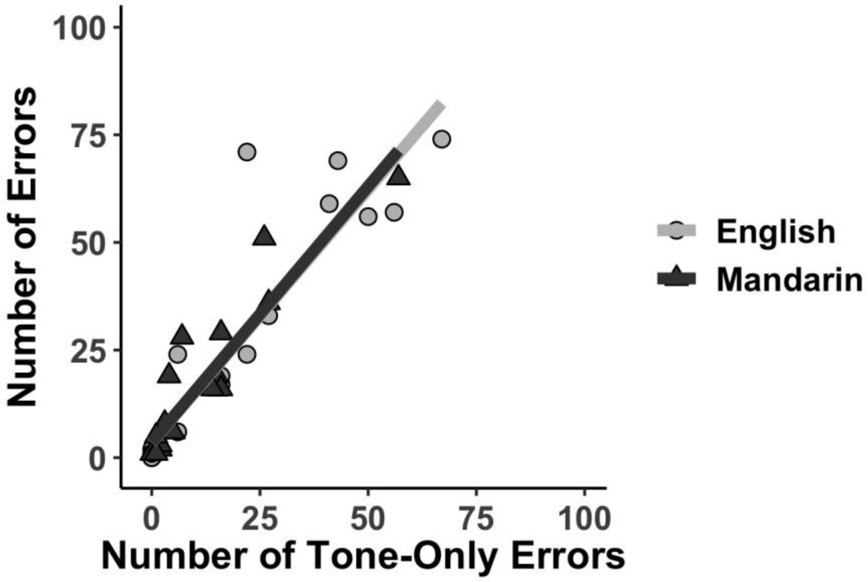
Counts are averaged over subject. Error bars = +/- 1 SE.

2.5.2.3 Tone-only error types in word identification

It is worth noting that on day 2 of the word identification task, the majority of errors were ‘tone-only errors’ (Wong & Perrachione, 2007), meaning that participants misidentified a word purely because of its tone, e.g., misidentifying /juɹ15/ as /juɹ22/. As reported earlier in Table 5, tone-only errors accounted for most errors. For visualization, Figure 15 plots the total number of errors in word identification against the total number of tone-only errors. Two simple linear regressions confirmed that the number of tone-only errors significantly predicted the total number of errors and explained a large portion of variance in both the English [$F(1,19) = 91.670$, $p < 0.001$, $R^2 = .8193$] and the Mandarin group [$F(1,18) = 100.300$, $p < 0.001$, $R^2 = .8393$]. This suggests that many participants had acquired the segmental, but not the tonal properties of the words at the end of the experiment.

Figure 15

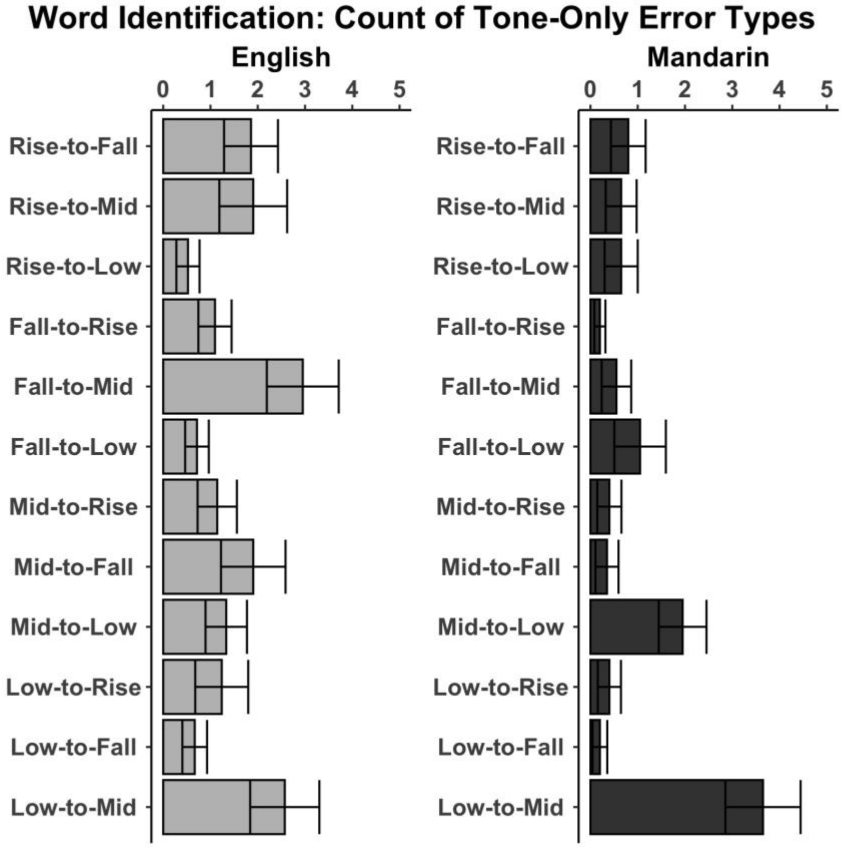
Word identification (day 2): Number of errors against number of tone-only errors.



To further investigate the nature of these tone-only errors, Figure 16 displays the distribution of tone-only error types. Similar to the error types in tone categorization (as presented before in Figure 14), it appears that English participants confused tones in words across the board, with no single error type particularly standing out. Mandarin participants however, seem to have made more low-to-mid errors in comparison to other errors. Mixed-effect models and multiple comparisons (Appendix 2.8) revealed that among the 12 possible error types, there was no indication of one particular error type occurring more often than others in the English group, although it is worth noting that fall-to-mid errors were likelier to occur in comparison to 5 other error types, and that low-to-mid errors were likelier to occur in comparison to 3 other error types. In the Mandarin group, there was a clear indication that the distribution of tone-only errors was skewed toward the low-to-mid type, which was significantly likelier to occur in comparison to almost all other 11 error types, except the mid-to-low error type. The mid-to-low error type was significantly likelier to occur in comparison to 2 other error types.

Figure 16

Word identification (day 2): Count of tone-only errors per L1 and error type.



Counts are averaged over subject. Error bars = +/-1 SE

2.5.2.4 L1:Musical Experience interaction

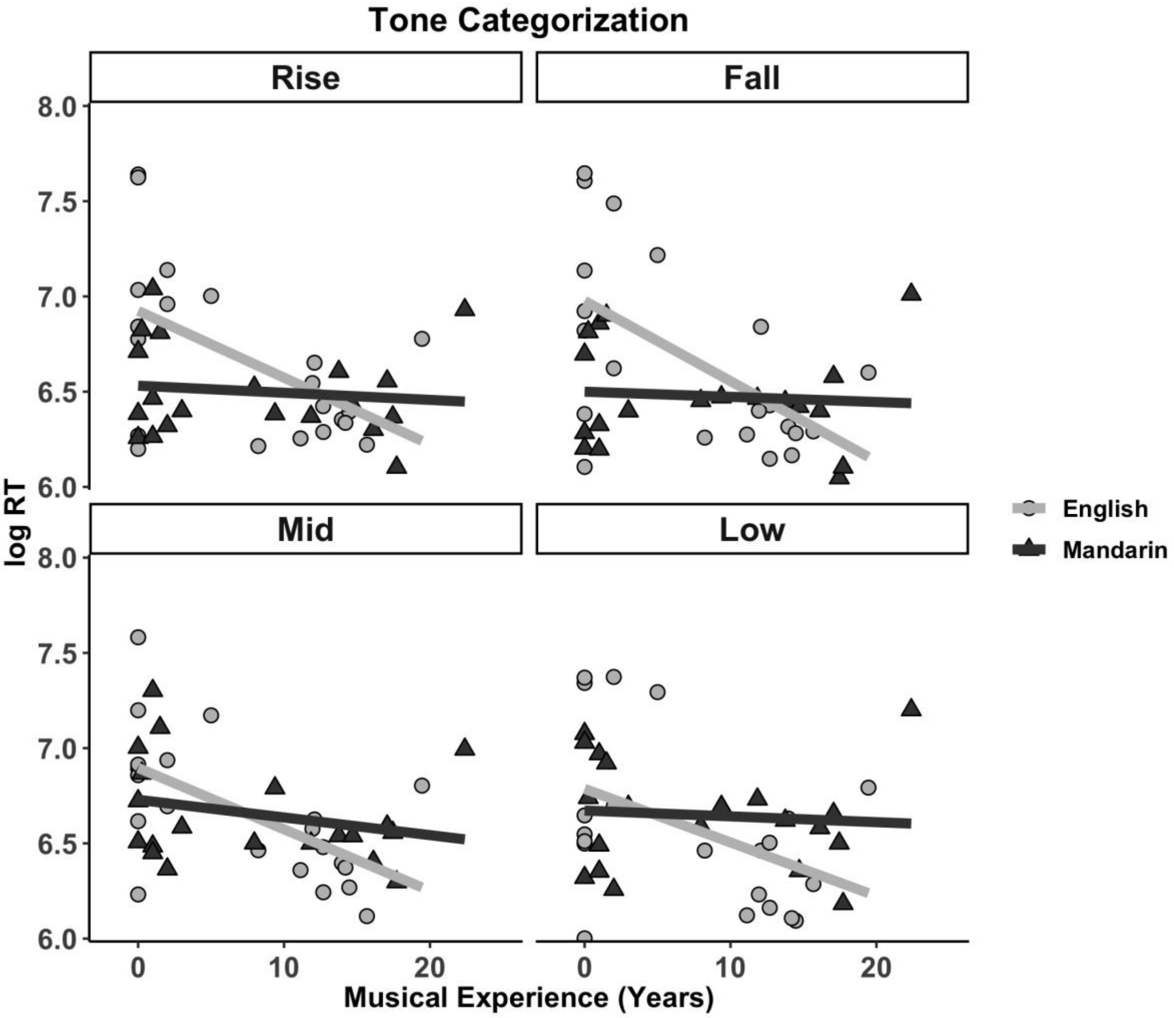
In tone categorization, *Musical Experience* led to faster log RTs in the English group (*b* = -0.28, *SE* = 0.08, *p* = 0.002), but not in the Mandarin group (*b* = -0.05, *SE* = 0.07, *p* = 0.699; full details in Appendix 2.9). Note that these are trends in the overall tone categorization task averaged over the four different tones: there was also a significant three-way *L1:Tone:Musical Experience* interaction, suggesting that the interaction between L1 and musical experience differed between tones.

To investigate the origin of this interaction, the effect of *Musical Experience* was analyzed per group and per tone. Multiple comparisons in Appendix 2.10 revealed that the effect for *Musical Experience* was significantly larger for the English group compared to the Mandarin for rising (*b* = -0.25, *SE* = 0.11, *p* = 0.019) and falling tones (*b* = -0.31, *SE* = 0.11, *p* = 0.005), but not for mid-level (*b* = -0.17, *SE* = 0.11, *p* = 0.106) and low-level (*b* = -0.19, *SE* = 0.11, *p* = 0.076) tones. A further post-hoc comparison revealed that the effect of *Musical Experience* was significantly larger for falling tones than for low-level tones within the English group (*b* = -0.11, *SE* = 0.03, *p* = 0.036).

This is illustrated in Figure 17, which plots tone categorization log RT against musical experience per tone. For the English group, it can be observed that the effect of *Musical Experience* is relatively strong (i.e., relatively steeper slopes) for rising and falling tones, and slightly less so for mid-level and low-level tones. For the Mandarin group, the flat slopes indicate that musical experience overall did not lead to faster log RTs, in none of the tones.

Figure 17

Tone categorization: log RT against musical experience per tone.

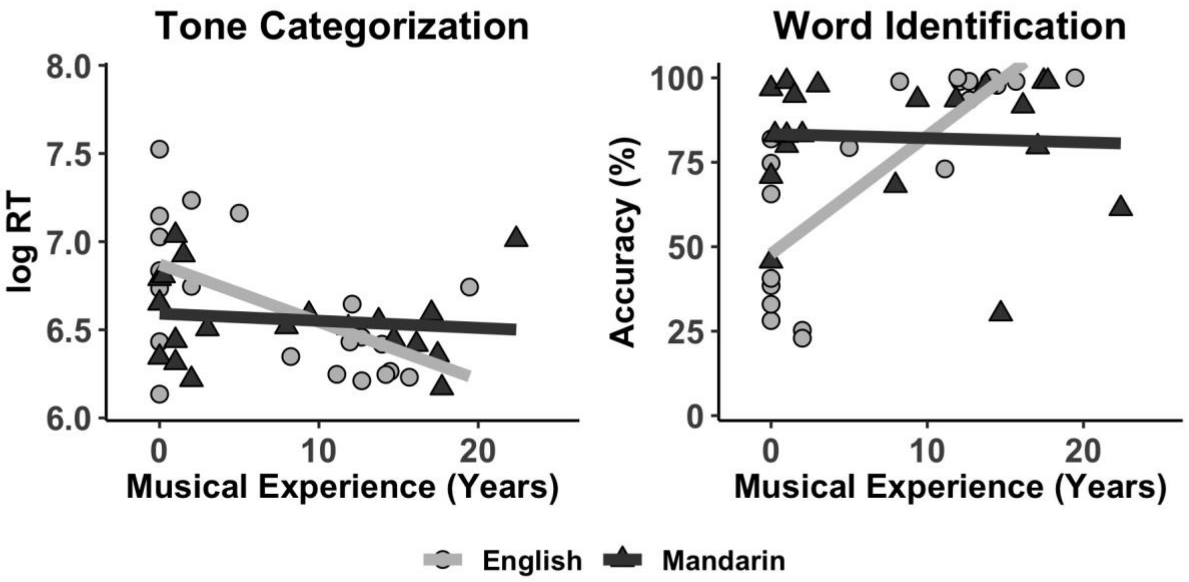


In the word identification task, *Musical Experience* significantly increased the likelihood of correct word identification in the English group (*b* = 2.21, *SE* = 0.45, *p* < 0.001), but not in the Mandarin group (*b* = 0.48, *SE* = 0.29, *p* = 0.183; full details in Appendix 2.11).

For visualization, Figure 18 illustrates the *L1:Musical Experience* interactions in tone categorization and word identification. It can be observed that whereas English participants appear to benefit from musical experience (resulting in faster RTs in tone categorization and higher accuracies in word identification), this trend is absent in the Mandarin participants.

Figure 18

Tone categorization and word identification: log RT and accuracy against musical experience.



2.5.2.5 L1:Working Memory interaction

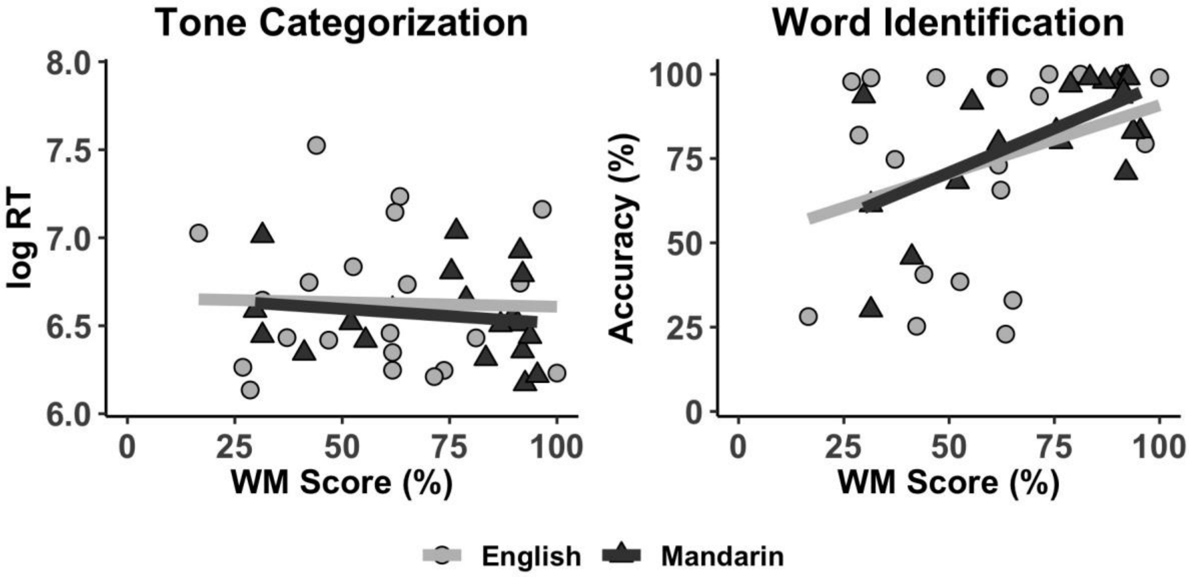
Working Memory did not predict performance in the tone categorization task for either group.

In the word identification task, *Working Memory* did not significantly increase the likelihood of correct word identification in the English group, but it did in the Mandarin group (*b* = 1.91, *SE* = 0.31, *p* < 0.001; full details in Appendix 2.11). This finding is illustrated in Figure 19. Note that although the trend line would suggest otherwise, there was no statistical confirmation that WM, alongside with our other predictors of interest, predicted

English participants' performance in the word identification task (*b* = 0.06, *SE* = 0.35, *p* = 0.982, 95% CI [-0.63 ; 0.75]).

Figure 19

Tone categorization and word identification: log RT and accuracy against WM.



2.5.2.6 Tone categorization reaction time as a predictor of word identification performance

Tone categorization log RTs did not predict word identification performance in either group in our model, however there was a significant *Tone:Tone Categorization* interaction. Post-hoc multiple comparisons revealed that for both groups together, the effect of *Tone Categorization* was largest for words with rising tones, however this effect on its own failed to reach significance (*b* = -0.63, *SE* = 0.27, *p* = 0.077; 95% CI [-1.16, -0.10]).

2.6 Discussion

This study's aim was to examine the combined effects of individual learners' L1-specific and extralinguistic factors as predictors of L2 tone perception and word learning facility. We will now discuss our findings in light of our research questions and previous research.

2.6.1 Effects of L1 tonal status and L1 tone types on tone categorization and word identification

RQ1 addressed how L1 tonal status and L1 tone type affect individual performance in both pre-lexical and lexical processing of tones. In the tone categorization task, which addressed pre-lexical tone perception, most participants attained near-ceiling performance in terms of accuracy, but they showed more individual variability in reaction times. This variability was not directly attributable to L1 tonal status, as Mandarin listeners were not significantly faster than English listeners in categorizing tones. Instead, as predicted, variability was explained by an interaction between L1 tonal status and L1 tone types.

Specifically, within the Mandarin group, mid-level and low-level tones yielded slower RTs than falling tones, and the error analysis further revealed that Mandarin participants predominantly miscategorized low-level as mid-level tones and vice versa. This suggests that telling apart low-level from mid-level tones constituted the real difficulty for the Mandarin participants in the tone categorization task. This finding is interpretable when considering Mandarin L1 tone types: in phonological-categorical terms, Mandarin listeners may have assimilated our low-level and mid-level tones to their L1 high-level tone, making the level distinction difficult. As pointed out by Francis et al. (2008, p. 284), any claims regarding categorical assimilation can only be "speculative in nature". This is especially the case in our study since we did not ask our participants to explicitly rate the similarity between target and L1 tones (J. Chen et al., 2020; Reid et al., 2015). Nevertheless, it is worth noting that, although purely anecdotal, many Mandarin participants did indicate that the mid-level and low-level tones were particularly difficult to categorize because they had no clear equivalents in Mandarin, unlike the rising and falling tones.

Alternatively, an acoustic-phonetic interpretation as to why Mandarin participants appeared to struggle with quickly categorizing level tone contrasts would be that they put relatively more weight on differences in F0 direction rather than in F0 height (Francis et al., 2008; Gandour & Harshman, 1978; Qin & Jongman, 2016). It is additionally possible that the categorization of low-level tones was complicated because of absence of phonation cues (creaky voice), which contributes to native speakers' perception of the low-dipping tone in Mandarin (R. Yang, 2015). Indeed, in real tone languages, acoustic cues such as phonation (Tsukada & Kondo, 2019) and duration (Liu & Samuel, 2004) can contribute to the overall salience of different tone types.

As to the English speakers, log RTs did not significantly differ across tones. The error analysis further revealed that English participants tended to confuse tone types with one another in every direction, incorrectly categorizing both contour as level tones (Fall-to-Mid) and level as level tones (Low-to-Mid) relatively often. Although again we cannot ascertain whether English listeners relied on L1 F0-based categories in their tone categorization, whatever reliance on intonational categories English participants may have had (for instance, assimilating rising and falling tones to LH and HL intonational types), it appears that these did not affect performance, as performance on individual tones was equal across the board. This resonates with Best's (2019) conclusion that assimilations of L2 tones to intonational distinctions may be "less categorical than are assimilations to another lexical tone system" (Best, 2019, p. 5). Although we had tentatively predicted that English speakers would categorize level tones faster than contour tones based on a phonetic-acoustic approach of tone type, this was not borne out by our data. Rather than being affected by tone type, English participants' performance appeared to be largely guided by their musical experience, as will be discussed in the next section.

Our findings from the word identification task suggest that L1 tonal status and L1 tone type modulated performance in a similar way as in the tone categorization task: differences between the English and Mandarin groups were not seen in overall performance (against our predictions), but in performance per tone. The error analysis showed that in both groups, most word identification errors were tone-only errors, suggesting that tonal rather than segmental distinctions were the hardest feature to memorize the pseudowords. However, *which* tonal distinctions were hardest to learn appeared to be strongly influenced by L1 tone type, as Mandarin participants were less likely to identify words with low-level tones compared to words with rising and falling tones, and even compared to English participants. Mandarin speakers predominantly misidentified low-level tone words as mid-level tone words, whereas the English participants had no clear confusion pattern that stood out and confused tones on words across the board.

In sum, our findings addressing **RQ1** show that L1 tone type not only interferes in pre-lexical tone processing, as has been shown widely in previous studies (Cooper & Wang, 2012; Hao, 2012; Qin & Jongman, 2016; So & Best, 2010; X. Wu et al., 2014), but also in lexical processing, and in remarkably similar ways. It is crucial to note that in our study, this effect appeared to be strong enough that Mandarin participants, who by virtue of their L1 tonal status would be expected to outperform non-tonal peers in L2 tone word learning (R. K. W. Chan & Leung, 2020; Poltrock et al., 2018), performed worse in recalling low-level tone words than non-tonal English participants. This highlights that L1 tonal status alone cannot fully account for individual differences in neither tone perception nor tone word learning facility, and that it is crucial to simultaneously factor in effects of L1 tone type. It is worth noting that if our pseudowords had contained the exact same tone types as in Mandarin, we would have expected Mandarin participants to outperform the English speakers, thereby indirectly showing an overall facilitative effect of L1 tonal status.

2.6.2 Combined effects of L1-specific and extralinguistic factors

In **RQ2,** we asked how musical experience and working memory affect individual performance in tone perception and tone word learning, and whether the effects of these extralinguistic factors are modulated by L1-specific factors.

We found that, in line with our predictions, musical experience significantly predicted tone categorization performance for English but not for Mandarin participants. Even for mid-level and low-level tones, which were relatively difficult for Mandarin participants, musical experience did not lead to faster RTs. The absence of a facilitative effect of musical experience on tone perception for Mandarin speakers in our study chimes in with earlier findings (W. Tang et al., 2016; Wong et al., 2020; H. Wu et al., 2015), although it is worth noting that finding such a facilitative effect may be task-dependent (D. Chang et al., 2016). For instance, Qin et al. (2021) tentatively suggest that musical *ability* (a different measure of musicianship) may in fact enhance perception (as measured by discrimination and identification accuracy) of Cantonese level tone contrasts for Mandarin-L1 speakers. We interpret however that in our tone categorization task, Mandarin participants' performance was largely guided by the effect of L1 tone type, and that this may have overridden any facilitative effect of musical experience on tone perception.

English participants did appear to benefit from musical experience, as musical experience led to significantly faster reaction times. In addition, the *L1:Tone:Musical Experience* interaction revealed that musical experience particularly facilitated categorization of falling tones as opposed to low-level tones. This suggests that English listeners, who have been found to pay less attention to F0 contour differences than to F0 height differences, particularly in falling contours (Jongman et al., 2017), may have benefited from additional pitch acuity derived from musical experience to quickly categorize 'difficult' falling tones.

In the word identification task, we similarly found that musical experience predicted performance for English but not for Mandarin participants. Our interpretation is similar to that reported in earlier work that showed a "differential in relevance of musicality depending on linguistic background" in tone word learning (Cooper & Wang, 2012, pp. 4765–4766). Namely, Mandarin participants, who are already familiar with the use of pitch for lexical purposes, may not benefit as much from enhanced pitch acuity gained through musical experience as English participants do.

In sum, these findings suggest a dynamic interplay of musical experience and L1 tonal status in L2 tone perception and word learning. We note that we only measured musicianship in terms of years of musical practice, and that more refined measures of *musicality* (Wallentin et al., 2010) might reveal different results.

As predicted, we did not find a significant facilitative effect of working memory on pre-lexical pitch processing in the tone categorization task for neither English nor Mandarin participants. Although this finding falls in line with existing literature that suggests that WM has a null, or limited effect on performance in relatively undemanding pre-lexical pitch perception tasks (Bidelman et al., 2013, p. 8; Goss, 2020; Goss & Tamaoka, 2019), we are aware that we only measured backwards digit span as a rough proxy of WM, and future studies could assess whether other cognitive measures, such as attentional resources or executive function, are linked to tone perception.

To the best of our knowledge, our study is the first of its kind that incorporates a measure of WM in assessing the combined effects of L1 tonal status, L1 tone type, and musical experience in tone word learning. We found that when considering all these factors together, WM significantly predicted word recall of tonal pseudowords for Mandarin but, unexpectedly, not for English participants, for whom musical experience was the only significant extralinguistic predictor. The finding for English participants resembles that of Bowles et al. (2016), who found that variance in English learners' performance in Mandarin tone word learning was only partially explained by domain-general memory skills, and most strongly by pitch-specific skills, suggesting that "mastery of a feature of a target language known to be particularly challenging for L2 learners – as a necessary component of learning the language at large – is predicted most successfully by behavioral measures that are most relevant to that feature" (Bowles et al., 2016, p. 775). In other words, our word identification task may have been particularly challenging for English participants because it involved *tone* words, and therefore individual participants with better pitch acuity (assumed to be derived from musical experience) would benefit from these skills to memorize words based on tonal distinctions. Mandarin participants, by virtue of their L1 tonal status, may not have found recalling our pseudowords particularly challenging because they contrasted in tone *per se* (except for the distinction between level tone words). This could explain why their ability to recall our pseudowords was mainly guided by WM capacity, rather than pitch-specific skills, as a general predictor of L2 vocabulary recall (Cheung, 1996; Kormos & Sáfár, 2008).

Finally, our models revealed that, when also accounting for other L1-specific and extralinguistic factors, performance in the tone categorization task (as measured by log RTs) did not independently predict performance in word identification. However, this does not imply that performance in the pre-lexical tone categorization task was completely unrelated to performance in the lexical word identification task. For instance, the tone error patterns largely mirrored one another across both tasks. Additionally, it is worth noting that in an alternative model of word identification in which we used tone categorization accuracy (i.e., *pitch aptitude*) instead of log RT as a proxy of pre-lexical tone perception ability, we did find a (marginally) significant main effect of pitch perception aptitude on word identification likelihood for both groups (b = 0.65, SE = 0.32, p = 0.043). Although we are cautious to derive strong conclusions from this alternative analysis given the near-ceiling accuracy scores in the tone categorization task, this may suggest a link between performance in pitch perception and lexical pitch processing in our tasks. Our general findings, in which we used log RTs as a proxy of pre-lexical processing, reveal that tone word learning performance in English participants was mainly facilitated by musical experience, and in Mandarin participants mainly by WM capacity, which may fill the gap when neither musical experience nor tone perception ability strongly facilitate tone word recall.

Thus, addressing **RQ2**, it appears that any facilitative effect of musical experience and working memory on pre-lexical and lexical tone processing is indeed modulated by L1 tonal status: for non-tonal English learners, musical experience appears to be facilitative for tone perception and word learning, whereas for tonal Mandarin learners, individual performance is guided by L1 tone type and working memory (the latter only for word identification). The findings from our study thus suggest that the ease with which L2 tones are perceived and learned depends on a dynamic interplay between L1 tonal status, L1 tone type, musical experience, and working memory. This provides a more refined account of the several factors that determine an individual learner’s aptitude to explain the large variability observed in L2 tone perception and word learning facility, beyond what has been described in previous studies that separately assessed the factors included in this study.

Future studies should examine the combined effects of L1-specific and extralinguistic factors in tone word learning in more naturalistic settings than our pseudoword identification task, for instance in tasks in which learners process tones in sentence contexts or in multi-speaker environments. As pointed out by a reviewer, the fact that we only modified F0 in our stimuli and kept other acoustic parameters constant may limit the applicability of our findings to real tone languages in which additional acoustic cues may modulate tone processing. Future studies should thus include a wider range of native and non-native tone systems to further refine our understanding of a dynamic interplay between L1-specific and extralinguistic factors in L2 tone learning.

2.7 Conclusion

This study aimed to account for individual differences in L2 tone perception and tone word learning by assessing the combined effects of L1-specific and extralinguistic factors, testing a combination of factors that were only addressed separately in earlier studies. We argue that none of the L1-specific and extralinguistic factors determine learning facility in and of themselves, but that both go hand-in-hand and dynamically affect tone perception and tone word learning performance in the individual and thereby shape the profile of learners who are expected to do relatively well, and learners who are expected to do relatively poorly in early-stage tone learning. Our findings suggest that a complete theoretical model of tone learning would ideally acknowledge this "dynamic" and "multisystemic" nature of L2 speech-learning (A. Li & Post, 2014). Although beyond the scope of this Chapter, a possible future theoretical description of tone learning could be framed within an 'L1-Modulated Domain-General Account' (this is elaborated on in Chapter 6). That is, our study shows that a comprehensive theory of L2 tone learning facility should not only be able to account for extralinguistic factors that shape individual performance in early-stage tone learning – such musical experience and working memory – but it should also be able to account for any L1-specific factors – L1 tonal status and L1 tone type, here – which interact with extralinguistic factors to modulate individual performance in complex ways.

2.8 Appendix to Chapter 2

2.8.1 Appendix tables

Appendix 2.1

Detailed participant demographics (English group).

ID	Age	L2s and self-reported level (0-10)	Currently Practicing	ME	WM	PP
EN-MU-F-1	21	German 3	Keyboard/Piano; Woodwind; Singing	14	47	100
EN-MU-F-2	19	Spanish 7, Portuguese 6	Drums; Keyboard/Piano; Singing	14	27	100
EN-MU-F-3	20	-	Keyboard/Piano; Strings; Woodwind	14	74	100
EN-MU-F-4	19	French 5, Spanish 4	Woodwind; Choral Singing	11	62	99
EN-MU-F-5	20	-	Guitar; Choral Singing; Singing	12	31	100
EN-MU-F-6	20	Gujarati* 5, Spanish 4, French 1	Keyboard/Piano; Strings; Choral Singing	13	71	100
EN-MU-M-1	20	-	Strings; Singing	13	61	99
EN-MU-M-2	20	-	Keyboard/Piano; Strings; Brass; Choral Singing	16	100	100
EN-MU-M-3	20	-	Keyboard/Piano; Woodwind; Choral Singing	12	81	96
EN-MU-M-4	25	Russian* 7, French 7, German 7, Spanish 7	Keyboard/Piano	19	91	97
EN-MU-M-5	22	Italian* 7, Spanish 4, French 1	Guitar; Keyboard/Piano; Singing	8	62	100
EN-NM-F-1	19	French 7, Spanish 2	-	-	17	76
EN-NM-F-2	22	French 5, Hindi 3	-	2	42	86
EN-NM-F-3	23	Spanish 4	-	-	53	65
EN-NM-F-4	21	-	-	-	65	94
EN-NM-F-5	21	Spanish 3	-	-	29	99
EN-NM-M-1	21	German 7	-	5	97	86
EN-NM-M-2	20	Hindi* 7	-	-	44	56
EN-NM-M-3	23	German 6	-	2	63	63
EN-NM-M-4	22	-	-	-	37	92
EN-NM-M-5	22	-	-	-	62	78

ME: Musical Experience (Years) WM: Working Memory Score (0-100); PP: Pitch Perception Aptitude Score (0-100)
** exposure to a (heritage) language before the age of 12 at home or in other surroundings.*

Appendix 2.2

Detailed participant demographics (Mandarin group).

ID	Age	L2s and self-reported level (0-10)	Currently Practicing	ME	WM	PP
MA-MU-F-1	20	English 8, Wu* 7, Japanese 6	Keyboard/Piano; Woodwind; Choral Singing	16	55	99
MA-MU-F-2	20	English 8, Italian 1, French 1	Strings	17	92	90
MA-MU-F-3	19	English 6	Guitar; Keyboard/Piano; Singing; Guzheng	17	62	97
MA-MU-F-4	29	English 8, Cantonese* 7, French 2	Guzheng	22	31	82
MA-MU-F-5	24	English 8, Cantonese* 7, French 1	Keyboard/Piano	14	90	99
MA-MU-M-1	24	English 10, French 1	Erhu	18	93	99
MA-MU-M-2	23	English 8, Cantonese* 7	Guitar	8	52	96
MA-MU-M-3	28	English 8, Wu* 7, German 5, Cantonese 2	Keyboard/Piano; Strings; Choral Singing	15	31	94
MA-MU-M-4	19	English 8	Strings; Singing	9	30	100
MA-MU-M-5	19	English 8, French 1	Guitar; Keyboard/Piano	12	91	97
MA-NM-F-1	29	Italian 10, English 8, Wu* 7, French 7, Japanese 2, Persian 2	-	3	87	99
MA-NM-F-2	23	English 8	-	2	95	90
MA-NM-F-3	25	English 8	-	-	41	97
MA-NM-F-4	22	English 8	-	-	92	97
MA-NM-F-5	24	English 8, Japanese 7	-	-	79	99
MA-NM-M-1	18	English 7	-	1	77	94
MA-NM-M-2	19	English 8	-	1	94	99
MA-NM-M-3	20	English 8	-	1	83	100
MA-NM-M-4	23	English 8	-	2	91	92
MA-NM-M-5	23	Kunming Chinese* 7, English 7	-	-	75	86

ME: Musical Experience (Years) WM: Working Memory Score (0-100); PP: Pitch Perception Aptitude Score (0-100)
** exposure to a (heritage) language before the age of 12 at home or in other surroundings.*

Appendix 2.3

Tone categorization: Mixed model ANOVA table for log RT results (Type III Wald Chisquare tests).

TONE CATEGORIZATION			
lmer(logRT ~ L1*Tone*Musical Experience + L1*Tone*Working Memory + (1|Subject) + (1|Item))			
Effect	χ^2	df	p
L1	0.51	1	0.087
Tone	4.35	3	0.226
Musical Experience	11.74	1	0.000
Working Memory	0.26	1	0.606
L1:Tone	61.02	3	0.000
L1:Musical Experience	5.97	1	0.014
L1:Working Memory	2.75	1	0.097
Tone:Musical Experience	6.52	3	0.089
Tone:Working Memory	3.41	3	0.332
L1:Tone:Musical Experience	12.11	3	0.007
L1:Tone:Working Memory	4.88	3	0.180

Appendix 2.4

Word identification: Mixed model ANOVA table for accuracy results (Type III Wald Chisquare tests).

WORD IDENTIFICATION (DAY 2)			
glmer(correct ~ L1*Tone*Musical Experience + L1*Tone*Working Memory + L1*Tone*Tone Categorization + (1|Subject) + (1|Item))			
Effect	χ^2	df	p
L1	1.59	1	0.207
Tone	8.87	3	0.031
Musical Experience	25.18	1	0.000
Working Memory	7.01	1	0.008
Tone Categorization	1.64	1	0.200
L1:Tone	11.10	3	0.011
L1:Musical Experience	10.49	1	0.001
L1:Working Memory	5.76	1	0.016
L1:Tone Categorization	0.01	1	0.896
Tone:Musical Experience	2.01	3	0.570
Tone:Working Memory	4.44	3	0.217
Tone:Tone Categorization	11.20	3	0.011
L1:Tone:Musical Experience	2.01	3	0.570
L1:Tone:Working Memory	6.30	3	0.097
L1:Tone:Tone Categorization	1.41	3	0.701

Appendix 2.5

Tone categorization: Significant multiple comparisons for tone.

TONE CATEGORIZATION				
Contrast	Estimate	std. Error	t	p
English				
(No sig. comparisons)	-	-	-	-
Mandarin				
Fall-Mid	-0.20	0.06	-3.32	0.027
Fall-Low	-0.18	0.06	-2.92	0.047

Appendix 2.6

Word identification: Significant multiple comparisons for tone.

WORD IDENTIFICATION (DAY 2)				
Contrast	Estimate	std. Error	t	p
Eng-Man | Low	1.11	0.47	2.36	0.018
English				
(No sig. comparisons)	-	-	-	-
Mandarin				
Rise-Low	1.15	0.35	3.31	0.005
Fall-Low	1.23	0.34	3.59	0.002

Appendix 2.7

Tone categorization: Significant multiple comparisons for count of error types.

TONE CATEGORIZATION					
Contrast		Estimate	std. Error	t-value	p
English					
Rise-to-Fall	Fall-to-Mid	-1.42	0.39	-3.60	0.026
Fall-to-Mid	Mid-to-Rise	1.70	0.44	3.84	0.010
	Mid-to-Fall	1.42	0.39	3.60	0.026
	Low-to-Rise	2.80	0.73	3.85	0.010
	Low-to-Fall	3.50	1.02	3.45	0.045
Mid-to-Rise	Mid-to-Low	-1.58	0.45	-3.51	0.035
	Low-to-Mid	-1.64	0.45	-3.68	0.019
Mid-to-Fall	Low-to-Mid	-1.35	0.40	-3.42	0.049
Mid-to-Low	Low-to-Rise	2.67	0.73	3.66	0.021
Low-to-Rise	Low-to-Mid	-2.74	0.73	-3.76	0.014
Mandarin					
Rise-to-Mid	Low-to-Mid	-2.44	0.60	-3.79	0.013
Fall-to-Rise	Low-to-Mid	-3.58	1.00	-3.47	0.042
Fall-to-Mid	Mid-to-Low	-2.59	0.80	-3.44	0.046
	Low-to-Mid	-2.90	0.70	-3.87	0.009

The counts of error types were subjected to a zero-inflated general linear mixed effect model (Brooks et al., 2017)*, with* Confusion Type *(12 levels: Rise-to-Fall, Rise-to-Mid, etc.) as fixed factor, and* Subject *as a random intercept. Because not all models would converge on the full data sets, the models were fitted on data subsets per group.* glmmTMB(value ~ ErrorType + (1|subject), ziformula=~1, family=poisson)

Appendix 2.8

Word identification: Significant multiple comparisons for count of tone-only error types.

WORD IDENTIFICATION (DAY 2)					
Contrast		Estimate	std. Error	t-value	p
English					
Rise-to-Mid	Rise-to-Low	1.34	0.36	3.76	0.014
	Fall-to-Low	1.09	0.31	3.51	0.035
Rise-to-Low	Fall-to-Mid	-1.72	0.34	-5.04	0.001
	Mid-to-Fall	-1.26	0.35	-3.56	0.029
	Low-to-Mid	-1.51	0.34	-4.40	0.001
Fall-to-Rise	Fall-to-Mid	-1.03	0.25	-4.15	0.003
Fall-to-Mid	Fall-to-Low	1.47	0.29	5.06	0.000
	Mid-to-Rise	1.01	0.24	4.13	0.003
	Low-to-Fall	1.45	0.31	4.64	0.000
Fall-to-Low	Low-to-Mid	-1.26	0.29	-4.31	0.002
Low-to-Fall	Low-to-Mid	-1.25	0.32	-3.94	0.007
Mandarin					
Rise-to-Fall	Low-to-Mid	-1.32	0.32	-4.17	0.003
Rise-to-Mid	Low-to-Mid	-1.38	0.38	-3.63	0.023
Rise-to-Low	Low-to-Mid	-1.54	0.34	-4.51	0.001
Fall-to-Rise	Mid-to-Low	-2.11	0.55	-3.86	0.009
	Low-to-Mid	-2.76	0.54	-5.13	0.000
Fall-to-Mid	Low-to-Mid	-1.69	0.37	-4.63	0.000
Fall-to-Low	Low-to-Mid	-1.08	0.28	-3.86	0.009
Mid-to-Rise	Low-to-Mid	-2.03	0.41	-4.93	0.001
Mid-to-Fall	Low-to-Mid	-2.10	0.45	-4.69	0.000
Mid-to-Low	Low-to-Fall	2.06	0.56	3.70	0.018
Low-to-Rise	Low-to-Mid	-2.04	0.41	-5.03	0.000
Low-to-Fall	Low-to-Mid	-2.71	0.55	-4.95	0.000

Appendix 2.9

Tone categorization: Multiple comparisons and estimates per L1 for extralinguistic factors.

TONE CATEGORIZATION					
Predictors	Estimate	std. Error	t	p	95% C.I.
Multiple Comparisons (Bonferroni-corrected)					
Eng-Man | Musical Experience	-0.23	0.10	-2.26	0.028	-
Eng-Man | Working Memory	0.16	0.11	1.53	0.131	-
English					
Musical Experience	-0.28	0.08	-3.58	0.002	[-0.43 ; -0.12]
Working Memory	0.11	0.08	1.40	0.307	[-0.05 ; 0.26]
Mandarin					
Musical Experience	-0.05	0.07	-0.70	0.979	[-0.18 ; 0.09]
Working Memory	-0.06	0.08	-0.76	0.699	[-0.21 ; 0.09]

Appendix 2.10

Tone categorization: Estimates of musical experience per L1 per tone.

TONE CATEGORIZATION THREE WAY INTERACTION					
Predictors	Estimate	std. Error	t	p	95% C.I.
Multiple Comparisons for effect of Musical Experience per Tone (Bonferroni-Corrected)					
Eng-Man (Rise)	-0.25	0.11	-2.41	0.019	-
Eng-Man (Fall)	-0.31	0.11	-2.91	0.005	-
Eng-Man (Mid)	-0.17	0.11	-1.64	0.106	-
Eng-Man (Low)	-0.19	0.10	-1.80	0.076	-
English					
Rise	-0.29	0.08	-3.60	0.001	[-0.45 ; -0.13]
Fall	-0.33	0.08	-4.14	0.000	[-0.50 ; -0.17]
Mid	-0.26	0.08	-3.29	0.002	[-0.42 ; -0.10]
Low	-0.23	0.08	-2.89	0.007	[-0.39 ; -0.07]
Mandarin					
Rise	-0.03	0.07	-0.50	0.616	[-0.17 ; 0.10]
Fall	-0.03	0.07	-0.37	0.711	[-0.16 ; 0.11]
Mid	-0.09	0.07	-1.31	0.196	[-0.23 ; 0.05]
Low	-0.04	0.07	-0.52	0.601	[-0.17 ; 0.10]

Appendix 2.11

Word identification: Multiple comparisons and estimates per L1 for extralinguistic factors.

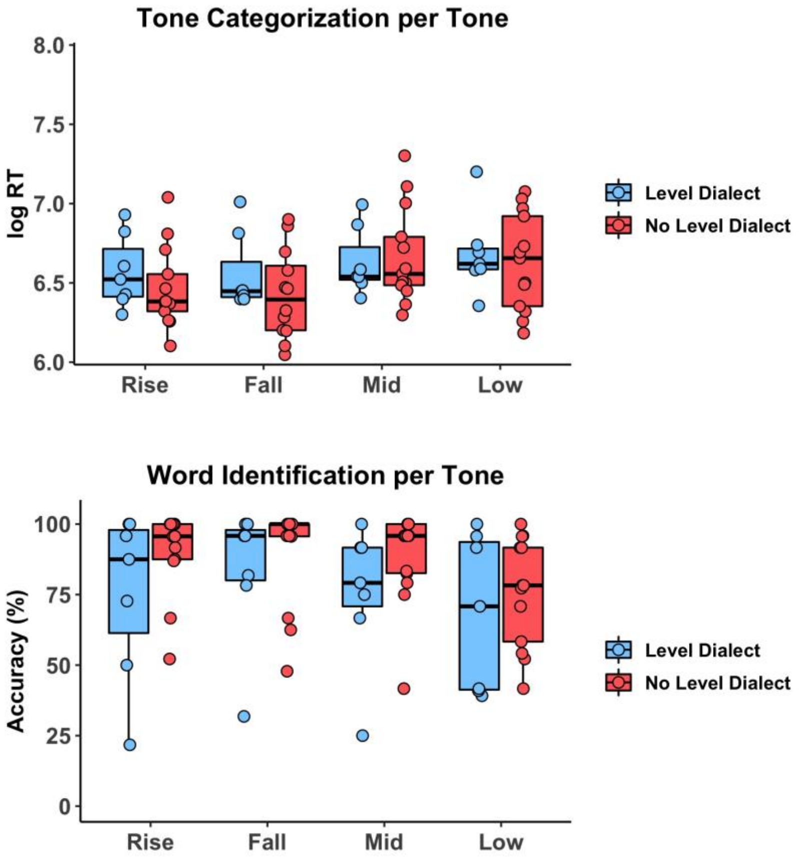
WORD IDENTIFICATION (DAY 2)					
Predictors	Estimate	std. Error	z	p	95% C.I.
Multiple Comparisons (Bonferroni-corrected)					
Eng-Man | Musical Experience	1.73	0.53	3.24	0.001	-
Eng-Man | Working Memory	-1.13	0.47	-2.40	0.016	-
Eng-Man | Tone Categorization	-0.06	0.48	-0.13	0.896	-
English					
Musical Experience	2.21	0.45	4.90	0.000	[1.32 ; 3.09]
Working Memory	0.06	0.35	0.17	0.982	[-0.63 ; 0.75]
Tone Categorization	-0.34	0.29	-1.17	0.424	[-0.91 ; 0.23]
Mandarin					
Musical Experience	0.48	0.28	1.66	0.193	[-0.09 ; 1.04]
Working Memory	1.19	0.31	3.79	0.000	[0.58 ; 1.81]
Tone Categorization	-0.28	0.38	-0.72	0.720	[-1.03 ; 0.47]

2.8.2 Details on performance for Mandarin speakers with knowledge of other Chinese languages

We note that some (varieties) of the Chinese L2s (Cantonese, Wu, Kunming Chinese) reported by our Mandarin speakers have level tone contrasts unlike Mandarin, which may have affected performance on our mid- and low-level tones. However, a visual inspection of performance and error types by participants who reported a L2 with level tone contrasts (‘Level Dialects’, in blue) versus participants who did not (‘No Level Dialect’, in red), did not reveal notable differences (Appendix 2.12–13). In addition to the fact that all participants reported that Mandarin was their L1 and the language they used the most, we therefore deemed it fit to group these participants together.

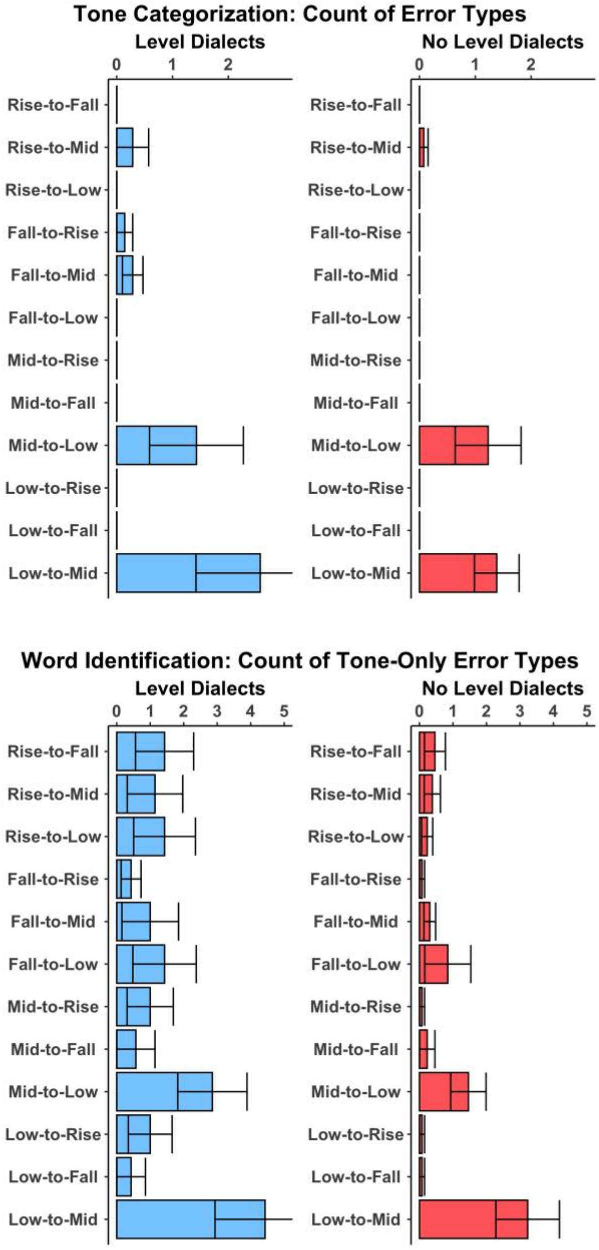
Appendix 2.12

Performance per Mandarin dialect group.



Appendix 2.13

Count of error types per Mandarin dialect groups.



Counts are averaged over subject. Error bars = +/- 1 SE.

Chapter 3 Tone imitation and image-naming[8]

Lexical tones are known to be a challenging aspect of speech to acquire in a second language, but factors such as L1 tonal status (whether a learner's L1 is tonal or not), tone type (the shape of the tones to be acquired) and individual pitch perception aptitude (the ability to accurately perceive pitch in non-lexical settings) are known to facilitate tone learning. Crucially, most of our knowledge of the effect of these factors is based on evidence from perception. The production side of tone learning and the origins of individual variability in learning facility remain relatively understudied. To this end, this study investigated accuracy in non-native tone production, both in phonetic-acoustic terms at a pre-lexical level and in phonological terms at a lexical level, in English-L1 and Mandarin-L1 speakers who participated in an imitation and an image-naming task of tonal pseudowords. Results show that L1 tonal status and tone type dynamically affected both pre-lexical and lexical tone production, revealing specific accuracy patterns for the English and Mandarin groups. Production accuracy was further facilitated by individual pitch aptitude. This study's findings add to the currently limited literature on how both language-specific and individual extralinguistic factors modulate non-native tone production.

[8] Adapted from: **Laméris, T.J.** (n.d.). 'L2 Phonetic and Phono-lexical Tone Production Accuracy by English-L1 and Mandarin-L1 speakers'. Under review (second round) at *Language and Speech*.

3.1 Introduction

Lexical tones are a relatively difficult aspect of speech to acquire in a second language for adult learners, and although learners may overcome difficulties in processing tones devoid of lexical meaning, for instance in tone identification or discrimination tasks (Tsukada & Kondo, 2019; X. Wang, 2013), it appears that lexical processing of tones, for instance in word learning tasks (Ling & Grüter, 2020; Pelzl et al., 2019, 2020), may be particularly challenging. Yet, some individuals learn tones more easily than others do, and previous studies have identified how tone learning facility may be affected by individual factors such as L1 tonal status, which refers to whether a learner's L1 is tonal or not (R. K. W. Chan & Leung, 2020; Cooper & Wang, 2012; Francis et al., 2008; Wayland & Guion, 2004), but also tone type, which refers to the shape of F0-based units (either tonal or intonational) in the learner's L1 and their similarity to L2 target tones (J. Chen et al., 2020; Hao, 2012; So & Best, 2010; K. Yu et al., 2017). In addition, extralinguistic factors such as pitch perception aptitude, which refers to the individual ability to perceive pitch pre-lexically (Dong et al., 2019) have been found to influence the ease with which non-native tones are processed (Ling & Grüter, 2020; Perrachione et al., 2011; Wong & Perrachione, 2007). Chapter 2 (Laméris & Post, 2022) assessed the combined effects of L1-specific and extralinguistic factors on individual tone learning performance in tone categorization and word identification.

Crucially, most of our knowledge of the effects of individual factors on L2 tone learning is based on evidence from perception. The combined effects of L1 tonal status, tone type, and pitch aptitude on tone production have received much less attention, as will be discussed in section 3.2. To this end, this Chapter investigates how these factors affect the learning of tones in a non-native tone system, viewed through the lens of the speaking modality. Specifically, it assesses tone production of tonal pseudowords by English-L1 and Mandarin-L1 speakers in terms of their phonetic production accuracy in a pre-lexical imitation task, and their phono-lexical production accuracy in a lexical image-naming task.

An imitation task can be seen as "a production task adopting auditory instead of orthographic prompts" (Hao & de Jong, 2016, p. 152) in which upon presentation of an auditory signal, speakers are asked to repeat that sound out loud and as accurately as possible. An imitation task was deemed to be a suitable tool to investigate individual differences in *phonetic accuracy* of L2 tone production. This is because participants are expected to perform relatively well in imitation tasks, and mostly differ between one another in terms of measurable F0-based values rather than in overt phonological categories (Dong et al., 2019; Hao, 2012). For instance, participants who hear a pseudoword stimulus /lɔn/ with a rising tone are all expected to reproduce that stimulus relatively accurately and produce a rising pitch, but the fine-grained acoustic nature of that pitch movement – that is, its phonetic accuracy – may differ between participants. The first aim of this study is thus to investigate whether any individual differences in such phonetic accuracy at a pre-lexical level can be attributed to an individual's L1 tonal status, tone type, or pitch aptitude.

Although imitation tasks may reveal part of the production side of tone learning, L2 learners in real life rarely have constant access to target sounds that they can then imitate. Instead, as active language users, they are likely to be involved in lexical production, which involves the breaking down of a lexical item into phonemes, or "sub-lexical representations" and subsequently into speech (Schmitz et al., 2018, p. 529). In this study, such lexical production was measured by means of an image-naming task, in which participants are presented with an image of a lexical item that they are then asked to name in an L2 (Barcroft & Sommers, 2014; M. Li & Dekeyser, 2017; A. C. L. Yu et al., 2021). Because it is the overt, abstract phonological categories rather than the phonetic properties of tone productions that are of interest here, *phono-lexical accuracy* was determined based on auditory labeling by two raters, following similar earlier studies (Dong et al., 2019; Shih & Lu, 2010). An image-naming task lends itself well to the investigation of individual differences in lexical tone production, as participants are expected to differ in their ability to link tonal categories to lexical items. For instance, upon presentation of an image of a chair, which participants in the present study were trained to associate with the pseudoword /lɔn/ with a rising tone, some participants may correctly produce the target word, whereas others may mispronounce it in terms of its tonal properties, for instance by producing /lɔn/ with a falling tone. This study's second aim is thus to investigate whether the ease with which individuals retain and associate tonal categories to lexical items in production is in any way modulated by L1 tonal status, tone type, and individual aptitude.

3.2 Background

3.2.1 Effects of L1 tonal status on pre-lexical production of non-native tones

The effect of L1 tonal status on pre-lexical processing of tones has mainly been investigated in perception (I. L. Chan & Chang, 2019; Schaefer & Darcy, 2014; Wayland & Guion, 2004) but there are a few studies that looked at production.

For instance, Y. Wang et al. (2003) evaluated tone production accuracy in a read-aloud task of Mandarin words written in pinyin[9] with tone marks by English-L1 beginner-level learners. An acoustic analysis comparing L2 and native speakers' pitch curves showed that even though L2 learners' phonetic tone accuracy improved after a training session, their productions still deviated from native norms in terms of pitch height and pitch contour. Similarly, in a read-aloud task of Mandarin pinyin words with tone marks by native speakers and German-L1 beginner-level learners, Ding et al. (2011) report that German speakers employed a narrower pitch range and produced less defined pitch changes than native Mandarin speakers. Recently, Kirby and Giang (2021) investigated phonetic production of Vietnamese tones by speakers of Khmer Krom, a non-tonal Austroasiatic language. They operationalized phonetic production accuracy by calculating two overall proxies of similarity between learner and native F0 curves, namely Dynamic Time Warping (Müller, 2007) and the Fréchet distance (Chambers et al., 2010). Their acoustic analyses revealed that in general, Khmer productions had a more compressed pitch range and lacked clear distinctions between two complex contour tones in comparison to native Vietnamese productions.

The abovementioned studies suggest that non-tonal L1ers may have difficulty in fully exploiting the phonetic properties of non-native tones. It should be noted, however, that these studies all compared *non-native* tone production by *non-tonal* L1ers to *native* tone production by *tonal* L1ers. It is thus unclear whether the reported L2 learners' poorer production is an effect of their L1 tonal status (i.e., their inexperience with tones leads to poorer L2 tone production) or a general effect of production in a non-native system, which may naturally

[9] Pinyin: System to transcribe Mandarin Chinese in the Roman alphabet.

lead to phenomena such as reduced pitch range (Grazia Busà & Urbani, 2011; Zimmerer et al., 2014). To the best of my knowledge, there are no studies that have explicitly compared non-tonal and tonal L1ers in terms of their L2 phonetic tone production in a tone system that is entirely unknown to both groups.

3.2.2 Effects of tone type on pre-lexical production of non-native tones

In addition to the effect of L1 tonal status, this Chapter also considers the effect of tone type. The term ‘tone type’ will henceforth be used as an overarching expression to describe the nature of F0-based units (which can be either lexical or intonational tones) in terms of 1) phonetic-acoustic and 2) phonological-categorical properties, following Chapter 2. It is important to consider tone type in addition to L1 tonal status because rather than an individual’s L1 tonal status alone, L1 tone types and their interaction with L2 tone types also often determine how easily specific L2 tones are learned.

The effect of tone type has been widely studied in pre-lexical tone perception, but there are only a handful of studies on pre-lexical production: Hao (2012) investigated production accuracy of Mandarin tones in an imitation task, and in a read-aloud task of Mandarin pinyin words with tone marks by Cantonese-L1 and English-L1 intermediate learners of Mandarin. Production accuracy (determined by two native raters) did not differ between the two groups, although error patterns in both the imitation and read-aloud tasks revealed some effects of the Cantonese tone types on Mandarin tone productions. More direct evidence for the effect of tone type on production accuracy comes from a read-aloud task with tone marks of Cantonese tones by native Cantonese and naïve Mandarin speakers by K. Zhang and Peng (2017), who suggest that mismatches between Mandarin and Cantonese tone inventories modulated phonetic production accuracy. They calculated the relative distance of Mandarin productions from native norms and showed that productions of the Cantonese high-level tone (which is similar to the Mandarin high-level tone) deviated least from the native norm, whereas productions of mid-level and low-level tones (which have no equivalent in Mandarin) deviated more. These findings could be taken as evidence for an effect of tone type in pre-lexical production, as Mandarin speakers either pay phonetically less attention to pitch height differences, and/or phonologically assimilate Cantonese level contrasts to their single L1 level tone category.

3.2.3 Effects of L1 tonal status and tone type on lexical production of non-native tones

Although the lexical processing of tones (i.e., the association between tone and meaning) and the individual factors that modulate it have been studied in perception (Cooper & Wang, 2012; Pelzl et al., 2020; Poltrock et al., 2018; Wong & Perrachione, 2007), there appear to be only a handful of studies that zoom in on lexical production. A study by Shih and Lu (2010) involved a read-aloud task of digits by six English-L1 intermediate learners of Mandarin[10]. Although they did not carry out a statistical analysis to confirm their findings, they note that some learners tended to produce Mandarin tones with English-like pitch patterns based on English intonation, suggesting an influence of tone type. Similarly, C. Yang (2019) investigated lexical production by means of a read-aloud task of Mandarin characters by Thai and Yoruba speakers. Overall production accuracy (assessed by a native rater) and analyses of error patterns suggested that Thai learners benefited from categorical assimilation from Mandarin to Thai tone types and thereby outperformed Yoruba speakers. However, the Thai speakers had more exposure to Mandarin than the Yoruba speakers, which may have influenced performance.

A recent study by A. C. L. Yu et al. (2021) examined Cantonese lexical production accuracy in a group of Hong Kong Cantonese native speakers and L2 speakers who spoke Urdu, Punjabi, or English as their dominant language. Production was elicited by an image-naming task, making the nature of the productions inherently lexical, but production accuracy was measured phonetically by a direct acoustic comparison between native and non-native productions. The results of these analyses revealed potential effects of L1 tonal status and tone types. For instance, speakers of Punjabi (a tone language) may have assimilated the two

[10] Note that the reading tasks by Shih and Lu (2010) and C. Yang (2019) can be deemed lexical because digits and Mandarin characters contain no direct cues of how tones should be pronounced. Participants thus need to make a conscious effort to break down a lexical item into sound. This is unlike the reading tasks by Ding et al. (2011); Hao (2012); and Y. Wang et al. (2003) which can be deemed pre-lexical as stimuli included only orthographic or acoustic cues (for instance by presenting the stimuli in pinyin with tone marks).

rising Cantonese tones to the Punjabi high tone and the lower Cantonese tones to the Punjabi mid tone, which may have caused difficulty in producing these tone distinctions in the lexical task. In a similar way, it is suggested that interference with English and Urdu rising intonation tone types may have been at the root of the same Cantonese tone production difficulties in English-dominant and Urdu-dominant participants

Following up from the available evidence on lexical L2 tone production, which is mainly based on learners with prior knowledge of the target L2, the present study investigates how *ab initio* learners learn to associate non-native tones to word meaning in the speaking modality, and how their L1 tonal status and tone types may affect the ease with which they do so.

3.2.4 Individual pitch aptitude and L2 tone production

To explain how individual performance in L2 tone production is modulated by not only L1-specific factors (such as L1 tonal status and tone types) but also by individual extralinguistic factors, this Chapter also investigated the effect of pitch perception aptitude. Earlier tone production studies have suggested that such pitch processing skills may affect both pre-lexical and lexical production.

For instance, English-L1 individuals with musical experience, which is assumed to enhance pitch processing abilities (Patel, 2011), have been found to be better than non-musicians at imitating Mandarin tones (Gottfried et al., 2004), and English-L1 speakers with enhanced musical sensitivity (measured by perception of pitch change and memory tests) have been found to sound more native-like in Mandarin pre-lexical and lexical tone production tasks (M. Li & Dekeyser, 2017).

In Dong et al. (2019), 60 English-L1 naïve learners were trained to learn a set of tone words in Mandarin Chinese. Individual pitch perception aptitude was measured through a tone categorization task. Participants took part in a tone word imitation as well as an image-naming task. Production accuracy in both tasks was assessed by two native raters. Findings from both tasks revealed a small effect of individual learners’ aptitudes, as pitch perception aptitude significantly predicted performance in tone imitation, and marginally so for performance in image-naming.

I deemed it necessary to also consider in this study to what extent individual factors other than L1 tonal status and tone type affect pre-lexical and lexical production. In this Chapter, I therefore additionally assess the effect of individual pitch perception aptitude on tone production. Here, I only investigate the effect of pitch aptitude and not of other measures (such as musical experience and working memory) to allow for a comparison with Dong et al. (2019) who similarly assessed the effect of pitch aptitude (measured by accuracy in tone categorization) on non-native tone imitation and image-naming. In Chapter 4, I address the effects of individual musical experience and working memory on imitation and image-naming accuracy.

3.3 Research questions

Following on from previous production studies that have in part addressed this Chapter's research aims, I formulate the following research questions.

RQ1: *How do L1 tonal status and tone types determine phonetic production accuracy in a pseudoword imitation task by English-L1 and Mandarin-L1 speakers?*

RQ2: *How do L1 tonal status and tone types determine phono-lexical production accuracy in a pseudoword image-naming task by English-L1 and Mandarin-L1 speakers?*

RQ3: *How does an individual's pitch perception aptitude, in addition to L1 tonal status and tone type, further modulate phonetic and phono-lexical production accuracy?*

These research questions aim to address two main gaps in the existing literature on non-native tone production. First, this study aims to provide a more direct comparison between tonal and non-tonal speakers by observing tone production in a tone system that is unknown to all participants. This enables us to more directly observe whether mere L1 tonal status is associated with better tone production in a non-native tone system at the earliest stages of learning. To the best of the author's knowledge, there are only two published studies that compared non-native tone production by tonal and non-tonal L1ers (Hao, 2012; A. C. L. Yu et al., 2021), however participants in these studies had previous knowledge of the target L2, thereby potentially influencing production accuracy.

Second, this study investigates not only the effect of L1 tonal status, but simultaneously the effects of tone type and individual aptitude. Although the individual effects of these factors have been studied separately in previous production studies, there appear to be no studies that investigated their combined effect. By simultaneously factoring in L1 tonal status, tone type, and individual aptitude, this study thus aims to provide a more comprehensive overview of the factors that modulate non-native tone production. I also highlight here that, as reported in Chapter 2, measures of individual participants' musical experience and working memory were obtained, and there was no difference in musical experience or WM between the English and Mandarin groups. In all the aforementioned cross-linguistic studies (Hao, 2012; Y. Wang et al., 2003; C. Yang, 2019; A. C. L. Yu et al., 2021), it appears that individual musical experience or WM was not controlled for. The present study thus presents a more controlled environment to study the effects of L1 tonal status, tone type, and pitch aptitude on non-native tone production, both at a pre-lexical and at a lexical level.

3.4 Methods

3.4.1 Participants

Participants were the same as in Chapter 2 (section 2.3.1, page 39).

3.4.2 Stimuli

Stimuli were the same as in Chapter 2 (section 2.3.2, page 40).

3.4.3 Procedure

The procedure was the same as described in Chapter 2 (section 2.3.3, page 45). A battery of eight tasks was conducted over two consecutive days. An overview of the tasks is given in Table 3. Note that in addition to the tasks reported in this Chapter (imitation and image-naming), participants also completed a tone categorization task to measure tone categorization accuracy (i.e., pitch aptitude), a backwards digit span task to measure working memory, and a word identification task. These tasks were described in detail in Chapter 2.

All experiments were administered in a sound-attenuated booth and run on a *DELL Inspiron 13 5000 Series* touchscreen tablet laptop through the *OpenSesame* software (Mathôt et al., 2012). Participants listened to audio stimuli over *Beyerdynamic DT 990* headphones at a comfortable listening level. Voice recordings were made using a *Sennheiser* cardioid microphone and *SoundDevices mixpre6* recorder.

Table 7

Overview of tasks.

DAY 1	
Description	Duration (minutes)
*Tone categorization (pitch aptitude)**	*5*
Imitation	10
Image-naming	5
*Word identification**	*15*
DAY 2	
Description	Duration (minutes)
Working memory	*5–10*
Imitation	10
Image-naming	5
Word identification	*15*

**Reported in Chapter 2.*

3.4.3.1 Imitation task

As described in Chapter 2, the imitation task served to prepare participants to learn a set of 16 pseudowords through a listen-and-repeat paradigm. It also served as a measure of phonetic tone production accuracy (Hao, 2012, p. 277)[11]. In the imitation task, participants were asked to repeat the words out loud and pronounce these as accurately as possible, whilst simultaneously trying to memorize them for the subsequent image-naming task. No feedback

[11] Although the nature of an imitation task is inherently acoustic-phonetic, as it requires a participant to perceive an auditory stimulus and accurately reproduce it immediately after, it is not unthinkable that this imitation task also involved some phono-lexical processing. This is because participants also saw the image that represented the meaning of a word when they heard the auditory stimulus. I will discuss the effect that any potential top-down processing may have had on imitation in more detail in Chapter 4 (section 4.6.1.2, page 154). However, the imitation task still lends itself to investigate differences in phonetic tone production accuracy, given that participants were all expected to reliably reproduce tones according to their broad phonological properties, and only show differences in fine-grained phonetic properties.

was given regarding their production and all productions were voice-recorded.

After a familiarization with the images and their meanings in participants' native language to ensure that participants considered the images to be analogous to a word in their L1, each of the 16 pseudowords was audiovisually presented four times, resulting in 64 trials in total. Participants had 5000 ms to repeat the word before the next audiovisual stimulus was presented. The first two trials were in a pseudorandomized order for all participants: each audiovisual stimulus was presented twice in a row (e.g., the word for 'cat', followed by the participant's imitation, followed by one more trial (presentation + imitation) for 'cat'), and the order was such that no segmental or tonal minimal pair followed one another. The last two presentations were fully randomized for each participant individually.

The same imitation task was conducted on day 2. The only difference was that the image familiarization was not conducted, and that the pseudorandomized presentation order was the reverse of that of day 1.

3.4.3.2 Image-naming task

The image-naming task, which replicates L1-to-L2 recall abilities (Barcroft & Sommers, 2014), served as a measure of phono-lexical production accuracy. The image-naming task was held directly after the imitation task. Each of the 16 visual stimuli was presented one by one, in a randomized order, for 5000 ms through the *OpenSesame* software. Participants were asked to pronounce the pseudoword to the best extent of their memory. Their productions were voice-recorded. The exact same task was repeated on day 2.

3.4.4 Determining tone production accuracy

3.4.4.1 Phonetic production accuracy (imitation)

The imitation task served to measure phonetic tone production accuracy, which was computed by calculating the Fréchet distance, as "a one-number summary" of phonetic similarity between produced and target pitch curves (Kirby and Giang, 2021: 255). To calculate the Fréchet distance, participants' imitations (only from the first block in which each word was repeated twice one after another) were first manually labelled in *Praat*

(Boersma & Weenink, 2019), after which F0 values were extracted using *ProsodyPro* (Xu, 2013) at 20 equidistant points for each imitated pseudoword. Because of a relatively high proportion of pitch excursions at the edges of the measurements, the first and last two measured points from the obtained curves were truncated. Any other pitch excursions or missing values within the remaining 16-point curves were manually corrected by interpolation or by re-obtaining the F0 values using the *Pitch Listing* command in *Praat*. The trimmed pitch curves were then converted to semitones for each participant to normalize for differences between speakers, and Fréchet distance per token was obtained by comparing participants' normalized pitch curves to those of the target stimuli, following the methodology and script described in Kirby and Giang (2021). 39 out of 2642 data points (2 productions x 16 words x 41 participants x 2 days) were excluded due to recording errors or non-responses, resulting in a final total of 2603 data points for analysis.

3.4.4.2 Phono-lexical accuracy (image-naming)

The image-naming task served to measure phono-lexical accuracy. Recall that phono-lexical accuracy in this Chapter is defined as the ability to produce overt phonological contrasts (segmental and tonal) to distinguish word meaning, for instance in contrasting /a/ and /u/ and a rising and a falling tone to respectively distinguish 'mountain' /jaɹ15/ from 'guitar' /juɹ51/. Phono-lexical accuracy was calculated based on manual labeling by myself and an external rater with over six years of experience in phonetics and tone and intonation research. Labels for segments and tones were assigned based on auditory observations and inspection of formants and F0 tracks in *Praat.* Inter-rater reliability was determined using the *irr* package (Gamer et al., 2019) in *R* (R Core Team, 2021). *Kappa* statistics were 0.96 for segmental and 0.86 for tonal labels, indicating a very strong level of agreement. In case of disagreement, the label indicated by the external rater was applied.

Although it was relatively unequivocal to determine whether a production should be labeled 'rise' or 'fall' based on an auditory impression and on F0 track observations, determining whether a production should be labeled as 'mid-level' or 'low-level' was slightly less clear in some cases. To disambiguate these tone productions, the *Contour Clustering* Graphical User Interface (Kaland, 2021) was used to determine in a more objective way whether participants produced distinct mid-level and low-level tone categories. The *Contour Clustering* GUI carries out an automated data-driven cluster analysis that groups together F0 contours based on their acoustic similarities into a defined number of clusters (Kaland, 2021).

F0 curves of productions (per participant and per day) that were labeled as level tones were imported in the *Contour Clustering* GUI and reduced to two clusters to determine whether participants made a categorical distinction between mid-level and low-level tones. The results revealed that 91.22% of initially assigned labels corresponded to mid- and low-level clusters generated by the GUI. The 52 non-corresponding labels (e.g., a word that was initially labeled as a mid-level tone by the raters but that was clustered in the low-level category by the GUI) were changed accordingly in line with the categorization proposed by the *Contour Clustering* GUI.

Finally, there were a few instances (2.80% of all productions) in which segmental productions were clearly deviant from the target words, such as [weɹ] or [na]. These productions were labeled accordingly as incorrect non-target-like responses. There were only two instances in which tone productions were not clearly categorizable according to any of the four categories, and these were removed from the analysis. 4 out of 1312 data points (16 productions x 41 participants x 2 days) were removed from analysis due to bad recording quality or due to non-responses, resulting in a final total of 1308 data points.

3.4.5 Statistical procedures

All analyses were performed in *R 4.1.1* (R Core Team, 2021). Figures were generated with the *ggplot2* package (Wickham, 2016). To measure the effects of L1 tonal status, tone type, and individual pitch aptitude on phonetic and phono-lexical accuracy, (generalized) linear mixed effects models were built in the *lme4* package (Bates et al., 2015). Models contained fixed effects and interactions of interest to this Chapter's research questions. Models were fitted with the *bobyqa* optimizer where applicable. Model diagnosis (observation of residual QQ plots) was carried out with the *DHARMa* package (Hartig, 2020). To avoid multicollinearity of fixed effects, a maximum Variance Inflation Factor (VIF) threshold of 5 was set for all models (O'Brien, 2007). None of the models showed multicollinearity.

For the phonetic accuracy results of the imitation task, a linear mixed effects model (dependent variable = Fréchet distance) was built with *Day* (Day 1, Day 2; contrast-coded), *L1* (English, Mandarin; contrast-coded), *Tone Type* (Rise, Fall, Mid-level, Low-level; contrast-coded), *Aptitude* (Expressing accuracy score in the tone categorization task; centered and scaled) as fixed effects, and a four-way *L1:Tone:Aptitude:Day* interaction. The model contained by-subject random slopes for *Day* and for *Tone*, and a random intercept for *Item*. Fréchet distances were z-transformed to improve normality of model residuals.

For the phono-lexical accuracy results of the image-naming task, a generalized mixed effects model (dependent variable = correct/incorrect) was built. As will be described in more detail in section 3.6.2, only the data of day 2 were analyzed. The final model contained *L1* (English, Mandarin; contrast-coded), *Tone Type* (Rise, Fall, Mid, Low; contrast-coded), *Aptitude* (Expressing accuracy score in the tone categorization task; centered and scaled) as fixed effects and a three-way *L1:Tone:Aptitude* interaction. The model contained random intercepts for *Subject* and *Item* (random slope models did not converge).

To investigate the nature of the interactions between fixed effects in the models in more detail, post-hoc tests were carried out using Bonferroni-corrected multiple comparisons in the *emmeans* package (Lenth, 2020).

3.5 Predictions

I formulate the following predictions with regard to this Chapter's research questions and production tasks:

P1: Any differences between English-L1 and Mandarin-L1 participants in phonetic accuracy in the imitation task will be due to an interaction between L1 tonal status and tone type. Specifically, it is predicted that English participants may be better than Mandarin participants in accurately imitating mid-level and low-level tones, but worse than Mandarin participants in imitating rising and falling tones.

P2: An effect of L1 tonal status is predicted for the image-naming task as Mandarin participants, by virtue of their L1 tonal status, may be better than English speakers at linking tones to lexical meaning. Mandarin speakers may thus be better prepared to retain tonal information as part of the overall phono-lexical information, and overall outperform English speakers in phono-lexical tone production.

P3: Finally, an overall positive effect of pitch perception aptitude is expected in both phonetic and phono-lexical tone production accuracy.

3.6 Results

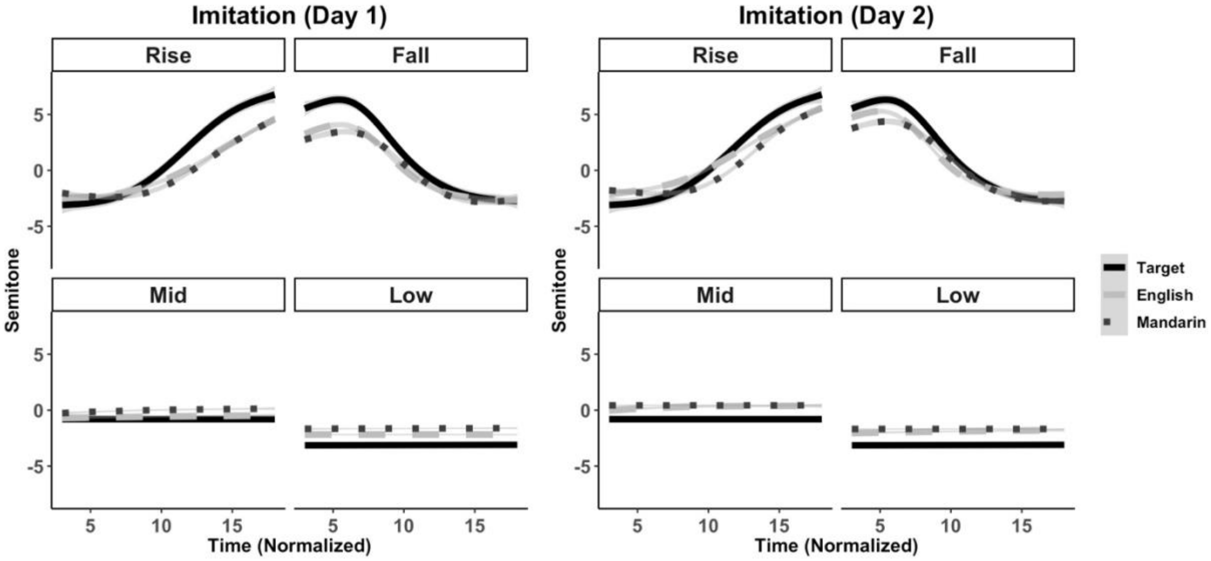
3.6.1 Imitation task (phonetic accuracy)

3.6.1.1 Visualization of productions

Figure 20 shows group-averaged normalized pitch curves per tone. Individual pitch curves are shown in Figure 21. A visual comparison between participants' pitch curves and the target curves reveals two main deviations from the target. First, pitch range for the rising and falling tones is relatively compressed, with participants not attaining high pitch values at the extremes of these contour tones (at least on day 1). Second, productions for the mid-level and low-level tones appear to be relatively high in comparison to the target values.

Figure 20

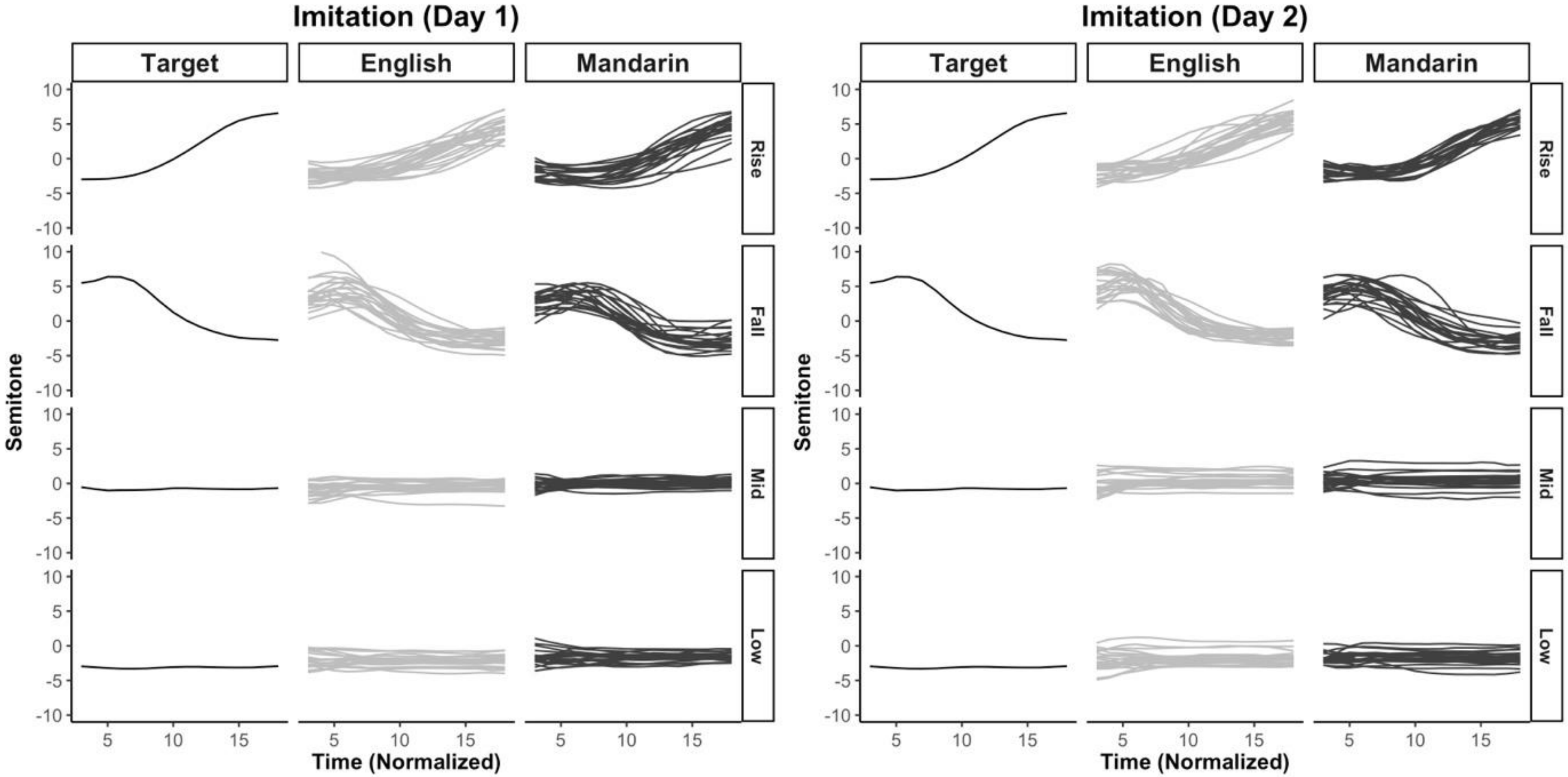
Imitation: Group-averaged smoothed normalized pitch curves.



Shading ribbons, where present, indicate a 95% Confidence Interval

Figure 21

Imitation: Individual normalized F0 traces.

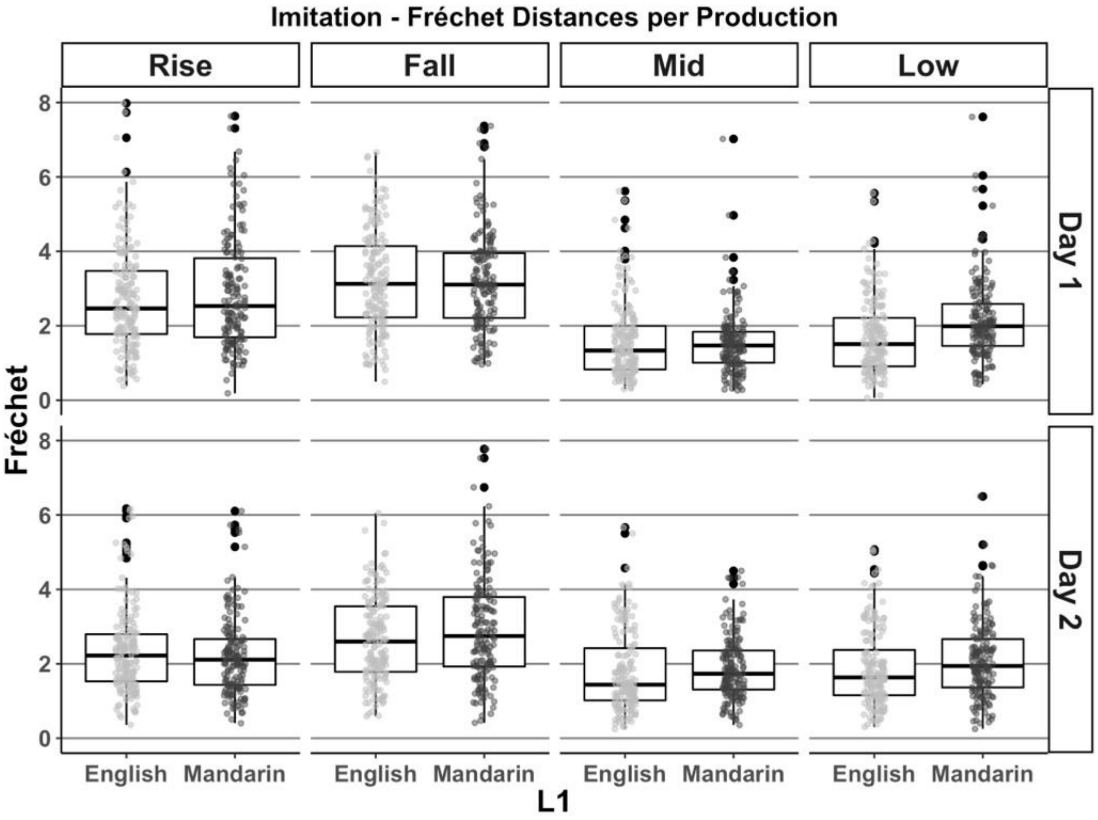


3.6.1.2 Overview of performance and individual variability

Figure 22 shows Fréchet distances for each imitation trial. Mean values are shown in Table 8. Recall that higher Fréchet distance values indicate greater deviation from the target curve and therefore less phonetically accurate tone productions. A visual inspection suggests that on day 1, participants' imitations were least phonetically accurate for words with rising and falling tones (i.e., highest Fréchet distance values), and most accurate for mid-level and low-level tones. On day 2, overall phonetic accuracy appeared to improve slightly (i.e., Fréchet distances decreased), but the difference between contour and level tones persists.

Figure 22

Imitation: Fréchet distances per tone, per L1, and per day.



Dots indicate individual productions per trial.

Table 8

Imitation: Mean Fréchet distances.

Day 1	Rise	Fall	Mid-level	Low-level
English	2.72	3.23	1.60	1.71
Mandarin	2.86	3.20	1.53	2.14
Day 2	Rise	Fall	Mid-level	Low-level
English	2.35	2.65	1.79	1.88
Mandarin	2.26	2.92	1.89	2.08

The linear mixed-effects model revealed a significant *L1:Tone:Aptitude:Day* main interaction, suggesting that Fréchet distances (i.e., phonetic accuracy) differed significantly according to L1, tone, and day, and that the effect of *Aptitude* differed according to these conditions. Full details on significant main effects and interactions are reported in Appendix 3.1.

I investigated the nature of this four-way interaction in two steps. First, in section 3.6.1.3, I conducted multiple comparisons for Fréchet distances between L1, tone, and day. Full details on these multiple comparisons are reported in Appendix 3.2–3 . Recall that estimates in multiple pairwise comparisons should be interpreted with reference to the latter element in the pair. For instance, the first row in Appendix 3.2 shows the Rise-Fall comparison on day 1 within the English group (*b*= -0.39, *SE* = 0.12, *p* = 0.015). This indicates that mean Fréchet distances for rising tones were significantly lower than falling tones in this condition, suggesting that these productions were significantly more accurate in phonetic terms. Because of the many comparisons, I will summarize the main observations in the text, and refer to Appendix 3.2–3 for the full statistical details.

In section 3.6.1.4, I report the effect of *Aptitude* per L1, tone, and day. Full details of the effect of *Aptitude* per condition is reported in Appendix 3.4.

3.6.1.3 Model results: L1:Tone:Day interaction

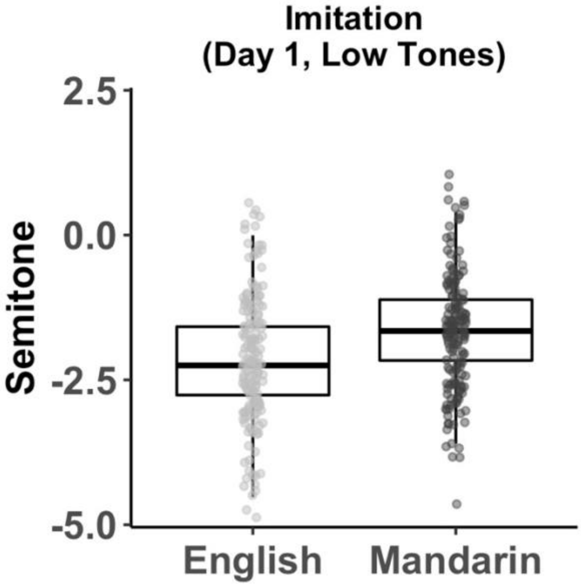
Full details on the multiple comparisons are reported in Appendix 3.2–3. Within-group multiple comparisons revealed that on day 1, phonetic accuracy for imitation of contour tones (i.e., rising and falling tones) was lower than that of level tones for both English and Mandarin participants (except Mandarin participants' rising tone accuracy compared to low-level tone accuracy, for which no significant difference in Fréchet distance was found). In addition, English imitations of falling tones were less accurate than imitations of rising tones, and Mandarin imitations of low-level tones were less accurate than imitations of mid-level tones.

On day 2, contour tones were imitated less accurately than level tones by both groups, with the exception of rising tones, which were not significantly less accurate than low-level tones. Mandarin imitations of falling tones were less accurate than imitations of rising tones.

The between-group comparisons revealed that Mandarin imitations of low-level tone words were phonetically less accurate than English imitations of low-level tone words, but only on day 1. To investigate the nature of this difference, a separate linear mixed-effects model (dependent variable: Mean semitone value per production) with *L1* (English, Mandarin; reference = English) as fixed effects and *Subject* and *Item* as random intercepts was fitted to the data of low-level tone word imitations on day 1. This model indicated that Mandarin speakers' imitations of low-level tone words were significantly higher in terms of semitone value than English imitations of low-level tone words (b = 0.55, SE = 0.21, p = 0.014). This is visualized in Figure 23.

Figure 23

Imitation (day 1): Semitone values for low-level tones per L1.



Dots indicate individual productions.

3.6.1.4 Pitch aptitude as a predictor of imitation performance

Full details of the effect of *Aptitude* per condition is reported in Appendix 3.4. *Aptitude* did not significantly predict phonetic accuracy (i.e., it did not lower Fréchet distance), except for falling tone imitations on day 2 by Mandarin participants (*b* = -0.69, *SE* = 0.25, *p* = 0.009).

3.6.2 Image-naming (phono-lexical accuracy)

3.6.2.1 Overview of performance and individual variability

Figure 24 shows accuracy for the image-naming task on days 1 and 2. Descriptive statistics are reported in Table 9. Like the word identification task in Chapter 2, participants improved their image-naming accuracy on day 2, and attained relatively high accuracy levels, but a large degree of individual variability was observed in both groups. Similar to Chapter 2, in the following I only analyze the data of day 2 of the lexical task, since on day 1, participants had not yet completed the word training.

Figure 24

Image-naming: Accuracy (% correct) per day and L1.

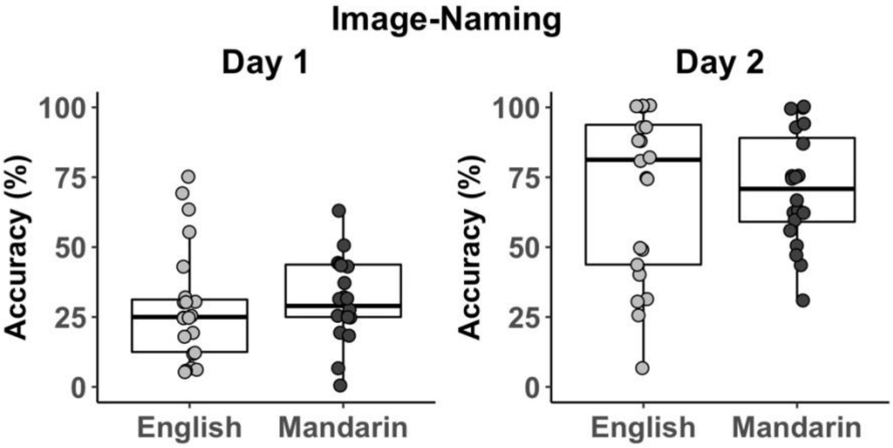


Table 9

Image-naming: descriptive statistics.

	English	Mandarin
Day 1 accuracy (%)	29.5 (20.8)	30.7 (14.7)
Day 1 % of tone-only errors	26.7 (14.2)	35.8 (15.5)
Day 2 accuracy (%)	68.6 (29.4)	70.9 (20.3)
Day 2 % of tone-only errors	53.5 (30.3)	49.8 (33.1)

Values are means with standard deviations in brackets.

3.6.2.2 Model results: L1:Tone interaction

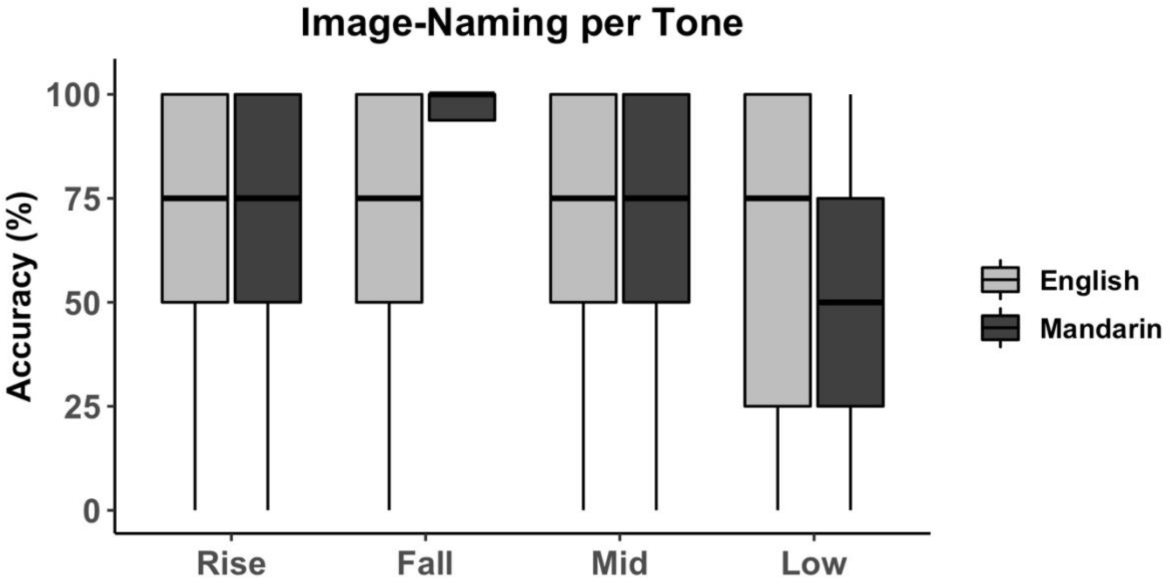
Appendix 3.5 reports main effects and interactions of the model that investigated the effects of L1 tonal status, tone type, and pitch aptitude on the likelihood of correct image-naming. The model revealed a significant *L1:Tone* interaction and a significant main effect of *Pitch Aptitude*.

To investigate the nature of the *L1:Tone* interaction, multiple comparisons for tone are presented in Appendix 3.6. Within-group comparisons revealed that English participants' phono-lexical tone accuracy did not differ depending on the tone. Mandarin participants were less likely to accurately name rising tone words than falling tone words (*b* = -1.61, *SE* = 0.58, *p* = 0.034). They were also less likely to name low-level tone words than falling (*b* = -2.89, *SE* = 0.61, *p* = 0.001) and mid-level tone words (*b* = 1.44, *SE* = 0.53, *p* = 0.039). The comparisons with regard to falling tone words should be interpreted with caution given that Mandarin participants performed at ceiling here.

Comparisons between groups per tone revealed that Mandarin participants were significantly less likely than English participants to correctly name words with a low-level tone (*b* = -1.76, *SE* = 0.60, *p* = 0.003). To visualize the *L1:Tone* interaction, Figure 25 shows accuracy per tone.

Figure 25

Image-naming (day 2): Accuracy per tone and L1.



3.6.2.3 Tone-only error types in image-naming

An analysis of error types revealed that on day 2, participants predominantly made 'tone-only errors'[12], although the proportion of tone-only errors in image-naming appeared to be lower than in word identification (cf. Table 5, page 53, with Table 9, page 104)[13]. Two simple linear regressions confirmed that the number of tone-only errors significantly predicted the total number of image-naming errors and explained a large portion of variance in both the English [$F(1,19) = 78.676$, $p < 0.001$, $R^2 = .7954$] and the Mandarin group [$F(1,18) = 22.43$, $p < 0.001$, $R^2 = .5300$]. This suggests that many participants had acquired the segmental, but

[12] I recall here that a 'tone-only error' refers to an error that indicates that a participant had retained the segmental, but not the tonal properties of a word. An example of a tone-only error in image-naming would be a participant who incorrectly names the image for /lon15/ as /lon22/.

[13] Although the proportion of tone-only errors in image-naming was still high, there were also relatively many image-naming errors that were simply due to a participant not naming the correct segmental properties of the word (e.g., when a participant would incorrectly name the image for /lon15/ as /juɹ22/). I therefore considered excluding any errors that were *not* tone-only errors from the analysis to zoom in on lexical tone production in particular. However, an analysis on this subset of data revealed the same main effects and interactions as an analysis on the full dataset, and I therefore conducted the analysis on the full dataset in line with Chapter 2.

not the tonal properties of the words at the end of the experiment. For visualization, Figure 26 plots the number of image-naming errors against the number of tone-only errors.

Figure 26

Image-naming (day 2): Number of errors against number of tone-only errors.

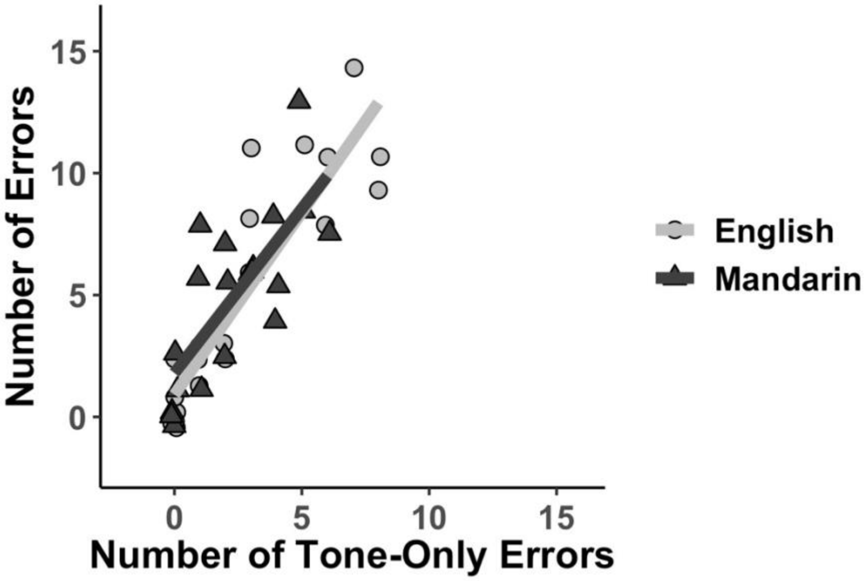
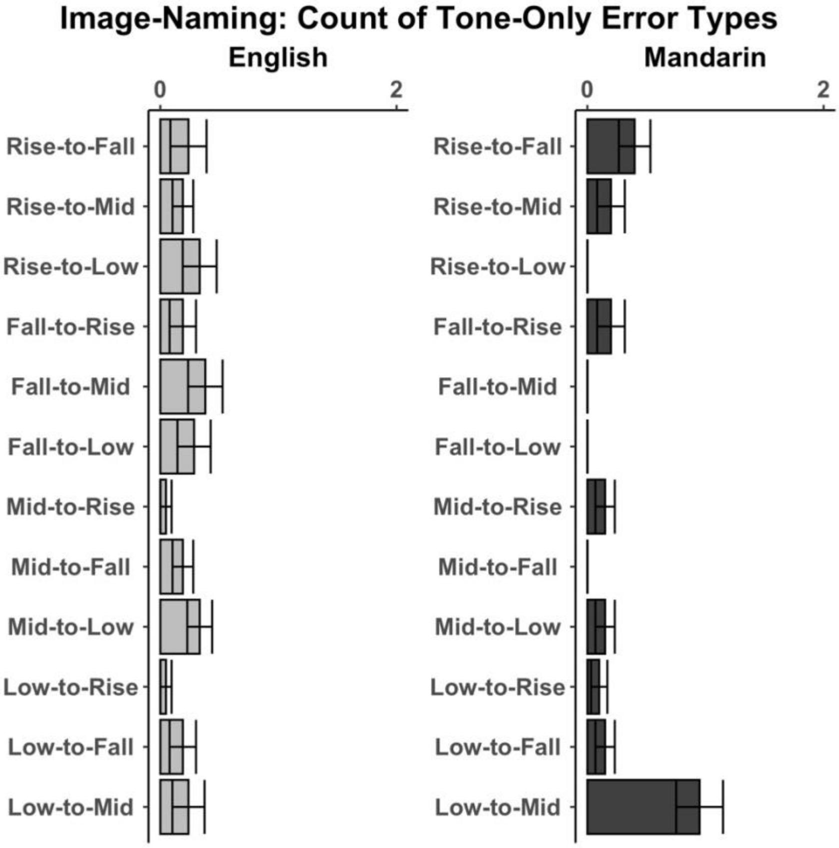


Figure 27 shows the distribution of the average count of tone-only error types. A visual inspection suggests that English participants mispronounced tones on words across the board, whereas it appears that Mandarin participants predominantly made low-to-mid errors. Similar to the approach in Chapter 2, a zero-inflated generalized linear mixed-effects model was fitted on the counts of the tone-only error types, using the *glmmTMB* package (Brooks et al., 2017). Models were fitted separately on each participant group with *Error Type* (Rise-to-Fall, Rise-to-Mid, etc.; contrast-coded) as a fixed effect and a random intercept for *Subject*. A significant main effect of *Error Type* was only found in the model for the Mandarin group (χ^2 = 30.559, *df*(11), *p* = 0.001), suggesting that some error patterns occurred more often than others, but post-hoc Bonferroni tests did not reveal any significant comparisons (all comparisons *p* > 0.05).

Figure 27

Image-naming (day 2): Count of tone-only errors per L1 and error type.



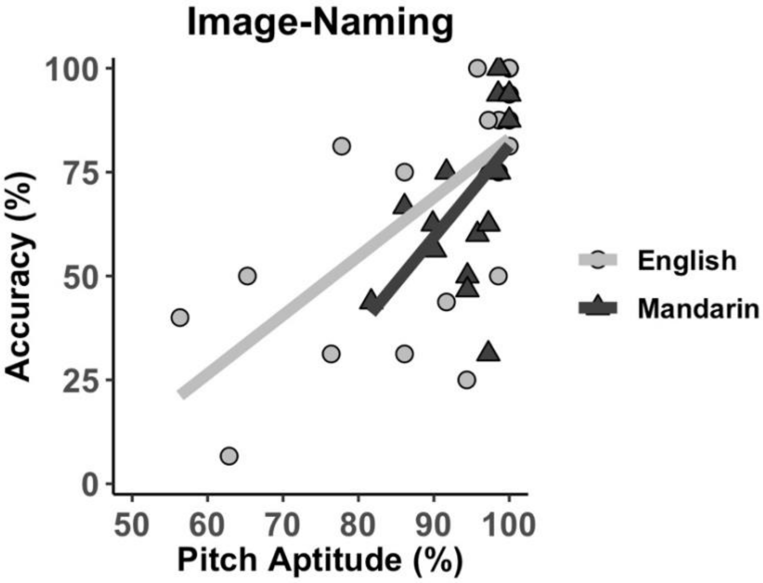
Counts are averaged over subject. Error bars = +/- 1 SE.

3.6.2.4 Pitch aptitude as a predictor of image-naming accuracy

Finally, the model revealed that *Aptitude* positively predicted image-naming performance for both groups ($b = 1.23$, $SE = 0.36$, $p < 0.001$). This effect is shown visually in Figure 28.

Figure 28

Image-naming (day 2): Accuracy against pitch aptitude.



3.7 Discussion

3.7.1 Effects of L1 tonal status and tone types on imitation

The present study explored the effects of L1 tonal status, L1 tone type, and individual pitch perception aptitude on non-native tone production in an imitation and image-naming task. Overall, it was found that, in terms of phonetic accuracy, both English and Mandarin participants were less accurate in imitation of contour tones (rising and falling) than level tones (mid-level and low-level), as Fréchet distances were larger in both groups for all contour tones on both days 1 and 2 of the imitation task. Although this finding may be in part due to the nature of the tones involved (contour tones are inherently more complex than level tones and thus more prone to deviation from a target curve), an observation of participants' pitch traces revealed that they did not fully exploit the pitch range required to accurately imitate the pseudoword rising and falling tones. The fact that this was observed in both L1 groups may be indicative of a general phenomenon of speakers operating in an L2 and not fully using the available phonetic information (Grazia Busà & Urbani, 2011; Zimmerer et al., 2014). Although the development of phonetic accuracy over time was not the focus of this Chapter, it is worth noting that participants slightly improved their imitation accuracy over the two sessions, as shown by the lower Fréchet distances on day 2. (The development of imitation accuracy over time will be discussed in detail in Chapter 4).

Crucially, a significant interaction between L1 tonal status and tone type was observed, and multiple comparisons revealed that on day 1, Mandarin imitations of low-level tones were significantly less accurate than English imitations of low-level tones. Mandarin participants' imitation of low-level tones was, further, significantly less accurate than that of mid-level tones. A follow-up analysis showed that Mandarin speakers imitated low-level tones with significantly higher pitch than did English speakers. This may have been at the root of the observed difference in phonetic accuracy relative to the target production between English and Mandarin speakers. This falls in line with the prediction made for **RQ1**: Mandarin speakers, who are known to struggle phonetically with pitch height distinctions relatively more than do English speakers, and who in addition may have phonologically assimilated the present study's pseudoword level tones to their single L1 high-level tone category, may have had a particular difficulty in accurately imitating low-level tones. The particular reason why only low-level tones and not both mid-level and low-level tones were produced less accurately may chime in with earlier findings that even though when tonal speakers assimilate L2 contrasts to a single L1 tone category, they are still sensitive to a "phonetic residual" (J. Chen et al., 2020). That is, this study's low-level tone (11) is more deviant from the Mandarin high-level tone (55) than the mid-level (22) tone, and this difference could explain why low-level tones were particularly difficult to imitate accurately for Mandarin speakers (Zhang and Peng, 2017). It is however noted that, as pointed out by Francis and colleagues (Francis et al., 2008: 284), any claims regarding assimilation between native and non-native tones can only be "speculative in nature". This is especially the case in the present study since Mandarin participants were not asked to rate the similarity between their native tones and the non-native tones, as was done for instance in the perception studies by J Chen et al. (2020) and Reid et al. (2015).

Contrary to the predictions, English participants were not less accurate in phonetically producing contour tones as opposed to Mandarin speakers.

All in all, and readdressing **RQ1**, the results from the imitation task suggest that L1 tonal status and tone type do affect phonetic production accuracy of L2 tones, but in relatively subtle ways. The main finding from the imitation task was the interaction between L1 tonal status and tone type, which was driven by Mandarin speakers' slightly high-pitched – and thus, less target-like – imitation of low-level tone words compared to English speakers. However, it should be emphasized that this was a relatively small effect, and the difference in relative pitch height was – although statistically significant – rather subtle and only observed on day 1. Apart from this between-group difference, all participants, regardless of their L1 tonal status, thus performed relatively uniformly in the imitation task.

3.7.2 Effects of L1 tonal status and tone types on image-naming

For **RQ2**, it was predicted that in the more cognitively demanding image-naming task, Mandarin participants would overall outperform English participants by correctly producing both the segmental and tonal phonological elements of the target words. However, the results suggested that Mandarin participants' L1 tonal status did not have an overall facilitative effect on lexical production of non-native tones. Mandarin participants did not significantly outperform English participants in the image-naming task, nor did they produce fewer tone-only errors than did English participants. Overall, both groups thus performed similarly in phono-lexical accuracy in the image-naming task. This finding suggests that linking tones to meaning in an L2 in the speaking modality may not necessarily be easier for tonal L1ers than for non-tonal L1ers in early stages of word learning. It is important to note that this conclusion is drawn from comparisons of *ab initio* tonal and non-tonal L1ers learning words in a novel language, unlike previous studies that only compared non-tonal (English) listeners to native Mandarin participants who performed lexical tasks in Mandarin (Ling & Grüter, 2020; Pelzl et al., 2019).

Although the image-naming task did not reveal a main effect of L1 tonal status on lexical production of non-native tones, there were clear effects of tone type, which resembled some of the results found in the imitation task. Namely, Mandarin participants were less likely to correctly name low-level tone words than English participants. A visual inspection of the distribution of error types (Figure 27, page 108) further suggested that Mandarin speakers appeared to predominantly mispronounce low-level as mid-level tones. This resonates with the findings of the imitation task in which low-level tones were produced with a relatively high pitch by Mandarin participants. This may indicate that tone type, both in the phonetic-acoustic way (i.e., a general difficulty with pitch height contrast), but also in the phonological-categorical way (i.e., level tone contrasts that may assimilate to the single Mandarin high-level tone) caused a relative difficulty with low-level tones for Mandarin speakers. Crucially, this effect was strong enough that Mandarin participants underperformed in comparison to English participants in lexical production of low-level tone words in the image-naming task. This suggests that L1 tone type not only interferes with pre-lexical tone processing, as has been shown widely in previous pre-lexical perception and production studies (Cooper and Wang, 2012; Hao, 2012; Qin and Jongman, 2016; So and Best, 2010; Wu et al., 2014), but that it continues to affect lexical processing of tones. This appears to be the case both in the listening modality (as shown by the word identification results in Chapter 2), but also in the speaking modality, as was shown recently by A. C. L. Yu et al. (2021) and now also here in the present study.

3.7.3 Effects of pitch aptitude on imitation and image-naming accuracy

The only effect of individual pitch perception aptitude on phonetic production accuracy in imitation was observed in the Mandarin group and for falling tones on day 2. This was the single condition in which individual pitch perception significantly predicted lower Fréchet scores, indicative of more target-like productions and better phonetic accuracy. Although the direction of this effect is interpretable, the fact that a positive effect of pitch aptitude on phonetic accuracy was only observed in one specific condition is puzzling and, does not confirm the predictions made for **RQ3** that pitch aptitude would facilitate imitation overall. Overall, the findings from the present study's imitation task thus yield limited evidence for the facilitative effect of individual pitch aptitude on phonetic accuracy in tone imitation. Although a more overall facilitative effect of pitch aptitude on tone imitation was predicted, following Dong et al. (2019) who employed the same measure of pitch aptitude and carried out a similar tone imitation task, one explanation for the lack of a clear facilitative effect pitch perception aptitude in the present study could be that perceptual skills are not necessarily strongly indicative of production skills, as proposed by the "skill-specificity hypothesis" (Li & Dekeyser, 2017). It is also noted however, that different measures of pitch processing skills, such as standardized musicality tests (Peretz et al., 2003; Wallentin et al., 2010) as well as different measures of production accuracy, such as assessment by raters as was done by Dong et al. (2019), may yield different results.

For image-naming, the prediction made for **RQ3** regarding the effect of pitch aptitude was borne out by the data. I observed a clear effect of individual pitch aptitude on phono-lexical production, as participants with higher pitch aptitude tended to be better at correctly linking tones to meaning in production. This chimes in with earlier findings from perception studies that suggest a continuity between pre-lexical and lexical pitch processing (Ling & Grüter, 2020; Wong & Perrachione, 2007). The reason why the effect of individual aptitude was clearer in the lexical image-naming task than in the pre-lexical imitation task may have to do with the cognitive load of the image-naming task. That is, individuals who are generally good at identifying pitch categories in a pre-lexical setting may benefit from those skills to facilitate the relatively demanding challenge of associating tone to meaning in the lexical image-naming task. The fact that linking tone to meaning was indeed challenging was shown by the high proportion of tone-only errors on day 2 of the image-naming task, indicating that participants had retained and were able to produce the segmental properties of the words, but not the tonal properties, which may have constituted the final hurdle to word learning.

Some limitations to the current study must be acknowledged. First, the use of pseudoword stimuli may limit its applicability to real-life tone learning. However, the pseudoword stimuli allowed me to make a direct comparison between a group of tonal and non-tonal speakers to investigate the effect of L1 tonal status and specific tone types (contrasting in contour and in level) on *ab initio* production in a tone system that is unknown to either group, unlike many previous production studies in which participants had prior knowledge of the target language.

Second, alternative methods of defining phonetic tone accuracy could have been used, such as measures of a pitch curve's slope and curvature (see the recent work by A. C. L. Yu et al. (2021) for such measures). The Fréchet distances obtained in the present studied served as an overarching proxy to indicate whether a production was target-like or not, but they are limited in revealing *why* a production was target-like or not. However, the Fréchet distances did allow me to reveal overall significant between-group differences, which could then further be addressed by follow-up analyses, as was done in the analysis of relative pitch height in low-level tone imitations by English and Mandarin speakers.

Finally, the fact that participants in the present study engaged in both pre-lexical and lexical tasks in the listening modality (as presented in Chapter 2) as well as in the speaking modality (as presented here), raises the question as to whether performance in one modality affected performance in the other. Chapter 4 deals with this perception-production link in detail, and also places the findings in the framework of theoretical models of speech perception and production.

3.8 Conclusion

The present study investigated the effects of L1 tonal status, tone type, and individual pitch perception aptitude on phonetic and phono-lexical tone production accuracy in an imitation and an image-naming task. The aim was to get a better insight into the factors that affect tone learning not only in the listening modality in perception, but also in the spoken modality in production. Results from an imitation and image-naming task revealed no clear effect of L1 tonal status, as Mandarin participants did not overall outperform English participants in tone production at either level of processing. Instead, tone production accuracy in both the pre-lexical and lexical tasks was mostly guided by the specific tone types, which were in turn produced with various degrees of accuracy depending on participants' L1. In particular, Mandarin-L1 participants appeared to struggle with level tone contrasts in the present study's pseudowords, both in immediate phonetic-acoustic imitation as well as more phono-lexical image-naming. English participants on the other hand, appeared to be less influenced by tone type in both the imitation and image-naming tasks. Individual pitch aptitude was not strongly associated with tone imitation, but more strongly with image-naming. Unlike many previous production studies which investigated tone production by comparing performance by non-tonal L1ers to native speakers of the target tone language (Ding et al., 2011; Kirby & Giang, 2021; Y. Wang et al., 2003) without considering individual extralinguistic factors, this study provides a more neutral and controlled setting to investigate the factors that facilitate early-stage non-native tone production because both non-tonal and tonal L1ers were exposed to non-native pseudowords; participants were matched for their musical experience and working memory capacity; and because the effect of individual pitch aptitude was taken into account.

3.9 Appendix to Chapter 3

Appendix 3.1

Imitation: Mixed ANOVA table for phonetic tone accuracy results (Type-III Wald Chisquare tests).

IMITATION			
lmer(Fréchet ~ L1*Tone*Aptitude*Day + (Day + Tone | Subject) + (1|Item))			
Effect	χ^2	df	p
L1	0.43	1	0.512
Tone	103.59	3	0.000
Aptitude	0.08	1	0.768
Day	2.79	1	0.279
L1:Tone	3.83	3	0.015
L1:Aptitude	0.28	1	0.594
Tone:Aptitude	3.87	3	0.276
L1:Day	0.00	1	0.950
Tone:Day	29.00	3	0.000
Aptitude:Day	0.33	1	0.565
L1:Tone:Aptitude	1.65	3	0.647
L1:Tone:Day	17.34	3	0.001
L1:Aptitude:Day	2.66	1	0.103
Tone:Aptitude:Day	43.20	3	0.000
L1:Tone:Aptitude:Day	50.16	3	0.000

Appendix 3.2

Imitation: Significant multiple comparisons between tones per L1.

IMITATION					
	Contrast	Estimate	std. Error	t	p
Day 1					
English	Rise-Fall	-0.39	0.12	-3.18	0.015
	Rise-Mid	0.78	0.14	5.60	0.001
	Rise-Low	0.69	0.13	5.30	0.001
	Fall-Mid	1.16	0.14	8.19	0.001
	Fall-Low	1.08	0.13	8.30	0.001
Mandarin	Rise-Mid	0.95	0.16	6.05	0.001
	Fall-Mid	1.19	0.16	7.36	0.001
	Fall-Low	0.60	0.15	4.08	0.001
	Mid-Low	-0.59	0.15	-3.84	0.002
Day 2					
English	Rise-Mid	0.38	0.14	2.76	0.047
	Fall-Mid	0.65	0.14	4.61	0.000
	Fall-Low	0.60	0.13	4.67	0.000
Mandarin	Rise-Fall	-0.52	0.14	-3.76	0.002
	Rise-Mid	0.49	0.16	3.13	0.016
	Fall-Mid	1.02	0.16	6.28	0.001
	Fall-Low	0.84	0.15	5.65	0.001

Appendix 3.3

Imitation: Significant multiple comparisons between L1s per tone.

Day/Tone	Contrast	Estimate	std. Error	t	p
Day 1					
Low	English-Mandarin	-0.42	0.18	-2.31	0.025*
Day 2					
	(No sig. comparisons)				

Appendix 3.4

Imitation: Estimates of aptitude per L1 and per tone.

IMITATION FOUR-WAY INTERACTION						
Predictors		Estimate	std. Error	t	p	95% C.I.
Day 1						
English	Rise	-0.03	0.09	-0.37	0.714	[-0.22 ; 0.15]
	Fall	-0.08	0.08	-0.26	0.080	[-0.23 ; 0.08]
	Mid	0.04	0.09	0.41	0.684	[-0.14 ; 0.21]
	Low	0.07	0.09	0.73	0.470	[-0.11 ; 0.25]
Mandarin	Rise	0.16	0.27	0.58	0.559	[-0.39 ; 0.72]
	Fall	0.18	0.25	-0.33	0.694	[-0.33 ; 0.69]
	Mid	-0.03	0.26	-0.13	0.901	[-0.55 ; 0.49]
	Low	-0.42	0.27	-1.58	0.123	[-0.95 ; 0.11]
Day 2						
English	Rise	-0.02	0.09	-0.25	0.806	[-0.21 ; 0.17]
	Fall	0.06	0.09	0.64	0.525	[-0.11 ; 0.22]
	Mid	0.13	0.10	1.39	0.171	[-0.05 ; 0.32]
	Low	0.07	0.10	0.63	0.531	[-0.95 ; 0.11]
Mandarin	Rise	-0.47	0.28	-1.71	0.094	[-1.03 ; 0.08]
	Fall	-0.69	0.25	-2.72	0.009	[-1.19 ; -0.17]
	Mid	0.37	0.28	1.35	0.183	[-0.18 ; 0.93]
	Low	0.16	0.30	0.53	0.597	[-0.44 ; 0.76]

Appendix 3.5

Image-naming: Mixed model ANOVA table for tone accuracy results (Type III Wald Chisquare tests).

IMAGE-NAMING (DAY 2)			
glmer(correct ~ L1*Tone*Aptitude + (Day + Tone | Subject) + (1|Item))			
Effect	χ^2	df	p
L1	1.18	1	0.276
Tone	10.62	3	0.013
Aptitude	11.83	1	0.001
L1:Tone	18.19	3	0.001
L1:Aptitude	0.77	1	0.377
Tone:Aptitude	2.16	3	0.538
L1:Tone:Aptitude	1.53	3	0.673

Appendix 3.6

Image-naming: Significant multiple comparisons for tone.

IMAGE-NAMING (DAY 2)				
Contrast	Estimate	std. Error	t	p
Eng-Man | Low	1.76	0.60	2.91	0.003
English				
(No sig. comparisons)	-	-	-	-
Mandarin				
Rise-Fall	-1.61	0.58	-2.75	0.034
Fall-Low	2.89	0.61	4.73	0.000
Mid-Low	1.44	0.53	2.72	0.039

Chapter 4 The perception-production link in non-native tone learning

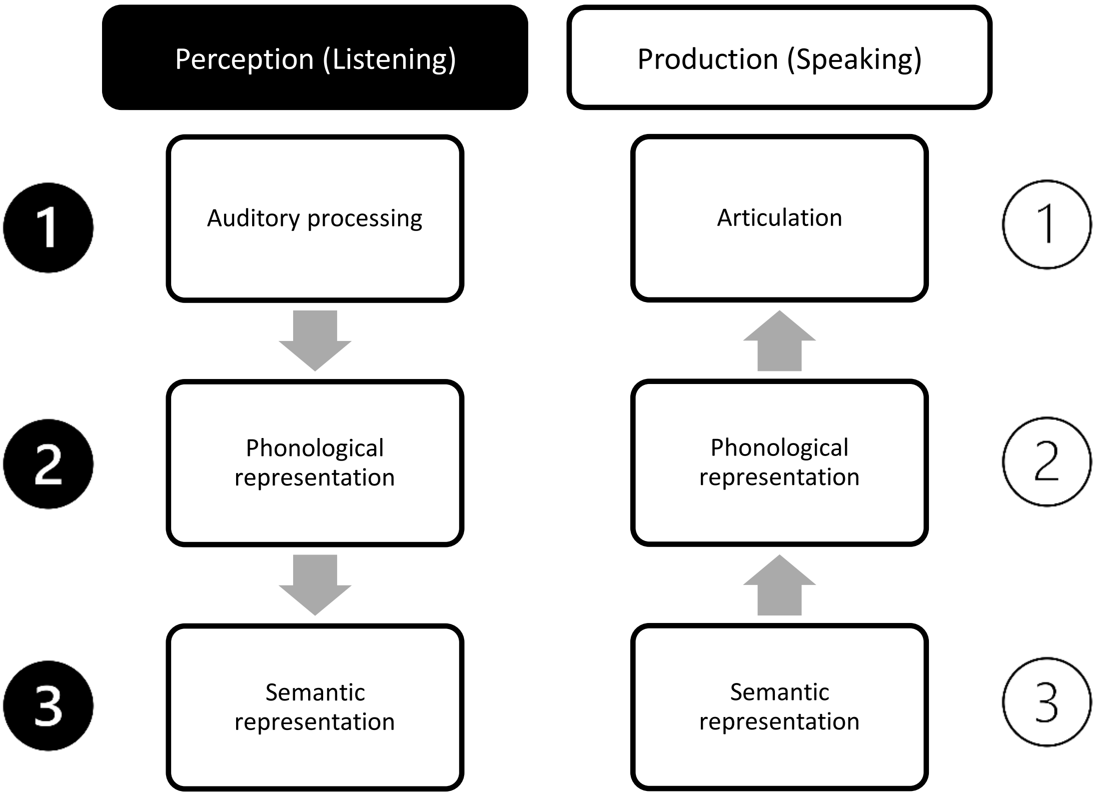
Although it is commonly agreed that perception and production in second language speech are closely intertwined, performance in one modality does not always mirror performance in the other. This Chapter presents new evidence for the perception-production link by looking at a relatively understudied feature of non-native speech, namely lexical tone. It presents a simultaneous investigation of performance, improvement over time, and error patterns in a tone categorization and word identification task (perception) and in a tone imitation and image-naming task (production) by English-L1 and Mandarin-L1 speakers to study non-native tone processing in the listening and speaking modalities. In addition, this Chapter compared the effect of extralinguistic factors (musical experience and working memory) on performance at different levels of processing and in the two modalities. This Chapter finds evidence suggesting that the perception-production link is relatively weak for phonetic and phonological processing of tones in pre-lexical tasks, but that once learners are required to link tones to meaning at higher levels in lexical tasks, perception and production mirror each other in remarkably similar ways.

4.1 Background

Human speech communication requires perception (listening) and production (speaking). Although perception and production have traditionally been studied separately, and have been described in separate psycholinguistic models, such as the TRACE model of speech perception (McClelland & Elman, 1986) and Bock and Levelt's model of speech production (Bock & Levelt, 1994), recent literature has started to study both speech modalities simultaneously. The motivation for the simultaneous investigation of perception and production is rooted in the observation that, essentially, the two modalities involve similar processes which simply run in opposite directions (Baese-Berk, 2019; Flege & Bohn, 2021, p. 12; Schmitz et al., 2018, p. 529). In one view, perception can be described as a process that starts with 1) auditory processing of sounds, followed by 2) the mapping of these sounds onto phonetic and phonological representations, and 3) the mapping onto lexical and semantic representation. In the same view, production can be described as "being the nearly same process in reverse" (Baese-Berk, 2019, p. 981). Figure 29 visualizes these processes. Although the processes described here are generalizations of the true nature of perception and production, the apparent similarity between the two raises the question to what extent an individual's perception is indicative of their production, and vice versa.

Figure 29

Visual representation of word perception and production, adapted from Baese-Berk (2019).



It is important to first clarify the exact definitions of perception and production, as both may refer to either pre-lexical or lexical processes, which involve different mechanisms and which are measured by different linguistic tasks. As I outlined earlier in Chapter 1 (section 1.3.1, page 7), pre-lexical perception and production involve processing devoid of lexical meaning. Pre-lexical perception can be described as a process in which a listener perceives a sound sequence without activating any semantic representation (i.e., word meaning). An example of a pre-lexical perception task is a categorization task, in which a listener hears an acoustic signal, (e.g., the sound sequence [da], [ta], or [t^{h}a]) and is asked to categorize the sound by indicating whether they perceived [d], [t], or [t^{h}]. On the other hand, pre-lexical production can be described as a process in which a speaker converts phonetic and phonological representations to acoustic realizations using the speech organs, again without activating any word meaning. An example of a pre-lexical production task is an imitation task or a read-aloud task of spoken or written words that are not associated to semantic representations, for instance if the meanings of the words are unknown. Broadly speaking, these pre-lexical processes correspond to steps 1 and 2[14] of the processes described in Figure 29. As will be described below, most studies target pre-lexical perception and pre-lexical production, and for ease of reading I will use the terms ‘perception’ and ‘production’ to refer in principle to these process that do not involve word meaning.

By contrast, I will use the term ‘lexical perception’ and ‘lexical production’ to refer to speech processes that do involve word meaning. These processes constitute the full mechanism of steps 1, 2, and 3 described in Figure 29. An example of a lexical perception task is a word identification task, which requires a listener to actively link sound to a semantic representation. An example of a lexical production task is an image-naming task, which requires a listener to access a semantic representation in the mind, and subsequently convert this to speech.

Despite the superficially similar processes of (lexical) perception and (lexical) production, it is not entirely clear whether an individual’s performance in one modality mirrors performance in the other. In the case of non-native speech, empirical studies that assess the perception-production link yield mixed results. Some show strong correlations between the two modalities, while others show only weak correlations or no correlation at all (See Schmitz et al. (2018, p. 529) for an overview). Importantly, most of the work on the perception-production link in non-native speech learning focuses on vowels and consonants. Much less studied are suprasegmental features of speech, such as lexical tone (Gut, 2009, p. 39). Therefore, this Chapter provides an account of the perception-production link in non-native tone learning. Specifically, it compares overall performance, improvement over time, error patterns, and facilitative factors on performance from the perception tasks reported in Chapter 2 to the production tasks reported in Chapter 3. Before describing the methodology and research questions, I will first review the literature on the perception-production link in

[14] It can be argued that some types of perception and production only constitute step 1 of the processes described in Figure 1. In perception, a distinction can be made between phonetic-acoustic perception (step 1) that requires listeners to only pay attention to fine-grained phonetic differences between stimuli, and phonological-categorical perception (step 2) that requires listeners to pay attention to categorical differences between stimuli (F. Chen & Peng, 2018). Some authors also suggest that production tasks such as imitation tasks do not necessarily require phonological encoding and only constitute phonetic processing, i.e., step 1 (Hao, 2012). For purposes of the present discussion, and as flagged in the General introduction (Chapter 1), I group together the phonetic and phonological processes and describe these collectively as ‘pre-lexical processing’ (i.e., steps 1 and 2), to emphasize the contrast with fully lexical processes (i.e., steps 1–3), but I will return the differences between steps 1 and 2 in the discussion of this Chapter (section 4.6).

non-native tone learning.

A theoretical framework that describes the perception-production link in non-native speech is Flege's Speech Learning Model (SLM; Flege 1995), which was recently revised to "SLM-r" (Flege & Bohn, 2021). As a model concerned with the ultimate attainment of non-native pronunciation, it postulates that the ease with which non-native sounds are perceived depends on the degree of similarity between native and non-native sounds. This principle is akin to that of the Perceptual Assimilation Model (PAM; Best 1995; Best & Tyler 2007), which predicts that non-native sound units that map onto separate native language units in a one-to-one fashion (two category assimilation) are easy to perceive, and units that map onto a single native language unit in a two-to-one fashion (single category assimilation) are difficult to perceive. Empirical studies on non-native tone perception tend to find support for PAM's predictions, particularly in perception by listeners whose L1 is tonal (J. Chen et al., 2020; Hao, 2012; So & Best, 2014). For instance, in Chapter 2, Mandarin listeners had difficulty in accurately perceiving pseudoword mid-level and low-level tone types, which do not exist as separate categories in Mandarin and which are hypothesized to assimilate in a two-to-one fashion to the Mandarin high-level tone type. PAM, however, is a model uniquely designed for non-native speech perception, and therefore cannot be applied to non-native tone production. Flege's Speech Learning Model forms a bridge between the two modalities. Flege specifies that in perception, non-native sounds may be stored as "identical", "similar", or "new" phonetic categories (Flege, 1987), either by undergoing categorical assimilation as predicted by PAM, by going through an L1 filter that discards phonetic or phonological properties that are irrelevant in the L1, or both (Flege, 1995, p. 238). Crucially, the SLM further postulates that this perceptual reorganization of non-native sounds determines the nature of non-native production, although motoric output constraints may further modulate non-native pronunciation. The SLM further factors in the effects of L2 exposure and age, as it is designed to model speech development over a speaker's lifetime. Since this Chapter considers early-stage tone learning in naïve adult learners, it will not consider the SLM predictions vis-à-vis the effects of long-term L2 exposure and usage over time. I will instead focus on the SLM's major prediction that non-native production performance converges with perception performance, or –according to the revised SLM-r – that they "coevolve" (Flege & Bohn, 2021, pp. 28–29). The following paragraphs provide a summary of empirical studies that simultaneously assessed non-native tone perception and production to assess the degree to which this theoretical prediction has been borne out by empirical data.

Evidence in support of a strong link between non-native tone perception and production is found in a study by Ding et al. (2011). They assessed perception and production of Mandarin tones by a group of German-L1 intermediate learners and found that accuracy in tone perception (in a tone categorization task) and tone production (in a read-aloud task with tone marks) were highly correlated. Furthermore, the type of errors that participants made in perception appeared to be the same in production (although this observation was not tested statistically). However, the authors note that some articulatory difficulties weakened the correlation between perception and production, as German speakers "displayed different pitch contours and pitch heights" than native Mandarin speakers (Ding et al., 2011, p. 515), which in some cases led to tones being incorrectly produced even if they were correctly perceived. These findings appear to fit neatly with the SLM prediction that perception and production mirror one another, but that not all production errors are perceptually motivated (Flege, 1995, p. 238).

The findings from Ding et al. (2011) appear to coincide with earlier data from Y. Wang et al. (2003, p. 1031), who found "strikingly similar patterns" in error types in perception (in a tone categorization task) and production (in a read-aloud task with tone marks) by English learners of Mandarin. However, they highlight that some error types differed in the two modalities in terms of the direction of confusion. Specifically, rise-to-dip tone errors were more common than dip-to-rise errors in perception, but dip-to-rise errors were more common than rise-to-dip errors in production. The authors suggest that this may be because the Mandarin dipping tone was inherently difficult to articulate for English learners and highlight that "not all aspects of perceptual learning can be incorporated in production" (Y. Wang et al., 2003, p. 1041).

Other studies emphasize that the correlation between perception and production may depend on the task type and on participants' L1 tonal status. In a series of experiments of Mandarin tone perception and production by English and Cantonese speakers, Hao (2012) found that although perception performance (in a tone categorization task) and production performance (in a read-aloud task with tone marks) were "moderately correlated", correlation between tone categorization and tone *imitation* was "relatively weak" (p. 277). Hao found that participants performed relatively well in imitation, whereas accuracies in the identification and reading tasks were much lower. Hao suggests that the different levels of processing required by the tasks explain these differences. In particular, she suggests that an imitation task involves surface acoustic perception and production of tone contours, (i.e., step in 1 in the process described in Figure 29). By contrast, tone categorization and read-aloud tasks involve “meta-linguistic skills” as they require linking pitch to a tonal label (i.e., steps 1 and 2), and these processes may be inherently more challenging. As for differences between the English (non-tonal) and Cantonese (tonal) groups, Hao found no clear advantage for Cantonese learners, either in perception or in production. She notes that Cantonese learners had to “suppress pre-learned categories”, highlighting interference from L1 tone types as predicted by PAM. Furthermore, it appeared that when assessing the perception-production link per language group, perception and production were more strongly correlated for English learners than for Cantonese learners, who showed no or even negative correlations (that is, for some tones, better perception was associated with worse production). Hao concludes that this observation requires more investigation to be explained (Hao, 2012, pp. 276–277).

Yet other empirical data provide insight into how the perception-production link may change over time. K. Zhang & Peng (2017) showed that Mandarin learners’ perception (in a tone categorization task) and production (in a read-aloud task with tone marks) of Cantonese tones were significantly correlated with one another. This correlation became even stronger after a two-week tone training session. The authors suggest that the perception-production correlation may strengthen with an “increase in experience of tone processing” (K. Zhang & Peng, 2017, p. 1802). They also looked at individual improvement from pre-training to post-training, and showed that improvement in perception did not significantly predict improvement in production. In particular, many learners improved in perception, but fewer did so in production. This mirrors the SLM tenet that perception may precede production, but that not all non-native production originates in non-native perception.

In the study by Kirby & Giang (2021), production of Vietnamese tones (elicited by verbal prompts) by speakers of Khmer Krom was compared to their tone perception (in an AX discrimination task). Correlations between production accuracy (operationalized by the Fréchet distance, as a one-number summary of similarity to target contours) and discrimination accuracy revealed a weak correlation, but in the expected direction, as smaller

Fréchet distances (i.e., more target-like productions) correlated with higher discrimination accuracies. The authors also correlated individual participants' discrimination accuracy per tonal contrast with individual Fréchet distances between productions of that same contrast. Here too, a weak correlation was found, as better discrimination of a tonal contrast correlated with larger Fréchet distances between productions of that same contrast (i.e., more distinct productions within-subject). Overall, all participants attained near native-like performance in perception, whereas their performance in production varied, and the authors suggest that high performance in tone perception may not necessarily facilitate tone production. They highlight that even if speakers' tone productions are not native-like, they may still be acoustically distinct within each speaker's tone inventory, and this ability to constitute distinct tone categories in one's own production may be related to the ability to discriminate separate tone categories in perception (Kirby & Giang, 2021, pp. 262–263).

A similar analysis of within-speaker perception and production accuracy was carried out by A. C. L. Yu et al. (2021), who compared perception (measured by an AX discrimination task) to lexical production (measured by an image-naming task) of Cantonese tones by native speakers and L2 speakers with different dominant languages (Urdu, Punjabi, or English). Largely similar to the findings by Kirby & Giang (2021), they found that a greater acoustic distance between tones within an individual speaker's tone repertoire significantly predicted tone discrimination accuracy in the Urdu and Punjabi-dominant groups, however this relation was not found in English-dominant groups and Cantonese L1 speakers. The lack of a clear relation between perception and production distinctiveness in the Cantonese L1 speakers is attributed to the fact that the AX discrimination task may not have revealed sufficient individual variability within speakers (A. C. L. Yu et al., 2021, p. 21).

Finally, some studies highlight that beyond L1-specific factors (such as an individual's L1 tonal status, or the interaction between tone types in the L1 and the L2), certain extralinguistic factors may facilitate tone perception and production in similar ways. For instance, Li & Dekeyser (2017) showed that English learners' musical ability facilitated performance in both (lexical) tone perception as well as (lexical) production of Mandarin tone-words. Kirby & Giang (2021) also accounted for extralinguistic factors that may influence performance in Vietnamese tone perception by Khmer speakers, namely age, education level, and L2 Vietnamese usage, however none of these factors additionally predicted performance.

All in all, the available work on the perception-production link in non-native tone learning hints that performance in the two modalities may be correlated, as shown by studies that have separately investigated the perception-production link in terms of overall performance (Ding et al., 2011; Kirby & Giang, 2021; A. C. L. Yu et al., 2021), improvement over time (K. Zhang & Peng, 2017), and specific error types (Y. Wang et al., 2003). However, the strength of this correlation appears to depend on the type of task and the level of processing that that task taps into (Hao, 2012), a learner's L1 tonal status (Hao, 2012; A. C. L. Yu et al., 2021), and their degree of experience with non-native tones or the amount of training and exposure (K. Zhang & Peng, 2017). There are also indications that extralinguistic factors such as musicianship facilitate tone perception and production (M. Li & Dekeyser, 2017), but whether these factors are equally facilitative to tone processing in both modalities is rather unclear.

4.2 Research aim and questions

Bringing together aspects of non-native tone learning that were studied separately in the abovementioned studies, the present study's aim is to provide a comprehensive account of the perception-production link in tone learning in terms of general performance, improvement over time, and error types. It also explores whether extralinguistic factors (musicianship and working memory, here) facilitate tone perception and production in similar ways.

To explore the perception-production link in non-native tone learning, this Chapter compares the pre-lexical "tone categorization" and the lexical "word identification" perception tasks in Chapter 2 with the pre-lexical "imitation" and lexical "image-naming" production tasks in Chapter 3. For ease of reading, I will occasionally refer to these specific tasks using more general terms, as introduced before in the General introduction (Figure 2, page 5), and as shown again here Table 10.

Table 10

Naming convention of tasks used in this Chapter.

Names in Chapters 2–3	Also referred to as
Tone categorization	Pre-lexical perception
Word identification	Lexical perception
Imitation	Pre-lexical production
Image-naming	Lexical production

The following research questions are addressed:

RQ1: *How are L2 perception and production of tones at the pre-lexical level correlated in terms of performance, improvement over time, and error patterns?*

RQ2: *How are L2 perception and production of tones at the lexical level correlated in terms of performance, improvement over time, and error patterns?*

RQ3: *How do L1-specific and extralinguistic factors facilitate L2 (lexical) tone perception and production?*

4.3 Methods

4.3.1 Participants

Participants were the same as in Chapters 2 and 3 (section 2.3.1, page 39).

4.3.2 Stimuli

Stimuli were the same as in Chapters 2 and 3 (section 2.3.2, page 40).

4.3.3 Procedure

Procedures were the same as in Chapters 2 and 3 (section 2.3.3, page 45, and section 3.4.3, page 91, respectively).

4.3.4 Statistical procedures

All analyses were performed in *R* (R Core Team, 2021). Figures were generated with the *ggplot2* package (Wickham, 2016). To investigate the perception-production link in terms of overall performance, Pearson bivariate correlation analyses were conducted, following earlier perception-production studies (Hao, 2012; Schmitz et al., 2018; Y. Wang et al., 2003; K. Zhang et al., 2018). Correlations were conducted on the performance measures in each of the four tasks (accuracy and log RT in tone categorization; accuracy in word identification; Fréchet distance in tone imitation; accuracy in image-naming). Terminology by Evans (1996) is used to describe correlation strength[15]. Correlation analyses are deemed to be a suitable tool for assessing the perception-production link in terms of overall performance since they assume no directionality on whether perception facilitates production, or vice versa. It is acknowledged, however, that correlations do not imply causation. Correlations were carried out per L1 on both overall performance as well as on performance per tone, similar to the approach by Hao (2012).

To compare the perception-production link in terms of improvement over time, (generalized) linear mixed-effects models were fitted on the data of day 1 and 2 of the word identification, imitation, and image-naming tasks. (Improvement over time for the tone categorization task could not be assessed because it was only carried out on day 1). For each task, the dependent variables (Fréchet distance in the imitation task, and correct/incorrect for the word identification and image-naming tasks) were subjected to (generalized) linear mixed-effects models with fixed effects for *L1, Tone,* and *Day,* and three-way and two-way interactions between these effects. The random effects structure consisted of a random intercept for *Item,* and by-subject slopes for *Tone*. Where applicable, post-hoc Bonferroni-corrected comparisons using the *emmeans* package were conducted to investigate significant interactions in more detail. Statistical details are reported throughout the text. Model output (main effects and interactions) and relevant multiple comparisons are provided in the

[15] Where <0.20 is “very weak”, 0.20~0.39 is “weak”, “0.40~0.59” is “moderate”, 0.60~0.79 is “strong”, and >0.80 is “very strong”.

appendix (section 4.8).

To assess the perception-production link in terms of error types, the results from the error patterns reported in Chapters 2 and 3 were compared.

Finally, to investigate the perception-production link in terms of facilitative extralinguistic factors, (generalized) linear mixed effects models with *Musical Experience* (a continuous variable expressing years of formal practice, centered and scaled) and *Working Memory* (a continuous variable expressing performance in the backwards digit span task reported in Chapter 2, centered and scaled) as fixed effects were fitted to the subset data of each task per participant group. Each model contained a random intercept for *Subject* and *Item* (random-slope models did not always converge). The choice was made to fit identical models to different data subsets, as has been done in previous studies on non-native tone processing across different tasks (Dong et al., 2019) and participant groups (Wiener & Lee, 2020). This was motivated by the fact that models fitted on the entire dataset for each task would require relatively complex structures (including four-way interactions between L1, tone, day, and working memory and musical experience) that in some cases led to convergence issues. In addition, fitting identical models to different data subsets made it possible to address more directly whether musical experience and working memory facilitate non-native tone processing in different ways across levels of processing (pre-lexical or lexical) and across different modalities (perception and production).

4.4 Predictions

Based on the literature reviewed in section 4.1, the following predictions can be made with regard to this study’s research questions:

P1: It is predicted that tone perception and production at a pre-lexical level are moderately correlated. Following the tenets of the Speech Learning Model, certain difficulties in tone production may not be perceptually based and may instead have their origins in inherent articulatory difficulties.

P2: It is predicted that tone perception and production at a lexical level is strongly correlated, especially under the SLM’s assumption that lexical processing of tones (either in perception or production) constitutes later stages of tone learning, and that therefore perceptual and productive performance are expected to converge.

P3: It is predicted that extralinguistic factors (musical experience and working memory) will affect tone perception and production in similar ways.

4.5 Results

The following sections address the perception-production link in performance, improvement over time, and error patterns, and address this link separately for the pre-lexical and lexical levels.

4.5.1 The perception-production link in performance

4.5.1.1 The perception-production link at a pre-lexical level

Correlations in terms of performance are summarized in Table 11. Within the English group, there was no clear evidence for a perception-production link at a pre-lexical level. All correlations between performance in tone categorization (both in terms of log RT and accuracy) and performance in imitation (in terms of Fréchet distance) were insignificant.

Within the Mandarin group, there was limited evidence for a perception-production link at a pre-lexical level. Specifically, there was a moderate correlation between low-level tone categorization accuracy and low-level Fréchet distance on day 1 of the imitation task, and a moderate correlation between falling tone categorization accuracy and falling tone Fréchet distance on day 2 of the imitation task. The sign of the correlation indicates that higher tone categorization accuracy scores were associated with shorter Fréchet distances, indicating more target-like productions.

Table 11

Pearson correlations for performance between perception and production tasks.

	Pearson correlation r	All tones	Rising	Falling	Mid-level	Low-level
English	Pre-lexical perception (accy) - Pre-lexical production (day 1)	.017	.092	-.042	.077	-.156
	Pre-lexical perception (accy) - Pre-lexical production (day 2)	.166	-.096	.212	.204	-.063
	Pre-lexical perception (accy) - Lexical production (day 1)	**.451***	.031	.224	.427(.)	.294
	Pre-lexical perception (accy) - Lexical production (day 2)	**.670*****	.170	.226	**.524***	**.603****
	Pre-lexical perception (log RT) - Pre-lexical production (day 1)	-.134	-.053	-.026	-.222	-.045
	Pre-lexical perception (log RT) - Pre-lexical production (day 2)	-.194	.016	-.303	-.291	.013
	Pre-lexical perception (log RT) - Lexical production (day 1)	**-.517***	-.172	-.372.	**-.513***	-.406(.)
	Pre-lexical perception (log RT) -Lexical production (day 2)	**-.553****	-.415(.)	-.275	-.349	**-.592****
	Lexical perception (day 1) - Lexical production (day 1)	**.812*****	**.555****	**.643****	**.603****	**.593****
	Lexical perception (day 2) - Lexical production (day 2)	**.926*****	**.820*****	**.614****	**.828*****	**.866*****
	Pre-lexical production (day 1) - Lexical perception (day 1)	-.162	-.060	-.087	-.129	-.140
	Pre-lexical production (day 2) - Lexical perception (day 2)	-.127	-.169	.060	-.043	-.097
Mandarin	Pre-lexical perception (accy) - Pre-lexical production (day 1)	-.002	.033	.016	.018	**-.593****
	Pre-lexical perception (accy) - Pre-lexical production (day 2)	-.237	.112	**-.592****	.287	-.217
	Pre-lexical perception (accy) - Lexical production (day 1)	.150	**-.471***	.170	.089	.265
	Pre-lexical perception (accy) - Lexical production (day 2)	**.537***	-.212	.105	.345	.425(.)
	Pre-lexical perception (log RT) - Pre-lexical production (day 1)	.164	-.035	.250	.269	.203
	Pre-lexical perception (log RT) - Pre-lexical production (day 2)	.272	.212	.442.	-.067	.303
	Pre-lexical perception (log RT) - Lexical production (day 1)	.131	.038	.176	.225	-.116
	Pre-lexical perception (log RT) -Lexical production (day 2)	-.231	-.188	-.060	-.121	-.132
	Lexical perception (day 1) - Lexical production (day 1)	**.583***	.324	**.474***	.267	.285
	Lexical perception (day 2) - Lexical production (day 2)	**.778*****	**.868*****	**.465***	.428(.)	**.700*****
	Pre-lexical production (day 1) - Lexical perception (day 1)	.307	.050	.400(.)	.347	-.026
	Pre-lexical production (day 2) - Lexical perception (day 2)	-.030	-.150	.108	.172	.047

*Signif. codes 0 '***' 0.001 '**' <0.05 '*' =0.05 '(.)'*

4.5.1.2 The perception-production-link at a lexical level

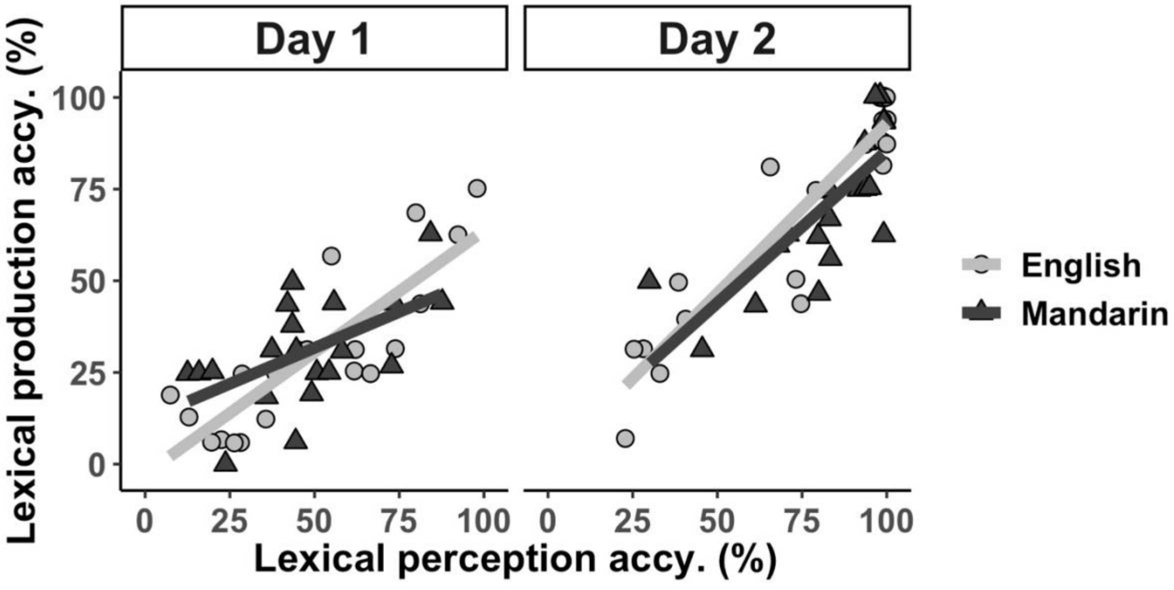
Within the English group, there was evidence for a strong perception-production link at the lexical level. Word identification accuracy was very strongly correlated with performance in image-naming accuracy, on both days 1 and day 2. The correlations per tone further revealed significant correlations (all moderate or stronger) for all conditions.

Within the Mandarin group, there was evidence for a moderate to strong perception-production link at the lexical level. Word identification accuracy was moderately correlated with image-naming accuracy on day 1, and strongly on day 2. The correlations per tone revealed a moderate correlation for performance in falling tones on day 1, and significant correlations (all moderate or stronger) for all tones except mid-level tones, although the correlation was marginally significant and moderate.

Overall, the positive sign of all the correlations indicates that higher accuracy scores in word identification were associated with higher accuracy scores in image-naming. This positive relation is visualized in Figure 30.

Figure 30

Lexical production (image-naming) against lexical perception (word identification) accuracy.



4.5.1.3 The perception-production link across pre-lexical and lexical levels

This section highlights significant correlations between perception and production *across* pre-lexical and lexical levels of processing.

Within the English group, there was evidence for a moderate link between pre-lexical tone perception and lexical tone production. There was a moderate correlation between overall tone categorization accuracy and image-naming on day 1, and a strong correlation on day 2. Tone categorization log RT was also moderately correlated with image-naming accuracy on days 1 and 2.

Within the Mandarin group, there was also evidence for a moderate link between pre-lexical tone perception and lexical tone production, but this was limited to specific conditions. There was a moderate correlation between tone categorization accuracy and image-naming accuracy on day 1 but only for rising tones. There was a moderate correlation between tone categorization accuracy and overall image-naming accuracy on day 2. Tone categorization log RT was not significantly correlated with image-naming.

Overall, the signs of the correlations suggest that higher tone categorization accuracy and faster tone categorization log RTs were associated with higher image-naming accuracy.

Finally, there was no evidence for a link between pre-lexical production and lexical perception. There were no significant correlations between imitation accuracy (in terms of Fréchet distance) and word identification accuracy.

4.5.1.4 The link between pre-lexical and lexical performance in each modality

Table 12 reports correlation coefficients that assess the link between pre-lexical and lexical performance *within each modality*. For perception tasks, English participants' performance in the tone categorization task (both in terms of tone categorization accuracy and log RTs) was significantly correlated with performance in the word identification, both on days 1 and 2, and all correlations were moderate or stronger. For the production tasks, there were no significant correlations between English participants' performance in the imitation task and the image-naming task.

For the Mandarin participants, there was a strong correlation between tone categorization accuracy and word identification, but only for low-level tones on day 2. For the production tasks, there was a moderate correlation between imitation and image-naming of falling tones on day 1, and a moderate correlation between imitation and image-naming of mid-level tones on day 2. Counterintuitively, the sign of these correlations suggests that larger Fréchet distances (i.e., less target-like imitations) were correlated with more accurate image-naming.

Table 12

Pearson correlations for performance within perception and production tasks.

	Pearson correlation r	All tones	Rising	Falling	Mid-level	Low-level
English	Pre-lexical perception (accy) - Lexical perception (day 1)	**.599****	**.514***	**.596****	.273	**.466***
	Pre-lexical perception (accy) - Lexical perception (day 2)	**.781*****	**.620****	**.614****	**.712*****	**.691*****
	Pre-lexical perception (log RT) - Lexical perception (day 1)	**-.606****	**-.611****	**-.547***	-.406.	**-.519***
	Pre-lexical perception (log RT) -Lexical perception (day 2)	**-.665****	**-.659****	**-.704*****	**-.553****	**-.563****
	Pre-lexical production (day 1) - Lexical production (day 1)	-.330	-.261	-.061	-.060	-.280
	Pre-lexical production (day 2) - Lexical production (day 2)	-.211	-.153	-.097	-.137	.002
Mandarin	Pre-lexical perception (accy) - Lexical perception (day 1)	.208	-.134	.061	.196	.077
	Pre-lexical perception (accy) - Lexical perception (day 2)	.263	-.157	-.008	-.017	**.619****
	Pre-lexical perception (log RT) - Lexical perception (day 1)	-.087	-.188	-.03	-.067	.081
	Pre-lexical perception (log RT) -Lexical perception (day 2)	-.108	-.195	-.062	.042	-.076
	Pre-lexical production (day 1) - Lexical production (day 1)	.090	.187	**.471***	.255	-.193
	Pre-lexical production (day 2) - Lexical production (day 2)	-.218	-.177	-.139	**.556***	.012

*Signif. codes 0 '***' 0.001 '**' <0.05 '*'*

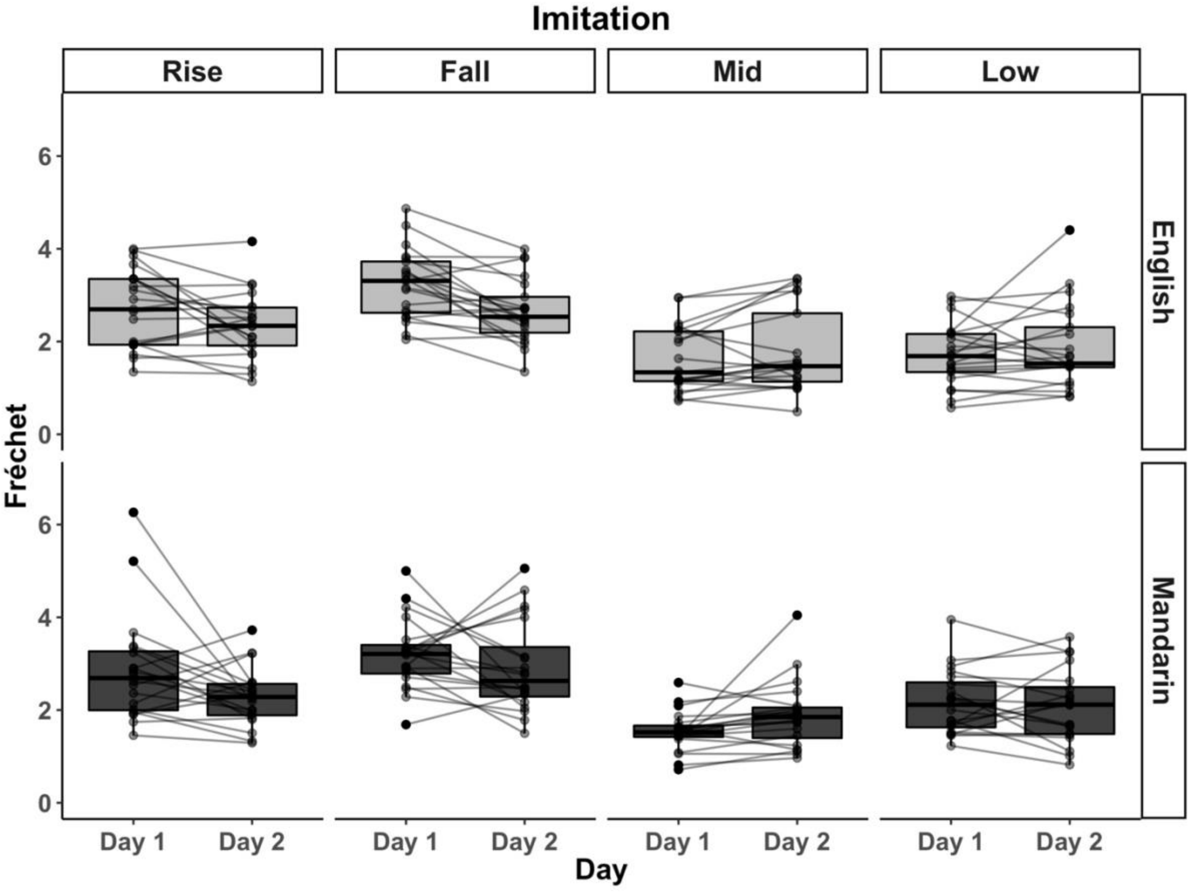
4.5.2 The perception-production link in improvement

4.5.2.1 Improvement in imitation

The model for the imitation task revealed a significant *L1:Tone:Day* interaction (Appendix 4.1). Multiple pairwise comparisons between days per tone and group (Appendix 4.2) revealed that imitation accuracy improved in all conditions (i.e., there were statistically significant differences between day 2 and day 1 Fréchet distances), except in the following conditions: For the English participants, there was no evidence that imitations improved for mid-level (*b* = 0.15, *SE* = 0.08, *p* = 0.058) or low-level tones (*b* = 0.13, *SE* = 0.08, *p* = 0.087). For the Mandarin participants, there was no evidence that imitation accuracy improved for low-level tones (*b* = -0.05, *SE* = 0.08, *p* = 0.459). In addition, there was evidence that imitation accuracy worsened from day 1 to day 2 for mid-level tones, as Fréchet distances significantly increased (*b* = 0.30, *SE* = 0.08, *p* = 0.007). For visualization, pre-lexical production performance in the imitation task across the two days is shown in Figure 31.

Figure 31

Improvement over session in the imitation task (pre-lexical tone production).



Dots represent mean Fréchet distance per subject.

4.5.2.2 Improvement in word identification

The model for the word identification task revealed significant main interactions between *L1:Tone, L1:Day* and *Tone:Day* (Appendix 4.3). Multiple comparisons following the *L1:Tone* interaction revealed that when averaging across the two days, Mandarin speakers were less likely to identify words with low-level tones compared to words with falling tones (*b* = -1.54, SE = 0.38, *p* < 0.001). No statistically significant differences between groups per tone were found. Note that these are results averaged over the two days: Chapter 2 showed that on day 2, Mandarin speakers were less likely than English speakers to identify words with low-level tones.

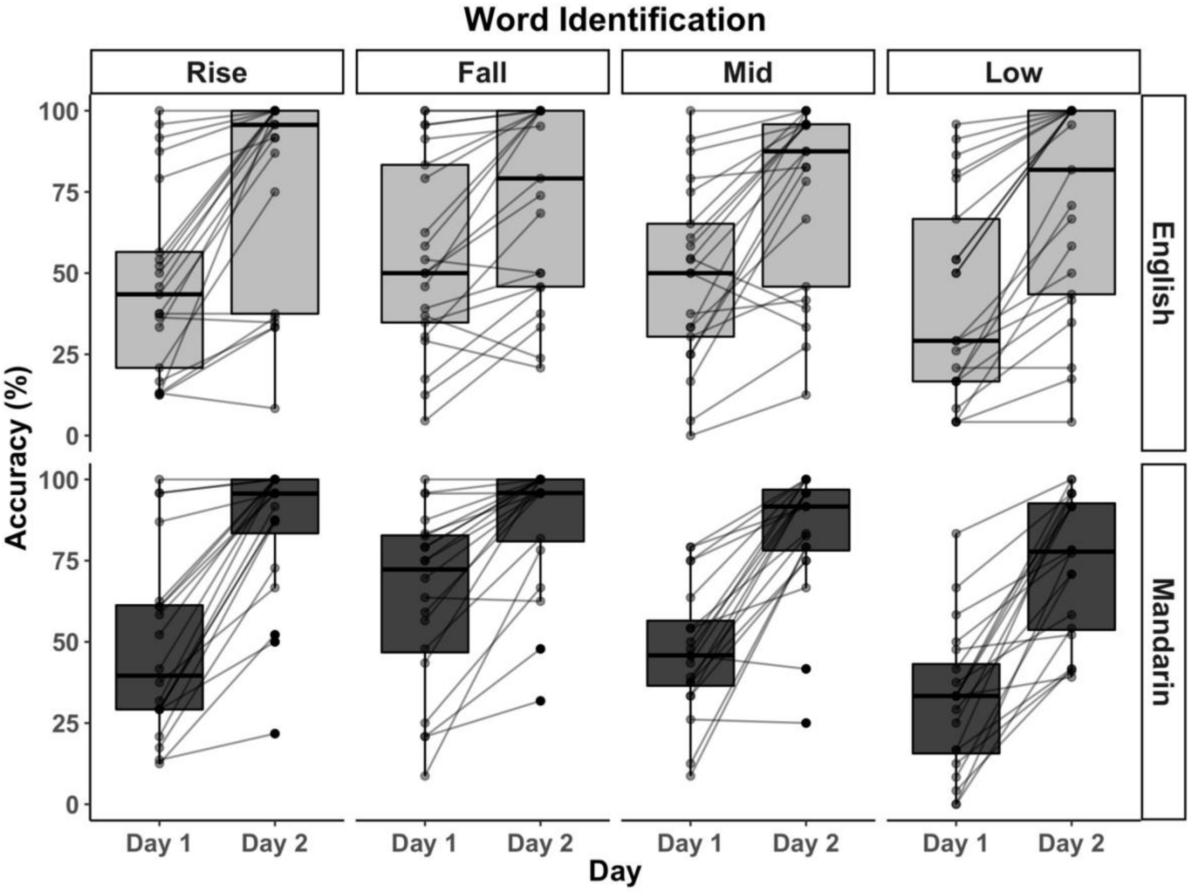
Multiple comparisons following the *L1:Day* interaction (Appendix 4.4) revealed a significant difference between word identification likelihood on day 2 as opposed to on day 1 for the English (*b* = 1.74, *SE* = 0.09, *p* < 0.001) and the Mandarin group (*b* = 2.24, SE = 0.09, *p* < 0.001). It is possible that the *L1:Day* interaction emerged as significant because of the difference in estimate size between the English and Mandarin group (respectively 1.74 versus 2.24), suggesting that the likelihood for correct word identification increased more strongly in the Mandarin group relative to the English group.

As for the *Tone:Day* interaction (Appendix 4.4), there was a significant difference between day 2 and day 1 word identification likelihood for rising tones (*b* = 2.36, *SE* = 0.14, *p* < 0.001), falling tones (*b* = 1.55, *SE* = 0.13, *p* < 0.001), mid-level tones (*b* = 1.13, *SE* = 0.12, *p* < 0.001), and low-level tones (*b* = 2.20, *SE* = 0.09, *p* < 0.001). The origin of the *Tone:Day* interaction appears to be in multiple comparisons between tones per day averaged across the two groups: on day 1, low-level tones were significantly less likely to be correctly identified than falling tone words (*b* = -1.39, *SE* = 0.36, *p* < 0.001), whereas no significant difference was found on day 2 (*b* = -0.73, *SE* = 0.37, *p* = 0.295).

For visualization, lexical perception performance in the word identification task across the two days is shown in Figure 32.

Figure 32

Improvement over session in the word identification task (lexical tone perception).



Dots represent mean accuracy per subject.

4.5.2.3 Improvement in image-naming

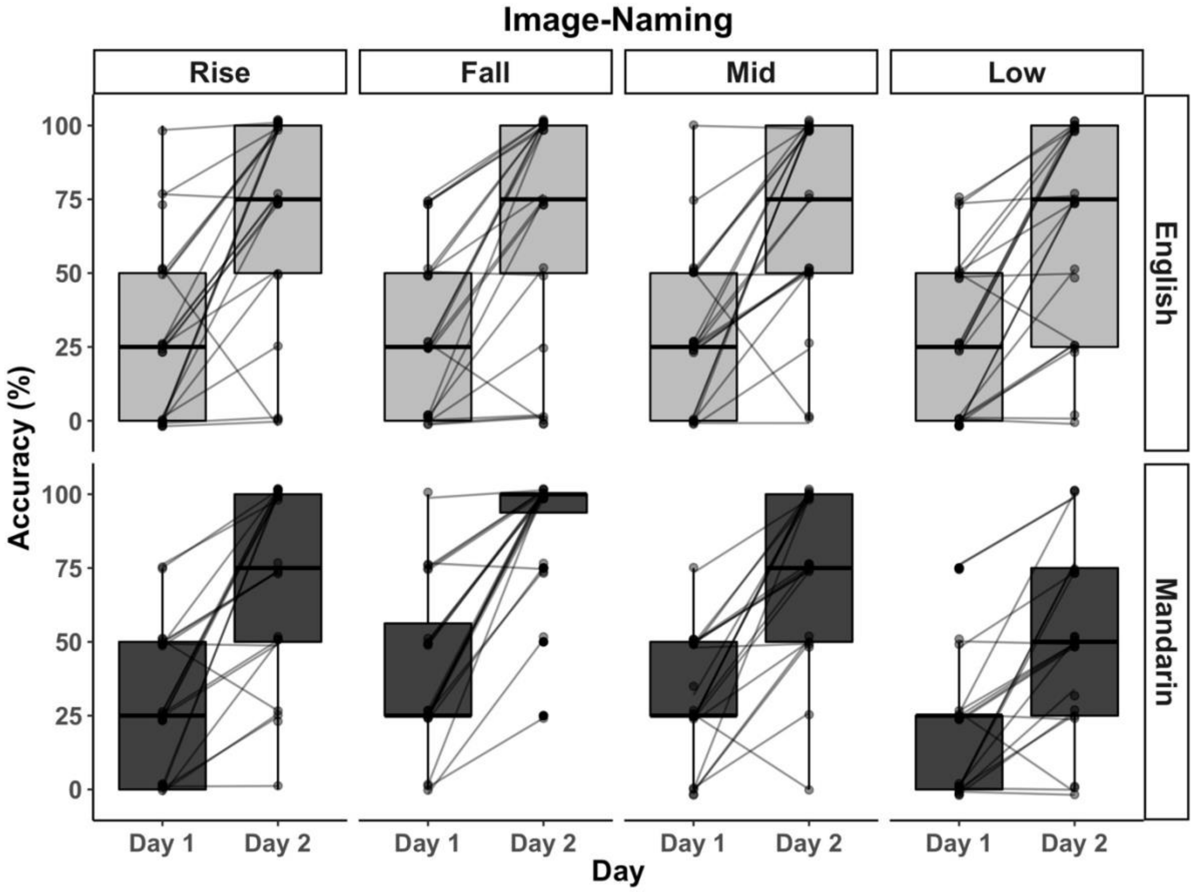
The model for the image-naming task revealed a significant main effect of *Day* and a significant *L1:Tone* interaction (Appendix 4.5). Multiple comparisons revealed that likelihood for correct image-naming was significantly larger on day 2 than on day 1 (b = 2.35, SE = 0.16, p < 0.001) across both groups.

Multiple comparisons following the *L1:Tone* interaction revealed that, averaging over the two days, Mandarin speakers were less likely to correctly name images of low-level tone words compared to images of falling tone words (b = -1.95, SE = 0.36, p < 0.001). There were no other significant comparisons.

For visualization, lexical production performance in the image-naming task across the two days is shown in Figure 33.

Figure 33

Improvement over session in the image-naming task (lexical tone production).



Dots represent mean accuracy per subject.

4.5.3 The perception-production link in terms of error patterns

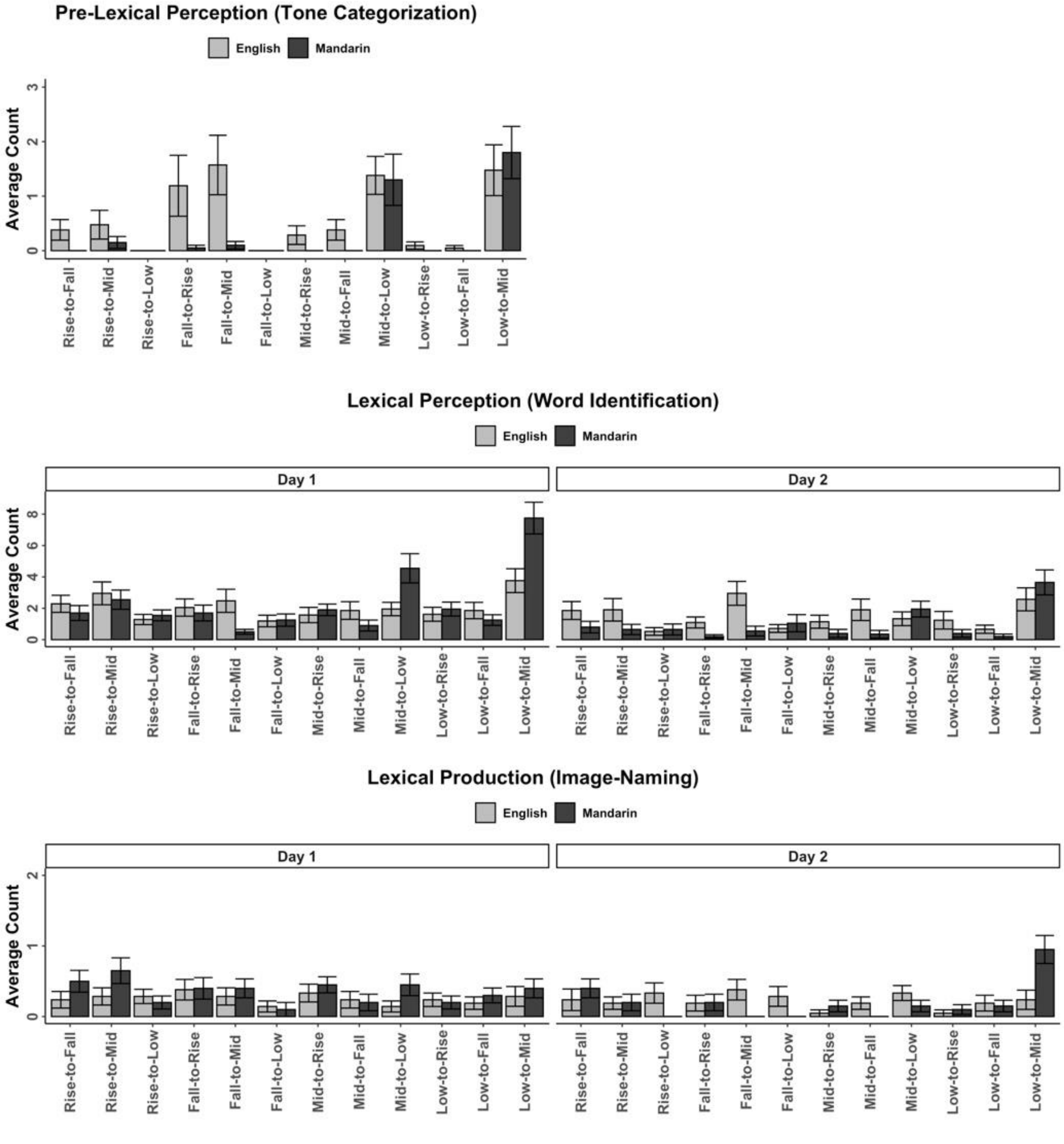
The error analyses in Chapters 2 and 3 provided insight into the occurrence of specific error types in pre-lexical perception, lexical perception, and lexical production. This section summarizes the findings to examine similarities and differences across the modalities. Figure 34 shows the distribution of the different error types across the tone categorization, word identification, and image-naming tasks.

As discussed in Chapter 2 (section 2.5.2.2, page 56), a major observation in both pre-lexical and lexical perception was that English participants tended to confuse tones across the board, misidentifying both contour tones as level tones and vice versa, whereas Mandarin participants predominantly confused low-level tones and mid-level tones. Further, on day 2 of the word identification task, there was statistical evidence that within the Mandarin group, low-to-mid error types were significantly more likely to occur in comparison to 10 out of 11 other possible error types. Within the English group, no such clear pattern of a dominant error type in lexical perception was found.

The findings from the lexical production task (image-naming) revealed a largely similar picture of the prevalence of low-to-mid errors in the Mandarin group. As can be seen in Figure 34, low-to-mid errors appeared to occur most often in image-naming, especially on day 2. However, as discussed in Chapter 3 (section 3.6.2.3, page 106), there was no statistical confirmation that low-to-mid errors occurred significantly more often than other error patterns.

Figure 34

Overview of tone error types across tasks.



Pre-Lexical Perception (Tone Categorization)

Lexical Perception (Word Identification)

Lexical Production (Image-Naming)

Counts are averaged over subject. Error bars = +/- 1 SE.

4.5.4 The perception-production link in terms of facilitative factors

This section presents an analysis of the effect of extralinguistic factors (musical experience and working memory capacity) on performance in pre-lexical and lexical perception and production. Appendix 4.6–9 summarize the estimates for each task. Significant effects observed in each task are listed hereunder.

Musical experience facilitated English participants' pre-lexical perception, as it led to higher likelihood of correct tone categorization (b = 9.42, SE = 4.26, p < 0.001) and faster log RTs (b = -0.28, SE = 0.09, p = 0.002). It facilitated lexical perception (word identification), but only for English participants on day 1 (b = 2.99, SE = 0.85, p < 0.001) and day 2 (b = 12.02, SE = 4.56, p < 0.001). Musical experience did not facilitate pre-lexical production (imitation) in either group. Musical experience facilitated lexical production (image-naming), but only for English participants on day 1 (b = 2.00, SE = 0.44, p < 0.002) and day 2 (b = 4.69, SE = 1.75, p < 0.001).

Working memory did not facilitate pre-lexical perception (tone categorization) in either group. It facilitated lexical perception (word identification), but only for Mandarin participants on day 2 (b = 3.26, SE = 1.00, p < 0.001). Working memory facilitated pre-lexical production (imitation), as it led to smaller Fréchet distances, but only for English participants on day 1 (b = -0.28, SE = 0.09, p = 0.003) and day 2 (b = -0.26, SE = 0.11, p = 0.015). Working memory facilitated lexical production (image-naming), but only for Mandarin participants and only on day 2 (b = 1.90, SE = 0.59, p = 0.040).

All in all, the models for the effect of musical experience and WM reveal a dynamic nature of the effect of extralinguistic factors on non-native tone processing. That is, the relative effects of musical experience and working memory appear to be modulated by a participant's L1. This replicates the findings shown earlier in Chapter 2. Crucially, this dynamic interaction appears to be largely maintained across levels of processing (pre-lexical and lexical) as well as modality (perception and production). Musical experience facilitated English participants' pre-lexical and lexical tone perception, as well as lexical tone production. WM facilitated Mandarin participants' lexical tone perception and production.

One unexpected exception to this trend however, is the observation that WM facilitated English participants' pre-lexical production in the imitation task. Given that

Chapter 3 did not investigate the effect of WM on pre-lexical production, a separate, more complex model was fitted to the imitation task data to investigate the nature of the relationship between WM and pre-lexical tone production in more detail. This model contained fixed effects for *L1, Tone, Day, and WM,* and a four-way interaction with these fixed effects. The model revealed a marginally significant *L1:Tone:Day:WM* interaction (b = -0.05, SE = 0.03, p = 0.049)[16]. Subsequent Bonferroni-corrected multiple comparisons revealed that WM significantly reduced Fréchet distances in the following conditions: For English participants, WM reduced Fréchet distances for rising (b = -0.32, SE = 0.11, p = 0.008), falling (b = -0.32, SE = 0.11, p = 0.008), and low-level tone words (b = -0.26, SE = 0.11, p = 0.030) on day 1, and for low-level tone words on day 2 (b = -0.30, SE = 0.11, p = 0.013). For Mandarin participants, WM reduced Fréchet distance only for low-level tone words on day 1 (b = -0.28, SE = 0.12, p = 0.022). Given the marginal significance on the overall interaction, these estimates should be interpreted with caution.

[16] This marginally significant four-way interaction was obtained through the summary() function in R. It emerged in the default model output for the *L1(English):WM_2:tone(Falling):session(Day 1)* interaction, referenced against the grand mean for fixed effects. It is worth noting that terms of Type III Wald Chisquare tests (which assumes all variables to be contrast coded), the main interaction failed to reach significance χ^2 = 6.653, *df*(3), p = 0.083. Given this discrepancy, I report the multiple comparisons following the four-way interaction, but emphasize that they should be interpreted with caution.

4.6 Discussion

4.6.1 The perception-production link at a pre-lexical level

RQ1 addressed how non-native tone perception and production are correlated at the pre-lexical level in terms of overall performance, improvement over time, and error patterns. The following sections discuss the each of the three areas.

4.6.1.1 Pre-lexical perception and production performance

In terms of overall performance, there was only limited evidence for a strong perception-production link at a pre-lexical level. English participants' performance in the tone categorization task was not significantly correlated with performance in the imitation task. For Mandarin participants, pre-lexical perception performance and pre-lexical production performance were only significantly correlated for low-level tones and for falling tones.

The absence of a convincing, overall correlation in pre-lexical tone perception and production in the present study does not fall in line with previous studies that did find clearer links between pre-lexical perception and production performance (Ding et al., 2011; K. Zhang & Peng, 2017) Although differences in methodology may in part explain this discrepancy, it is worth noting that results from Kirby & Giang (2021), whose measures of tone perception and production closely resembled that of the present study (accuracy in an AX discrimination task and Fréchet distances in an elicitation task) only revealed a weak perception-production correlation. Similarly, Hao (2012) only found weak correlations between perception and production performance in a tone categorization and a tone imitation task.

It is possible that only limited evidence was found for a link between tone categorization and tone imitation because even though both tasks can be broadly described as pre-lexical since they do not necessarily involve word meaning, they may in fact tap into different levels of processing. As highlighted by Hao (2012), imitation may only require surface acoustic perception and production of tones (i.e., step 1 of the process described in Figure 29), whereas tone categorization requires phonological processing (i.e., steps 1 and 2).

This difference in processing levels may result in a relatively weak link between performance in imitation versus performance in tone categorization. This notion was investigated in more detail in a later study (Hao & de Jong, 2016), in which English-L1 intermediate learners of Mandarin tones participated in a tone categorization, a read-aloud, and an imitation task. It was found that participants were highly accurate in tone imitation (as determined in terms of phonological accuracy by native raters), but performed worse in tone categorization and reading. It is indeed worth recalling that a visual inspection of individual productions in the present study's imitation task (as shown earlier in Chapter 3, Figure 21, page 99) suggests that participants were highly accurate in imitation in terms of phonological accuracy. That is, a participant would typically not confuse tones categorically and imitate a rising tone as a falling tone. By contrast, participants did make such confusions in tone categorization (Figure 34, page 147). This fits with the conclusion by Hao & de Jong (2016) that "L2 imitation can bypass some of the difficulties of phonological categorization" (p. 164). Thus, an imitation task may be inherently easy, whereas a tone categorization task may be more demanding, and this difference in cognitive demand may explain why there is no clear perception-production link in terms of performance between the two tasks.

Despite the absence of an overall correlation between tone categorization and tone imitation in the present study, it is worth noting that there were two instances in which tone categorization and imitation were significantly correlated, namely for Mandarin speakers' performance on low-level tones (on day 1) and falling tones (on day 2). In both cases, the correlation direction was intuitively plausible and replicates earlier findings (Kirby & Giang, 2021). Namely, higher tone perception accuracy was associated with higher phonetic production accuracy (i.e., smaller Fréchet distances). It is interesting that this significant correlation was observed in the one condition in which there was evidence that tone imitation was in fact relatively difficult: On day 1, Mandarin speakers' phonetic accuracy of low-level tone imitation was significantly worse than that of English speakers. In tone perception too, low-level tones were relatively difficult for Mandarin speakers, as they yielded relatively slower categorization RTs compared to other tones (Chapter 2, Figure 12, page 55). This could mean that the perception-production link in pre-lexical tone processing is particularly strong when the tone contrast to be perceived or produced is known to be relatively difficult. However, this cannot be claimed conclusively since a significant correlation was also observed for falling tones, which were perceived well by Mandarin speakers in tone categorization, but which were also produced relatively well in imitation.

4.6.1.2 Pre-lexical perception and production improvement

Second, **RQ1** addressed how perception and production at a pre-lexical level are related in terms of improvement over time. The tone categorization task was only conducted on day 1, and it is therefore not possible to discuss improvement in pre-lexical perception. Instead, I highlight some observations of improvement in imitation accuracy over time, and discuss how this may have been affected by the other perception and production tasks in the experiment.

The results from the imitation task showed that for rising and falling tones, participants' imitations improved over time. Fréchet distances for these contour tones significantly decreased from day 1 to day 2, suggesting that participants' tone productions became more target-like. However, there was no statistical confirmation for improvement in mid-level and low-level tones. Moreover, Mandarin participants' productions of mid-level tones appeared to have become less target-like on day 2, as Fréchet distances significantly increased.

Baese-Berk (2019, p. 998) proposes two routes along which speakers may improve in imitation. The first is an acoustic-phonetic route, which assumes that participants improve their imitations by more accurately matching their productions to the target tokens. In the present study, it is plausible that, as a function of increased practice, participants had simply become better at fine-grained phonetic imitation of rising and falling tones, which involves precise control over F0 height, the timing of a change in F0, and the velocity with which this F0 change takes place. The reason why no significant improvement was observed in phonetic production of mid-level and low-level tones may be because these tones were relatively easy to imitate from the onset, as they do not involve the complexity of contour tones. That is, there may have been room for improvement in phonetic accuracy of contour tones, but not for level tones.

The second possible route of improvement in imitation is a phonological one, which assumes that speakers acquire new categories from which they can select exemplars to use in production (Baese-Berk, 2019, p. 998). This scenario would explain the (perhaps surprising) finding that Mandarin speakers' phonetic productions of mid-level tones were in fact less accurate on day 2 than on day 1. Namely, it is possible that because Mandarin participants struggled in forming the phonological contrast between mid-level and low-level tones in perception – as evidenced by their relatively poor performance on this contrast in the tone categorization task – they became hyperaware of this contrast and treated the phonetic imitation task as a phonological task instead. Thus, it may have been the case that on day 2 of the imitation task, Mandarin participants tried to match their imitations of mid-level tones to (often incorrectly) established phonological representations of their level tone categories instead of to the acoustic signal presented to them. This could result in less target-like imitations. Indeed, it is worth observing the larger degree of inter-speaker variability in realization of level tone productions on day 2 compared to day 1 (Chapter 3, Figure 21, page 99). This could also explain the counterintuitive finding in section 4.5.1.4 that imitation accuracy for mid-level tones was negatively correlated with image-naming accuracies of mid-level tones for Mandarin speakers. That is, if Mandarin participants indeed started imitating mid-level tones in the imitation task on day 2 based on their established phonological representation of the level tones, then this may have been detrimental to their phonetic accuracy in imitation, but could have aided them in their phono-lexical accuracy in image-naming. Although this is purely hypothetical, it may thus be that in this way, improvement over time in pre-lexical production was affected by perception and that there was a top-down effect on Mandarin participants' imitations.

4.6.1.3 Pre-lexical perception and production error patterns

The third area in which **RQ1** addressed the perception-production link at a pre-lexical level was in terms of error patterns. Recall that in tone categorization, Mandarin participants predominantly misidentified low-level tones as mid-level tones. Although the nature of the imitation task did not allow for a similar error pattern analysis since accuracy was defined by phonetic proximity to target tones and not by a binary correct-incorrect scale, it is worth recalling that Mandarin participants' inaccurate phonetic production of low-level tones on day 1 appeared to be caused by the fact that participants produced these tones with a relatively high pitch, as discussed in Chapter 3 (section 3.6.1.3, page 102). Thus, it seems that the tendency for Mandarin speakers to mistake low-level tones for mid-level tones was reflected in both their perception and their production. An observation of English error patterns suggested that English speakers mistook tones with one another across the board, and that they did so both in perception and production.

4.6.1.4 Summary of the pre-lexical perception-production link

In sum, a comparison between performance, improvement over time, and error patterns in tone categorization and tone imitation reveals limited evidence for a perception-production link at a pre-lexical level, although symmetries were observed in certain conditions, particularly in categorization and imitation of low-level tones by Mandarin speakers. In general, performance in tone categorization was not indicative of performance in tone imitation. I tentatively conclude that this is mainly because although both tasks can be described as a pre-lexical, the imitation task – both in terms of its nature and in terms of the measure of performance (Fréchet distances) – was more phonetic than the tone categorization task, which was inherently phonological. Indeed, it is possible that stronger correlations with tone categorization would have been observed if pre-lexical production were operationalized by a read-aloud task with orthographic prompts, and if production accuracy were defined in less fine-grained phonetic terms by relying on raters' accuracy judgments (Hao, 2012; Hao & de Jong, 2016).

However, the areas in which correlations and symmetries were found (namely, perception and production of level tones by Mandarin speakers) appear to support the notion that phonological representations of certain sounds in perception can trickle down to shape acoustic-phonetic representations of those sounds in production, as proposed by the phonological route of production development by Baese-Berk (2019, p. 998), as well as the SLM (Flege, 1995, p. 238). All in all, and readdressing **RQ1,** the findings from the tone categorization and imitation tasks show that L2 perception and production of non-native tones at the pre-lexical level are moderately correlated, and it further appears that, in line with theoretical accounts, perception may modulate production. However, the strength of this correlation and the possible symmetries between pre-lexical tone perception and production depend to a large extent on the type of task and the measure of performance.

4.6.2 The perception-production link at a lexical level

RQ2 addressed the perception-production link in terms of performance, error patterns, and improvement over time, but at the lexical level by comparing the word identification and image-naming tasks. To the best of my knowledge, this is one of the first direct comparisons between tone perception and production at a lexical level within the same study. Although Li & Dekeyser (2017) also conducted a word identification and an image-naming task which were methodologically similar to the present study, they did not explicitly examine the perception-production link. I will therefore discuss evidence for the tone perception-production link at the lexical level predominantly in terms of the present study's findings, but they will be compared with Li & Dekeyser's findings where appropriate given the similarities between the tasks.

4.6.2.1 Lexical perception and production performance

In terms of performance, I found strong to very strong correlations between overall accuracy in word identification and image naming, and the correlations per tone were mostly significant and moderate to strong. The reason why I observed more convincing evidence for a perception-production link between the lexical tasks than the pre-lexical tasks may be because the word identification and image-naming tasks were methodologically more alike than the tone categorization and imitation tasks. More importantly, the fact that the word identification and image-naming occurred at the same level of processing (i.e., at levels 1-2-3 of the process described in Figure 29), whereas the tone categorization and imitation tasks occurred at potentially different levels (i.e., respectively at levels 1-2 and levels 1) may have further strengthened this link. In particular, since both the word identification and image-naming tasks required participants to memorize a connection between a tonal representation and a lexical-semantic representation, it could be said that the (in)ability to form such a tone-meaning connection was the strongest driver of performance in both tasks, regardless of what modality those tasks tap into.

4.6.2.2 Lexical perception and production improvement

Analyses on improvement over time revealed that across days 1 and 2 of the word identification and image-naming tasks, participants significantly improved their accuracy. No significant three-way *L1:Tone:Day* interactions were found, suggesting that the degree to which participants improved was similar for each tone (unlike what was found in the imitation task). The only difference in improvement between the two groups was found in word identification, where it appeared that Mandarin participants improved their accuracy scores relatively more than did English participants. A visual inspection of Figure 32 (page 144) suggests however that broadly speaking, improvement was relatively equal for both English and Mandarin participants, and this is also reflected in their final accuracy scores (which are shown again here in Table 13 for reference).

Table 13

Accuracy in word identification and image-naming.

	English	Mandarin
Word identification (day 1)	48.4 (26.5)	47.5 (20.9)
Word identification (day 2)	73.8 (29.5)	82.4 (19.0)
Image-naming (day 1)	29.5 (20.8)	30.7 (14.7)
Image-naming (day 2)	68.8 (29.4)	70.9 (20.3)

Values are means with standard deviations in brackets.

In terms of the actual accuracy scores shown in Table 13, it is worth noting the slight discrepancy between image-naming and word identification accuracy on day 1. Image-naming accuracy was around 30%, whereas word identification was well above 45% on average. It is possible that this is due to an unequal amount of training in perception and production at this point in the experiment. Recall that at this stage, participants had completed six trials in the tone categorization task (which may have served as perceptual training), but only four trials of the productive imitation task to learn the pseudowords. After this, participants completed the image-naming task, in which they scored around 30%. For the subsequent word identification task, they received two additional training trials in the perception modality (consisting of word identification with feedback), and their accuracy reached 45%. Indeed, in the study by Li & Dekeyser (2017), which specifically investigated the effect of training modality on final performance in a word identification and an image-naming task, it was found that participants performed much better when they were trained in the same modality that they were tested on. It may thus be that the higher amount of perceptual training (in the tone categorization task and the feedbacked word identification trials) resulted in relatively higher scores in word identification on day 1 in this study.

However, on day 2, accuracy levels started to converge for word identification and image-naming, both reaching approximately 75% and 70%, respectively, on average. I surmise that this is thanks to a roughly equal amount of training in both the listening and speaking modalities at this stage in the experiment. That is, by this stage participants had received eight training trials targeting production (in the imitation task), and eight training trials targeting perception in the feedbacked word identification trials. It is additionally possible that the imitation task facilitated both perception and production abilities since the task involves both listening and speaking skills (Hao & de Jong, 2016, p. 152). Overall, it appears that lexical perception and production improve in tandem when the amount of training in both modalities is relatively similar. This is in line with the most recent Speech Learning Model's prediction that perception and production "coevolve" (Flege & Bohn, 2021).

4.6.2.3 Lexical perception and production error patterns

I observed similar tone-only error patterns across lexical perception and production. The general observation here was that English participants mistook tones on words across the board, whereas Mandarin participants predominantly mistook low-level tones as mid-level tones. This replicates the findings by Y. Wang et al. (2003) who found "strikingly similar patterns" between error patterns in a pre-lexical perception (tone categorization) and production (read-aloud) task. Unlike Y. Wang et al. (2003), however, I did not observe any differences in the direction of confusion across the listening and speaking modalities. As commented earlier in Chapter 3 (section 3.7.1, page 110), it is possible that the direction of confusion was predominantly low-to-mid (and not vice versa), because Mandarin participants were sensitive to a phonetic residual from their native tone system. That is, the Mandarin high-level tone (55) is acoustically more similar to the present study's mid-level (22) than to the low-level (11) tone in terms of pitch, and therefore potentially easier to perceive and produce at the lexical level. The processing of the pseudoword low-level tones may have been further complicated by the lack of secondary acoustic cues that contribute to salience of the low-dipping tone in Mandarin (R. Yang, 2015). Indeed, the fact that Mandarin participants' inaccurate imitations of low-level tones on day 1 were due to these tones being produced with a too high pitch could imply that Mandarin participants were rather consistent in their processing of low-level tones as if they were higher-level tones. It thus appears that this tendency to confuse low-level tones with mid-level tones occurred in both modalities, and at a pre-lexical as well as a lexical level.

4.6.2.4 Summary of the lexical perception-production link

Results from overall accuracy, improvement over time, and error patterns in the word identification and image-naming tasks show that tone perception and production at a lexical level are very strongly linked. I found evidence that – regardless of their L1 tonal status – participants were equally good (or bad) at linking tone to meaning, improved equally in doing so over time, and made the same types of errors in both the listening and speaking modality. If we assume that within the microcosm of this experiment, participants had reached the later stages of speech learning on day 2 of the word identification and image-naming tasks, this remarkable similarity between the listening and speaking modalities confirms both previous empirical findings (K. Zhang & Peng, 2017), as well as theoretical predictions by the original Speech Learning Model (Flege, 1995) that perceptual performance converges with productive performance over time.

4.6.3 The perception-production link in terms of facilitative factors

Finally, **RQ3** asked how extralinguistic factors (musical experience and working memory) facilitate performance in pre-lexical and lexical tone perception and production. The combined effects of L1-specific and extralinguistic factors were extensively covered in Chapter 2, which showed that musical experience facilitated tone perception and word identification in English participants, whereas working memory only facilitated word identification in Mandarin participants. Following earlier findings that investigated the effects of musical experience and working memory on Mandarin tone word learning by English-L1 speakers (Bowles et al., 2016), it was suggested that English speakers may have benefited most from pitch processing skills (derived from musical practice) rather than domain-general skills (working memory) because pitch was arguably the most challenging feature of the word learning task given English speakers' unfamiliarity with lexical tone. Therefore, behavioral measures most relevant to lexical pitch (musical experience, here) would best predict tone word learning performance for English participants. By contrast, Mandarin participants, by virtue of their L1 tonal status, may not have found the tonal features of the pseudowords challenging *per se* (except in the level tone distinctions). Therefore, their performance may have been better predicted by general behavioral measures relevant to word learning at large (WM, here). Chapter 2 tentatively formulated an 'L1-Modulated Domain-General Account' to describe this dynamic interplay between L1-specific and extralinguistic factors (section 2.7, page 68). The exact details of this account will be outlined in detail in Chapter 6 (section 6.3, page 271).

Chapter 3 did not investigate the effects of musical experience and working memory on production. In the present study, I therefore investigated for the first time the effects of these factors on imitation and image-naming accuracy, and compared this with the dynamic effects of these factors on tone categorization and word identification observed in perception in Chapter 2.

Musical experience did not predict imitation accuracy. It is possible that the measure of musical experience was unable to reveal any individual differences in pitch acuity that would lead to individual differences in fine-grained phonetic production accuracy. Indeed, Chapter 3 showed that another measure of pitch acuity, namely individual pitch aptitude (as measured by accuracy in the tone categorization task) did predict imitation accuracy, although this was limited to Mandarin participants' imitations of falling tones on day 2. This partially reproduces the finding by Li & De Keyser (2017) that musical ability (in their study measured by a standardized musicality test) can predict native-likeness of tone production in pre-lexical imitation tasks. Yet overall, the evidence for a facilitative effect of pitch-related skills on pre-lexical production accuracy was limited. One explanation could be that, as discussed in section 4.6.1.1, the imitation task was relatively easy, thereby making the benefits from musical experience less relevant. Indeed, when looking at the more demanding, perceptive counterpart of the imitation task (tone categorization), performance was significantly predicted by musical experience, but only for English participants. Similarly, in the higher-level lexical tasks (word identification and image-naming), musical experience significantly predicted performance, again only for English participants.

Working memory facilitated imitation, but only for English participants. This goes against the intuition that English individuals would benefit in particular from pitch-related skills (musical experience), and less so from domain-general skills (WM) in non-native tone learning, whereas Mandarin speakers would benefit mostly from domain-general skills.

The follow-up analysis of the effect of WM on imitation accuracy suggested that WM facilitated imitation of rising, falling, and low tones on day 1, and low tones on day 2 for English speakers, and imitation of low tones on day 1 for Mandarin speakers. These findings should be interpreted with caution, however, given the marginal significance of the main interaction in the model. Overall, it thus seems that indeed, English speakers did benefit quite substantially from WM in imitation, whereas Mandarin speakers only benefited partially. The reason why precisely WM appeared to be more relevant for English than for Mandarin speakers in imitation is difficult to evaluate and would require further investigation. Yet, the fact that WM did facilitate some tone imitations in both groups is theoretically plausible when considering memory models that link the ability to recall digit sequences to the ability to accurately listen and repeat sound sequences (Gupta, 2003).

For image-naming, working memory had the same facilitative effect as word identification. It predicted performance for Mandarin, but not for English speakers, for whom musical experience predicted performance. This suggests that for tone word learning in both the listening and listening modalities, musical experience is relatively beneficial for English speakers, whereas working memory is relatively beneficial for Mandarin speakers.

In sum, it appears that the predictions of the L1-Modulated Domain-General Account of tone learning, which was introduced in Chapter 2 and specifies that L1 tonal status reduces the relevance of pitch-related skills (derived from musical experience) and in turn increases the relevance of working memory capacity, are largely borne out by the data from both the listening and speaking modalities, but only in tasks that require processing at a more abstract, phonological level (i.e., level 2 or higher). For tasks that require processing of information uniquely at the phonetic level (i.e., level 1), such as imitation, the relative contribution of pitch-specific skills and WM seems to be less dependent on a participant’s prior experience with lexical tone.

4.7 Conclusion

In conclusion, this study has provided new insights into the perception-production link in non-native tone learning at a pre-lexical and a lexical level. Although perception and production may not mirror one another strongly at lower-lying levels of processing, as shown by the comparisons between performance in tone categorization and tone imitation, this study found very strong links in terms of overall performance, improvement, and error types in word identification and image-naming, which require the processing of tone at a higher, lexical level. It was also shown that, some L1-specific differences aside, extralinguistic factors such as musical experience and working memory facilitate tone perception and production in similar ways, although the parallels here are again stronger at a pre-lexical than at a lexical level.

4.8 Appendix to Chapter 4

Appendix 4.1

Imitation: Mixed ANOVA table for improvement (Type-III Wald Chisquare tests).

IMITATION			
lmer(Fréchet ~ L1*Tone*Day + (Tone | Subject) + (1|Item))			
Effect	χ^2	df	p
L1	0.367	1	0.545
Tone	106.420	3	< 0.001
Day	10.835	1	0.001
L1:Tone	3.500	3	0.321
L1:Day	0.105	1	0.746
Tone:Day	69.477	3	0.000
L1:Tone:Day	10.092	3	0.018

Appendix 4.2

Imitation: Multiple comparisons for day per L1 and tone.

IMITATION THREE-WAY INTERACTION					
	Contrast	Estimate	std. Error	t	p
English					
Rise	Day 2-Day 1	-0.23	0.08	-2.97	0.003
Fall	Day 2-Day 1	-0.38	0.08	-4.88	<.0001
Mid	Day 2-Day 1	0.15	0.08	1.89	0.058
Low	Day 2-Day 1	0.13	0.08	1.71	0.087
Mandarin					
Rise	Day 2-Day 1	-0.43	0.08	-5.31	<.0001
Fall	Day 2-Day 1	-0.22	0.08	-2.69	0.007
Mid	Day 2-Day 1	0.30	0.08	3.80	0.000
Low	Day 2-Day 1	-0.06	0.08	-0.74	0.460

Appendix 4.3

Word identification: Mixed ANOVA table for improvement (Type-III Wald Chisquare tests).

WORD IDENTIFICATION			
glmer(correct ~ L1*Tone*Day + (Tone | Subject) + (1|Item))			
Effect	χ^2	df	p
L1	0.14	1	0.709
Tone	9.44	3	0.024
Day	870.24	1	0.000
L1:Tone	12.92	3	0.005
L1:Day	14.26	1	0.000
Tone:Day	20.59	3	0.000
L1:Tone:Day	1.26	3	0.739

Appendix 4.4

Imitation: Multiple comparisons for day per tone and per L1.

WORD IDENTIFICATION TWO-WAY INTERACTION					
	Contrast	Estimate	std. Error	z	p
Rise	Day 2-Day 1	2.36	0.14	16.45	0.000
Fall	Day 2-Day 1	1.55	0.13	11.24	0.000
Mid	Day 2-Day 1	1.13	0.12	14.93	0.000
Low	Day 2-Day 1	2.20	0.09	16.57	0.000
English	Day 2-Day 1	1.74	0.09	18.85	0.000
Mandarin	Day 2-Day 1	2.24	0.09	23.14	0.000

Appendix 4.5

Image-naming: Mixed ANOVA table for improvement (Type-III Wald Chisquare tests).

IMAGE-NAMING			
glmer(correct ~ L1*Tone*Day + (Tone | Subject) + (1|Item))			
Effect	χ^2	df	p
L1	0.27	1	0.603
Tone	6.71	3	0.082
Day	197.41	1	0.000
L1:Tone	9.41	3	0.024
L1:Day	0.00	1	0.974
Tone:Day	4.28	3	0.233
L1:Tone:Day	3.31	3	0.346

Appendix 4.6

Estimates of effect of extralinguistic factors on pre-lexical perception (tone categorization).

		Estimate	std. Error	Statistic*	p	95% C.I.
English						
Accuracy						
	WM	0.48	0.20	-1.76	0.078	[0.21 ; 1.09]
	Musical Experience	9.42	4.26	4.96	<0.001	[3.88 ; 22.86]
Log RT						
	WM	0.11	0.09	1.23	0.217	[-0.06 ; 0.28]
	Musical Experience	-0.28	0.09	-3.17	0.002	[-0.45 ; -0.11]
Mandarin						
Accuracy						
	WM	1.04	0.30	0.13	0.897	[0.59 ; 1.82]
	Musical Experience	0.86	0.22	-0.58	0.564	[0.52 ; 1.42]
Log RT						
	WM	-0.06	0.06	-0.90	0.368	[-0.18 ; 0.07]
	Musical Experience	-0.05	0.06	-0.81	0.420	[-0.16 ; 0.06]

**z-statistic for accuracy, t-statistic for log RT.*

Appendix 4.7

Estimates of effect of extralinguistic factors on pre-lexical production (imitation).

		Estimate	std. Error	t	p	95% C.I.
English						
Day 1						
	WM	-0.28	0.09	-2.96	0.003	[-0.47 ; -0.10]
	Musical Experience	0.02	0.10	0.18	0.859	[-0.17 ; 0.21]
Day 2						
	WM	-0.26	0.11	-2.43	0.015	[-0.46 ; -0.05]
	Musical Experience	0.05	0.11	0.48	0.628	[-0.16 ; 0.26]
Mandarin						
Day 1						
	WM	-0.10	0.09	-1.02	0.307	[-0.28 ; 0.09]
	Musical Experience	0.01	0.09	0.11	0.910	[-0.16 ; 0.18]
Day 2						
	WM	-0.17	0.11	-1.59	0.112	[-0.37 ; 0.04]
	Musical Experience	-0.01	0.10	-0.12	0.905	[-0.20 ; 0.18]

Appendix 4.8

Estimates of effect of extralinguistic factors on lexical perception (word identification).

		Estimate	std. Error	z	p	95% C.I.
English						
Day 1						
	WM	0.88	0.25	-0.46	0.642	[0.51 ; 1.52]
	Musical Experience	2.99	0.85	3.84	<0.001	[1.71 ; 5.24]
Day 2						
	WM	0.96	0.34	-0.10	0.920	[0.48 ; 1.94]
	Musical Experience	12.02	4.56	6.55	<0.001	[5.71 ; 25.30]
Mandarin						
Day 1						
	WM	1.64	0.42	1.94	0.053	[0.99 ; 2.71]
	Musical Experience	1.24	0.29	0.93	0.350	[0.79 ; 1.95]
Day 2						
	WM	3.26	1.00	3.85	<0.001	[1.79 ; 5.96]
	Musical Experience	1.59	0.45	1.65	0.099	[0.92 ; 2.76]

Appendix 4.9

Estimates of effect of extralinguistic factors on lexical production (image-naming).

		Estimate	std. Error	z	p	95% C.I.
English						
Day 1						
	WM	1.03	0.22	0.15	0.882	[0.68 ; 1.57]
	Musical Experience	2.00	0.44	3.14	0.002	[1.30 ; 3.07]
Day 2						
	WM	0.89	0.32	-0.31	0.753	[0.44 ; 1.81]
	Musical Experience	4.69	1.75	4.15	<0.001	[2.26 ; 9.74]
Mandarin						
Day 1						
	WM	1.33	0.24	1.62	0.106	[0.94 ; 1.89]
	Musical Experience	1.05	0.16	0.30	0.764	[0.77 ; 1.42]
Day 2						
	(Intercept)	2.92	1.12	2.81	0.005	[1.38 ; 6.18]
	WM	1.90	0.59	2.05	0.040	[1.03 ; 3.51]
	Musical Experience	1.14	0.32	0.46	0.646	[0.65 ; 1.98]

Chapter 5 Tone categorization and word learning across a spectrum of L1s[17]

Some L2 learners acquire tone more easily than others do. Such inter-learner variability has been attributed to both linguistic factors, (such as L1 tonal status or tone types) and extralinguistic factors (such as musical experience, working memory, and pitch perception aptitude). However, the relative importance of all these factors when taken together is not well understood. Therefore, this study investigated tonal pseudoword tone perception and word learning by 80 native speakers of languages on a spectrum of L1 tonal statuses: Dutch (stress), Japanese and Swedish (pitch accent), and Thai (tonal). Participants were matched for musical experience and working memory capacity. Tone perception was measured by means of a task requiring categorization of tones from a three-way contrast (level, falling, peak). Tone word learning was measured by a word generalization task in which participants were trained and tested on their ability to indicate the meaning of nine pseudowords with a three-way segmental (/lala/ /lɛlɛ/ /lili/) and the same three-way tonal contrast.

Results from Bayesian inference revealed that tone perception was predominantly facilitated by musical experience. This individual tone perception performance (i.e., pitch aptitude) was in turn the strongest facilitator of successful tone word learning. L1 tonal status appeared to modulate performance only slightly. The findings of this study are discussed in the light of the "Functional Pitch Hypothesis" of non-native tone processing (Schaefer & Darcy, 2014), and highlight the importance of accounting for extralinguistic individual aptitudes in speech learning.

[17] Adapted from: **Laméris, T.J.** (2022) The Effect of L1 Pitch Status and Extralinguistic Factors on L2 Tone Learning. *Proceedings of. Speech Prosody 2022*, 708-712, https://doi.org/10.21437/speechprosody.2022-144

5.1 Introduction

Languages differ in the extent to which F0 (pitch) can be used as a primary cue to signal lexical meaning. In stress languages like Dutch, pitch has a limited lexical function, and is used together with other acoustic elements such as duration, vowel quality, and intensity to distinguish word meaning based on stress, so that first-syllable stressed ['vɔːr.ko.mə(n)] means ‘to occur’ and that second-syllable stressed [vɔːr.'ko.mə(n)] means ‘to prevent’. Pitch has a higher functional load in so-called pitch-accent languages like Swedish and Japanese, in which pitch can be used as a primary device to differentiate meaning between a limited set of words. For instance, the Japanese [bɯdo:] can either mean ‘martial art’ or ‘grape’ depending on the relative pitch height assigned on each mora. The functional load of pitch is highest in tone languages like Thai, in which the syllable [k^{h}a:] can have five different meanings, from ‘galangal root’ to ‘leg’, depending on the pitch pattern it is produced with (Kaan et al., 2008, p. 2).

In the discussion that follows, I will describe these cross-linguistic differences in terms of the functional load of pitch as differences in *L1 tonal status*, following the terminology defined in Chapter 1. As such, languages like Dutch can be described as having a low tonal status, languages like Swedish and Japanese as having an intermediate tonal status, and languages like Thai as having a high tonal status.

It may seem intuitively plausible that a high L1 tonal status contributes to the ease with which an adult acquires the tonal system in a second language. Indeed, it may be hypothesized that because tone language speakers are familiar with “the process of mapping a change in pitch to a lexical semantic change” (Cooper & Wang, 2012, p. 4763), they would learn how to link pitch to meaning in a second language without too much difficulty. In comparison, non-tone language speakers may find this more challenging because their L1s do not specifically prepare them for the process of linking pitch to meaning.

This notion of a facilitative effect of L1 tonal status on non-native tone processing was explored in a study by Schaefer & Darcy (2014), who investigated the effect of “L1 pitch functionality” on non-native tone perception. They describe pitch functionality as a combination of (1) the *exclusivity* of pitch to signal lexical contrast (relevant to other acoustic cues such as duration, intensity, and vowel quality), (2) the *functional load* of pitch (indicating the extent and number of pitch-based minimal pairs), and (3) the *inventory* of pitch patterns (Schaefer & Darcy, 2014, p. 514). For purposes of the present discussion, I will use the term 'L1 tonal status' to describe typological differences between languages in terms of the lexical function of pitch, but I note that this term neatly coincides with Schaefer and Darcy's definition of L1 pitch functionality.

Schaefer & Darcy (2014) recruited a group of speakers on a spectrum of different L1 tonal statuses, namely Mandarin (high tonal status), Japanese (high-intermediate), English (low-intermediate), and non-pitch accent Standard Korean (low). These participants were tested on their ability to perceive tonal contrasts in a Thai tone discrimination task. Their results support the intuition that L1 tonal status facilitates non-native tone processing, and that it does so in an incremental way. They showed that speakers whose L1 tonal status is highest (Mandarin) where faster and more accurate in Thai tone perception than speakers with an intermediate L1 tonal status (Japanese), who in turn outperformed speakers with a low L1 tonal status (English and Korean).

Based on these findings, they propose a "Functional Pitch Hypothesis" (henceforth: "FPH"), which posits that L1 tonal status shapes perception of a non-native tone system. The FPH suggests that in addition to the functionality of pitch, the prosodic domain at which pitch variations are realized in the L1 determines perception facility in a tonal L2, in which pitch variations are typically realized at the syllable level. The FPH predicts that applying sensitivity to pitch variations from a given domain in the L1 to a *smaller* domain in the L2 may be particularly challenging. For instance, English speakers, who are sensitive to pitch variations at a phrasal level (intonation) are expected to struggle with applying this sensitivity to a smaller syllable level as required in a tonal language like Thai. However, Mandarin speakers, who are sensitive to pitch variations at the syllable level (lexical tone) are expected to benefit from this sensitivity in perceiving syllable-level pitch variations in a tonal L2.

Although the predictions of the FPH are supported by a number of studies that investigated non-native tone perception between speakers of different L1 tonal statuses (R. K. W. Chan & Leung, 2020; Peng et al., 2010; Wayland & Guion, 2004), there exist many cross-linguistic studies that fail to find a facilitative effect of L1 tonal status on tone perception (Cooper & Wang, 2012; Francis et al., 2008; Gandour & Harshman, 1978; So & Best, 2010). Different methodologies may in part explain this discrepancy. Additionally, it may be difficult to detect an effect of L1 tonal status on *overall* L2 tone perception because the ease with which listeners perceive tones often depends on *which* tone type they perceive. It has been widely shown that speakers from different L1 backgrounds differ in the ease with which they perceive specific tone types in an L2. For instance, Francis et al. (2008) showed that Mandarin and English speakers performed similarly in terms of overall perception accuracy of Cantonese tones, but differed in their perception per tone type: Mandarin participants were relatively good at perceiving contour tones, whereas English participants were relatively good at perceiving level tones. Similarly, in Chapters 2 and 3 I showed that Mandarin speakers did not outperform English speakers in tone perception, tone production, and word learning. As an explanation for this finding, I proposed that Mandarin speakers had a particular difficulty with the mid-level and low-level tones because these tone types are incompatible with Mandarin tone types. I suggested that Mandarin speakers might have in fact outperformed English speakers if the target tonal system had been more compatible with the Mandarin tone system. In other words, a potential facilitative effect of L1 tonal status may be masked in some studies because of the effect of specific tone types.

Although the study by Schaefer and Darcy (2014) has made important empirical and theoretical contributions to the tone learning literature, there are two aspects of the original study that deserve further examination. First, their study only investigated tone perception and not tone word learning, thereby only partially replicating true-life L2 tone acquisition. Although an examination of tone perception at a pre-lexical level may establish a “baseline” for acquisition (Schaefer & Darcy, 2014, p. 489), a direct examination of tone perception at the lexical level by means of a word identification task will provide a more complete account of tone acquisition.

Second, Schaefer & Darcy did not consider the potential effects of extralinguistic factors in addition to the effects of L1 tonal status on individual performance. There is increasing evidence that extralinguistic factors, such as an individual’s musical experience (Wong et al., 2020; H. Wu et al., 2015) and working memory (WM) capacity (Bowles et al., 2016; Goss, 2020; Laméris & Post, 2022) modulate individual performance in non-native tone perception and word learning.

To this end, the present study widens the scope of Schaefer and Darcy’s study from perception to word learning, and additionally factors in the effects of musical experience and working memory. This will provide a more representative account of the various factors that determine individual performance in non-native tone learning. This study also builds on from Chapter 2, which already investigated effects of L1 tonal status, musical experience and working memory on tone perception and word learning, but only in English and Mandarin speakers, and in a tone system that was specifically designed to contain difficult tone contrasts for Mandarin speakers.

The present study zooms in on the effect of L1 tonal status, and attempted to control for additional effects of tone type, musical experience, and working memory. By examining non-native tone learning in participants that represent a spectrum of L1 tonal statuses, namely Dutch (low), Swedish and Japanese (intermediate), and Thai (high), the following research question is formulated:

RQ: *When controlling for other factors (tone type, musical experience, and working memory), does L1 tonal status facilitate non-native tone perception and word learning?*

5.2 The present study

5.2.1 A spectrum of L1 tonal statuses

Inspired by Schaefer & Darcy (2014) and building on from Chapter 2, the present study expands on earlier study design in three ways to further investigate the effect of L1 tonal status on non-native tone learning.

First, unlike Chapter 2, which involved a group of non-tonal (English) and tonal (Mandarin) language speakers, the present study re-examines the effect of L1 tonal status on non-native tone perception and word learning by not only including a group of non-tonal (Dutch) and tonal speakers (Thai), but in addition a group of pitch-accent language speakers (Swedish and Japanese), for whom L1 tonal status is intermediate. The purpose here is to assess whether one can observe a hierarchy in tone processing based on L1 tonal status, as found by Schaefer & Darcy (2014), while also accounting for extralinguistic factors, unlike Schaefer & Darcy (2014) who did not account for these factors and only investigated tone perception.

Second, Chapters 2–3 showed that Mandarin speakers' performance in perception, production, and lexical processing was strongly affected by the target tone type, as they had more difficulty than English speakers in mastering the distinction between level tones. Indeed, the level contrast was included in those studies to specifically address the effect of tone type on Mandarin tone perception and production. Because the present study focuses on the overall effect of L1 tonal status rather than the L1-specific effects of tone type, care was taken to design a non-native tone system with tonal contrasts that should be equally easy or difficult regardless of a learner's L1 tone types. Specifically, it involved a contrast between a low-level (11), a falling (51) and a peaking (141) tone. As will be described in more detail in section 5.2.3, it is hypothesized that the static-dynamic contrasts (i.e., level-fall, level-peak) should be equally easy, and dynamic-dynamic contrasts (i.e., fall-peak) should be equally difficult for speakers of the four different L1 backgrounds. The aim here is to eliminate as much as possible any strong interference from L1 tone types on the processing of specific tone types, which allows us to address more directly whether L1 tonal status in and of itself facilitates non-native tone processing across the board.

Finally, the present study included a word generalization task at the end of the word training phase on day 2 to test whether participants were able to identify the meaning of a tonal pseudoword when spoken by a new speaker. One of the limitations of the word identification task in Chapter 2 was that participants were trained with and tested on words spoken by the same speaker. This leaves it unclear whether participants had truly learned a word by establishing a broad phono-lexical representation, or whether they had simply memorized the combination between a specific acoustic signal and a lexical item. To rule out the latter possibility, participants in the present study were tested on their ability to identify the meaning of tonal pseudowords that were spoken by a speaker they had not heard before, which is an ability that is deemed to be more indicative of real-life word learning.

5.2.2 L1s

The present study examined the effect of different degrees of L1 tonal status on tone perception and word learning facility of tonal pseudowords, factoring in the additional effects of musical experience and working memory. It included participants from languages with different degrees of L1 tonal status: Dutch, Swedish, Japanese, and Thai. The differences between their tonal statuses are summarized in Table 14.

Table 14

Respective L1 tone statuses according to language type and domain, adapted from Schaefer & Darcy (2014).

Language	Domain	L1 tonal status
Non-tonal/word stress (Dutch)	Lexical, word	Low
Pitch accent (Swedish)	Lexical, word	Intermediate
Pitch accent (Japanese)	Lexical, word	Intermediate
Tonal (Thai)	Lexical, syllable/word	Maximal

Standard Dutch, like English, is a non-tonal language in which pitch alone is typically not used for lexical distinctions (Ramachers et al., 2017, p. 2). Previous studies have shown that in certain pre-lexical tone perception tasks, Dutch speakers perform comparably to speakers of tone languages such as Mandarin (A. Chen et al., 2016) and Cantonese (Cutler & Chen, 1997). This is arguably because Dutch speakers can process pitch contrasts in a psychoacoustic manner when these are not associated to any linguistic information (Braun & Johnson, 2011, p. 593). To the best of my knowledge, there are no studies that involve tone word learning by Dutch speakers, but it is expected that they may find it relatively difficult to establish a phono-lexical representation in tonal pseudowords compared to speakers from tonal and pitch-accent language backgrounds, based on tone word learning studies involving non-tonal and tonal L1ers (Cooper & Wang, 2012; Poltrock et al., 2018).

Central Swedish is a pitch-accent language in which words can be distinguished in meaning by an “acute” Accent I and a “grave” Accent II, which in citation form or focus position are typically described as a rise-fall pitch pattern and a peak-peak pitch pattern, respectively (Bruce, 1977; Engstrand, 1997; Ota, 2006). Although only a relatively small number of minimal pairs are distinguished by pitch alone (Köhnlein, 2020, pp. 154–156), pitch carries a higher lexical functionality in Swedish than in non-tonal languages like English or Dutch. Cross-linguistic perception data involving Swedish speakers suggest that this intermediate L1 tonal status may benefit Swedish speakers’ non-native tone processing. In the large-scale experiments by Burnham et al. (2015), Swedish speakers were found to perform similarly to, or slightly worse than tone language speakers (Mandarin and Cantonese), but better than non-tonal (English) speakers in perception of Thai tones. This supports the notion that pitch-accent language speakers may perform somewhere in between non-tonal and tonal speakers in certain non-native tone perception tasks (Schaefer & Darcy, 2014). However, studies with speakers of Norwegian (which has a similar pitch accent system) suggest that Norwegian and non-tonal (English) listeners perceive non-native tone similarly (van Dommelen & Husby, 2009). Given the paucity of cross-linguistic data on non-native tone perception by Swedish speakers and the apparent absence of any studies involving tone word learning, I make the tentative prediction that Swedish speakers will show enhanced non-native tone processing compared to Dutch speakers.

Standard Tokyo Japanese, like Swedish, is a pitch-accent language in which pitch has an intermediate lexical function. Japanese prosodic words can carry a pitch accent, which in the Japanese context refers to a sharp drop in pitch realized over one mora (the minimal tone-bearing unit) onto the subsequent mora (Kawahara, 2015). Words in Japanese carry predefined pitch patterns depending on the presence and location of the pitch accent (Laméris & Graham, 2020, p. 110), and different pitch patterns on otherwise segmentally identical words can be used for lexical distinctions. Unlike Swedish, more cross-linguistic research has been carried out with Japanese speakers to test whether an intermediate L1 tonal status facilitates non-native tone processing. Some studies suggest that this intermediate L1 tonal status indeed does, so that Japanese speakers' non-native tone perception accuracy sits somewhere in between that of non-tone language and tone language speakers (Schaefer & Darcy, 2014), or even approximates that of tone language speakers in some perceptual tasks (Zhu et al., 2021). As to non-native tone word learning, a study by Braun et al. (2014) showed that Japanese speakers were less accurate than Mandarin speakers in determining word meaning based on tonal distinctions, but also less accurate than German (non-tonal) speakers. The authors suggest that Japanese speakers failed to reliably process pitch lexically because of a combination of a low functional load and a "poverty of sentence-level pitch events in their L1" (p. 341), and argue that the large inventory size of intonational tone types in German may have aided these speakers in outperforming Japanese speakers in a lexical tone task. Given these mixed findings, I predict here that Japanese speakers, like Swedish speakers, may show some enhanced non-native tone processing compared to Dutch speakers in both non-native tone perception and word learning.

Central Thai is a tone language with two dynamic (rising and falling) and three static tones (high, mid, and low) which contrast on a single syllable. In citation form, the respective

Chao notations of these tones are 315 (rising), 51/241[18] (falling), 45 (high), 33 (mid) and 21 (low) (Burnham et al., 2015, p. 1460; X. Wu et al., 2014, p. 90). Pitch functionality in Thai is maximal (Schaefer & Darcy, 2014), and this maximal L1 tonal status appears to modulate non-native tone perception and word learning. In what appears to be the only study involving Thai and non-tonal L1ers' non-native tone perception and word learning, Cooper & Wang (2012) showed that Thai-L1 speakers without musical experience did not outperform English-L1 counterparts in pre-lexical tone perception of Cantonese tones, but they did in word learning. This suggests that Thai speakers' maximal L1 tonal status facilitates non-native tone processing at a lexical level. This agrees with other studies that show advantages for tonal L1ers in tone processing beyond the pre-lexical level (R. K. W. Chan & Leung, 2020; Poltrock et al., 2018). Based on these findings, it is expected that Thai speakers will outperform Dutch speakers, particularly in tone word learning. Thai speakers may also outperform Japanese and Swedish speakers in word learning, and potentially also outperform the other L1 groups in pre-lexical tone perception.

In sum, previous studies with Dutch, Swedish, Japanese, and Thai speakers, which predominantly address pre-lexical tone perception, show mixed evidence of an incremental effect of L1 tonal status on non-native tone processing. There is no consistent evidence that shows a default facilitative effect of L1 tonal status. This is often because the effect of tone type may mask any effect of L1 tonal status, as specific tone types are relatively easy or difficult to perceive depending on the L1 (Laméris & Post, 2022; So & Best, 2010; Zhu et al., 2021). To mitigate the effect of tone type and to focus on whether L1 tonal status in and of itself facilitates non-native tone processing, the present study examined perception and word learning of pseudowords that include contrasts that are hypothesized to be equally challenging to acquire for all speakers. This tone system consists of a low-level (11), a falling (51), and a peaking (141) tone. The static low-level tone contrasts with the dynamic falling and peaking tones in both height and direction, and such "static-dynamic" contrasts are

[18] This tone has been described as both as rising-falling 241 (J. Chen et al., 2020, p. 6) or high-falling 51 (X. Wu et al., 2014). To avoid confusion, I will henceforth refer to this tone as 'falling' given that this is the typical phonological description (I. L. Chan & Chang, 2019; J. Chen et al., 2020, p. 4). In addition, naïve Mandarin listeners of Thai appear to categorize this tone as either a falling or high-level tone (J. Chen et al., 2020, p. 9), consistently as a high-level tone (Reid et al., 2015), or consistently as a falling tone (X. Wu et al., 2014), but not as rising-falling tones.

expected to be inherently easy to perceive. The fall-peak contrast constitutes a “dynamic-dynamic” contrast, which may be inherently difficult to perceive, regardless of L1 background (Burnham et al., 2018; Schaefer & Darcy, 2014).

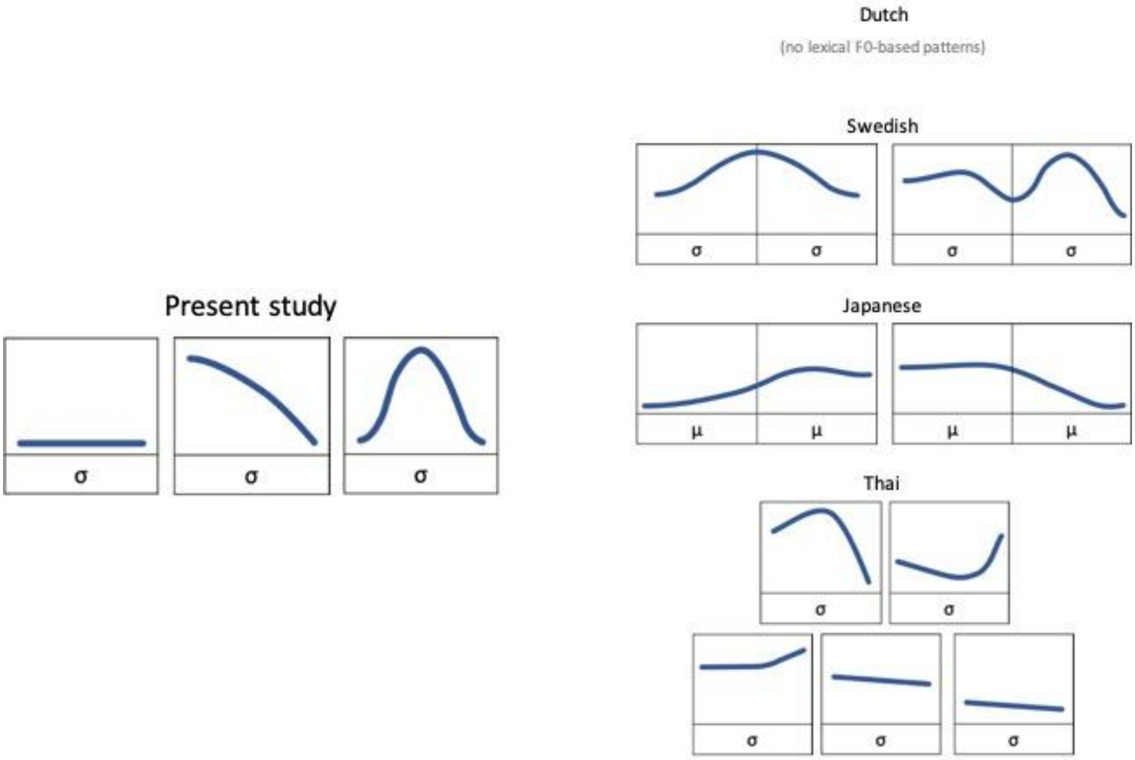
5.2.3 L1-specific predictions

Specific predictions of the possible effect of tone type per L1 are made as follows. An overview of the lexical tone types in the respective L1s is provided in Figure 35. Note that all these predictions are speculative in nature and based on the principles of the Perceptual Assimilation Model (Best, 1995). That is, L2 tone categories that map onto native tone categories in a one-to-one fashion (“two category assimilation”) are relatively easy to distinguish, and L2 categories that map onto native categories in a one-to-many fashion (“single category assimilation”) are difficult to distinguish (Best, 2019). In addition, I make the same assumption as Schaefer & Darcy (2014, p. 513) and predict that assimilation to tone types in a non-tonal or pitch-accent language (Dutch, Swedish, Japanese) will be relatively weak as opposed to assimilation to tone types in a tonal language (Thai).

I will further assume here that participants process the pseudoword tonal contrasts in both the tone categorization and word identification task as monosyllabic. This is because in the tone categorization task, tones were realized on monosyllabic vowels. The word identification task involved disyllabic words, but the crucial lexical tone contrast was on the first syllable only, and the second syllable was assigned a default low-level tone. I will consider the possibility that participants may have in fact processed the tones as disyllabic in the Discussion (section 5.5).

Figure 35

Overview of present study's tone system and lexical tone types in L1s.



The tone type visualisations were adapted from Köhnlein (2020 pp. 154–155) for Swedish, Laméris & Graham (2019, p. 108) for Japanese, and Burnham et al. (2015, p. 1461) for Thai.

Dutch contains no lexical F0-based categories, and only makes exclusive use of pitch in sentence-level intonation. The inventory of Dutch intonational types has been described to contain "level", "fall", as well as a 'downstepped rise-fall' categories, amongst others (Gussenhoven, 2005). If tone-to-intonation assimilation occurs in a single-category fashion, this could make the present study's three tones equally easy or difficult for Dutch speakers. However, the primary assumption is that Dutch speakers do not strongly assimilate non-native tones to intonational tone types (Best, 2019, p. 5; Francis et al., 2008, p. 269). Instead, Dutch speakers may process tones in a more psychoacoustic manner (A. Chen et al., 2018; Peng et al., 2010; X. Wang, 2013; K. Yu et al., 2019). This being the case, based on their inherent acoustic properties, the static-dynamic contrast should be relatively easy, and the dynamic-dynamic contrast should be relatively difficult to perceive.

Swedish contains two disyllabic lexical tone types in citation form or focus position. It is expected that the difference between the pseudowords and Swedish word tone types in terms of the prosodic domain on which the crucial tonal contrast occurs (i.e., one versus two syllables) would limit the effect of assimilation of L2 onto L1 tone types (Schaefer & Darcy, 2014, p. 513). Influence of Swedish tone types on perception of the fall-peak contrast is thus expected to be limited, and Swedish speakers are expected to perceive static-dynamic contrasts relatively well, and the dynamic-dynamic contrast relatively poorly due to the inherent acoustic properties of these tones.

It should be noted however, that in non-focal positions, the realization of Swedish lexical tone types differs from what is presented in Figure 35. In such contexts, Accent I words tend to be realized with a gradual pitch fall over two syllables, whereas Accent II words tend to be realized with a peak on the first syllable. Such realizations are in fact similar to the fall-peak distinction in the pseudoword tones. It has further been suggested that the truly consistent distinction between Accent I and Accent II words is the peak on the first syllable in Accent II words, which is retained under focus and non-focus positions (Bruce, 1977; Ota, 2006). If Swedish speakers perceive the pseudoword tone distinctions by assimilating the fall-peak contrasts to both focal and non-focal realizations of Swedish lexical tone types, it may be in fact that the pseudoword fall-peak contrast would be quite easy to perceive, given that a similar contrast exists on first syllables of non-focused Accent I and Accent II words in Swedish. I will for the moment assume that, if assimilation to Swedish pitch accent types does take place, that that assimilation takes place on canonical citation forms as presented in Figure 35.

Japanese lexical tone types consist of low and high pitch values carried on individual morae. As in Swedish, the influence of Japanese tone types on perception of the pseudoword contrast is expected to be limited because of a difference in the prosodic domain of lexical tone types (Schaefer & Darcy, 2014, p. 513). Instead, based on the acoustic properties of the tones and on research that suggests that Japanese listeners attend more to pitch height than to pitch contour differences (Zhu et al., 2021), it is predicted that the static-dynamic contrasts will be relatively easy, whereas the dynamic-dynamic contrast may be relatively difficult to perceive.

Thai is a tonal language in which tonal contrasts are realized on a single syllable, and it is possible that there is a strong interference from Thai tone types on perception of the pseudoword tones (Best, 2019, p. 5; Schaefer & Darcy, 2014). The fall and peak tones may assimilate in a many-to-one fashion to the Thai falling tone, which is sometimes described as a rising-falling, i.e., peaking tone (J. Chen et al., 2020, p. 6). Consequently, the fall-peak contrast in our pseudowords may be relatively difficult to distinguish. On the other hand, the presence of static-dynamic tone contrasts in Thai would imply that the low-level tone will be easily distinguished from the falling and peaking tones.

In sum, and although it cannot be ruled out that any of the language groups would in fact be relatively better or worse in perception and word learning of specific tones because of interference from L1-based tone types, the present study's pseudoword tone system is expected to be equally challenging for all speakers. In doing so, this system allows us to investigate more directly whether different degrees of L1 tonal status in and of itself facilitate non-native tone perception and word learning.

Finally, the following predictions are made for the effects of musical experience and working memory on non-native tone perception and word learning. As to musical experience, previous studies suggest that Dutch listeners benefit from musicianship more than Mandarin listeners in pre-lexical tone processing, potentially because Dutch listeners process tones in a psychoacoustic and not a categorical manner, similar to the psychoacoustic processing that is involved in music (A. Chen et al., 2016, 2018). This differential in the effect of musical experience between non-tonal and tonal language speakers has also been attested in word learning (Cooper & Wang, 2012; Laméris & Post, 2022). Although there appear to be no cross-linguistic studies that investigate effects of extralinguistic factors in speakers of languages on a spectrum of L1 tonal statuses, it is expected that musical experience will be most beneficial for non-tonal language speakers, relatively beneficial for pitch-accent language speakers, and only partially beneficial for tonal language speakers. The same differential is expected for the effect of pitch aptitude (i.e., tone categorization accuracy) on word learning accuracy, cf. Cooper & Wang (2012).

Finally, based on Chapter 2, it is predicted that working memory will be relatively facilitative for tone word learning and less so for tone perception. Further, this effect is expected to be relatively strong in tonal language speakers (Thai), less so in pitch-accent language speakers (Swedish and Japanese), and the least strong in non-tonal language speakers (Dutch).

5.3 Methodology

5.3.1 Participants

The study was approved by Research Ethics Committee of the Faculty of MMLL at the University of Cambridge. A total of 115 participants took part. Participants were recruited through university networks and social media, participated voluntarily, and received a small token fee upon completion of the experiment. All were native speakers of Dutch (Netherlands), Standard Swedish, Tokyo Japanese, or Central Thai, and had grown up in the respective countries of origin but were resident in the UK as students or young professionals at the time of the study. Participants first filled out a linguistic and musical background questionnaire before being included in the main study. Because of an imbalance in the number of musicians and non-musicians across groups in the original participant pool, the data presented here focus on a subset of 80 participants (22 Dutch, 15 Swedish, 23 Japanese, and 20 Thai participants) who were matched for their degree of musical experience, measured in years of formal training, and working memory (WM), measured by a backwards digit span task. Equivalence tests (Lakens et al., 2018) revealed no significant difference between the groups in terms of their musical experience or WM. Following exclusion criteria of previous studies (Burnham et al., 2015; Schaefer & Darcy, 2014), native speakers of a non-tonal or pitch-accent language with knowledge of a tone language were not included in the subset[19]. Some Thai speakers in the subset reported knowledge of Mandarin, but it was deemed appropriate to include these speakers given evidence that knowledge of a second tone language does not appear to substantially facilitate non-native tone learning for tonal L1ers

[19] The exclusion of participants based on their L2 knowledge of a tone language, as well as the exclusion of participants to balance out the number of musicians in each group, did not substantially change the results. For reference, I provide an overview of the main analyses based on the full participant pool (n = 114; one participant was excluded because they scored below chance) in Appendix 5.10–13.

(Maggu et al., 2018), and given that knowledge of the Mandarin tone system should not contribute to better perception of the fall-peak contrast in the present study's pseudowords. However, Thai speakers who reported knowledge of Northern Thai were excluded. This is because Northern Thai contains a fall-peak contrast (Katsura, 1969) which could facilitate perception of the fall-peak contrast included in the study. An overview of the participant demographics is provided in Table 15 (see Appendix 5.1–4 for further details).

Table 15

Participant demographics.

	Dutch		Swedish		Japanese		Thai	
	(*n* = 22)		(*n* = 15)		(*n* = 23)		(*n* = 20)	
Age (years)	26.30 (3.64)		27.50 (4.95)		29.30 (5.04)		24.90 (5.31)	
Backwards digit span	6.00 (1.54)		5.73 (1.22)		6.83 (1.34)		6.15 (1.14)	
Pitch aptitude (%)	81.6 (13.3)		79.7 (12.0)		88.0 (10.2)		79.3 (17.3)	
Musical experience	MU	NM	MU	NM	MU	NM	MU	NM
(years)	(n=12)	(n=10)	(n=7)	(n=8)	(n=12)	(n=11)	(n=10)	(n=10)
MU= musicians, NM=	10.20	1.10	11.10	0.38	8.92	0.81	8.10	0.60
non-musicians	(4.63)	(3.48)	(3.85)	(1.06)	(5.21)	(1.60)	(5.07)	(0.96)

Values are means with standard deviations in brackets.

5.3.2 Stimuli

The audio stimuli consisted of set of meaningless vowels ([a] [ɛ] [i]) for the tone categorization task and a set of pseudowords(/lala/ /lɛlɛ/ /lili/; see Table 16) for the word identification task. Each of these stimuli carried either a low-level, a falling, or a peaking tone, resulting in nine vowel stimuli and nine pseudoword stimuli for each task.

Visual stimuli in the tone categorization task consisted of tiles representing the level, falling, and peak contours, as shown earlier in Figure 35. In the word identification task, each pseudoword (which was only presented aurally) was linked to an image representing its meaning (Figure 36).

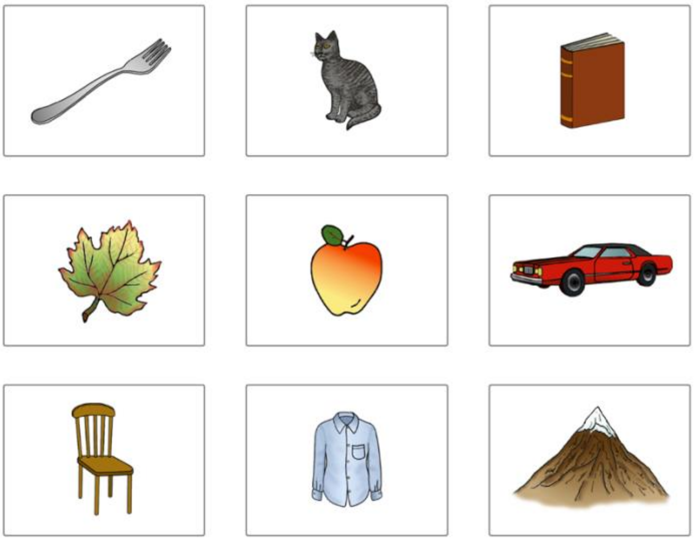
Table 16

Pseudowords.

	Tone 1 (Low-level 11)	Tone 2 (Falling 51)	Tone 3 (Peak 141)
Segment 1	/la11.la11/	/la51.la11/	/la141.la11/
Meaning	*leaf*	*fork*	*television*
Segment 2	/lɛ11.lɛ11/	/lɛ51.lɛ11/	/lɛ141.lɛ11/
meaning	*chair*	*apple*	*hair*
Segment 3	/li11.li11/	/li51.li11/	/li141.li11/
meaning	*book*	*shirt*	*cat*

Figure 36

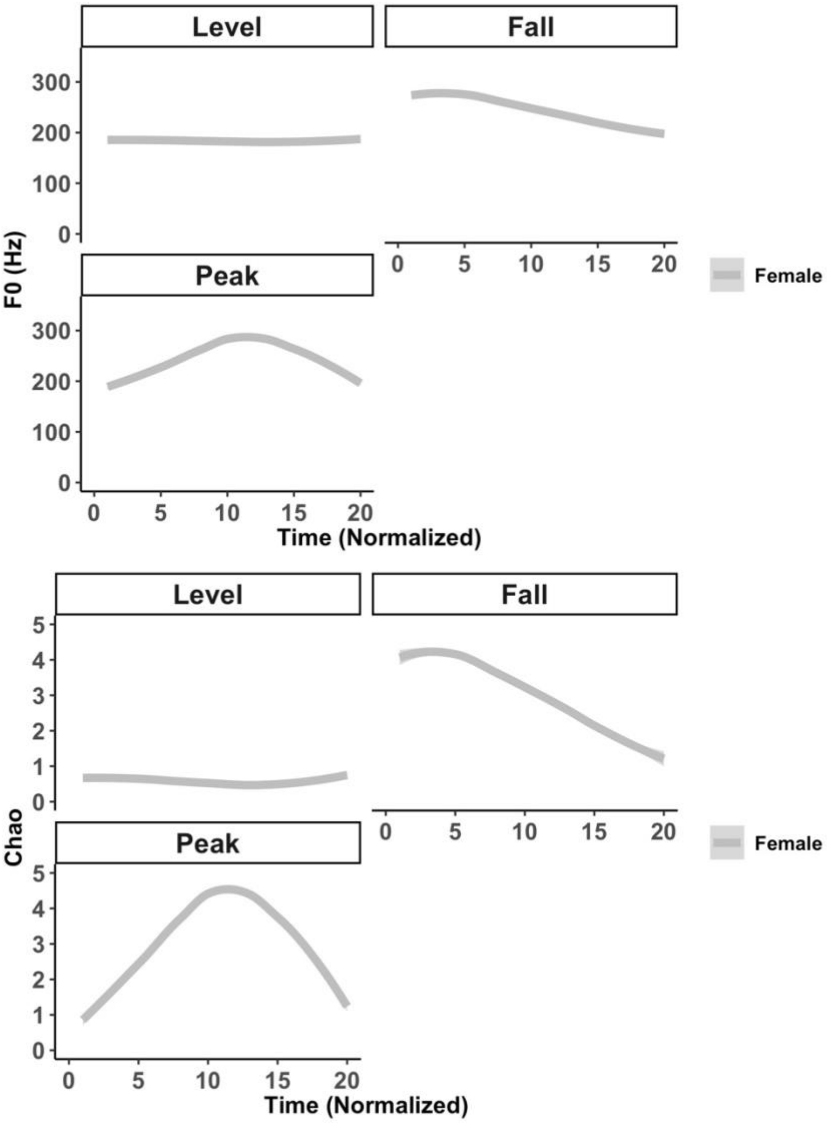
Visual stimuli for pseudowords.



Stimuli were recorded in a sound-attenuated booth and produced by two native speakers of Italian (one male, one female). The baseline stimuli were produced with a flat (mid-level) tone. Stimuli with the low-level, fall, and peak tones, of which the contours were based on natural productions, were synthesized using Pitch Synchronous Overlap (PSOLA) in *Praat* (Boersma & Weenink, 2019). This ensured that tone minimal triplets only differed in F0 and not in other acoustic cues. Both the male and female tones had the same relative tone values in terms of Chao numerals, and the stimuli in the tone categorization and word identification tasks were deemed to belong to the same three tone categories, namely low-level (11); fall (51); and peak (141). For visualization, the F0 and Chao-normalized contours of the tone and word stimuli are shown in Figure 37–38.

Figure 37

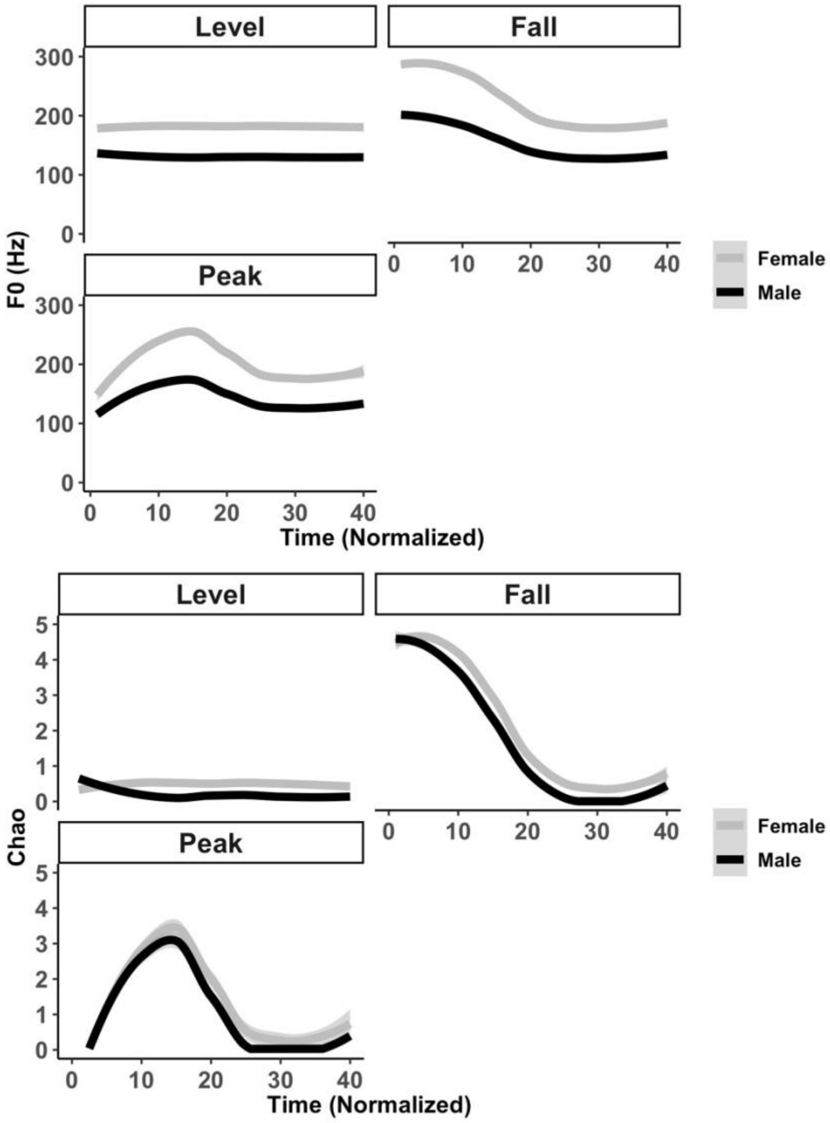
Smoothed F0 and Chao numeral traces for the three tones in the tone categorization task.



Shading ribbons, where present, indicate a 95% Confidence Interval.

Figure 38

Smoothed F0 and Chao numeral traces for the three pseudoword tones.



Shading ribbons, where present, indicate a 95% Confidence Interval.

The choice for disyllabic instead of monosyllabic pseudoword stimuli, unlike in Chapter 2, was motivated by observations by Pelzl et al. (2020, p. 4) that monosyllabic tone word stimuli may have limited generalizability to real-word tone learning. Tone contrasts only occured on the first syllable of the word and not on the second (for which the tone was kept constant as a low-level tone) to avoid tonal contrasts being associated with intonational contrasts for Dutch listeners (Braun & Johnson, 2011, p. 589) and to make participants focus primarily on the tonal contrast occurring on one syllable, similar to the tone categorization task.

5.3.3 Procedure

Due to the coronavirus pandemic, research facilities were closed during the data collection period (January-September 2021). Therefore, this study was entirely carried out online using the *Gorilla Experiment Builder* (Anwyl-Irvine et al., 2020). The study consisted of a battery of eight tasks in total (including training blocks), carried out in two sessions over two days (Table 17). Each session took approximately 25 minutes to complete. Written instructions were in the participants' respective L1s. Headphone screening before each session ensured that participants were in a silent room and were using headphones (Woods et al., 2017). Participants were told that they were taking part in a study that investigated vocabulary learning in a non-native language. After signing a consent form, participants completed the tasks individually. A debriefing was included to ensure that participants had no technical issues during the experiment. Participants who reported technical issues or distractions that were deemed to significantly affect performance in the experiment were excluded.

Table 17

Overview of tasks.

DAY 1	
Description	Duration (minutes)
Tone categorization	5
Word training (imitation)	5
Word training (word identification with feedback)	5
Word identification	10
DAY 2	
Description	Duration (minutes)
Backwards digit span (WM)	5
Word training (imitation)	5
Word training (word identification with feedback)	5
Word generalization	10
Debriefing	5

5.3.3.1 Tone categorization task

On day 1, participants completed a tone categorization task to measure pre-lexical tone perception accuracy and to operationalize pitch perception aptitude (Dong et al., 2019; Laméris & Post, 2022; Wong & Perrachione, 2007). Participants heard one of the vowels with a level, a falling, or a peaking tone, and were instructed to categorize the tone by clicking with their mouse on the tile representing the pitch contour. They were encouraged to make their choice as quickly as possible and to guess if unsure. Time-out was 5000 ms after presentation of the audio stimulus. Only the female voice was used for the tone categorization audio stimuli.

Once practice session with 9 trials (3 presentations per tone) including feedback was held at the beginning. The feedback consisted of the visual presentation of a green circle if the response was correct or a red cross if the response was incorrect, followed by the audiovisual presentation of the audio stimulus and the corresponding tone contour. In the practice session, the vowel [o] was used, which was not used in the main session. The practice session was followed by a main session with 54 trials (6 presentations per stimulus) without feedback in a randomized order.

5.3.3.2 Word training

The tone categorization task was followed by tone word training. As in Chapter 2, the training consisted of imitation and a feedbacked word identification task, which both proved to be efficient ways to stimulate retention of pseudowords. However, given that the present study was carried out remotely, it was technically impossible to verify whether participants faithfully imitated the words all the time. In addition, it was estimated that participants might lose concentration during imitation and would remain more focused in a more interactive task like word identification with feedback. Therefore, there were two imitation trials (instead of four) per item, and four feedbacked trials (instead of two) per item.

In the imitation block, participants were presented with the individual pseudowords (the audio stimuli, male voice) and their meaning (the images). They were asked to repeat the words out loud and pronounce them as accurately as possible, whilst simultaneously trying to memorize the word. No feedback was given regarding their pronunciation and productions were not recorded. Participants had 5000 ms to repeat the word before the next audiovisual stimulus was presented. Each audiovisual stimulus was presented twice in a row (e.g., the word for ‘apple’, followed by the participant’s imitation, followed by one more trial (presentation + imitation) for ‘apple’), and the presentation order was such that no segmental or tonal minimal pair followed one another. The debriefing revealed that 87% of the time, participants repeated out loud “all of the words” and 11% of the time “about half of all the words”. The exact same imitation block was repeated on day 2, with the only difference that the order of presentation of the stimuli was reversed.

In the feedbacked word identification block, participants heard a pseudoword and were then prompted to identify the meaning of that word by clicking the corresponding tile from a 9-way choice answer board, as shown in Figure 36. Participants were encouraged to make their choice as quickly as possible and to guess if unsure. Time-out per trial was set to 10 s. The feedback consisted of a green circle if the response was correct or a red cross if the response was incorrect, followed by the correct sound-image combination. Each stimulus was presented 4 times, totaling 36 trials, in a randomized order. There was a break halfway through, after which the images’ positions on the answer board were shuffled. The exact same feedbacked word identification task was repeated on day 2, with the only difference being that the positions of the images on the answer boards were again shuffled for each half of the block.

5.3.3.3 Word identification task

The feedbacked block was followed by the word identification task without feedback. The set-up was the same as the feedbacked block, but there were 6 trials per stimulus, totaling 54 trials. There was a small break after the participants had completed two-thirds of the task, and the images’ positions on the answer boards were shuffled after the break. After having completed the word identification task, participants received instructions to resume the experiment after 18 to 30 hours.

5.3.3.4 Working memory task

WM was operationalized through a backwards digit span task, based on Chapter 2. Participants were instructed to type in backward order a sequence of digits that was presented to them on the screen. Each of the digits was presented one by one for 750 ms with an ISI (inter-stimulus interval) of 250 ms. After the sequence was presented, participants could type their answer, for which they had 10 s. After a practice session, they were presented with a block of five 4-digit sequences (e.g., 1-7-5-8; 6-3-4-1; 2-5-1-5; 8-4-1-4; 9-5-7-8). The first block consisted of 4-digit sequences instead of 2-digit sequences like in Chapter 2 to limit the amount of time spent on the task and because most participants were expected to be able to attain a 4-digit digit span. Participants would move onto a next block of five n+1-digit sequences (e.g., 5-8-2-5-2; 6-9-4-2-4; etc.) and continue to do so if they correctly typed in at least three sequences per block. If participants did not reach this threshold, the task was aborted at the end of a block. The maximum attainable block consisted of five 8-digit sequences.

Unlike Chapter 2, in which working memory score was calculated by dividing the total number of digits from fully correctly recalled sequences by the maximum attainable score, working memory score here was defined by the highest attained digit span. This was to allow for the possible error margin that may have arisen for participants who had correctly retained the backward digit span, but whose answers were automatically detected as incorrect because they accidentally mistyped the digits. Because in Chapter 2, participants repeated the digit spans out loud and were allowed to correct themselves (if applicable), it was deemed appropriate to allow for this error margin.

5.3.3.5 Word identification (generalization)

On day 2, a word generalization task was conducted. The set-up was identical to the word identification task on day 1, except that the female voice was used instead of the male voice for the audio stimuli. After the word generalization task, participants filled out a debriefing form on which they responded to questions about their performance, their concentration, and their general experience during the experiment. A selection of the debriefing questions and their responses is provided in Appendix 5.8–10.

5.3.4 Statistical procedures

All analyses were performed in *R 4.1.1* (R Core Team, 2021). Figures were generated with the *ggplot2* package (Wickham, 2016). I present descriptive statistics and results from Bayesian inference to assess the effects of L1 tonal status and extralinguistic factors on performance in the tone categorization and word identification tasks. Null responses and responses with unnaturally fast reaction times (< 250 ms) were removed, excluding 0.62% and 0.45% from each task, respectively. I report accuracy and reaction time data for the tone categorization task, but only accuracy data from the word generalization task, as this was at the end of the word training phase, following Chapter 2. Reaction times were analyzed for correctly categorized trials only, and outliers (2.5 SDs from the mean of each subject) were removed.

Models were fitted using the *brms* package (Bürkner, 2018). Bayesian inference offers an alternative to frequentist analyses in that it includes a prior specification of assumed beliefs of a model parameter. The output of a Bayesian model is a posterior distribution, which contains updated model parameters after having been fitted on the data. This posterior distribution generates 95% Credible Intervals (CrIs), which indicate the range of parameter values within which one can be 95% certain that the true parameter value lies. The posterior also generates maximum probabilities of effect, which describe the probability that a parameter is positive or negative. I use the guidance by Nicenboim et al. (2018, p. 1079) to interpret Bayesian model results. Namely, I assume ‘compelling evidence’ for an effect if zero lies outside the 95% CrI. I assume ‘weak evidence’ for an effect if zero is included in the 95% CrI but the maximum probability of effect is relatively high. Finally, I assume ‘no evidence’ for an effect if the maximum probability of effect is near 50%.

The choice for Bayesian analysis over frequentist methods in the present study was motivated by several aspects of the data and research questions for which Bayesian analysis offers a (more appropriate) alternative over frequentist methods. In particular, Bayesian analysis is appropriate for complex models with a relatively small sample size, it limits the chance of model convergence failure and of Type I (false positive) and Type II (false negative) errors, and by focusing on probability distributions, it allows for more varying degrees of result interpretation than binary outcomes (i.e., p-values) as is the case in frequentist analyses (Haendler et al., 2020).

Following common practice (Haendler et al., 2020; Vasishth et al., 2018), models were constructed using weakly informative (regularizing) priors with the mean centered around zero and a standard deviation of 10 for all population- and group-level regression coefficients and LKJ(2) for correlation priors. Four sampling chains with 3000 iterations each were run, with 1500 warm-up iterations. Models for accuracy (dependent variable = correct/incorrect) were fitted with a Bernoulli distribution and models for reaction time (dependent variable = RT in milliseconds) with a shifted lognormal distribution. The use of shifted lognormal distribution in Bayesian models for reaction time data has been suggested to be a preferable distribution to normal distribution, as reaction times are rarely normally distributed. Shifted lognormal distributions make it possible to work with the raw reaction time data and should therefore produce more accurate estimates (Nicenboim et al., 2018, pp. 1078–1079). Model diagnosis was carried out by observing Rhat values (ensuring these were close to 1), and by inspecting posterior draws using the pp_draws() command of the *brms* package.

The model for tone categorization (accuracy and RT) contained fixed effects for *L1* (Dutch, Swedish, Japanese, Thai; contrast-coded), *Tone* (Level, Fall, Peak; contrast-coded), *Musical Experience* (Years of formal practice; centered and scaled), *Working Memory* (Digit span score; centered and scaled), and two-way interactions with *L1* and each of the fixed effects. The random effects structure contained a by-subject random slope for *Tone* and a random intercept for *Item*.

The model structure for word identification (accuracy) was the same as for the tone categorization task, but additionally contained a fixed effect of *Pitch Aptitude*, (Mean accuracy scores in the tone categorization task; centered and scaled) and an *L1:Pitch Aptitude* interaction to assess the effect of pre-lexical tone perception on tone word learning.

Planned comparisons between tones per group, between groups per tone, and for the effects of musical experience, WM, and pitch aptitude per L1 were carried out using the *emmeans* package (Lenth, 2020). I assume compelling evidence for differences between factors and for effects of factors per L1 if there is compelling evidence for the overall interaction, and if zero is not included in the 95% highest posterior density (HPD) as calculated by the *emmeans* package.

5.4 Results

In the results section, I report descriptive statistics, results from Bayesian inference, and error patterns in the tone categorization and word identification tasks.

5.4.1 Tone categorization

5.4.1.1 Descriptive statistics

Table 18 and Figure 39 show accuracies and mean reaction times (RTs) for the tone categorization task. A visual inspection reveals no stark difference between L1s, either in terms of accuracies or RTs. Unlike in Chapter 2, in which performance in tone categorization was at ceiling, there was more individual variability in accuracy scores in the present tone categorization task. Therefore, in the following sections I will report results from Bayesian inference based on both the accuracy and the RT results.

Table 18

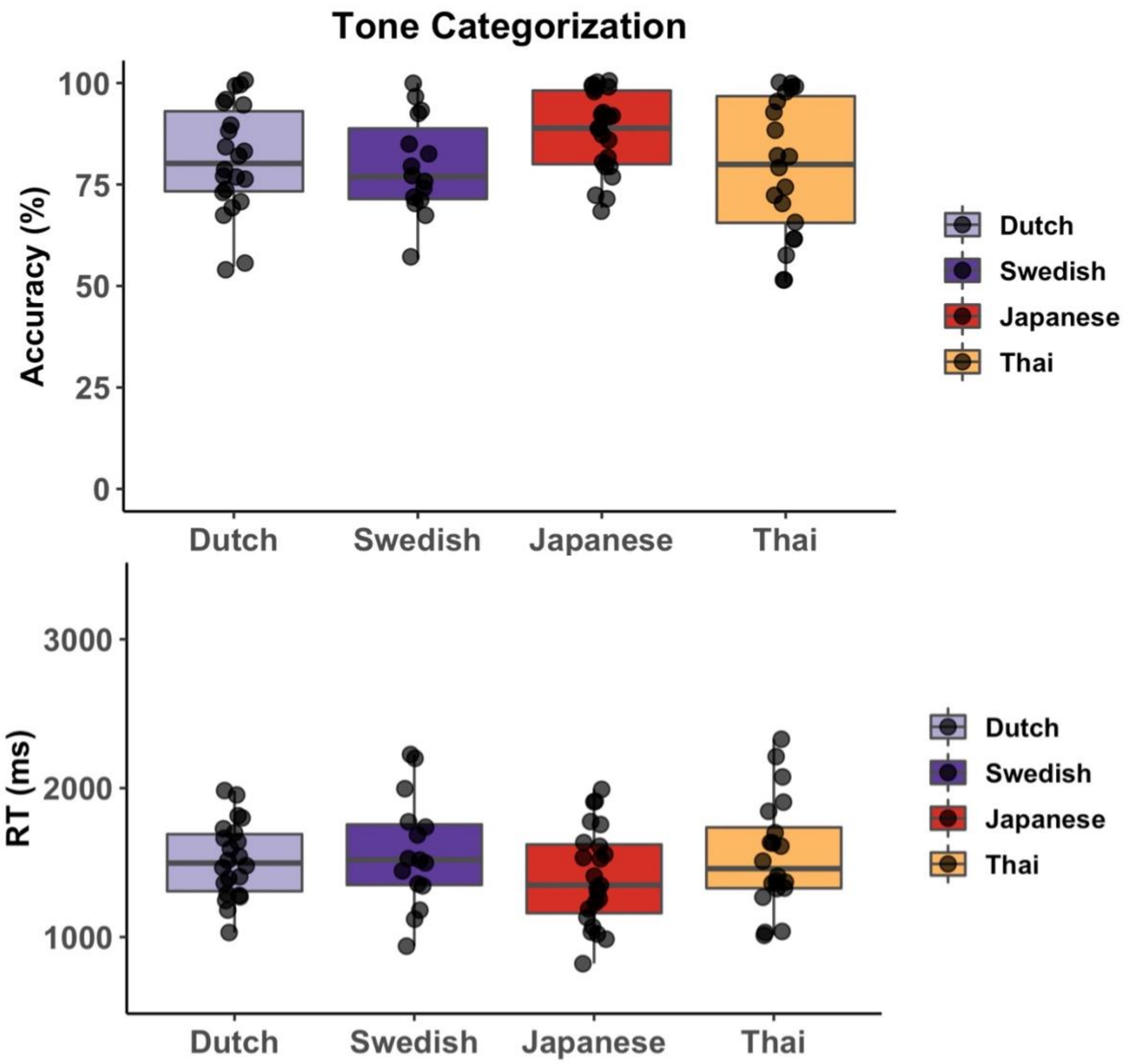
Descriptive statistics for tone categorization task.

	Dutch	Swedish	Japanese	Thai
Accuracy (%)	81.3 (13.3)	79.5 (12.0)	88.0 (10.1)	79.2 (17.3)
Reaction time (ms)	1520 (253)	1570 (375)	1410 (329)	1550 (377)

Values are means with standard deviations in brackets.

Figure 39

Accuracy and RT for tone categorization per L1.



5.4.1.2 Bayesian inference (accuracy)

The full posterior distribution for tone categorization accuracy is provided in Appendix 5.5. The model revealed compelling evidence for an effect of *Musical Experience* (*b* = 0.74 [0.37, 1.12]), and for *L1:Tone* (*b* = -0.58 [-1.17, -0.01]) and *L1:WM* (*b* = 0.70 [0.16, 1.26]) interactions. There was weak, though near-compelling evidence for an *L1:Musical*

Experience interaction, for which zero lay outside the 95% credible interval (*b* = -0.65 [-1.36, 0.02]) but for which the probability of direction was 97.02%.

To investigate these interactions in more detail, relevant multiple comparisons were conducted (Table 19–21). I summarize main findings in the text hereunder. For visualization of the effects and interactions, Figure 40 plots predicted tone categorization accuracy per tone and L1. Figure 41 plots predicted tone categorization accuracy against musical experience and WM.

Comparisons between tones per L1 revealed compelling evidence that in all groups except the Thai group, level tones were more likely to be accurately categorized than falling and peaking tones. Within the Thai group, there was compelling evidence that level tones were more likely to be accurately categorized than falling tones, and that peaking tones were more likely to be categorized than falling tones.

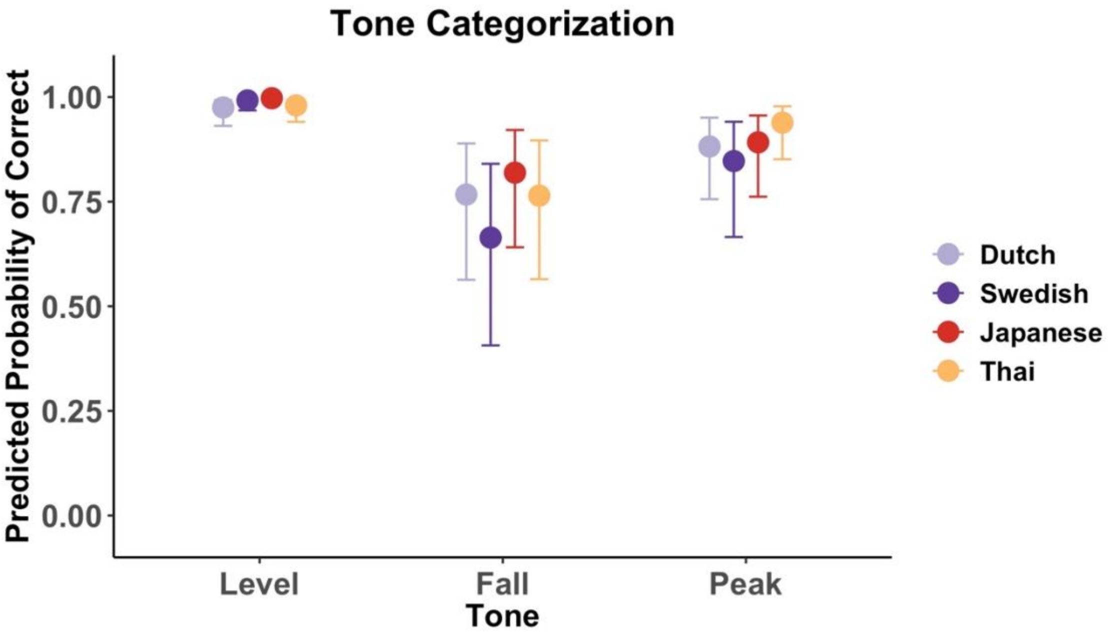
Comparisons between L1s per tone revealed compelling evidence that Japanese speakers were more likely than Dutch and Thai speakers to accurately categorize level tones. This comparison should be interpreted with caution, though, because Japanese speakers' performance for level tone categorization was at ceiling.

The model revealed compelling evidence that musical experience increased tone categorization likelihood across all groups, but there was weak evidence for a *L1:Musical Experience* interaction. The estimates of the effects of musical experience per L1 in Table 21 must therefore be interpreted with caution. Based on the estimate sizes and the visualization in Figure 41, it appears that the effect of musical experience was strongest in Thai and Dutch participants, and less so in Swedish and Japanese participants.

Comparisons for the effect of WM per L1 revealed compelling evidence that WM led to higher tone categorization likelihood for the Japanese group. There was only weak evidence for a facilitative effect of WM in the other groups, as zero was included in the 95% credible intervals.

Figure 40

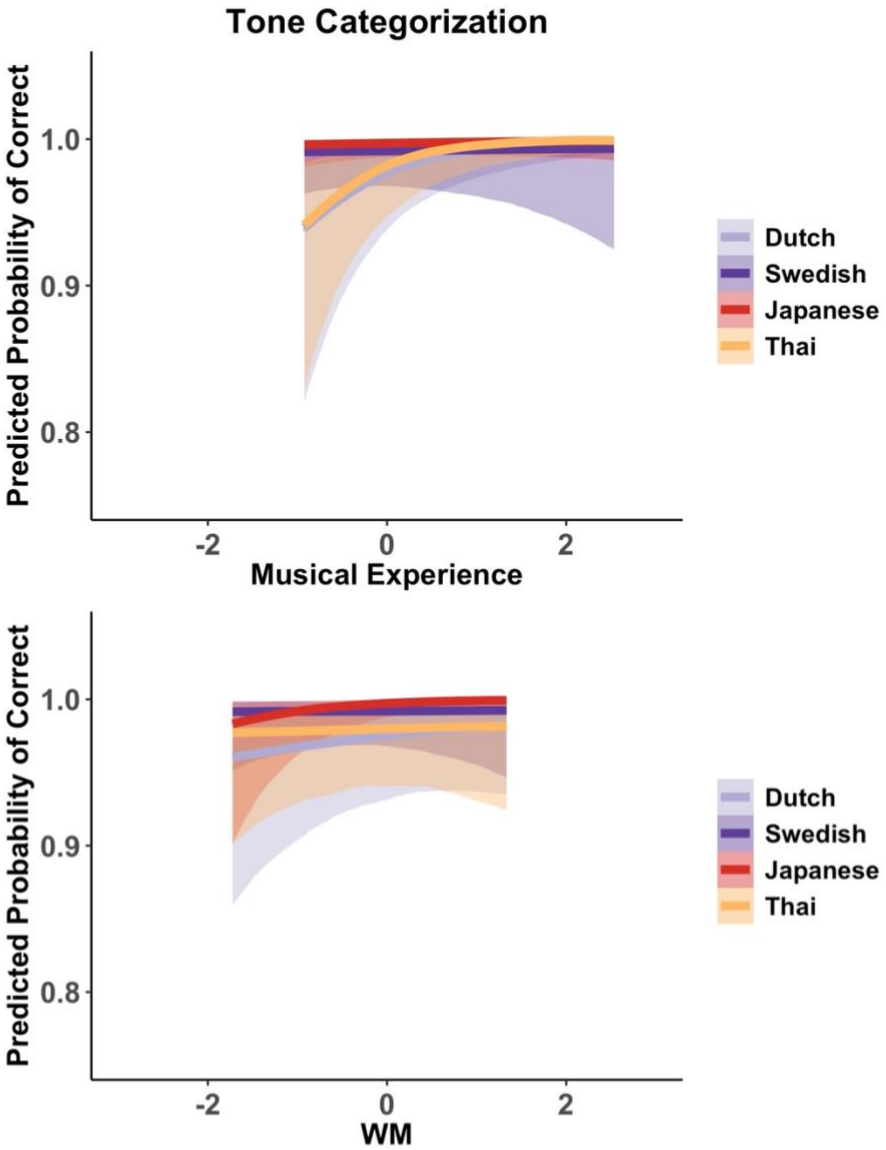
Predicted probability of tone categorization accuracy per tone and L1.



Bars represent 95% CrIs.

Figure 41

Predicted tone categorization accuracy against musical experience and WM (centered and scaled).



Shading ribbons represent 95% CrIs.

Table 19

Multiple comparisons between tones per L1 for tone categorization accuracy.

Tone Categorization (accuracy) Parameter	Estimate	95% Cr.I.
DUTCH		
Level-Fall	1.36	[1.36 ; 3.72]
Level-Peak	0.41	[0.41 ; 3.08]
Fall-Peak	-1.88	[-1.88 ; 0.22]
SWEDISH		
Level-Fall	2.66	[2.66 ; 5.66]
Level-Peak	1.48	[1.48 ; 4.84]
Fall-Peak	-2.20	[-2.20 ; 0.07]
JAPANESE		
Level-Fall	2.77	[2.77 ; 6.21]
Level-Peak	1.97	[1.97 ; 5.69]
Fall-Peak	-1.68	[-1.68 ; 0.46]
THAI		
Level-Fall	1.50	[1.50 ; 3.90]
Level-Peak	-0.29	[-0.29 ; 2.53]
Fall-Peak	-2.65	[-2.65 ; -0.40]

Table 20

Multiple comparisons between L1s per tone for tone categorization accuracy.

Tone Categorization (accuracy) Parameter	Estimate	95% Cr.I.
LEVEL TONES		
Dutch-Swedish	-1.16	[-2.77 ; 0.51]
Dutch-Japanese	-2.20	[-3.99 ; -0.44]
Dutch-Thai	-0.23	[-1.65 ; 1.08]
Swedish-Japanese	-1.05	[-3.14 ; 1.03]
Swedish-Thai	0.93	[-0.75 ; 2.62]
Japanese-Thai	1.96	[0.32 ; 3.86]
FALLING TONES		
Dutch-Swedish	0.51	[-0.60 ; 1.63]
Dutch-Japanese	-0.33	[-1.31 ; 0.72]
Dutch-Thai	0.01	[-1.06 ; 1.07]
Swedish-Japanese	-0.82	[-2.09 ; 0.24]
Swedish-Thai	-0.51	[-1.70 ; 0.65]
Japanese-Thai	0.32	[-0.72 ; 1.37]
PEAKING TONES		
Dutch-Swedish	0.32	[-0.84 ; 1.43]
Dutch-Japanese	-0.09	[-1.13 ; 0.94]
Dutch-Thai	-0.71	[-1.82 ; 0.38]
Swedish-Japanese	-0.39	[-1.56 ; 0.75]
Swedish-Thai	-1.02	[-2.24 ; 0.16]
Japanese-Thai	-0.61	[-1.75 ; 0.48]

Table 21

Effects of musical experience and WM on tone categorization accuracy.

Tone Categorization (accuracy) Parameter	Estimate	95% Cr.I.
Musical Experience		
Dutch	1.15	[0.44 ; 1.82]
Swedish	0.09	[-0.69 ; 0.88]
Japanese	0.32	[-0.43 ; 1.08]
Thai	1.38	[0.62 ; 2.29]
WM		
Dutch	0.28	[-0.28 ; 0.88]
Swedish	0.05	[-0.74 ; 0.90]
Japanese	1.06	[0.45 ; 1.69]
Thai	0.07	[-0.64 ; 0.78]

5.4.1.3 Bayesian inference (reaction times)

The full posterior distribution is provided in Appendix 5.6. The model revealed compelling evidence for *L1:Tone* (*b* = -0.04 [-0.07, 0.00]) and *L1:Musical Experience* (*b* = 0.09 [0.00, 0.18]) interactions. There was weak evidence for an *L1:WM* interaction for which zero lay outside the 95% credible interval (*b* = -0.05 [-0.14, 0.03]) and for which the probability of direction was 89.53%.

To investigate these interactions in more detail, relevant multiple comparisons were conducted (Table 22–24). I summarize main findings in the text hereunder. For visualization of the effects and interactions, Figure 42 plots predicted tone categorization RT per tone by L1. Figure 43 plots predicted tone categorization RT against musical experience and WM.

Comparisons between tones within groups revealed compelling evidence that in all groups, level tones were categorized faster than falling and peaking tones.

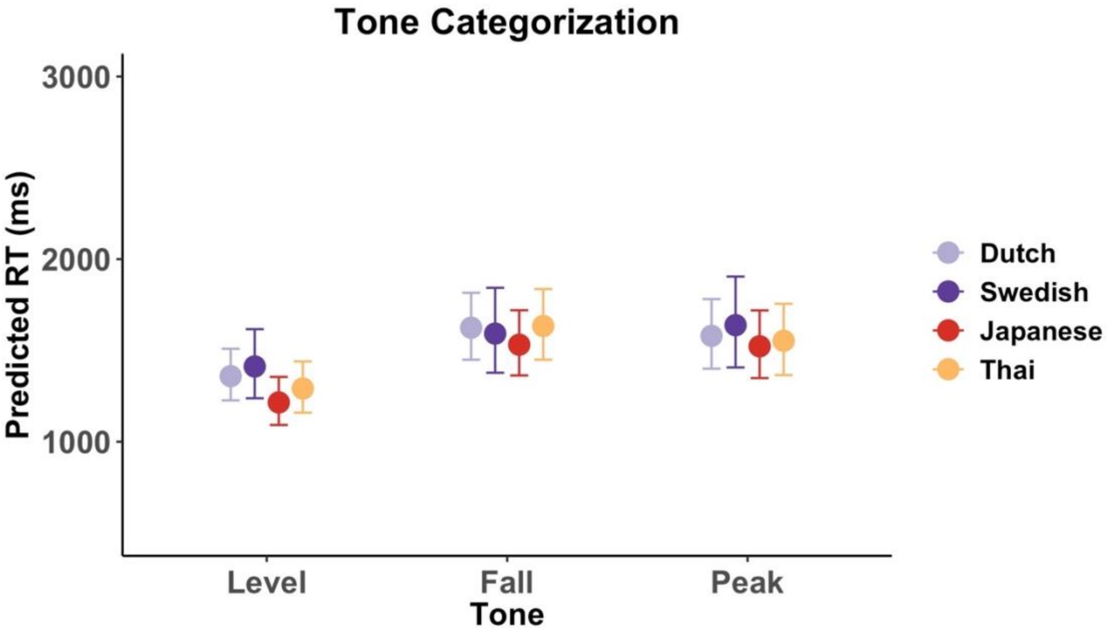
There was no compelling evidence for differences between L1s per tone, although there was weak evidence that Japanese speakers identified level tones faster than did Swedish speakers (*b* = 0.15 [-0.01, 0.31]).

An observation of the estimates in Table 24 and the plots in Figure 43 provides weak evidence that musical experience led to faster reaction times, except in the Japanese group, for which there was weak evidence that musical experience led to slower reaction times.

There was weak evidence for an *L1:WM* interaction, and the estimates per L1 for WM in Table 24 should therefore be interpreted with caution. Based on the estimate sizes and the plots in Figure 43, it appears that WM did not lead to faster RTs in tone categorization, although it may have done so for the Japanese group.

Figure 42

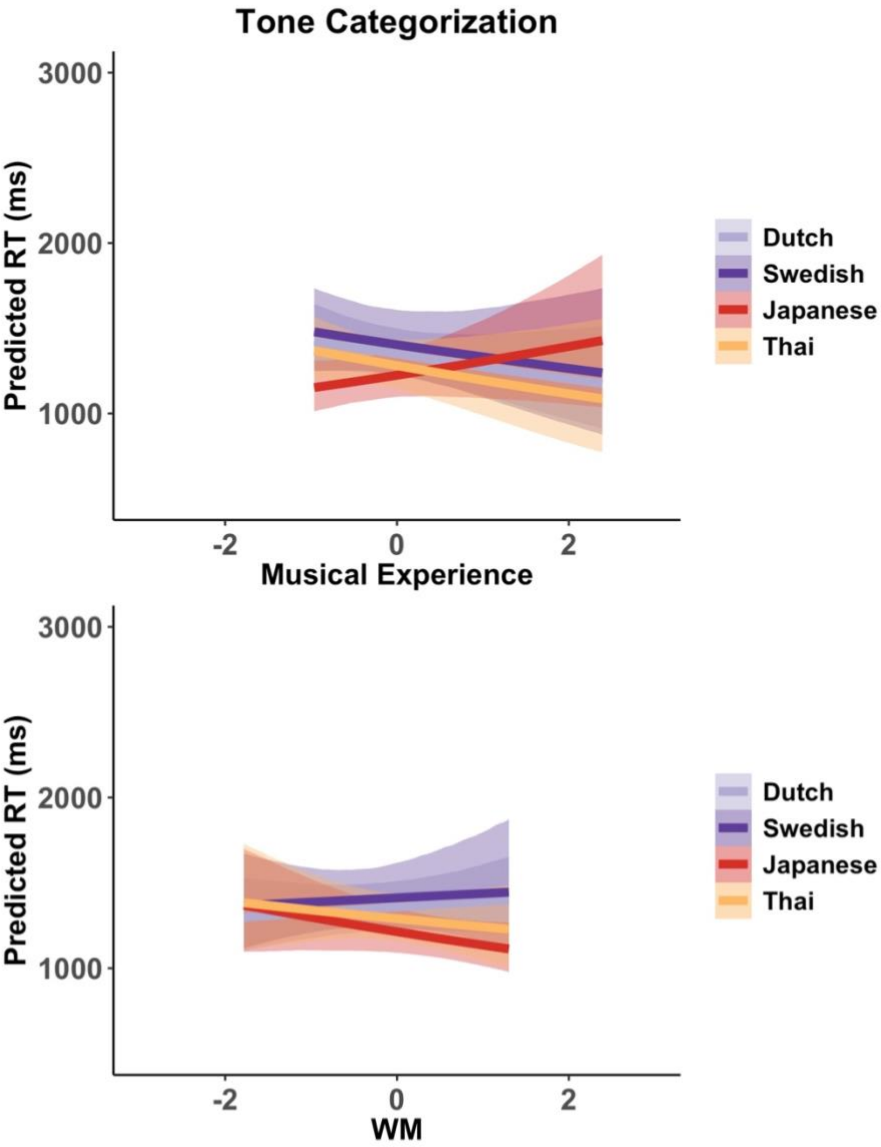
Predicted tone categorization RT per tone by L1.



Bars represent 95% CrIs.

Figure 43

Predicted categorization RT against musical experience and WM (centered and scaled).



Shading ribbons represent 95% CrIs.

Table 22

Multiple comparisons between L1s per tone for tone categorization RT.

Tone Categorization (RT) Parameter	Estimate	95% Cr.I.
DUTCH		
Level-Fall	-0.18	[-0.26 ; -0.09]
Level-Peak	-0.15	[-0.24 ; -0.06]
Fall-Peak	0.03	[-0.05 ; 0.11]
SWEDISH		
Level-Fall	-0.12	[-0.21 ; -0.02]
Level-Peak	-0.15	[-0.25 ; -0.05]
Fall-Peak	-0.03	[-0.12 ; 0.06]
JAPANESE		
Level-Fall	-0.23	[-0.31 ; -0.15]
Level-Peak	-0.23	[-0.32 ; -0.14]
Fall-Peak	0.01	[-0.07 ; 0.08]
THAI		
Level-Fall	-0.23	[-0.33 ; -0.15]
Level-Peak	-0.18	[-0.28 ; -0.09]
Fall-Peak	0.05	[-0.03 ; 0.13]

Table 23

Multiple comparisons between tones per L1 for tone categorization RT.

Tone Categorization (RT) Parameter	Estimate	95% Cr.I.
LEVEL TONES		
Dutch-Swedish	-0.04	[-0.20 ; 0.12]
Dutch-Japanese	0.11	[-0.03 ; 0.25]
Dutch-Thai	0.05	[-0.09 ; 0.20]
Swedish-Japanese	0.15	[-0.01 ; 0.32]
Swedish-Thai	0.09	[-0.07 ; 0.25]
Japanese-Thai	-0.06	[-0.20 ; 0.09]
FALLING TONES		
Dutch-Swedish	0.02	[-0.15 ; 0.19]
Dutch-Japanese	0.06	[-0.11 ; 0.20]
Dutch-Thai	0.00	[-0.17 ; 0.14]
Swedish-Japanese	0.04	[-0.13 ; 0.22]
Swedish-Thai	-0.03	[-0.20 ; 0.16]
Japanese-Thai	-0.06	[-0.23 ; 0.09]
PEAKING TONES		
Dutch-Swedish	-0.04	[-0.23 ; 0.14]
Dutch-Japanese	0.04	[-0.13 ; 0.19]
Dutch-Thai	0.02	[-0.15 ; 0.18]
Swedish-Japanese	0.07	[-0.11 ; 0.26]
Swedish-Thai	0.05	[-0.14 ; 0.24]
Japanese-Thai	-0.02	[-0.19 ; 0.14]

Table 24

Effects of musical experience and WM on tone categorization RT.

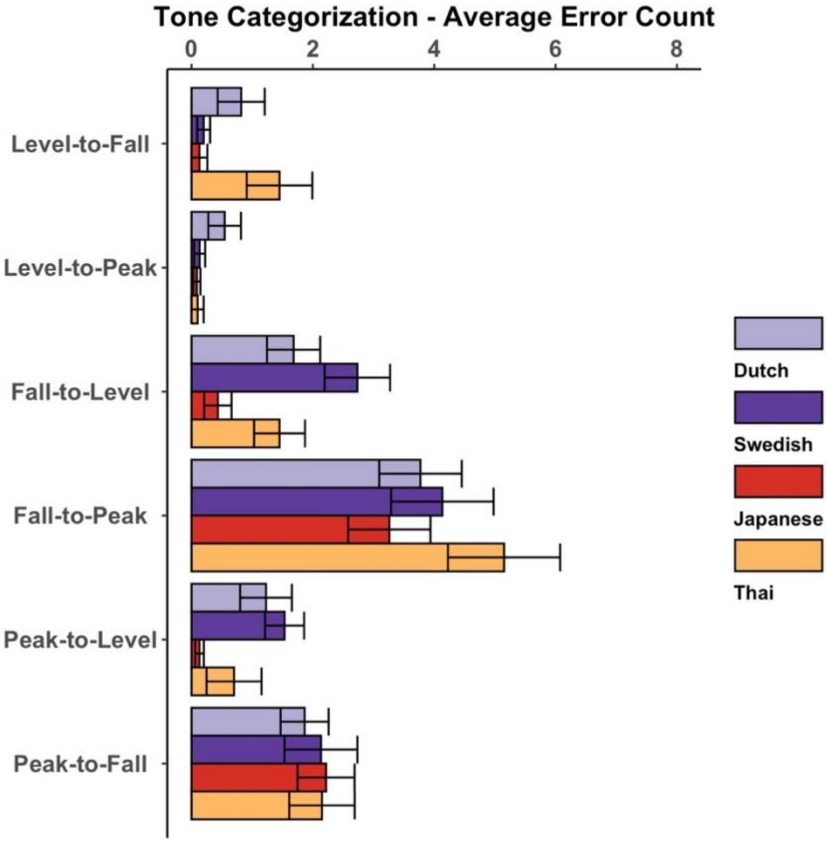
Tone Categorization (RT) Parameter	Estimate	95% Cr.I.
Musical Experience		
Dutch	-0.05	[-0.15 ; 0.04]
Swedish	-0.05	[-0.17 ; 0.06]
Japanese	0.06	[-0.04 ; 0.17]
Thai	-0.07	[-0.19 ; 0.05]
WM		
Dutch	0.03	[-0.05 ; 0.11]
Swedish	0.02	[-0.10 ; 0.14]
Japanese	-0.07	[-0.16 ; 0.02]
Thai	-0.04	[-0.16 ; 0.08]

5.4.1.4 Error type analysis

This section presents an analysis of error type to further explore the nature of tone categorization errors. Figure 44 shows the distribution of error types. A visual inspection suggests that participants predominantly miscategorized falling as peaking tones, and vice versa.

Figure 44

Error types in tone categorization.



Counts are averaged over subject. Error bars = +/- 1 SE.

Following the methodology in Chapters 2–3 , the counts of error types were subjected to a zero-inflated generalized linear mixed effect model with an *L1:Error Type* interaction as fixed effect and a random intercept for *Subject* to observe whether certain error types were significantly more likely to occur than others. This revealed a significant *L1:Error Type* interaction (χ^2 = 64.488, *df*(15), *p* < 0.001). Subsequent Bonferroni-corrected multiple comparisons showed that indeed, some error patterns were significantly more likely to occur than others, both within and across L1 groups. Because of the many multiple comparisons (92 in total), the significant comparisons ($p < 0.05$) are summarized in Table 25–26. The full output is provided in Appendix 5.13–15.

The comparisons between error types per L1 revealed that fall-to-peak errors were significantly more likely to occur than other error types in most groups. This was particularly the case for the Thai group, for which the fall-to-peak errors were more likely to occur in comparison to all other error types. Overall, and in line with the predictions, dynamic-dynamic errors (i.e., errors involving a confusion between a falling and a peaking tone) were significantly more likely to occur than static-dynamic errors (i.e., errors involving a confusion with a level tone). This confirms that participants predominantly confused falling and peaking tones with one another.

The comparisons between L1s per error type revealed no clear differences between groups, although it is noteworthy that Dutch and Swedish speakers were more likely than Japanese speakers to misidentify falling or peaking tones as level tones. The results from the level-to-fall error types (which Dutch speakers made more than did Japanese, and Thai speakers made more than did Swedish and Japanese) should be interpreted with caution because overall, participants did not very often incorrectly categorize level tones in the first place, as performance for level tone categorization was at ceiling.

Table 25

Significant count comparisons between error types per L1.

Dutch			Swedish			Japanese			Thai		
Level-to-Fall	>		Level-to-Fall	>		Level-to-Fall	>		Level-to-Fall	>	Level-to-Peak
Level-to-Peak	>		Level-to-Peak	>		Level-to-Peak	>		Level-to-Peak	>	
Fall-to-Level	>	Level-to-Peak	Fall-to-Level	>	Level-to-Fall	Fall-to-Level	>		Fall-to-Level	>	Level-to-Peak
					Level-to-Peak						
Fall-to-Peak	>	Level-to-Fall	Fall-to-Peak	>	Level-to-Fall	Fall-to-Peak	>	Level-to-Fall	Fall-to-Peak	>	Level-to-Fall
		Level-to-Peak			Level-to-Peak			Level-to-Peak			Level-to-Peak
		Peak-to-Level			Peak-to-Level			Peak-to-Level			Fall-to-Level
											Peak-to-Level
											Peak-to-Fall
Peak-to-Level	>		Peak-to-Level	>	Level-to-Fall	Peak-to-Level	>		Peak-to-Level	>	
					Level-to-Peak						
Peak-to-Fall	>	Level-to-Peak	Peak-to-Fall	>	Level-to-Fall	Peak-to-Fall	>	Level-to-Fall	Peak-to-Fall	>	Level-to-Peak
					Peak-to-Fall			Level-to-Peak			
								Fall-to-Level			
								Fall-to-Peak			

> indicates significantly higher likelihood to occur according to the glmmTMB model.

Table 26

Significant count comparisons between L1s per error type.

Level-to-Fall			Level-to-Peak	Fall-to-Level			Fall-to-Peak	Peak-to-Level			Peak-to-Fall
Dutch	>	Japanese		Dutch	>	Japanese		Dutch	>	Japanese	
Thai	>	Swedish		Swedish	>	Japanese		Swedish	>	Japanese	
		Japanese									

> indicates significantly higher likelihood to occur according to the glmmTMB model

5.4.2 Word generalization

5.4.2.1 Descriptive statistics

Figure 45 shows accuracy and RT for the word identification task across the different sessions. These show that over the course of time, participants improved their word identification accuracies. There appears to be no clear change in participants' reaction times. As mentioned earlier, I report accuracy scores of the word identification task at the end of the experiment (in the generalization task on day 2) to assess participants' performance in tone word learning.

Figure 45

Accuracy and RT for word identification per session by L1.

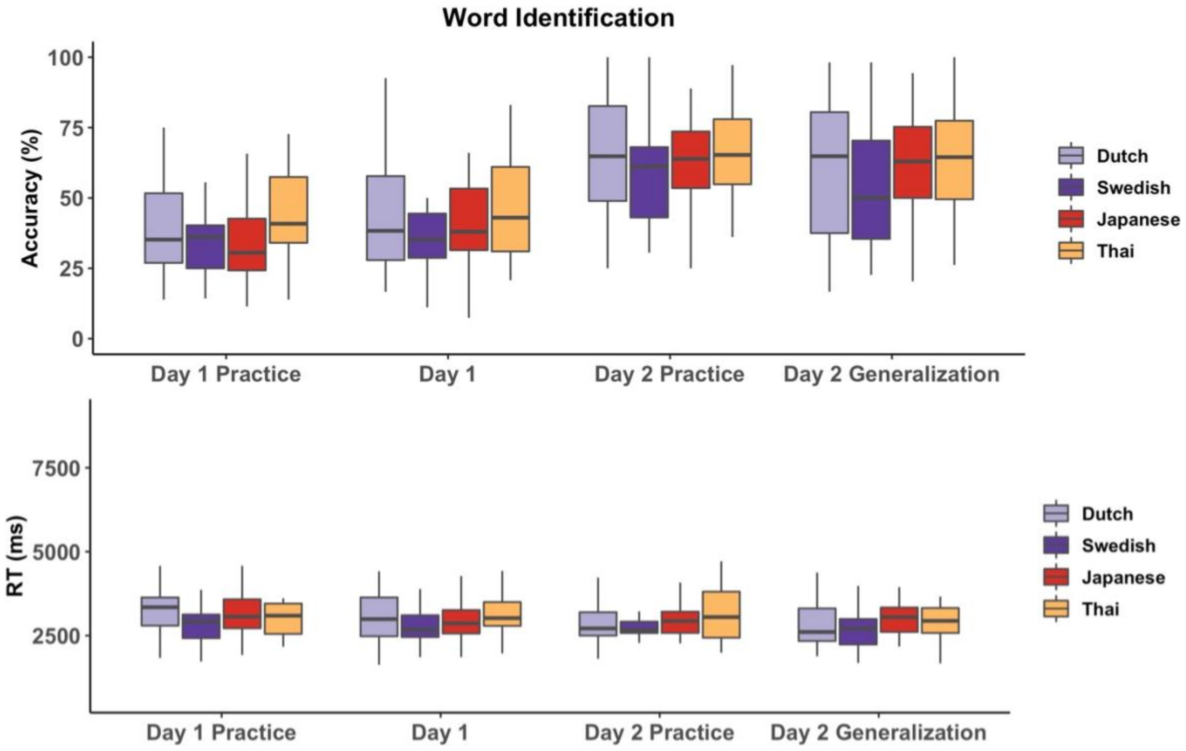


Table 27 summarizes the descriptive statistics for the word generalization task. There appears to be no stark difference in accuracy nor in RT between L1s. This is further visualized in Figure 46. In addition, most of the errors that participants made were "Tone-Only Errors", suggesting that they had retained the segmental, but not the tonal properties of the pseudowords. The proportion of tone-only errors did not appear to vary notably between participant groups.

Table 27

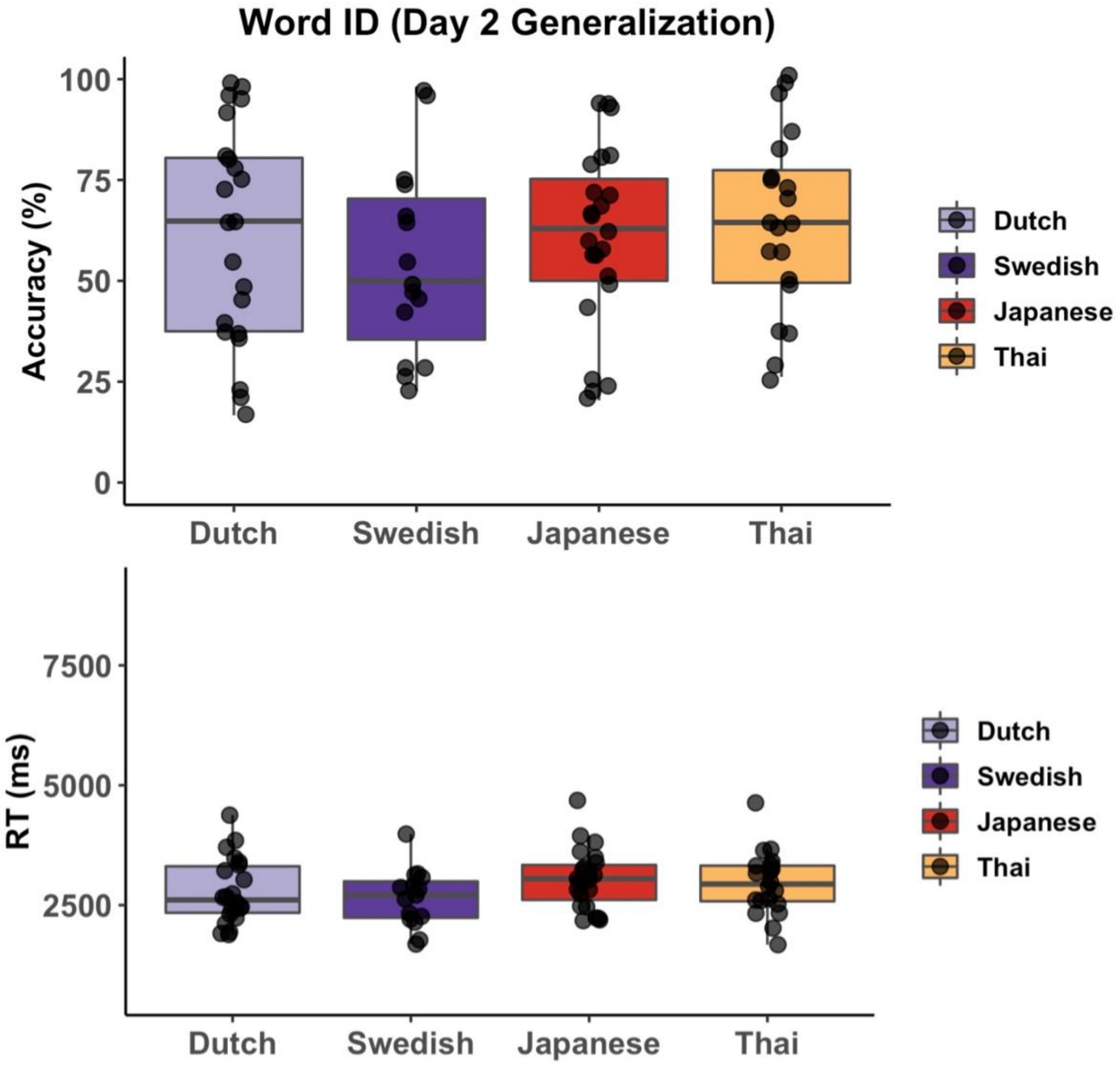
Descriptive statistics for the word identification task (day 2 generalization).

	Dutch	Swedish	Japanese	Thai
Accuracy (%)	61.4 (27.0)	54.9 (24.3)	60.5 (22.6)	64.7 (22.2)
Reaction time (ms)	2780 (681)	2650 (598)	3040 (622)	2950 (662)
% of tone-only errors	81.1 (21.6)	75.4 (22.5)	71.0 (22.5)	80.6 (25.8)

Values are means with standard deviations in brackets.

Figure 46

Accuracy and RT for word identification (day 2 generalization) per L1.



5.4.2.2 Bayesian inference (accuracy)

The full posterior distribution is provided in Appendix 5.7. The model revealed compelling evidence for an effect of *Pitch Aptitude* (*b* = 0.91 [0.44, 1.36]), and for *L1:Tone* (*b* = 0.45 [0.13, 0.78]) and *L1:Musical Experience* interactions (*b* = 0.68 [0.11, 1.28]). There was weak evidence for an *L1:WM* interaction (*b* = -0.36 [-0.96, 0.22]), and the probability of direction was 88.98%. There was weak evidence for an *L1:Pitch Aptitude* interaction (*b* = 0.60 [-0.10, 1.33]), and the probability of direction was 95.40%.

To investigate the interactions in more detail, relevant multiple comparisons were conducted (Table 28–30). I summarize main findings in the text hereunder. For visualization of the effects and interactions, Figure 47 plots predicted word identification accuracy per tone by L1. Figure 48 plots predicted word identification accuracy against musical experience, WM, and pitch aptitude.

Comparisons between tones within groups revealed compelling evidence that words with level tones were more likely to be correctly identified than words with falling and peaking tones on the first syllable. An exception to this was the Swedish group, for which there was only weak evidence that level tone words were more likely to be identified than falling tone words (*b* = 0.58 [-0.41, 1.62]), and virtually no evidence that level tones were more likely to be identified than peaking tone words, with zero falling roughly in the middle of the credible interval (*b* = -0.13 [-1.23, 1.10]). There was further compelling evidence that words with peaking tones were more likely to be correctly identified than words with falling tones, except in the Dutch group, for which zero lay in the credible interval (*b* = -0.59 [-1.20, 0.05]).

Comparisons between L1s per tone revealed compelling evidence that Thai speakers were more likely than Swedish speakers to correctly identify words with level tones, and that Dutch speakers were more likely than Japanese speakers to correctly identify words with falling tones.

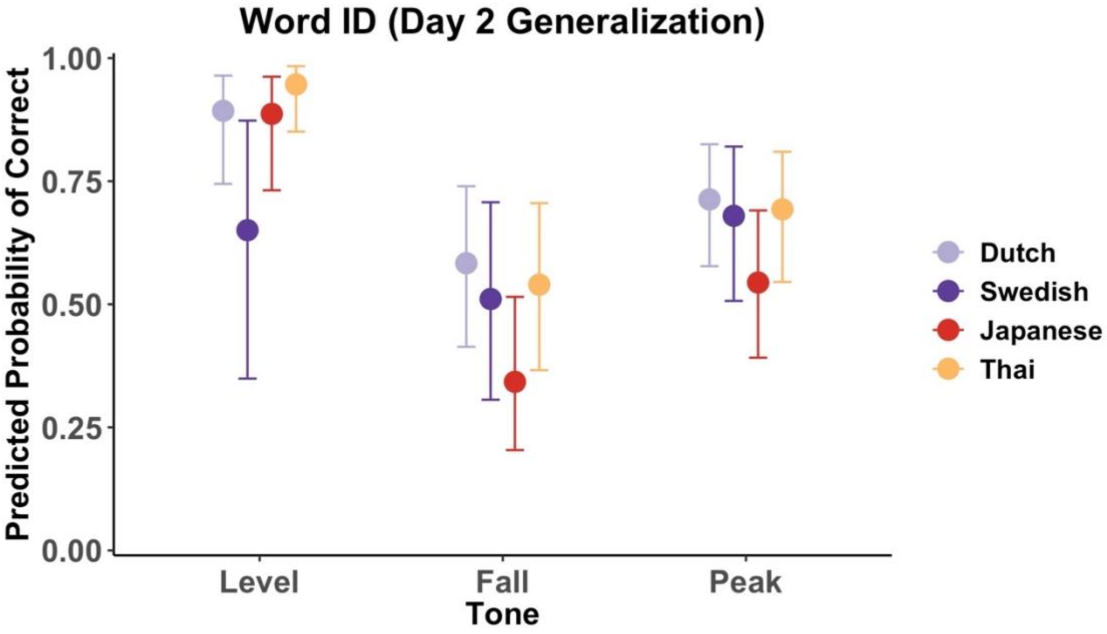
Estimates for the effect of musical experience per L1 revealed compelling evidence that musical experience facilitated word identification in the Swedish group.

There was only weak evidence for an *L1:WM* interaction, and effects of WM per L1 should therefore be interpreted with caution. Recall that there was no compelling evidence for an overall effect of WM on word identification (*b* = 0.10 [-0.27, 0.44]). Based on the estimates in Table 30 and the plots in Figure 48, it appears that WM did not facilitate word identification across the board, although it appears that it may have done so for Japanese and Thai speakers.

There was only weak evidence for a *L1:Pitch Aptitude* interaction, and effects of pitch aptitude per L1 should therefore be interpreted with caution. Recall that there was compelling evidence for an overall effect of pitch aptitude on word identification (*b* = 0.91 [0.44, 1.36]). Based on the estimates in Table 30 and the plots in Figure 48, it appears that pitch aptitude led to a higher likelihood of correct word identification in all groups, although this effect may have been particularly strong for Dutch, and less so for Thai participants.

Figure 47

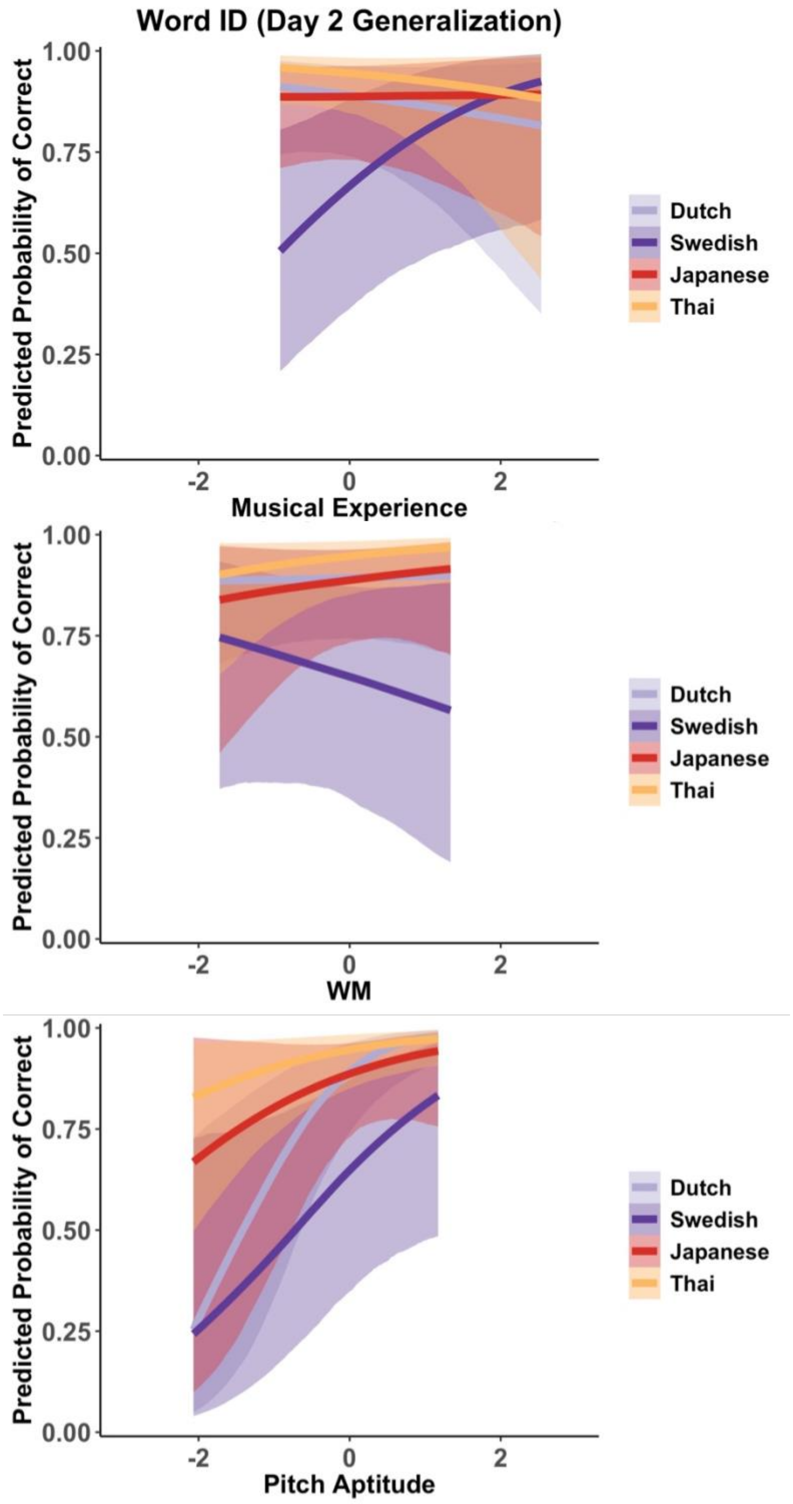
Predicted word identification accuracy per tone by L1.



Bars represent 95% CrIs.

Figure 48

Predicted word identification accuracy against musical experience, WM, and pitch aptitude (centered and scaled).



Shading ribbons represent 95% CrIs.

Table 28

Multiple comparisons between tones per L1 for word identification accuracy.

Word ID Day 2 Generalization (accuracy) Parameter	Estimate	95% Cr.I.
DUTCH		
Level-Fall	1.79	[0.88 ; 2.75]
Level-Peak	1.20	[0.14 ; 2.22]
Fall-Peak	-0.59	[-1.2 ; 0.05]
SWEDISH		
Level-Fall	0.58	[-0.41 ; 1.62]
Level-Peak	-0.13	[-1.23 ; 1.10]
Fall-Peak	-0.71	[-1.41 ; 0.00]
JAPANESE		
Level-Fall	2.72	[1.83 ; 3.65]
Level-Peak	1.89	[0.91 ; 2.97]
Fall-Peak	-0.83	[-1.40 ; -0.20]
THAI		
Level-Fall	2.71	[1.72 ; 3.69]
Level-Peak	2.06	[0.98 ; 3.22]
Fall-Peak	-0.65	[-1.31 ; -0.03]

Table 29

Multiple comparisons between L1s per tone for word identification accuracy.

Word ID Day 2 Generalization (accuracy) Parameter	Estimate	95% Cr.I.
LEVEL TONES		
Dutch-Swedish	1.48	[-0.18 ; 3.14]
Dutch-Japanese	0.06	[-1.43 ; 1.56]
Dutch-Thai	-0.76	[-2.32 ; 0.76]
Swedish-Japanese	-1.43	[-3.02 ; 0.26]
Swedish-Thai	-2.26	[-3.94 ; -0.61]
Japanese-Thai	-0.81	[-2.25 ; 0.81]
FALLING TONES		
Dutch-Swedish	0.29	[-0.78 ; 1.35]
Dutch-Japanese	1.00	[0.05 ; 1.94]
Dutch-Thai	0.18	[-0.72 ; 1.16]
Swedish-Japanese	0.69	[-0.35 ; 1.77]
Swedish-Thai	-0.12	[-1.16 ; 0.94]
Japanese-Thai	-0.81	[-1.76 ; 0.13]
PEAKING TONES		
Dutch-Swedish	0.16	[-0.73 ; 1.08]
Dutch-Japanese	0.74	[-0.09 ; 1.57]
Dutch-Thai	0.10	[-0.75 ; 0.89]
Swedish-Japanese	0.58	[-0.34 ; 1.53]
Swedish-Thai	-0.06	[-0.96 ; 0.87]
Japanese-Thai	-0.64	[-1.45 ; 0.20]

Table 30

Effects of musical experience, WM, and pitch aptitude on word identification accuracy.

Word ID Day 2 Generalization (accuracy) Parameter	Estimate	95% Cr.I.
Musical Experience		
Dutch	-0.25	[-0.95 ; 0.44]
Swedish	0.72	[0.08 ; 1.43]
Japanese	0.02	[-0.58 ; 0.60]
Thai	-0.43	[-1.23 ; 0.45]
WM		
Dutch	0.05	[-0.46 ; 0.53]
Swedish	-0.26	[-1.08 ; 0.42]
Japanese	0.23	[-0.60 ; 1.07]
Thai	0.39	[-0.27 ; 1.01]
Pitch Aptitude (Tone Categorization Accuracy)		
Dutch	1.51	[0.74 ; 2.35]
Swedish	0.84	[0.03 ; 1.76]
Japanese	0.65	[-0.58 ; 1.91]
Thai	0.62	[0.02 ; 1.18]

5.4.2.3 Tone-only error-analysis

As reported earlier in Table 27 (page 211), participants predominantly made tone-only errors on the day 2 generalization task. Four simple linear regressions confirmed that the number of tone-only errors significantly predicted the total number of errors and explained a large portion of variance in the Dutch [$F(1,20) = 49.490$, $p < 0.001$, $R^2 = .6978$], the Swedish, [$F(1,13) = 12.110$, $p = 0.004$, $R^2 = .4425$], the Japanese [$F(1,21) = 21.030$, $p < 0.001$, $R^2 = .4765$], and the Thai group [$F(1,18) = 17.190$, $p < 0.001$, $R^2 = .4600$]. This is further shown in Figure 49, which plots the number of errors against the number of tone-only errors. The distribution of tone-only error types is shown in Figure 50. A visual inspection suggests that participants predominantly confused falling with peaking tones and vice versa in word identification. This reflects the error types that were found in tone categorization.

Figure 49

Number of errors against number of tone-only errors in word identification.

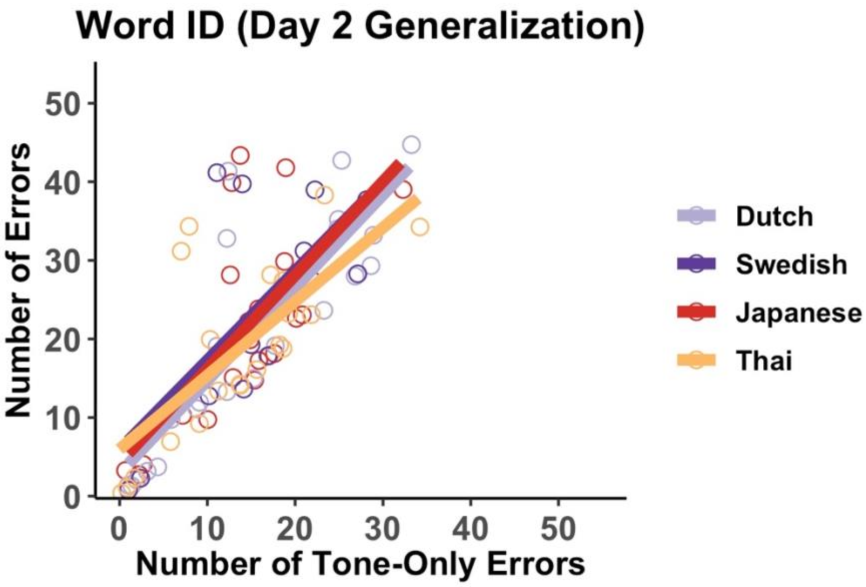
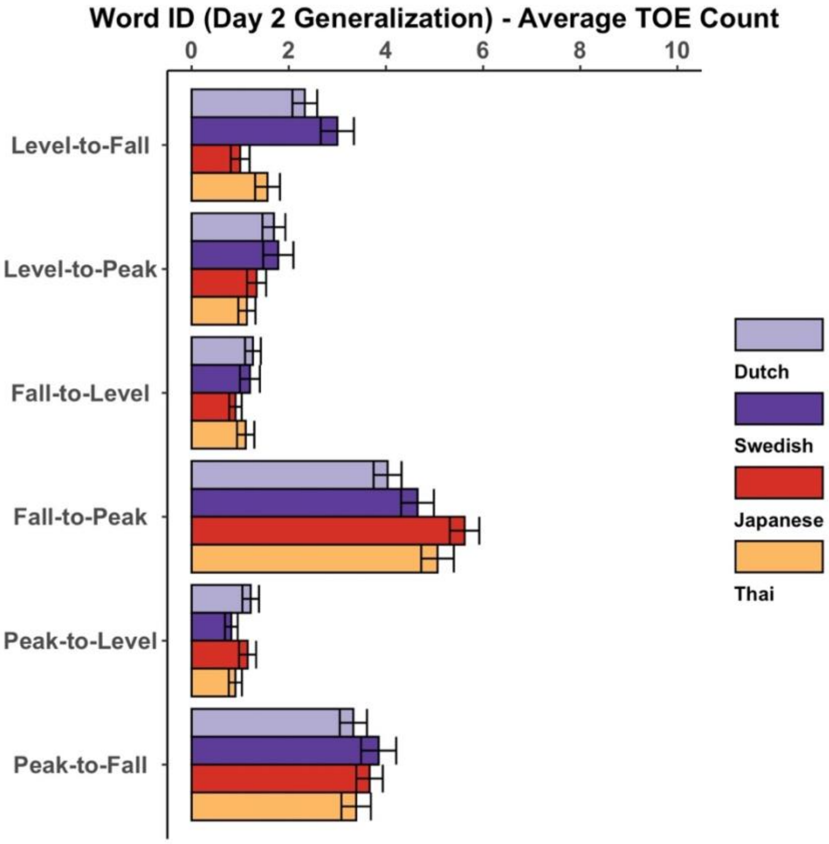


Figure 50

Tone-only error types in word identification.



Counts are averaged over subject. Error bars = +/- 1 SE.

To check for statistical significance of the occurrence of certain tone-only error types, a zero-inflated generalized linear mixed effects model with an *L1:Error Type* interaction was fitted on the counts of the tone-only error data. This revealed a significant *L1:Error Type* interaction (χ^2 = 54.162, *df*(15), *p* < 0.001). Subsequent Bonferroni-corrected multiple comparisons showed that indeed, some error patterns were significantly more likely to occur than others, both within and across L1 groups. Because of the many multiple comparisons (92 in total), the significant comparisons (*p* < 0.05) are listed in Table 31–32. The full output is provided in Appendix 5.15–16 .

The comparisons between error types per L1 revealed that fall-to-peak errors were significantly more likely to occur than other error types in most groups. This was particularly the case for the Japanese and Thai groups, for which the fall-to-peak errors were more likely to occur in comparison to all other error types (but for the Thai group, there was no statistical confirmation that fall-to-peak errors were more likely to occur than peak-to-fall errors). Overall, and in line with the predictions, and with what was found in tone categorization, dynamic-dynamic errors (i.e., errors involving a confusion between a falling and a peaking tone) were significantly more likely to occur than static-dynamic errors (i.e., errors involving a confusion with a level tone). This confirms that participants predominantly confused falling and peaking tones with one another at the word level.

The comparisons between L1s per error type revealed that both Dutch and Swedish speakers were more likely than Japanese and Thai speakers to misidentify words with level tones as words with peaking tones. Swedish and Thai speakers were more likely than Japanese speakers to misidentify words with falling tones as words with level tones.

Table 31

Significant count comparisons between tone-only error types per L1.

Dutch			Swedish			Japanese			Thai		
Level-to-Fall	>		Level-to-Fall	>	Peak-to-Level	Level-to-Fall	>		Level-to-Fall	>	
Level-to-Peak	>	Peak-to-Level	Level-to-Peak	>	Peak-to-Level	Level-to-Peak	>		Level-to-Peak	>	
Fall-to-Level	>		Fall-to-Level	>		Fall-to-Level	>		Fall-to-Level	>	
Fall-to-Peak	>	Level-to-Fall	Fall-to-Peak	>	Peak-to-Level	Fall-to-Peak	>	Level-to-Fall	Fall-to-Peak	>	Level-to-Fall
		Fall-to-Level						Level-to-Peak			Level-to-Peak
		Peak-to-Level						Fall-to-Level			Fall-to-Level
								Peak-to-Level			Peak-to-Level
								Peak-to-Fall			
Peak-to-Level	>	Level-to-Fall	Peak-to-Level	>		Peak-to-Level	>		Peak-to-Level	>	
Peak-to-Fall	>	Fall-to-Level	Peak-to-Fall	>		Peak-to-Fall	>	Level-to-Fall	Peak-to-Fall	>	Level-to-Peak
		Peak-to-Level						Level-to-Peak			Peak-to-Level
								Fall-to-Level			
								Peak-to-Level			

> indicates significantly higher likelihood to occur according to the glmmTMB model

Table 32

Significant count comparisons between L1 per tone-only error types.

Level-to-Fall			Level-to-Peak			Fall-to-Level			Fall-to-Peak			Peak-to-Level			Peak-to-Fall		
-	-	-	Dutch	>	Japanese	Swedish	>	Japanese	-	-	-	-	-	-	-	-	-
			Dutch	>	Thai	Thai	>	Japanese									
			Swedish	>	Japanese												
			Swedish	>	Thai												

> indicates significantly higher likelihood to occur according to the glmmTMB model.

5.5 Discussion

The present study explored whether L1 tonal status facilitates individual performance in non-native tone perception and word learning, whilst controlling for the effects of tone type, musical experience, and working memory. In the following, I will discuss the findings in light of the research question, predictions, and previous literature.

5.5.1 Facilitators in tone categorization

It was tentatively predicted that there would be a hierarchy in tone perception based on L1 tonal status, in which speakers of Thai (maximal tonal status) would outperform speakers of Swedish and Japanese (intermediate tonal status), who in turn would outperform speakers of Dutch (low tonal status). It was further predicted that this hierarchy would be observed across the board, given that the tone system of a low-level, a falling, and a peaking tone was hypothesized to be equally challenging for all speakers.

The findings from the tone categorization task lend no support for an overall facilitative effect of L1 tonal status on tone categorization performance, either in terms of accuracy or reaction times. The only compelling evidence that was found between groups was in categorization accuracy of level tones, which Japanese speakers identified better than did Swedish or Thai speakers. But this finding should be interpreted with caution given that participants performed at ceiling in the categorization of level tones. Overall, these findings suggest that participants' performance was relatively uniform and not strongly modulated by their L1.

When looking at the specific performance per tone type, it was found that the true difficulty of tone categorization lay in the dynamic-dynamic tone contrast between the falling and peaking tones. Across the board, participants identified falling and peaking tones less accurately, and even if they did categorize them correctly, they categorized them slower in comparison to level tones. The findings from the error pattern analysis further confirmed that participants predominantly misidentified falling tones as peaking tones, and vice versa. This confirms the hypothesis that static-dynamic contrasts in the tone categorization task were inherently easy, whereas dynamic-dynamic contrasts were inherently challenging, and that this was the case for Dutch, Swedish, Japanese, and Thai speakers.

There was however one observation that did suggest that L1 further modulated performance per tone. Unlike participants from other L1s, Thai participants were less likely to accurately categorize falling tones compared to peaking tones. Findings from the error pattern analysis further suggest that Thai participants were more likely to misidentify falling tones as peaking tones than vice versa. In other groups, there was no compelling evidence that falling tones were potentially more difficult to identify than peaking tones. Thai participants, in comparison to other speakers, may have had a stronger tendency to misidentify falling as peaking tones because the peaking tone shares some resemblance with the Thai falling tone type, which is sometimes characterized as a rise-fall (241) tone, although the rise in the Thai tone is less pronounced than in the present study's peaking tone. Another possibility is that the "pitch integral" , i.e., the perceived amount of pitch under the curve (Segerup & Nolan, 2004) played a role and that Thai participants expected more pitch under the curve for it to be perceived as falling, given that the Thai falling tone is relatively higher and starts falling later than the present study's falling tone, which starts falling from the onset. Therefore, one could hypothesize that Thai speakers were biased to categorize falling tonal contours as peaking tones tonal contours because they showed sensitivity to a "phonetic residual" from their L1 (J. Chen et al., 2020). This would result in a preference to categorize any falling contour in an L2 as a falling contour containing a small onset rise and a later fall, just like the Thai falling tone type.

It should be noted however, that there was also weak evidence in all other groups that falling tones were categorized less accurately than peaking tones (Table 19, page 198). Another explanation could therefore be that the peaking tone, due to its relative complexity, was somehow acoustically more salient so that participants identified it more easily than falling tones. It is further possible that a physiological bias towards registering tone contours containing F0 rises (as opposed to tone contours containing only F0 falls) strengthened this preference (Burnham et al., 2015, p. 1487; Krishnan et al., 2010).

The lack of evidence for a facilitative effect of L1 tonal status on tone categorization replicates findings from a wealth of other studies that similarly suggest that a higher degree of pitch functionality in the L1 does not necessarily translate into an advantage in non-native pitch processing at the pre-lexical level (A. Chen et al., 2016; Cooper & Wang, 2012; Cutler & Chen, 1997; Francis et al., 2008; Gandour & Harshman, 1978). Yet, the seemingly compelling evidence against an L1 tonal status advantage found in the present study contrasts with two previous studies that similarly involved tone categorization by speakers on a spectrum of L1 tonal statuses, but that did find an incremental effect of L1 tonal status (Burnham et al., 2015; Schaefer & Darcy, 2014). This warrants a closer comparison between the present study and those by Burnham et al. (2015) and Schaefer & Darcy (2014).

Burnham et al. (2015) found that Thai, Cantonese, and Mandarin speakers outperformed Swedish speakers, who in turn outperformed English speakers in discrimination accuracy of Thai tones. However, it is possible that this overall result was a product of differential discrimination accuracies depending on the specific tone contrast. Indeed, they only found significantly lower discrimination sensitivity by English speakers (which was statistically compared relative to the aggregate sensitivity of pitch-accent and tone language speakers) in four out of ten possible tonal contrasts, and these included tonal contrasts that are very similar to those that exist in Mandarin and Cantonese (Burnham et al., 2015, p. 1475). For the other tonal contrasts, no significant difference in discrimination was found between English speakers and speakers of the other languages. It is therefore possible that the overall superiority in tone perception by pitch-accent and tone language speakers observed in Burnham et al. (2015) was in fact driven by discrimination superiority on specific tonal contrasts which showed similarities to their L1 tone types.

However, Schaefer & Darcy (2014), who similarly investigated Thai tone perception (by Mandarin, Japanese, English, and Korean speakers) by means of an AXB discrimination task, found no clear evidence that superior tone discrimination by tone or pitch-accent language speakers was limited to specific tonal contrasts. Although they did not directly compare performance per tonal contrast between L1s, they found no significant interaction between L1 and contrast type, thereby suggesting an overall superiority for tonal and pitch-accent language speakers over non-tonal language speakers in tone discrimination.

Why did Burnham et al. (2015) and Schaefer & Darcy (2014) observe a hierarchy in tone perception according to L1 tonal status whereas the present study did not? Differences in methodology, such as the measure of tone perception (discrimination versus categorization), but also the control for musical experience, which was partially controlled for in Burnham et al. (2015) but not in Schaefer & Darcy (2014), and sample size (Burnham et al. (2015) had over 30 participants per group, but Schaefer & Darcy (2014) a maximum of 12) may in part explain this discrepancy. Yet, I make the tentative suggestion that the target tone to be perceived is in fact crucial to whether L1 tonal status is facilitative or not. The fact that Burnham et al. (2015, p. 1475) made direct statistical comparisons between discrimination on tonal pairs, and found that L1 tonal status was only facilitative in some conditions, would support this notion.

All in all, when placing the findings of the present study's tone categorization task in the light of previous studies that involved speakers on a spectrum of L1 tonal statuses, it appears that L1 tonal status does not facilitate tone categorization when the tonal system is equally challenging for all L1s involved.

Indeed, above and beyond L1 tonal status, the present study shows that it is instead musical experience that facilitates tone perception. There was compelling evidence that across the board, musicians were more likely than non-musicians to accurately categorize tones, and there was weak evidence that they were additionally faster at doing so. This replicates previous findings regarding the effect of long-term musical training on pre-lexical tone perception (Laméris & Post, 2022; Wong et al., 2020; H. Wu et al., 2015). Counterintuitively, there was also weak evidence that Japanese speakers' musical experience in fact slowed down tone categorization reaction times. One possibility is that Japanese musicians were somehow more attentive and hesitant to make tone categorization decisions, which may have slowed down their reaction times.

As to the effect of working memory, there was no clear evidence that digit span led to higher accuracy in tone categorization. This is in line with previous studies that found no, or only weak links between working memory capacity and pre-lexical tone processing (Bidelman et al., 2013; Hutka et al., 2015). However, for Japanese speakers there was in fact compelling evidence that digit span led to higher tone categorization accuracy. There was additionally weak evidence that digit span led to faster reaction times for Japanese speakers. Although there is some evidence that WM facilitates Japanese speakers' L1 pitch pattern categorization (Goss & Tamaoka, 2015), it is unclear why WM only clearly facilitated Japanese, but not other participants' tone categorization accuracy. It was suggested to me that, unlike other participants, Japanese listeners may have perceived the monosyllabic stimuli as multimoraic (Francis Nolan, *voce*) which could have resulted in a more string-like perception of the monosyllabic tones for which backwards digit span could be facilitative. To investigate this, I distributed a post-experiment survey six months after the original experiment in which I asked participants how many temporal units they perceived in the tone categorization stimuli[20]. Only five of the Japanese participants who had taken part in the experiment responded, and one of them indicated that she perceived two morae in the monosyllabic stimuli, namely in the falling and peaking tones. However, two out of seven Dutch respondents also perceived the peaking stimuli as disyllabic. Thus, the perceived temporal value of the stimuli does not seem to explain why WM appeared to facilitate tone categorization only for Japanese speakers either.

5.5.2 Facilitators in word generalization

Results from the word generalization task showed that participants were able to identify the meaning of tonal pseudowords spoken by a new speaker. However, they appeared to struggle with identifying the meaning of words that formed tonal minimal pairs, and an analysis of word identification errors showed that most errors were tone-only errors. This suggests that participants had acquired the segmental, but not the tonal properties of words. This finding highlights that lexical tones are indeed a relatively challenging feature of speech to acquire, as shown in earlier studies (Laméris & Post, 2022; Wong & Perrachione, 2007).

Crucially, accuracy levels across L1s were very similar. This suggests that L1 tonal status did not facilitate word generalization performance. Parallel to the findings of the tone categorization task, it again appeared that dynamic-dynamic tone contrasts (i.e., between falling and peaking tones) were the most difficult, whereas static-dynamic (i.e., between level and contour tones) were the easiest to process, and this was observed in all L1 groups.

However, there were some differences between groups in terms of performance per tone. These could be indicative of interactions between the L1 prosodic systems and the

[20] In this questionnaire, I presented the stimulus [a] with the three different tones and asked how many temporal units listeners perceived. I used the words for ‘syllable’ in Dutch, Swedish, and Thai, and the common word for ‘mora/beat’ in the Japanese version of the questionnaire. These were each accompanied by a real-life word that indicated the number of temporal units. I also included one disyllabic /lala/ stimulus to ensure that individuals understood the concept of a syllable/mora.

target tone types.

First, unlike in other groups, there was no compelling evidence that Swedish speakers were more likely to identify words with level tones than words with peaking tones. A closer inspection of the error patterns revealed that Swedish speakers tended to misidentify level tone words as peaking tone words relatively often, particularly in comparison to Japanese and Thai speakers. These confusion patterns do not mirror the findings of the tone categorization task. It is therefore unlikely that the inherent acoustic properties of these tones contributed to the relative difficulty of distinguishing them in words. Instead, it may have been that Swedish speakers somehow struggled with encoding the level-peak contrast lexically. One potential source of this difficulty could be that in Swedish, accent II words contain a small pitch peak in the first syllable, which has been suggested to be phonetically and cognitively marked (Roll et al., 2011, but cf. Köhnlein, 2020, p. 458). It is therefore possible that, due to influence from Swedish word prosody, Swedish speakers were relatively inattentive to words without any clear prosodic prominence (i.e., level tone words), and may have been biased towards identifying words that contained a pitch peak on the first syllable (i.e., peaking tone words). It is alternatively possible that disyllabic words with a level tone throughout constituted a phonotactic violation for Swedish speakers[21], and that they therefore "repaired" (Dupoux et al., 2011; Guevara-Rukoz et al., 2017) this apparent violation in their perception.[22]

Second, it was found that Japanese speakers were less likely than Dutch speakers to identify words with falling tones. Although falling tone words were generally the least likely to be identified in all groups (because they were often misidentified as peaking tone words) this appeared to occur relatively often in Japanese speakers. Here again, it is possible that

[21] Indeed, in the post-experiment questionnaire that I sent out a few months after the experiment, two Swedish speakers indicated that the pitch accent on the disyllabic /la11.la11/ was very dissimilar (1 on a scale to 7) to Swedish pitch accents, whereas /la51.la11/ and /la141.la11/ were perceived as more similar (respectively 4 and 5.5 on a scale to 7).

[22] One final possibility could be that Swedish speakers simply made more errors in word identification of level tone words because they had not sufficiently acquired the segmental properties, and therefore made many general word identification errors not owing to tonal confusions, such as misidentifying /la11.la11/ 'leaf' as /lɛ11.lɛ11/ 'chair'. This would result in a relatively low identification accuracy for level tone words. To rule out this possibility, the same model was fitted on the data that excluded any word identification errors that were not tone-only errors. However, even in this model, there was still no compelling evidence that Swedish speakers were more likely to identify level tone words than peaking tone words (b = 0.45 [-0.97, 1.90]). It thus appears that, relative to the other groups, level tone words were indeed difficult to identify for Swedish speakers.

interference with L1 lexical pitch patterns complicated the distinction between disyllabic words with different pitch patterns. Japanese first-mora accented words contain a high tone target on the first mora, followed by a low tone target on the subsequent mora. It is possible that Japanese speakers perceived both falling and peaking level tone words in the present study as if they were Japanese first-mora accented words. This may have led to a relative difficulty in encoding this tonal contrast lexically in the word generalization task. The reason why Japanese speakers were in addition less likely to identify falling tone words relative to Dutch speakers could be explained by findings from previous studies involving Japanese and Dutch speakers in the processing of disyllabic words with pitch-based contrasts on the first syllable. Specifically, Dutch speakers may be able to accurately detect differences in pitch contours on first syllables because they process them psychoacoustically (Braun & Johnson, 2011), whereas Japanese speakers may be limited in their sensitivity to non-native pitch contrasts due to a combination of their intermediate L1 tonal status, as well as a relatively limited utterance-level pitch pattern inventory, which may hinder psychoacoustic processing of complex pitch contours at a lexical level (Braun et al., 2014).

Despite these two specific observations, performance in the word generalization task was generally uniform across participants. It revealed no clear evidence that an increased L1 tonal status facilitates lexical encoding of tones, at least in pseudowords with tones that are relatively equally challenging for all L1s involved. It is important to note that the studies that did show an advantage for tonal L1 speakers in word learning relative to non-tonal speakers indeed involved L2s that were either typologically very similar to the L1, such as L1 Mandarin and L2 Cantonese (Poltrock et al., 2018), or that have overlapping tone types, such as L1 Thai and L2 Cantonese (Cooper & Wang, 2012).

As was the case in the tone categorization task, it was instead extralinguistic factors that more clearly predicted individual performance in word generalization. Musical experience was only predictive of higher word identification in the Swedish group. In other groups, there was no evidence that musicians were more likely than non-musicians to correctly identify the tonal pseudowords. This goes against the intuition that a differential relevance of musicianship would be observed across L1s, following previous word learning studies that showed that speakers from non-tonal L1 backgrounds benefited more from musicianship than did speakers from tonal L1 backgrounds (Cooper & Wang, 2012; Laméris

& Post, 2022). One possibility is that Swedish speakers benefited relatively more from musical experience because unlike the other L1 groups, Swedish speakers struggled more with contrasts between level and contour tones, and therefore would benefit more from musical experience, but the evidence is inconclusive[23]. Another possibility is that in this experiment, musical experience as measured by years of formal training was not a sufficiently fine-grained, reliable measure of musicianship as opposed to more direct measures such as musicality as measured by standardized tests (Wallentin et al., 2010).

As to working memory, there was no compelling evidence that participants with longer digit spans tended to be better at correctly identifying words than participants with shorter digit spans. Against the predictions, no differential in the relative contribution of WM was found per L1, unlike the findings in Chapter 2 that showed that WM only facilitated pseudoword learning for Mandarin-L1, but not English-L1 learners. The interpretation made in Chapter 2 was that Mandarin-L1 learners may not have experienced the task as a challenging *tone* word learning task as such (because, apart from the mid-level and low-level distinction, the pseudoword tone system was identical to the Mandarin one), but instead as a more general vocabulary learning task, for which WM may be more directly facilitative. However, in the present study, which employed a tone system that was equally challenging to all participants due to the presence of the falling-peaking contrast, most participants may have experienced the present task more as a *tone* word learning task rather than a general word learning task. Therefore, individual skills related to pitch processing, above and beyond measures of WM capacity, may have been more indicative of individual performance in the learning of the tonal pseudowords in the present study.

Indeed, it was individual pitch aptitude that was by far the strongest predictor of performance in word identification. Across the board, there was compelling evidence that participants who scored higher in tone categorization were more likely to correctly identify tonal pseudowords. This replicates findings by previous studies that show a strong link

[23] The possibility that musical experience facilitated Swedish speakers only for specific tones was explored in a model that contained a three-way interaction between L1, tone type, and the extralinguistic factors, but this revealed no compelling evidence that this was the case. There was weak evidence for a three-way *L1:Tone Type:Musical Experience* interaction (*b* = 0.37 [-0.05, 0.80], probability of direction: 95.93%). Subsequent multiple comparisons revealed compelling evidence that musical experience facilitated word identification of peaking tones for Swedish speakers (*b* = 0.91 [0.16, 1.68]), but this should be interpreted with caution given the weak evidence for the overall interaction.

between pre-lexical and lexical tone processing, specifically when pre-lexical pitch processing is assessed by accuracy in a tone categorization task (Bowles et al., 2016; Laméris & Post, 2022; Ling & Grüter, 2020; Wong & Perrachione, 2007). In particular, it is worth contextualizing the present study's findings with regard to the effect of musical experience, working memory, and pitch aptitude with those from Bowles et al. (2016), who investigated Mandarin tone word learning in 160 English-L1 participants and who measured the relative contribution of an array of extralinguistic factors, including tone identification (categorization) accuracy, nonword span, months of private music lessons, as well as indicators of general intelligence and foreign language experience. They found that measures of pitch processing, in particular tone identification accuracy, and to a limited extent the number of months of private music lessons, improved predictions of tone word learning beyond measures of general cognitive ability and foreign language experience, and conclude the following with regard to the relative weighting of these different factors:

> *"the large number of participants in the current study enabled a more sensitive correlational treatment of musicianship in terms of a continuum of individual differences. With this treatment, musical variables on the whole did not turn out to be powerful predictors of tonal word learning compared to tone-specific measures (..) it is clear that musical variables, which are only indirectly related to tone, are less effective predictors of tonal word learning than measures directly related to tone perception (..) a feature-specific approach to the prediction of L2 attainment (drawing on abilities/aptitudes that are most closely related to the linguistic challenge) is more powerful than a language-general approach".*
> (Bowles et al., 2016, pp. 798–799).

Indeed, in the present study, individual pitch aptitude also emerged as the strongest facilitator of individual word generalization performance compared to other individual measures of musical experience and working memory. Crucially, this study show that this effect holds even when examining tone word learning across learners from a spectrum of different L1 tonal statuses.

5.5.3 Final considerations

The findings from this study reveal that, when controlling as much as possible for the effects of tone type, musical experience, and working memory, an individual's L1 tonal status is not indicative of tone perception or tone word learning facility. Instead, the ease with which learners perceive non-native tonal contrasts devoid of any lexical meaning appears to be largely guided by individual pitch acuity derived from musical experience. This individual ability to pre-lexically perceive tones is in turn the strongest predictor of how easily learners then go on to learn how to use those tones in words. Although the interaction between tone types in the L1 and the to-be learned tonal contrasts may modulate the ease of perceiving and learning certain tonal contrasts, it appears that overall, pitch-specific abilities are the most reliable predictors of facility in early stages of tone perception and word learning. Individual working memory capacity may facilitate tone perception and word learning in some contexts, but it appears to have a secondary facilitative role overall. These findings suggest that individual variability in tone learning facility can be better captured by a domain-general account of pitch processing (which may be modulated by L1-specific effects) than by the Functional Pitch Hypothesis (Schaefer & Darcy, 2014).

There are some limitations of this study that must be acknowledged. First, despite including headphone checks, the fact that this study was carried out online may have contributed to a relatively larger amount of noise in the data, thereby masking some potential effects that could have been observed in a more narrowly controlled lab-based study. It is additionally possible that the fact that all participants were resident in the UK dampened effects from the L1 that could have been more easily detectable if participants had been monolingual speakers of their L1 and had been recruited from their respective home countries. However, the previous studies that investigated the effect of L1 tonal status which served as points of reference for the present study also included speakers living in English-speaking countries (Burnham et al., 2015; Schaefer & Darcy, 2014). Lastly, it is acknowledged that a word learning paradigm of nine pseudowords can only partially replicate the true nature of learning new words in a tonal language in which learners need to learn how to use lexical tones in more dynamic settings and modalities, and in addition employ other acoustic cues than F0, such as phonation and duration, to make tonal distinctions (Tsukada &

Kondo, 2019; Y. Zhang & Kirby, 2020).

Nevertheless, the present study provides compelling evidence that individual pitch aptitude is the main determiner of early-stage non-native tone learning, all else being equal. Further, these data, based on participants whose L1s represent a spectrum of L1 tonal statuses, replicate findings from previous studies that were carried out with only English-L1 learners (Bowles et al., 2016; Wong & Perrachione, 2007).

Finally, it is worth taking a closer look at how participants personally experienced the study, not least because these individual reports echo the larger picture that was inferred from the accuracy and reaction time data. A selection of comments from the debriefing (Appendix 5.8) reveals for instance that individuals found the identification of individual tones, especially the falling and peaking tones, to be the true challenge in the word learning task. Two Thai individuals also commented that knowledge of Thai did not help them in memorizing the pseudowords. Individuals also commented on their personal learning strategies, such as trying to group together words by vowel first and then focus on the tones. Others indicated that they would have benefited from writing down the words, or that they would have found it easier if words were presented as tonal minimal pairs. One individual indicated that musical experience must have helped her in learning the tone words. Some participants also commented that they might have performed better if they were learning more natural-sounding words. When plotting accuracy in the word generalization task against responses to questions from the debrief (Appendix 5.9), it indeed appeared that many participants found the audio stimuli relatively unnatural, although this did not appear to affect their performance. Participants' concentration throughout the experiment did not seem to be correlated to word generalization accuracy either. Interestingly, the degree to which participants rated the experiment as "fun" was in fact correlated to their performance, as was their awareness to the similarity between the tones in the tone categorization and the word generalization task. All in all, this suggests that motivation, and perhaps most importantly, having an 'ear' for language is a pivotal requirement for successful speech learning of the type explored here.

5.6 Appendix to Chapter 5

5.6.1 Detailed participant demographics

This appendix section contains detailed participant demographics.

- In each table, participants in shade were excluded from the analysis of the subset as reported in the main paper.
- * indicates exposure to a heritage language.
- ME: Musical experience (years)
- WM: Working memory (backwards digit span 4–8)
- PP: Pitch perception aptitude (tone categorization accuracy 0–100%)

Appendix 5.1

Detailed participant demographics (Dutch group).

ID	Age	L2s and self-reported level (0-10)	Currently Practicing	ME	WM	PP
NL-MU-F-01	25	English 10, German 5, French 2	Keyboard/Piano, Singing (other)	6	7	78
NL-MU-F-02	27	English* 10, French 4, German 4,Italian 4, Shona 3	Keyboard/Piano	5	6	73
NL-MU-F-03	27	English 10, Swedish 4, French 3	Keyboard/Piano	6	5	65
NL-MU-F-04	29	English 10, French 6, German 7, Italian 4, Russian 2	Keyboard/Piano	8	8	100
NL-MU-F-05	28	English 10, French 5, German 4	Keyboard/Piano	14	6	94
NL-MU-F-06	19	English* 10, German 7,Italian 1	Keyboard/Piano	6	6	85
NL-MU-F-07	29	English* 10, German 6, French 4	Woodwind, Choral singing	22	5	100
NL-MU-F-08	31	English 10, German 5, Spanish 2, French 1	Brass	10	6	94
NL-MU-F-09	30	English 9, French 2, German 3	Woodwind	13	5	85
NL-MU-F-10	26	English 10, French 6,German 6, Italian 5, Greek 2	Guitar, Singing (other)	9	8	76
NL-MU-F-11	29	English* 10, German 7, Mandarin 3, French 1	Strings , Choral singing	23	8	100
NL-MU-F-12	30	French* 10, German*, Luxembourgish*, English 10, Spanish 8,Italian 5	Woodwind, Choral singing	6	4	56
NL-MU-M-01	27	English 10, German 7, French 3	Guitar , Keyboard/Piano, Strings, Choral singing, Singing (other)	13	7	100
NL-MU-M-02	22	English* 10, Spanish 5	Guitar, Singing (other)	11	7	75
NL-MU-M-03	33	English* 10, Indonesian* 6, Japanese 8	Guitar, Keyboard/Piano	4	7	98
NL-MU-M-04	20	English* 10	Strings, Woodwind, Choral singing, Singing (other)	8	4	81
NL-MU-M-05	31	English 10, Swedish 8, Russian 8, German 8, Spanish 7	Guitar	6	6	96
NL-MU-M-07	25	English 10, German 5	Keyboard/Piano	16	7	100
NL-MU-M-08	28	English 9,French 4, German 7, Arabic 1	Other instrument/type of singing	10	8	100
NL-MU-M-09	30	English 10, German 8, Japanese 5, French 5,Mandarin 5	Keyboard/Piano, Strings, Choral singing, Singing (other)	23	8	100
NL-NM-F-01	22	English 8, German 4	-	0	7	89
NL-NM-F-03	23	English 9, French 5,German 6	-	0	8	71
NL-NM-F-04	26	English 10,Spanish 5,German 2, French 2	-	0	5	66
NL-NM-F-05	29	French* 10, English 10, German 6	-	0	4	79
NL-NM-F-07	25	English 10, German 7, French 2	-	0	6	55
NL-NM-F-08	24	German* 10, English 10, Japanese 9, French 9, Luxembourgish 7	-	0	5	56
NL-NM-F-09	32	English 8, French 5, German 5	Choral singing	11	5	96
NL-NM-F-10	28	English 6	-	0	4	56
NL-NM-F-13	28	English 8	-	0	4	78
NL-NM-M-02	26	English* 10, German 6, French 3, Spanish 3	-	0	8	74
NL-NM-M-03	23	English 10, French 6, German 6, Hebrew 4, Arabic 2	-	0	5	90

Appendix 5.2

Detailed participant demographics (Swedish group).

ID	Age	L2s and self-reported level (0-10)	Currently Practicing	ME	WM	PP
SE-MU-F-01	22	English 8	Woodwind	12	5	84
SE-MU-F-03	35	English 10, German 5	Guitar, Keyboard/Piano, Brass, Choral singing, Other instrument/type of singing	11	8	76
SE-MU-F-05	31	Norwegian 7, Cantonese 2	Guitar	6	5	56
SE-MU-F-06	25	English 10, Spanish 3	Choral singing	6	6	85
SE-MU-F-07	26	English 10, Spanish 6	Woodwind	8	5	70
SE-MU-F-08	20	English 10, Japanese* 6, French 5	Keyboard/Piano, Singing (other), Guitar	7	6	63
SE-MU-F-10	20	Italian* 10, English 9 Spanish 6	Keyboard/Piano	9	7	100
SE-MU-F-12	22	English 10, Spanish 4, Norwegian 4	Singing (other); Guitar (all types); Keyboard/Piano	16	4	96
SE-MU-M-03	31	English 10	Guitar	16	5	57
SE-NM-F-03	26	English 10, Norwegian 7	-	0	6	80
SE-NM-F-04	30	English 8	-	0	8	76
SE-NM-F-05	29	English 10, French 1	-	0	5	72
SE-NM-F-06	32	English, German, French	-	3	5	80
SE-NM-F-07	29	English* 10, French 1, Spanish 1	Keyboard/Piano	6	6	67
SE-NM-F-08	32	Bosnian* 10, English 10, German 2, French 1	-	0	6	76
SE-NM-F-09	20	English* 10, Spanish 6	-	0	5	67
SE-NM-M-01	27	English 10, Japanese 7	-	0	6	94
SE-NM-M-03	25	English 10	-	0	5	71
SE-NM-M-04	31	English 10, German 2, French 1	-	0	7	93
SE-NM-M-05	32	English 10, Persian 3, German 2, Spanish 1	-	0	5	92

Appendix 5.3

Detailed participant demographics (Japanese group).

ID	Age	L2s and self-reported level (0-10)	Currently Practicing	ME	WM	PP
JP-MU-F-01	19	English 8	Keyboard/Piano, Woodwind, Choral singing	12	7	100
JP-MU-F-02	30	English 8	Drums, Keyboard/Piano	11	8	80
JP-MU-F-03	29	English 9	Guitar, Keyboard/Piano, Singing (other)	3	8	98
JP-MU-F-04	30	English 5, Mandarin 1	Strings, Choral singing	14	7	98
JP-MU-F-05	26	English 8	Keyboard/Piano	12	5	98
JP-MU-F-07	34	English 8, Arabic 1	Keyboard/Piano	7	7	98
JP-MU-F-09	25	Mandarin* 8, English 8	Strings	6	6	96
JP-MU-F-10	21	Mandarin* 7, English 8	Keyboard/Piano, Woodwind	11	6	96
JP-MU-F-11	27	English* 10, Korean 1, Spanish 1	Singing (other), Guitar	19	8	100
JP-MU-F-12	36	English 3	Keyboard/Piano	6	8	92
JP-MU-F-13	30	English 5 Russian 5 Turkish 4	Guitar, Drums, Keyboard/Piano	26	5	98
JP-MU-F-14	34	English 8	Guitar, Drums, Keyboard/Piano, Other instrument/type of singing	10	5	71
JP-MU-F-15	31	English 5	Keyboard/Piano	7	5	76
JP-MU-M-01	30	English 7, Spanish 1	Strings, Choral singing	6	8	100
JP-MU-M-02	20	English 8	Guitar	5	8	100
JP-MU-M-04	25	English 8	Guitar, Strings	3	8	93
JP-MU-M-05	30	English 6	Guitar, Keyboard/Piano	13	6	88
JP-MU-M-06	31	English 8	Keyboard/Piano, Brass	23	5	80
JP-MU-M-07	22	English 7	Keyboard/Piano, Woodwind	17	8	98
JP-NM-F-01	22	English* 10	-	0	5	85
JP-NM-F-02	33	English 7	-	0	6	87
JP-NM-F-03	32	English 7	-	0	8	93
JP-NM-F-04	27	English 7	-	0	7	89
JP-NM-F-05	30	English 5	Keyboard/Piano, Strings	2	7	93
JP-NM-F-06	28	English 8 Italian 6	Singing (other)	0	8	100
JP-NM-F-07	33	English 4	-	0	6	79
JP-NM-F-08	27	English 3	-	0	4	67
JP-NM-F-09	28	English 7, German 4, Vietnamese 1	-	0	8	96
JP-NM-F-10	33	English 5	-	0	6	83
JP-NM-F-11	34	English 7	-	0	5	71
JP-NM-M-01	21	English 9, Mandarin 8	Drums	4	7	96
JP-NM-M-03	19	English* 10	Guitar, Piano	5	4	69
JP-NM-M-04	35	Dutch 2	Guitar, Keyboard/Piano,	2	8	82
JP-NM-M-05	21	English 10, French 9, Mandarin 8, Spanish 7, Korean 7	Choral singing	0	8	100

Appendix 5.4

Detailed participant demographics (Thai group).

ID	Age	L2s and self-reported level (0-10)	Currently Practicing	ME	WM	PP
TH-MU-F-01	22	English 10, Spanish 1	Guitar	1	5	58
TH-MU-F-02	21	English* 10, Mandarin 5,Isan 1	Keyboard/Piano, Strings, Choral singing, Singing (other)	16	7	100
TH-MU-F-04	27	English 7	Choral singing, Singing (other),	6	7	100
TH-MU-F-05	29	English 8	Choral singing	3	6	52
TH-MU-F-06	35	English 9, Mandarin 1, Japanese 1	Keyboard/Piano	3	8	50
TH-MU-F-07	19	English* 10, Mandarin 8, German 4, Japanese 2	Guitar , Other instrument/type of singing	14	5	89
TH-MU-F-10	21	English* 10, German 4, French 2, Spanish 1	Guitar, Piano, Strings	10	7	83
TH-MU-F-15	20	English 9, Dutch 7, Japanese 2	Keyboard/Piano, Choral singing, Other instrument/type of singing	12	7	100
TH-MU-M-01	28	Northern Thai* 10, English 10, Mandarin 1	Guitar , Keyboard/Piano, Strings, Choral singing, Singing (other), Other instrument/type of singing,	5	5	96
TH-MU-M-03	18	English* 10, French 5, Italian 3, Danish 2, Spanish 2	Guitar, Keyboard/Piano, Strings, Woodwind, Choral singing, Singing (other), Other instrument/type of singing	10	4	98
TH-MU-M-04	34	Northern Thai* 10, English 7	Guitar, Keyboard/Piano, Strings , Singing (other), Other instrument/type of singing	10	5	100
TH-MU-M-05	26	English 7, Chinese 1	Keyboard/Piano	6	5	72
TH-NM-F-01	32	English 10	Choral singing	2	6	96
TH-NM-F-02	22	English 8	-	0	6	93
TH-NM-F-03	28	Northern Thai* 10, English 8, Mandarin 2	Choral singing, Other instrument/type of singing	5	8	61
TH-NM-F-04	21	English 7, Japanese 1	Keyboard/Piano, Strings	2	7	80
TH-NM-F-06	26	English 8	-	0	5	62
TH-NM-F-07	22	English 7	-	0	6	100
TH-NM-F-08	36	English 8	-	0	5	66
TH-NM-F-10	30	Isan* 10, English 8, Mandarin 2	-	0	5	31
TH-NM-F-12	26	English 9, Cantonese 3	-	0	7	67
TH-NM-F-13	22	Isan* 10, English 7	-	0	8	66
TH-NM-F-15	30	English* 10, German 7, Japanese 6, Mandarin* 3, Cantonese* 2	-	0	8	48
TH-NM-F-17	26	English 9, Mandarin 5	Keyboard/Piano	2	8	83
TH-NM-F-18	27	Southern Thai* 10, English 7	Piano, Flute	3	8	51
TH-NM-F-19	28	English 10	-	0	7	61
TH-NM-F-20	22	Northern Thai* 10, English 7	Other instrument/type of singing	2	5	88
TH-NM-M-01	21	English 8, Korean 1	-	0	7	74
TH-NM-M-03	31	English 8	-	0	5	70
TH-NM-M-04	21	English 9, Isan 7	-	0	7	61

Appendix 5.5

Posterior distribution of tone categorization accuracy model.

Parameter	Estimate	95% Credible interval	Probability of direction	Rhat	ESS
(Intercept)	2.70	[2.18 ; 3.21]	100%	1.001	1973
L11	-0.29	[-0.87 ; 0.30]	84.17%	1.001	1724
L12	-0.28	[-0.99 ; 0.39]	78.77%	1	2528
L13	0.52	[-0.13 ; 1.17]	94.52%	1.001	2162
target_tone1	1.95	[1.32 ; 2.64]	100%	1.002	2164
target_tone2	-1.48	[-2.00 ; -0.95]	100%	1.001	2304
mu_y_training_2	0.74	[0.37 ; 1.12]	100%	1.001	2380
WM_2	0.36	[0.03 ; 0.71]	98.30%	1.002	1959
L11:target_tone1	-0.58	[-1.17 ; -0.01]	97.35%	1.001	2433
L12:target_tone1	0.47	[-0.26 ; 1.19]	90.37%	1.001	2708
L13:target_tone1	0.75	[-0.01 ; 1.58]	98.00%	1.003	2171
L11:target_tone2	0.38	[0.02 ; 0.72]	98.10%	1.001	2795
L12:target_tone2	-0.25	[-0.69 ; 0.15]	88.37%	1.002	2781
L13:target_tone2	-0.18	[-0.65 ; 0.26]	77.80%	1.002	2575
L11:mu_y_training_2	0.40	[-0.25 ; 1.06]	89.87%	1	2217
L12:mu_y_training_2	-0.65	[-1.36 ; 0.02]	97.02%	1.001	2195
L13:mu_y_training_2	-0.41	[-1.08 ; 0.27]	88.40%	1.001	2292
L11:WM_2	-0.08	[-0.66 ; 0.44]	61.05%	1.003	1650
L12:WM_2	-0.32	[-1.01 ; 0.35]	82.27%	1.002	1921
L13:WM_2	0.70	[0.16 ; 1.26]	99.50%	1.002	2005

Appendix 5.6

Posterior distribution of tone categorization RT model.

Parameter	Estimate	95% Credible interval	Probability of direction	Rhat	ESS
(Intercept)	7.26	[7.21 ; 7.33]	100%	1.001	2554
L11	0.01	[-0.07 ; 0.10]	63.62%	1.001	1627
L12	0.03	[-0.07 ; 0.14]	73.70%	1.001	2572
L13	-0.04	[-0.13 ; 0.05]	81.37%	1.002	2251
target_tone1	-0.12	[-0.16 ; -0.08]	100%	1	3894
target_tone2	0.07	[0.03 ; 0.10]	100%	1.001	4140
mu_y_training_2	-0.03	[-0.09 ; 0.03]	84%	1.001	2945
WM_2	-0.02	[-0.07 ; 0.04]	72.10%	1.001	2671
L11:target_tone1	0.01	[-0.02 ; 0.05]	76.37%	1.001	3982
L12:target_tone1	0.03	[-0.01 ; 0.08]	93.83%	1.001	3707
L13:target_tone1	-0.03	[-0.07 ; 0.01]	93.85%	1	3866
L11:target_tone2	0.00	[-0.03 ; 0.03]	51.63%	1	5430
L12:target_tone2	-0.04	[-0.07 ; 0.00]	98.07%	1.001	5633
L13:target_tone2	0.01	[-0.02 ; 0.04]	75.98%	1	5385
L11:mu_y_training_2	-0.03	[-0.12 ; 0.06]	73.58%	1.001	2684
L12:mu_y_training_2	-0.02	[-0.13 ; 0.07]	68.08%	1.002	2931
L13:mu_y_training_2	0.09	[0.00 ; 0.18]	97.43%	1	2869
L11:WM_2	0.04	[-0.04 ; 0.12]	84.37%	1	2262
L12:WM_2	0.03	[-0.07 ; 0.13]	74.77%	1	3148
L13:WM_2	-0.05	[-0.14 ; 0.03]	89.53%	1	2325

Appendix 5.7

Posterior distribution of word identification accuracy model.

Parameter	Estimate	95% Credible interval	Probability of direction	Rhat	ESS
(Intercept)	0.85	[0.48 ; 1.23]	100%	1.002	2629
L11	0.24	[-0.31 ; 0.81]	80.02%	1.001	2180
L12	-0.32	[-0.99 ; 0.33]	83.20%	1.001	2014
L13	-0.32	[-0.90 ; 0.30]	85.58%	1.001	2307
target_tone1	1.07	[0.71 ; 1.48]	100%	1.001	2671
target_tone2	-0.88	[-1.16 ; -0.63]	100%	1.002	3351
mu_y_training_2	0.04	[-0.31 ; 0.41]	58%	1.003	2238
WM_2	0.10	[-0.27 ; 0.44]	71.62%	1	2174
tt_accy_2	0.91	[0.44 ; 1.36]	100.00%	1.001	2498
L11:target_tone1	-0.08	[-0.59 ; 0.44]	60.87%	1	2284
L12:target_tone1	-0.92	[-1.45 ; -0.33]	99.97%	1	2380
L13:target_tone1	0.47	[-0.02 ; 1.00]	96.48%	1.001	2654
L11:target_tone2	0.09	[-0.23 ; 0.38]	72.22%	0.999	3764
L12:target_tone2	0.45	[0.13 ; 0.78]	99.70%	1	3311
L13:target_tone2	-0.30	[-0.59 ; 0.01]	97.38%	1	3952
L11:mu_y_training_2	-0.30	[-0.91 ; 0.31]	83.35%	1.001	2393
L12:mu_y_training_2	0.68	[0.11 ; 1.28]	98.93%	1	3053
L13:mu_y_training_2	-0.02	[-0.57 ; 0.51]	53.37%	1	2494
L11:WM_2	-0.05	[-0.52 ; 0.44]	58.73%	1	2586
L12:WM_2	-0.36	[-0.96 ; 0.22]	88.98%	1.002	2689
L13:WM_2	0.13	[-0.54 ; 0.83]	64.88%	1.002	2069
L11:tt_accy_2	0.60	[-0.10 ; 1.33]	95.40%	1	2702
L12:tt_accy_2	-0.06	[-0.82 ; 0.68]	57.48%	1.002	2702
L13:tt_accy_2	-0.25	[-1.29 ; 0.71]	69.42%	1.001	2254

5.6.2 Findings from debriefing

This appendix section presents a selection of individual comments and responses to the debriefing that participants filled out after the experiment. These represent a selection of responses that were deemed relevant to the discussion and to the wider interpretation of the findings.

Appendix 5.8

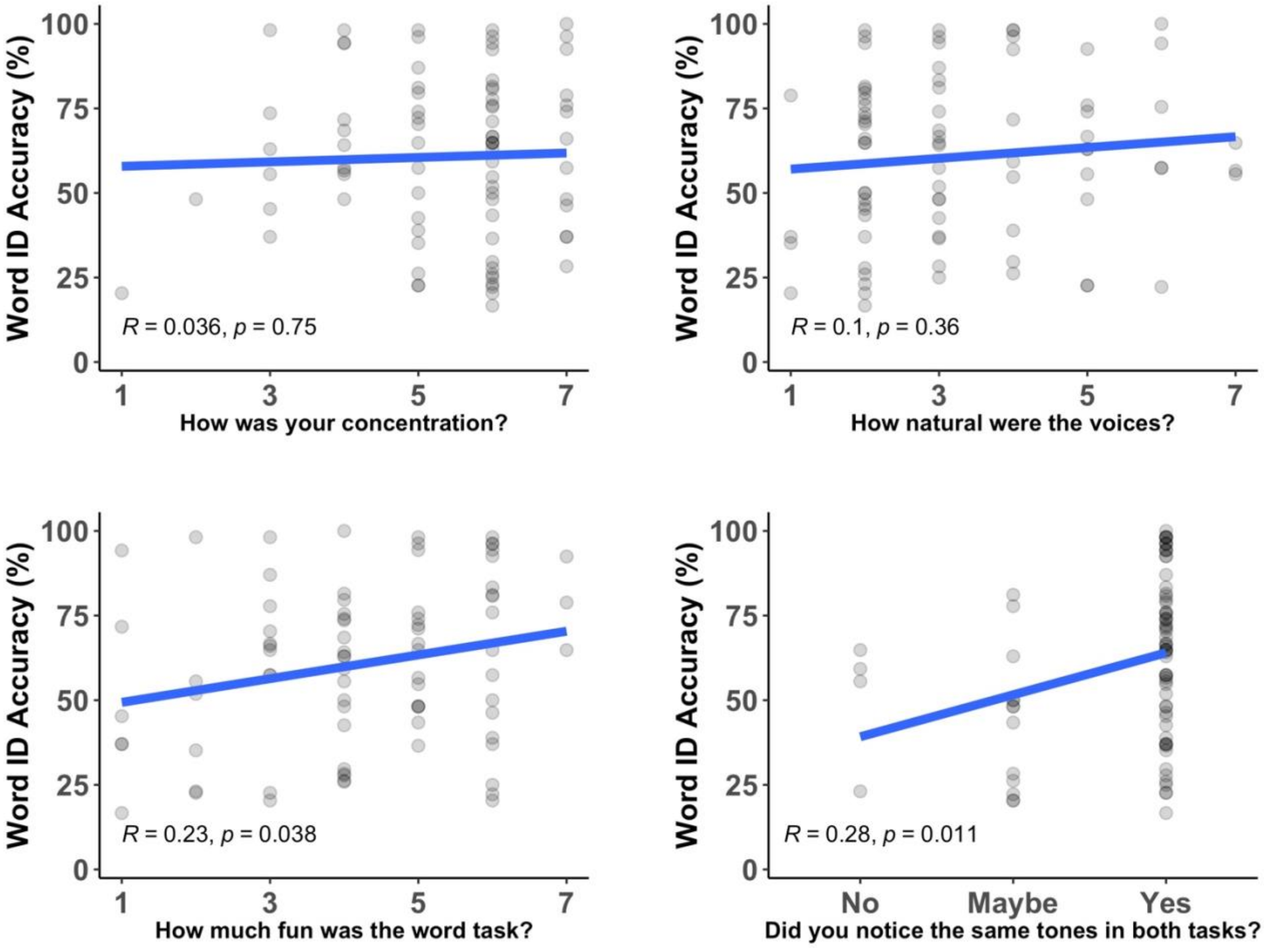
Selection of comments from the debriefing.

Comment	Participant	Word Gen. (%)
On difficulty of learning tones		
Only after having "learned" each word did I notice that they differed in tone. **My first instinct was to listen to the vowels** and the complete word, and not the melody.	NL-MU-F-1	79
I noticed that the words differed in vowels but because the words were so similar it was still difficult to learn them. For me, the practice round was too short, and I require more time to learn languages.	NL-NM-F-13	37
It was **difficult to tell the difference in tones** on the word test.	SE-MU-F-1	42
During the word recognition part, **I had a much easier time hearing flat sounds than rising-falling and falling.**	SE-NM-F-4	46
Until the end, it was difficult for me to get the difference between words, **except those with the flat tone** (chair, book, etc.). For instance, the words for car, apple, shirt, and cat all sounded similar, and as much as I tried, I couldn't tell them apart.	JP-MU-F-2	66
I was able to tell that the sounds differed in tones and in 'la', 'le', and 'li', but the task of combining [the sounds] and the images, and **in addition to tell apart "↗↘" and "↗↘"** made it extremely difficult.	JP-MU-M-1	92
I understand Thai is one of the tonal languages, but I find the items difficult to differentiate. As a native speaker of Thai, I spent a long time learning words.	TH-MU-F-6	73
The two tones are too similar (..) I can't connect with any knowledge of an existing language. For example, **it cannot be compared with Thai or Chinese** tones. This makes it difficult to memorize the words in the word test.	TH-MU-F-7	70
Other factors that made the task easy or difficult		
I noticed that learning the words on the second day was easier, because I had remembered three from yesterday (mountain, book, cat), and so I only had to learn six words. **(..) I found it easier to memorize the words as soon as I was able to divide them up in groups of three (based on their vowels),** after which I could start trying to memorize the tonal pattern.	NL-MU-F-10	72
It was extremely difficult to hear the different **tones because the vowels weren't presented in order**.	NL-NM-F-7	35
It was difficult to get used to an **artificial intelligence-like voice.** I think it would have been a real voice it would have been easier to learn.	JP-MU-F-14	20
I play a musical instrument, and I suppose **it must be easy for musicians to learn tone languages.**	JP-MU-M-6*	94
Memorizing words is challenging. But at the same time if the **pronunciation of the system is more natural**, it might make the answer more correct.	TH-MU-F-15	83
Learning new words **without taking notes** is difficult.	TH-NM-M-4	37

Participant not included in subset of 80 participants. The comments written here are responses to the question "Do you have any other comments about this experiment" (original question and response in participants' L1s, and translated into English). I grouped the comments together in two categories "on the difficulty of learning tones" and "other factors that made the task easy or difficult". Addition of* ***emphasis* *and [words for context] are mine.*

Appendix 5.9

Word generalization accuracy against a selection of debrief questions. R = Pearson's correlation coefficient.



5.6.3 Analysis on full dataset

This appendix section provides an overview of the compelling effects and interactions that were found when running the same models on the full dataset of 114 participants. Any differences regarding the dataset presented in the main paper are highlighted in **bold**. For brevity, only effects and interactions for which compelling evidence was found are listed. Overall, the comparisons between the two datasets reveal largely the same compelling evidence for main effects, interactions, as well as subsequent multiple comparisons.

Appendix 5.10

Comparisons of model results for tone categorization (accuracy).

Model reported in paper (n=80)	Model on full dataset (n=114)
Main effects and interactions	
L1:Tone (-0.58 [-1.17, -0.01])	L1:Tone (-0.57 [-1.10, -0.04])
Musical Experience (0.74 [0.37, 1.12])	**L1:Musical experience (-0.91 [-1.53, -0.30])**
L1:WM (0.70 [0.16, 1.26])	L1:WM (0.71 [0.27, 1.16])
Multiple comparisons	
Tone	
JP > NL | Level (2.20 [0.44, 3.99]).	JP > NL | Level (1.96 [0.56, 3.54]).
JP > TH | Level (1.96 [0.32, 3.86])	JP > TH | Level (1.46 [0.02, -3.03])
	JP > SE | Fall (1.36 [-0.34, 2.39])
Musical Experience	*Musical Experience*
-	**SE (-0.06 [-0.76, 0.61]).**
	NL (1.35 [0.76, 1.97]).
	JP (0.50 [0.04, 1.02])
	TH (1.56 [0.80, 2.40])
WM	*WM*
JP (1.06 [0.45, 1.69]	JP (1.00 [0.50, 1.51])
	NL (0.53 [0.02, 1.03])

Appendix 5.11

Comparisons of model results for tone categorization (RT).

Model reported in paper (n=80)	Model on full dataset (n=114)
Main effects and interactions	
L1:Tone (-0.04 [-0.07, 0.00])	L1:Tone (0.04 [0.00, 0.08])
L1:Musical Experience (0.09 [0.00, 0.18])	L1:Musical experience (0.07 [0.01, 0.14])
Multiple comparisons	
Tone	*Tone*
-	**NL > JP | Level (0.11 [0.01, 0.23])** **SE > JP | Level (0.15 [0.02, 0.29])**
Musical Experience	*Musical Experience*
-	-
WM	*WM*
-	**-**

Appendix 5.12

Comparisons of model results for word generalization (accuracy).

Model reported in paper (n=80)	Model on full dataset (n=114)
Main effects and interactions	
Pitch Aptitude (0.91 [0.44, 1.36])	Pitch Aptitude (0.86 [0.55, 1.25])
L1:Tone (0.45 [0.13, 0.78])	L1:Tone (-0.88 [-1.41, -0.41])
L1:Musical experience (0.68 [0.11, 1.28])	L1:Musical experience (0.58 [0.04, 1.11])
Multiple comparisons	
Tone	
	NL > SE | Level (1.90 [0.39, 3.49])
TH > SE | Level (2.26 [0.61, 3.94])	TH > SE | Level (1.79 [0.27, 3.39])
NL > JP | Fall (1.00 [0.05, 1.94])	NL > JP | Fall (1.19 [0.31, 2.04])
	NL > JP | Peak (0.75 [0.04, 2.04])
Musical Experience	*Musical Experience*
SE (0.72 [0.08, 1.43])	SE (0.77 [0.15, 1.46])
WM	*WM*
-	-

5.6.4 Multiple comparisons for error types

This appendix section contains significant multiple comparisons for the model of the count of error types in the tone categorization and word generalization tasks. In the tables, the comparisons should be made with reference to the latter element in a pair.

Appendix 5.13

Significant multiple comparisons for counts of error types between L1s per error type.

TONE CATEGORIZATION					
Contrast		Estimates	std. Error	t-value	p
Significant Multiple Comparisons (Bonferroni-Corrected)					
Level-to-Fall					
Dutch	Japanese	1.92	0.70	2.73	0.039
Swedish	Thai	-2.01	0.70	-2.86	0.027
Japanese	Thai	-2.53	0.69	-3.68	0.002
Level-to-Peak					
-					
Fall-to-Level					
Dutch	Japanese	1.50	0.48	3.13	0.011
Swedish	Japanese	1.83	0.49	3.76	0.001
Fall-to-Peak					
-					
Peak-to-Level					
Dutch	Japanese	2.30	0.68	3.37	0.005
Swedish	Japanese	2.47	0.70	3.54	0.003
Peak-to-Fall					
-					

Appendix 5.14

Significant multiple comparisons for counts of error types between error types per L1.

TONE CATEGORIZATION					
Contrast		Estimates	std. Error	t-value	p
Significant Multiple Comparisons (Bonferroni-Corrected)					
Dutch					
Level-to-Fall	Fall-to-Peak	-1.40	0.29	-4.76	0.000
Level-to-Peak	Fall-to-Level	-1.25	0.37	-3.38	0.012
Level-to-Peak	Fall-to-Peak	-1.84	0.34	-5.50	0.000
Level-to-Peak	Peak-to-Fall	-1.27	0.36	-3.54	0.007
Fall-to-Peak	Peak-to-Level	1.03	0.25	4.10	0.001
Swedish					
Level-to-Fall	Fall-to-Level	-2.54	0.62	-4.13	0.001
Level-to-Fall	Fall-to-Peak	-2.94	0.61	-4.83	0.000
Level-to-Fall	Peak-to-Level	-1.93	0.63	-3.06	0.035
Level-to-Fall	Peak-to-Fall	-2.34	0.62	-3.76	0.003
Level-to-Peak	Fall-to-Level	-2.96	0.74	-4.01	0.001
Level-to-Peak	Fall-to-Peak	-3.35	0.73	-4.59	0.000
Level-to-Peak	Peak-to-Level	-2.34	0.75	-3.13	0.028
Level-to-Peak	Peak-to-Fall	-2.75	0.74	-3.71	0.004
Japanese					
Level-to-Fall	Fall-to-Peak	-3.16	0.60	-5.24	0.000
Level-to-Fall	Peak-to-Fall	-2.68	0.61	-4.41	0.000
Level-to-Peak	Fall-to-Peak	-3.59	0.73	-4.95	0.000
Level-to-Peak	Peak-to-Fall	-3.11	0.73	-4.27	0.000
Fall-to-Level	Fall-to-Peak	-1.93	0.36	-5.30	0.000
Fall-to-Level	Peak-to-Fall	-1.45	0.37	-3.89	0.002

Fall-to-Peak	Peak-to-Level	3.17	0.60	5.27	0.000
Peak-to-Level	Peak-to-Fall	-2.69	0.61	-4.44	0.000
Thai					
Level-to-Fall	Level-to-Peak	2.88	0.75	3.84	0.002
Level-to-Fall	Fall-to-Peak	-0.90	0.24	-3.75	0.003
Level-to-Peak	Fall-to-Level	-2.64	0.75	-3.54	0.007
Level-to-Peak	Fall-to-Peak	-3.78	0.73	-5.21	0.000
Level-to-Peak	Peak-to-Fall	-2.99	0.74	-4.06	0.001
Fall-to-Level	Fall-to-Peak	-1.14	0.22	-5.14	0.000
Fall-to-Peak	Peak-to-Level	1.81	0.31	5.81	0.000
Fall-to-Peak	Peak-to-Fall	0.79	0.19	4.22	0.000
Peak-to-Level	Peak-to-Fall	-1.02	0.33	-3.05	0.036

Appendix 5.15

Significant multiple comparisons for counts of tone-only error types between L1s per error type.

WORD GENERALIZATION					
Contrast		Estimates	std. Error	t-value	p
Significant Multiple Comparisons (Bonferroni-Corrected)					
Level-to-Fall					
-					
Level-to-Peak					
Dutch	Japanese	1.14	0.40	2.88	0.025
Dutch	Thai	1.49	0.43	3.49	0.003
Swedish	Japanese	1.43	0.42	3.42	0.004
Swedish	Thai	1.78	0.45	3.98	0.001
Fall-to-Level					
Swedish	Japanese	1.39	0.42	3.34	0.006
Japanese	Thai	-1.16	0.44	-2.66	0.048
Fall-to-Peak					
-					
Peak-to-Level					
-					
Peak-to-Fall					
-					

Appendix 5.16

Significant multiple comparisons for counts of tone-only error types between error types per L1.

WORD GENERALIZATION					
Contrast		Estimates	std. Error	t-value	p
Significant Multiple Comparisons (Bonferroni-Corrected)					
Dutch					
Level-to-Fall	Fall-to-Peak	-0.60	0.20	-3.02	0.040
Level-to-Fall	Peak-to-Level	0.97	0.30	3.21	0.021
Level-to-Peak	Peak-to-Level	1.12	0.33	3.43	0.010
Fall-to-Level	Fall-to-Peak	-1.17	0.22	-5.44	0.000
Fall-to-Level	Peak-to-Fall	-1.06	0.22	-4.76	0.000
Fall-to-Peak	Peak-to-Level	1.57	0.27	5.84	0.000
Swedish					
Level-to-Fall	Peak-to-Level	1.35	0.38	3.58	0.006
Level-to-Peak	Peak-to-Level	1.49	0.41	3.61	0.005
Fall-to-Peak	Peak-to-Level	1.71	0.36	4.80	0.000
Peak-to-Level	Peak-to-Fall	-1.53	0.36	-4.23	0.000
Japanese					
Level-to-Fall	Fall-to-Peak	-1.76	0.43	-4.07	0.001
Level-to-Fall	Peak-to-Fall	-1.30	0.44	-2.96	0.048
Level-to-Peak	Fall-to-Peak	-1.91	0.31	-6.10	0.000
Level-to-Peak	Peak-to-Fall	-1.45	0.32	-4.50	0.000
Fall-to-Level	Fall-to-Peak	-2.32	0.33	-7.09	0.000
Fall-to-Level	Peak-to-Fall	-1.86	0.34	-5.55	0.000
Fall-to-Peak	Peak-to-Level	1.62	0.26	6.34	0.000
Fall-to-Peak	Peak-to-Fall	0.46	0.13	3.47	0.009
Peak-to-Level	Peak-to-Fall	-1.17	0.27	-4.36	0.000

Thai					
Level-to-Fall	Fall-to-Peak	-1.13	0.31	-3.65	0.004
Level-to-Peak	Fall-to-Peak	-2.05	0.35	-5.86	0.000
Level-to-Peak	Peak-to-Fall	-1.73	0.36	-4.85	0.000
Fall-to-Level	Fall-to-Peak	-0.96	0.28	-3.39	0.011
Fall-to-Peak	Peak-to-Level	1.48	0.34	4.38	0.000
Peak-to-Level	Peak-to-Fall	-1.16	0.35	-3.38	0.012

Chapter 6 General discussion

6.1 Revisiting the research question

In this dissertation I asked what explains individual variability in non-native tone learning facility. I approached this question by means of a series of studies which zoomed in on tone learning in the listening modality (Chapter 2), in the speaking modality (Chapter 3), tone learning across modalities and processing levels (Chapter 4), and tone learning across a spectrum of typologically different languages (Chapter 5). In each of these studies, I measured individual performance in different aspects of tone learning and identified the factors that explained why some individuals learned tones quite easily, whereas others learned tones with greater difficulty.

The expected outcome of this dissertation was a novel and integral empirical and theoretical account of tone learning. As I have argued throughout this dissertation, previous studies in the tone learning literature have addressed tone learning and the individual factors that affect it, but the scope of these studies was limited. Very often, they only separately assessed the effects of a limited number of L1-specific or extralinguistic factors, and only at a specific level of processing or in a specific modality. In this dissertation, I have collected these separate strings from previous literature, and woven them together into a whole investigation.

In this General discussion, I readdress the dissertation's key themes that I presented in the General introduction: pre-lexical and lexical tone processing, L1 tonal status, tone types, musicianship, working memory, and pitch aptitude. I will contextualize the findings of my dissertation according to each theme and present key conclusions. I will attempt to be as concise as possible in doing so, given that detailed interpretations of findings have already been covered in the discussion sections of each experimental chapter.

At the end of this General discussion, I will summarize my findings in a theoretical framework, present the wider implications of this dissertation, and identify avenues for future work.

6.2 Contextualization and conclusions

6.2.1 Pre-lexical tone processing

In line with the notion that pre-lexical processing of speech is an important steppingstone for speech learning at large, which has been described in the context of tones as a phonetic-phonological-lexical continuity (Wong & Perrachione, 2007), I first examined non-native tone learning at a pre-lexical level, both in perception by means of tone categorization tasks, and in production by means of an imitation task.

Based on accuracy and reaction time data in tone categorization, and phonetic accuracy data in imitation, I observed degrees of variability in the ease with which individuals processed non-native tones at a pre-lexical level. In the categorization task reported in Chapter 2, English and Mandarin speakers were able to perceive and categorize tones accurately and quickly, but some conditions appeared to be more challenging, resulting in more variability. Mandarin speakers struggled with level tone contrasts, which are known to be relatively challenging for Mandarin speakers as described in earlier studies (Qin et al., 2021; Zhu et al., 2021) and as predicted by the Perceptual Assimilation Model (Best, 1995). By contrast, they were very accurate and quick at perceiving rising and falling tone types, which are similar to tone types that exist in Mandarin. English speakers did not appear to perceive tones categorically, and perceived tones more psychoacoustically. However, they exhibited more individual variability in doing so, and this variability was driven by their musical experience.

Whereas the tone categorization task in Chapter 2 involved a tone system that was designed to be both easy and difficult for Mandarin-L1 learners, the tone system used in the study reported in Chapter 5 was designed to be equally easy and difficult for leaners from a spectrum of L1 tonal statuses (Dutch, Swedish, Japanese, and Thai). Here, all participants struggled with dynamic-dynamic tone contrasts (falling vs. peaking tones), regardless of their

L1. The ease with which tones were perceived was strongly driven by musical experience.

In the imitation task reported in Chapters 3–4, I found that individuals exhibited some degree of variability in their tone phonetic-acoustic tone production accuracy, but overall, their imitations were remarkably uniform. The only instance that appeared to create differences between individuals in tone production accuracy was found in the imitation of low-level tones by Mandarin speakers, who imitated these tones less accurately than did English speakers on day 1 because they produced them with a higher pitch than the target. This was potentially due to inference from the Mandarin tone system. This mirrors the difficulty of level tone processing that was found in the tone categorization task. These findings fall in line with the Speech Learning Model hypothesis that “many production errors have a perceptual basis” (Flege, 1995, p. 238) and highlight the parallels between perception and production. However, on day 2, it seems that this relative difficulty of imitating low-level tones had disappeared for Mandarin speakers, as their accuracy was similar to that of English speakers. Given that there was relatively little individual variability in the tone imitation task overall, I did not find clear evidence that imitation accuracy was guided by individual factors, although pitch aptitude and working memory improved imitation accuracy in some conditions. Where it did, it did so in the expected directions: higher pitch aptitude and higher working memory capacity were associated with higher production accuracy (Dong et al., 2019; Gupta, 2003).

As I highlighted in Chapter 4, a tone categorization task may tap into both phonetic and phonological aspects of pre-lexical processing, whereas an imitation task may only require phonetic processing. However, it cannot be excluded that there was in fact some top-down lexical influence (McClelland & Elman, 1986) on participants’ imitations given that they imitated sounds that represented pseudowords. Yet, I would still argue that the imitation task was predominantly phonetic in nature, given that participants were instructed to listen to the sound and repeat it as accurately as possible.

Therefore, although I observed that pre-lexical perception and production often went hand-in-hand, discrepancies in performance between the two modalities may have been due to a discrepancy in the more detailed phonetic or phonological levels of pre-lexical processing. Perhaps a better comparison between pre-lexical processing at a low-lying phonetic level could have been achieved by measuring performance in a tone *discrimination* task and an imitation task of meaningless vowels. Similarly, a more refined comparison between pre-lexical processing at a phonological level could have been achieved by measuring performance in tone categorization and a read-aloud task with orthographic prompts.

Generalizing over both modalities, I observed that, some specific conditions aside, participants were generally good at the pre-lexical perception and production of tones in a system that they had never encountered before. Although I did not assess perceptual performance over time, it is likely that, had I repeated the tone categorization task on day 2, participants would have improved their perception accuracy in the few areas that still had room for improvement, given previous studies that show that individuals can make considerable gains in tone perception accuracy with sufficient perceptual training (X. Wang, 2013; Y. Wang et al., 1999). The findings from the imitation task, which was repeated on day 2, did show that production performance generally improved.

Overall, I draw the following conclusion from this dissertation's findings on non-native pre-lexical tone processing:

C1: *At the very first stages of encountering a non-native tone system, individuals exhibit some degree of variability in the ease with which they process non-native tones. The ease of non-native processing appears to be guided by the shape of the tone to be processed, either in its phonetic-acoustic properties or in its phonological-categorical properties, and its similarity to tone types in the L1. An individual's musical experience, and to some extent their working memory, can facilitate the ease with which tones are processed. Overall, pre-lexical tone processing performance in the listening and speaking modalities mirror one another.*

6.2.2 Lexical tone processing (word learning)

In this dissertation, I examined tone word learning, which I argued to be more representative of real-life tone learning than just pre-lexical perception or production. After all, second language learners need to not only learn how to perceive and produce tones, but they also need to learn how to *use* those tones by linking them to words and their meanings.

Across the three word learning tasks reported in this dissertation (word identification in Chapter 2, image-naming in Chapter 3, and word generalization in Chapter 5), I found that, after two training sessions, participants were able to start associating spoken pseudowords with lexical meanings. However, unlike the pre-lexical tasks, in which individual variability was relatively small, the word learning tasks revealed a substantial degree of unevenness in learning facility. Towards the end of the word learning tasks on day 2, some individuals had established solid sound-meaning connections, whereas others continued to have trouble in encoding the relevant information at a lexical level.

Strikingly, the source of the difficulty of tone word learning lay not in linking sound to meaning as such, but instead in linking *tones* to meaning. This was illustrated by the fact that on day 2 of the word learning tasks, participants predominantly made tone-only errors. That is, they had learned that a specific lexical item such as ‘fork’ was defined by a segmental representation, e.g., /la.la/. However, many had yet to learn that that in addition, the meaning of ‘fork’ was defined by a tonal representation, e.g., /la51.la11/, and that this tonal representation was crucial to its meaning, because otherwise the word could also mean ‘leaf’ or ‘television’. Thus, it seems that for most learners, acquiring the tonal representations of a word formed the last hurdle in word learning. These observations echo the common claim that tones are indeed a particularly challenging aspect to acquire in a second language (Antoniou & Chin, 2018; R. K. W. Chan & Leung, 2020; Francis et al., 2008).

The individual variability in the ease with which learners fully acquired the tonal pseudowords appeared to be fueled primarily by their pitch aptitude (when represented by tone categorization *accuracy*, not by reaction times), but also to some extent by musical experience and working memory. However, the relative weighting of these extralinguistic factors was not uniform across participants. The origins of this differential were discussed in the separate experimental chapters, and will be briefly readdressed in sections 6.2.3.3–5 .

The observation that participants confused words with words that contrasted in tone alone implies that the word learning tasks represented word learning in a broad sense, involving both lexical configuration, which refers to the factual knowledge of a word, including its sound and meaning, and lexical engagement, which refers to the ability for a word to activate other lexical representations (Leach & Samuel, 2007). As evidenced by the predominance of tone-only errors, it appears that a word – either presented aurally (as /juɹ15/) or visually (as an image of a door) – activated all segmentally identical items, which competed with one another in lexical perception and production. I note that, in order to fully assert whether such lexical competition took place, other methods such as event-related potentials (Pelzl et al., 2020) or eye-tracking experiments (Qin et al., 2019) could be conducted, as these methods can more precisely record activation of lexical competitors at a neurological level and describe the time-course of a lexical decision, respectively. However, based on my behavioral data, it does appear that the word learning tasks demanded lexical engagement, and that participants had started to form sound-meaning connections. This was further confirmed by the findings from the word generalization task in Chapter 5, in which participants were able to identify the meaning of words when spoken by a new speaker.

Another reason for me to examine word learning was to verify whether I could observe a continuity between pre-lexical and lexical learning. I generally found that this was the case, since what happened at a pre-lexical level was indicative of what happened at a lexical level. More specifically, tones that were relatively easy or difficult to perceive or produce at the pre-lexical level were also relatively easy or difficult to use at a lexical level. However, this does not mean that pre-lexical performance perfectly mirrored lexical performance. For instance, whereas Swedish speakers did not appear to struggle greatly with contrasts involving level tones at a pre-lexical level in tone categorization, they did appear to struggle slightly with these contrasts at a lexical level. The same was observed in Japanese speakers, who appeared to struggle relatively more with fall-peak contrasts at a lexical level than at a pre-lexical level. In such a way, lexical tasks can reveal difficulties in tone processing that pre-lexical tasks cannot.

Therefore, although I agree that pre-lexical tasks can to some extent reveal what will eventually happen at a lexical level, and that therefore they can be good instruments to study speech learning at large, I argue that lexical tasks better represent the true nature of speech learning.

Summarizing the above, I draw the following conclusion from this dissertation's findings on lexical processing of tones:

C2: *In non-native tone word learning, individuals appear to first link segmental information to lexical-semantic representations, and only later add tonal information to fully acquire a tone word. This makes tones a particularly difficult speech feature to acquire in a second language, both in the listening and in the speaking modality. Individuals vary greatly in the ease with which they learn tone words. This facility appears to be driven by an individual's pitch aptitude, and to lesser extents their musical experience and working memory. The degree to which they rely on these extralinguistic factors may be modulated by their L1 tonal status. Overall, there is a continuity between pre-lexical and lexical processing of tones, but some learners may experience difficulties with the lexical processing of tones that they do not experience with the pre-lexical processing of tones.*

6.2.3 Facilitative factors in tone learning

In this section I return to each of the factors for which I had hypothesized that they would play a role in tone learning facility. Specifically, I will readdress to what extent each of the factors weighed in to determine the ease with which individuals learned tones.

6.2.3.1 L1 tonal status

As I indicated in the General introduction, L1 tonal status is often mentioned as one of the factors that influence non-native tone learning facility. This is rooted in the intuition that speakers who do not use pitch as a primary instrument for lexical distinctions would find it relatively difficult to do so in a non-native language. Throughout this dissertation, I examined the extent to which different degrees of L1 tonal status modulate the ease with which individuals learn tones.

Overall, my findings do not support the notion that mere experience with lexical tones in an L1 provides a default advantage in non-native tone learning, either at a pre-lexical or at a lexical level. In none of my experiments did I find that L1 tonal status in and of itself facilitated tone learning. It neither appeared that learners from an L1 tonal background were in any way better prepared than non-tonal speakers to acquire a novel tonal system, given that in all groups, associating tones to lexical meaning appeared to be inherently challenging, as illustrated by the high proportion of tone-only errors.

Yet, that does not mean that L1 tonal status is completely irrelevant to non-native tone learning. As I concluded in Chapters 2–4, L1 tonal status appears to dynamically interact with other factors such as tone type, musical experience, and working memory. Although the facilitative effect of L1 tonal status may not emerge directly, it can emerge indirectly through these factors. For instance, Mandarin speakers showed less variability than English speakers in their acquisition of rising and falling tones. The fact that rise-fall tonal contrasts occur in Mandarin is likely to be at the root of this finding. Indeed, I hypothesized in Chapter 2 that if the Mandarin speakers had been learning a tonal system that is similar to the Mandarin tonal system, they would have likely outperformed English speakers. Therefore, under that scenario a facilitative effect of L1 tonal status could have emerged. However, it appears that any potential facilitative effect of L1 tonal status was offset because the pseudowords contained level tone contrasts which were difficult to acquire for Mandarin speakers. Therefore, whether L1 tonal status facilitates non-native tone learning depends on the tone types that form the tone system to be learned.

Another way in which an advantage of L1 tonal status may have emerged for the Mandarin speakers was found when looking at the relative contributions of musical experience and working memory. Overall, Mandarin speakers relied less on musical experience to acquire tones than did English speakers. Instead, individual variability in Mandarin learners was explained by working memory capacity and by pitch aptitude, the latter being universally facilitative for tone word learning. The fact that, unlike English speakers, Mandarin speakers did not appear to rely strongly on additional pitch processing skills gained from musical experience to acquire tones could be seen as an indirect facilitative effect of L1 tonal status. That is, L1 tonal status may reduce the reliance on additional pitch processing skills to facilitate non-native tone learning, as has been suggested in earlier studies (Cooper & Wang, 2012).

However, the dynamic interaction between L1 tonal status and extralinguistic factors that was observed in Chapters 2–4 was not fully reproduced in Chapter 5. Overall, a higher degree of L1 tonal status did not appear to reduce individuals' reliance on musical experience. The only instance in which musical experience was differentially beneficial according to L1 tonal status was in word generalization. Here, there was compelling evidence that only Swedish participants benefitted from musical experience. A possibility is that, as discussed in Chapter 5, the tone system may have in fact been relatively more challenging for Swedish speakers to acquire at a lexical level because of potential negative interference from Swedish word prosody. Therefore, this added degree of difficulty may have made musical experience more relevant. Another possibility is that the measure of musical experience used here may not be the most accurate proxy of musicianship-derived pitch acuity, as will be discussed in more detail in section 6.2.3.3. In addition, the different experimental designs of the studies in Chapters 2–4 and Chapter 5 may not have allowed for full reproducibility of the dynamic interaction between L1 tonal status and extralinguistic factors. It is noteworthy however to recall the estimate sizes of the effect of pitch aptitude on word generalization (Table 30, page 217). Dutch individuals appeared to most benefit from pitch aptitude, whereas Thai participants did least. Although these differences were marginal (and there was no compelling evidence for an interaction between L1 and the effect of pitch aptitude), they could indicate that Thai speakers had some indirect advantage from their L1 tonal status by relying less on pre-lexical pitch processing skills to acquire tone words (cf. Cooper & Wang, 2012).

In sum, I conclude the following regarding the effect of L1 tonal status on non-native tone learning:

C3: *Whether or not L1 tonal status facilitates non-native tone learning is a question that dominates the tone learning literature, but that can only be answered with an opaque answer: it depends. If the tone system to be acquired assimilates neatly to the tone system in the L1, individuals whose L1 is tonal may learn those tones more easily than individuals whose L1 is non-tonal, thereby showing an indirect facilitative effect of L1 tonal status via tone types. However, this also implies that L1 tonal status can be detrimental to non-native tone learning when non-native tones do not assimilate clearly to native tonal contrasts. Another indirect way in which L1 tonal status may emerge as indirectly facilitative is by* *reducing dependence on extralinguistic factors related to pitch processing (such as musical experience or pitch aptitude).*

6.2.3.2 Tone types

The findings from this dissertation show that tone type is a pivotal factor that determines the ease with which an individual perceives, produces, and learns a tone. I found that tone types can be relatively easy or difficult in terms of phonetic-acoustic and phonological-categorical properties.

In Chapters 2–4, the clearest effect of tone type was found in Mandarin speakers' performance on mid-level and low-level tones. These tones form a contrast that does not exist in Mandarin and were therefore hypothesized to be relatively challenging to perceive because they constitute single-category assimilation to the Mandarin high-level tone. In addition, previous studies have shown that Mandarin speakers are relatively less sensitive to differences in pitch height, which would make level contrasts in a non-native tone system relatively difficult in phonetic-acoustic terms. All the tasks reported in Chapters 2–4, which targeted tone processing at pre-lexical and lexical levels and in the listening and speaking modalities, revealed that Mandarin speakers had difficulty learning the level tone contrast. This suggests that the predictions of the Perceptual Assimilation Model (PAM), which only concerns pre-lexical tone perception, can be extended to higher-level lexical tone processing, as well as to the speaking modality.

For the English participants, it appeared that tone type played a less important role to determine tone learning facility. There was no clear suggestion that English speakers relied on any L1-based intonational tone types in tone learning. This would therefore constitute a 'no assimilation' scenario under PAM, which accords with earlier accounts that speakers from a non-tone language background can process tones more psychoacoustically and less categorically than speakers from a tone language background (A. Chen et al., 2018; K. Yu et al., 2019).

The study reported in Chapter 5 involved a tonal pseudoword system that was designed to be equally challenging for all L1s involved, based on its phonetic-acoustic properties of both 'easy' static-dynamic and 'difficult' dynamic-dynamic tone contrasts. Here again, it was shown that tone type determines to a large extent the ease with which tones are perceived and learned. Generally, most participants struggled with the falling and peaking tone contrasts, whereas contrasts with level tones were relatively easy. This trend was largely the same at a pre-lexical level in tone categorization and at a lexical level in word generalization.

However, two observations from the word generalization task hint that there may have been an additional phonological-categorical effect of L1 tone types on the lexical processing of pseudoword tones. Compared to other speakers, Swedish speakers appeared to struggle with lexically encoding level tones, and Japanese speakers with falling tones on disyllabic words. As discussed in Chapter 5, it is possible that native pitch accent categories in fact affected performance here. This is unlike what was originally hypothesized based on PAM, as the effect of tone types was expected to be relatively weak for speakers of pitch-accent languages (Best, 2019). However, it is possible that a relative overlap in broader terms between the disyllabic pseudowords and Swedish and Japanese pitch accents strengthened the phonological-categorical effect of L1 tone type for Swedish and Japanese speakers. Namely, both Swedish and Japanese pitch accents can be respectively disyllabic (or dimoraic), just like the tones on the pseudowords of the experiment, although the crucial contrast only occurred on one syllable. The fact that in addition to the specific shape of the pitch movement, the specific temporal domain of that pitch movement determines whether L1 prosody affects L2 prosodic learning echoes the prediction from the Functional Pitch Hypothesis that the "specific prosodic domain in which pitch differentiates lexical items also constrains performance" (Schaefer & Darcy, 2014, p. 513).

Considering the abovementioned, I formulate the following regarding the effect of tone type:

C4: *The specific shape of a lexical tone in a non-native tone system, also known as tone type, greatly determines the ease with which it is perceived, produced, and encoded at a lexical level. Tone types can be inherently easy or difficult according to their phonetic-acoustic properties, and although these inherent properties affect learning facility of specific tones in a relatively universal way, learners from some L1 backgrounds exhibit L1-specific sensitivity to phonetic-acoustic properties of non-native tone types. The relative ease of* *learning non-native tone types is further determined by potential phonological-categorical assimilation to L1-based tone types. Categorical assimilation appears to take place when the L1 and the L2 share the temporal domain (such as the number of syllables) over which tone types are realized, and when they share the functionality (such as lexical or phrasal) of the tone type.*

6.2.3.3 Musicianship

Throughout this dissertation I have investigated the effect of musicianship (in particular that of musical experience) on tone learning facility.

In line with the OPERA model that predicts that musical experience leads to enhanced speech processing abilities (Patel, 2011), I found that overall, individuals with musical experience outperformed non-musician peers in tone learning.

However, as we saw in Chapters 2–4, the benefit of musical experience appeared to be subject to the absence of “relevant experience”, which has recently been suggested as an extra requirement to the OPERA model (Choi, 2021). I found that whereas English participants benefitted substantially from prior musical experience in non-native tone perception and word learning, Mandarin participants did not. These findings replicate earlier studies that show a differential in relevance of musicianship on non-native tone processing according to a learner’s L1 experience with lexical tones (Cooper & Wang, 2012).

This L1-modulated effect of musical experience was not replicated in the study reported in Chapter 5. Here, all learners, regardless of their L1, benefited from musical experience in pre-lexical tone categorization. In the word generalization task, only Swedish speakers benefited from musical experience. This goes against the claim that relevant L1-derived pitch experience would attune the effect of musical experience on non-native tone processing, as it would have been expected that musical experience would be most facilitative for Dutch speakers, less so for Swedish and Japanese speakers, and least for Thai speakers.

I tentatively suggest that the following factors may be behind the discrepancy between findings from Chapters 2–4 and Chapter 5. First, it is possible that in the tasks in Chapters 2–4, which involved a tonal system that was designed to be partially like that of Mandarin, Mandarin speakers may have mainly relied on their native tone inventory (i.e., their relevant experience) and perceived and lexically encoded the tones categorically. Therefore, the effect of L1 tone type may have masked or overridden any potential beneficial effect that musical experience could have had. By contrast, in the tone categorization task in Chapter 5, which involved a tonal system that was designed to be equally foreign to all participants, participants may have not relied on their L1 prosody, and instead more on pitch processing skills gained from musical experience. It is additionally possible that the fall-peak contrast in the tone categorization task elicited more psychoacoustic than categorical perception, for which musical experience may be particularly facilitative (Wayland et al., 2010).

Second, the finding that musical experience only facilitated Swedish speakers' word generalization could be because Swedish speakers overall struggled relatively more in the word generalization task than did other speakers. Therefore, musical experience may have been additionally beneficial for Swedish individuals. It should however be noted that in all other groups (including the Swedish), pitch aptitude was the most strongly predictive factor of performance in word generalization, above and beyond musical experience. Pitch aptitude may therefore be a more reliable measure of individual pitch-related skills than musical experience to explain individual variability in tone word learning (cf. Bowles et al., 2016),

This last point relates to the third possible origin behind the observed discrepancy. It may be that musical experience, as measured by the number of years of formal instruction, may have simply been too crude a measure of an individual's musicianship. Indeed, as I signaled in the General introduction, musical experience is independent of musicality, which is measured by standardized tests. It may be that in my studies, some individuals who reported musical experience had relatively low degrees of musicality, whereas some individuals who reported no musical experience had relatively high degrees of musicality. It is also possible that the reliability of the measure of musical experience was further compromised by the fact that the study reported in Chapter 5 was web-based, thereby yielding noisier data than the lab-based studies reported in Chapters 2–4.

I formulate the following conclusion regarding the effect of musical experience on non-native tone processing:

C5: *An individual's musical experience generally facilitates non-native tone processing. This is particularly the case in pre-lexical processing for individuals who have no* *other relevant experience and who do not rely on L1 prosody to facilitate tone processing. For those participants who do rely on L1 prosody, musical experience may be less relevant. Similarly, musical experience may facilitate lexical processing if a learner has no relevant L1 experience to rely on, but musical experience is not as accurate a predictor compared to more direct measures of pitch processing skills such as pitch aptitude.*

6.2.3.4 Working memory

Working memory, operationalized through backwards digit span tasks, revealed to be a facilitative factor of tone learning, but its effect was relatively moderate compared to other factors: pitch-related skills such as musical experience and pitch aptitude appeared to be more predictive of individual performance in tone learning.

There were three instances in which WM was facilitative. The first was in pre-lexical production (imitation), but only for English speakers. The finding that an individual's capacity to recall strings predicts their ability to accurately imitate sound sequences corresponds to theoretical accounts that describe the phonological loop and sequence memory and that suggest that the ability to retain and sub-vocalize digits is similar to the retention and sub-vocalization of speech (Baddeley & Hitch, 2019; Gupta, 2003). It is slightly puzzling why WM only facilitated English, but not Mandarin participants' imitation. One possible explanation is analogous to what I suggested to be one cause of the differential relevance of musical experience on non-native tone processing, as described in section 6.2.3.3. Specifically, as was discussed in Chapter 3, it appeared that the Mandarin tone inventory affected Mandarin participants' imitation accuracy of pseudowords. It could therefore be that the effect of L1-based prosody overrode any facilitative effect that WM may have had on Mandarin speakers' imitations.

The second instance in which WM was found to be facilitative was in lexical tone processing (word identification and image-naming) by Mandarin speakers. As was argued in Chapter 2, the reason why WM only facilitated word learning for Mandarin, but not for English speakers may be because the facilitative effect of musical experience had overridden any facilitative effect of WM in the English group. As argued by feature-specific accounts of tone word learning (Bowles et al., 2016), it may have been that whereas English participants particularly experienced the word identification and image-naming tasks as *tone* word learning tasks (for pitch-related skills such as musical experience would be facilitative), the Mandarin participants experienced the tasks as more general word learning tasks, for which WM would be facilitative.

Third, WM facilitated Japanese speakers' tone categorization. This finding was relatively surprising and remains slightly difficult to interpret, given the mixed empirical evidence of the facilitative effect of WM on pre-lexical perception (Bidelman et al., 2013; Goss, 2020; Goss & Tamaoka, 2019; Hutka et al., 2015). As argued in Chapter 5, it may have been that the moraic count made Japanese listeners tap into higher (lexical) levels of speech processing, for which WM would be expected to be facilitative. Another explanation for this finding could be that the web-based nature of the study reported in Chapter 5 made the single measure of working memory employed in this study (backwards digit span) a not sufficiently reliable measure of WM.

In the larger scheme of things, WM appears to facilitate non-native tone learning only moderately. However, the fact that it certainly does explain some of the individual variability in non-native tone processing justifies the inclusion of a measure of WM and advocates the need to control for individual WM capacity to account for individual variability.

I formulate the following conclusion about the effect of WM on non-native tone processing:

C6: *As predicted by theoretical accounts of the phonological loop and sequence memory, an individual's working memory capacity can facilitate certain aspects of non-native speech learning, including the non-native learning of tone, and thus explain individual variability in tone learning facility. However, the overall effect of WM on non-native tone learning is generally moderate, and its effect may be overridden if there is a strong influence from L1-based prosody, and/or if skills related to pitch processing are more relevant to the specificities of tone learning compared to other aspects of speech or vocabulary learning.*

6.2.3.5 Pitch aptitude

Finally, I investigated to what extent pitch aptitude predicted individual variability in pre-lexical production and in lexical perception and production of tones in a non-native tone system.

Pitch aptitude only partially predicted pre-lexical production in the imitation task. As discussed in Chapters 3 and 4, this is likely because pitch aptitude taps into phonological processing, while imitation taps into phonetic processing. In addition, the relatively weak link may be due to different methodologies employed by earlier studies that did find a clear and overall effect of pitch aptitude on tone imitation accuracy (Dong et al., 2019).

Across the other studies reported in this dissertation, pitch aptitude very clearly facilitated lexical processing of tones, in both the listening and speaking modalities. This strongly supports the notion of a phonetic-phonological-lexical continuity (Wong & Perrachione, 2007) and shows that pre-lexical tone processing skills predict lexical tone processing skills. Although there was some evidence that suggested that the effect of pitch aptitude was moderated by an individual's L1 tonal status (judging from the estimate sizes of pitch aptitude on word generalization in Chapter 5), I generally observed that individuals who were good at pre-lexical tone perception were also good at lexically encoding tones.

As mentioned in the General introduction, I did not investigate to what extent pitch aptitude was further modulated by individual auditory abilities, which can be measured, for instance, by Just Noticeable Differences or adaptive pitch tests (Jongman et al., 2017; Mandel, 2009). However, as discussed in section 6.2.3.3, it appears that musical experience was in turn a relatively reliable indicator of pitch aptitude. It is therefore plausible that if I had included other measures of individual auditory ability, these would have in turn explained individual variability in pitch aptitude. I reiterate further that individual variability in the tone categorization task reported in Chapters 2–4 was relatively limited, and that consequently, there was relatively little variability in degrees of pitch aptitude in these studies. Therefore, other measures of individual pitch acuity, such as adaptive pitch tests, could be better measures if tone categorization tasks reveal ceiling effects on performance.

I conclude the following regarding the effect of pitch aptitude:

C7: *Pitch aptitude, measured by accuracy in a pre-lexical tone categorization task, is a strong indicator of how individuals perform in the processing of tones at a lexical level. Pitch aptitude is also partially indicative of performance in the pre-lexical production of* *tones. It appears that the effect of pitch aptitude is relatively universal and only slightly modulated by L1 tonal status.*

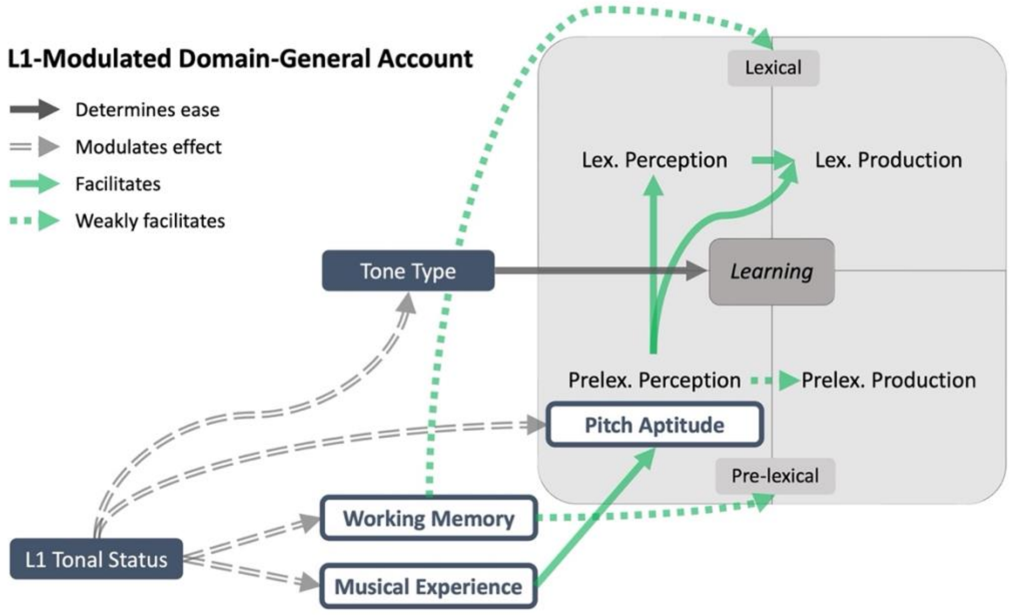
6.3 A novel and integral account of tone learning

Summarizing the conclusions of this dissertation, I here present a theoretical model that can account for individual differences in tone learning facility.

I propose an L1-Modulated Domain-General Account as a novel and integral account of tone learning. I schematically summarize this account in Figure 51, which describes how specific factors may determine the overall ease of tone learning (grey arrows), how specific factors may modulate the effect of other factors (dashed grey arrows), and how these factors may facilitate tone learning (green arrows).

Figure 51

L1-Modulated Domain-General Account of tone learning facility.



I propose that the ease with which adults learn tones in a non-native language is primarily determined by the specific shape of the tone to be learned. A tone type's inherent phonetic-acoustic properties, as well as its phonological-categorical properties, are primary factors that determine whether it is relatively easy or difficult to learn at both pre-lexical and lexical levels, and in the listening and speaking modalities.

However, an individual's L1 tonal status, and the specific tone types that occur in the L1, may modulate the effect of tone type. In phonetic-acoustic terms, speakers of some languages may exhibit better sensitivity to specific acoustic contrasts, such as differences in pitch height or contour. In phonological-categorical terms, speakers of some languages may learn certain L2 tones more easily if there is a neat categorical assimilation from L2 tone types to L1 tone types in a one-to-one fashion. This facilitative effect of categorical assimilation is strongest when the L2 and L1 tone types share functional properties (e.g., whether the tone types serve lexical or phrasal purposes) and temporal properties (e.g., whether the tone types occur over the same number of syllables). Conversely, speakers of other languages may learn certain L2 tones with greater difficulty if the categorical assimilation from L2 tones to L1 tone types follows a many-to-one assimilation pattern.

Pitch aptitude, which refers to the ability to perceive tones devoid of lexical meaning, as can be measured by a tone categorization task, is a strong facilitator of tone learning in lexical perception and lexical production, and a weak facilitator of tone learning in pre-lexical production. However, a speaker's L1 tonal status may modulate the relative effect of pitch aptitude, and the higher a tonal status, the weaker the facilitative effect of pitch aptitude on tone learning in other areas.

Pitch aptitude is in turn facilitated by musical experience, which refers to the years of musical practice an individual may have had. However, a speaker's L1 tonal status can modulate the relative effect of musical experience, and the higher a tonal status, the weaker the facilitative effect of musical experience on tone learning.

Working memory, which can be measured by a backwards digit span task, is a weak facilitator of pre-lexical and lexical tone processing. However, a speaker's L1 tonal status can modulate the relative effect of WM, and in the absence of facilitation from pitch aptitude or musical experience, WM can account for individual variability observed in tone learning facility.

The L1-Modulated Domain-General Account of non-native tone learning facility is a purely theoretical model, and although the findings from this dissertation lend empirical evidence for its tenets, more work will be needed to confirm its predictions. However, in the absence of other comprehensive models, I suggest that this account lays a foundation for future research into the study of individual variability in non-native tone learning facility.

6.4 Wider implications and avenues for future research

Throughout this dissertation I have shown that a language learner’s individual set of extralinguistic ‘tools’, such as musical experience or working memory capacity, affect performance in psycholinguistic tasks in considerable and quite complex ways. One of the wider implications of my research is therefore to acknowledge the role that these individual tools play in early stages of speech learning, and to control for these when carrying out experiments. Indeed, in an era in which experimental data are increasingly being collected remotely and outside the controlled environment of a laboratory, obtaining information on individual aptitudes and systematically varying or controlling for them is crucial to explain variability in the obtained data. Obtaining such additional information does not need to be too cumbersome or to the detriment of the core experiment: recently, template digit span tasks and musicality tests have become freely available on online experiment building platforms, and they can be easily incorporated within a researcher’s experimental battery (Anwyl-Irvine et al., 2020). If the implementation of such additional control measures is not feasible, at the very least a researcher should screen participants for musical background, especially when conducting experiments that require pitch and rhythm processing.

As I flagged in the General introduction, I did not extensively consider the efficacy of specific training methods on tone learning facility. All participants in my experiments underwent the same amount of training. For some individuals this appeared to work well, as they reached high accuracy scores in the word learning tasks, whereas other individuals would probably have benefited from more or different training to reach similar accuracy levels. However, the observation that an individual’s pitch aptitude in pre-lexical perception greatly facilitated tone learning in other areas could indicate that enhancing pitch aptitude may be a good starting point for effective tone learning at large. Enhancing pitch aptitude appears to be achievable: individuals who initially have poor tone perception skills in Mandarin can easily enhance these skills through repeated perceptual training (X. Wang, 2013). Therefore, the use of such perceptual training methods in language classroom settings, or on popular language learning applications may be of help to the overall tone learning process.

My research has also shown the evident benefits of musical practice. Indeed, musical practice does not only benefit tone learning, but also various other aspects of language learning and cognitive development (*The Importance of Music Education*, 2011). Given the benefits that musical practice can transfer to other areas of learning, this dissertation provides further support for the inclusion of musical education in school curricula.

Although the scope of my dissertation research was limited to the study of early-stage learning of tonal pseudowords by *ab initio* learners, future research should address how L1-specific and extralinguistic factors determine tone learning facility in real-life tone languages, and in settings beyond the lexical level, such as the phrasal level. Additionally, future research should look into advanced second language learners and investigate how multiple individual factors determine ultimate attainment in tone learning capacity, and how these factors may facilitate the incredibly challenging task of acquiring a second language's sound system. I have shown that behavioral experiments can shed light on the mechanisms of tone learning, but future examinations can employ different devices, such as eye-tracking and neuro-imaging to refine our understanding of these mechanisms. This will also enable me to see if the findings from this dissertation can be replicated, and will improve the generalizability of my results.

6.5 Conclusion

In this dissertation I asked what explains individual variability in non-native tone learning facility. I have presented data that show that such individual variability can be explained by a combination of individual pitch aptitude, musicianship, working memory capacity, and the shape of the tone and its relation to tone shapes in a learner's native language. These individual factors interact with one another in complex ways and explain why some individuals perform relatively well, whereas others perform relatively poorly at early stages of tone learning. I proposed an L1-Modulated Domain-General Account as a novel and integral theoretical account of these empirical findings.

The findings from this dissertation, which have brought together separate strands from previous studies, lay the foundation for an encompassing account of the linguistic and extralinguistic origins of individual variability in second language tone learning, and by extension, in speech learning at large.

References

Ajíbóyè, O., Déchaine, R. M., Gick, B., & Pulleyblank, D. (2011). Disambiguating yorùbá tones: At the interface between syntax, morphology, phonology and phonetics. *Lingua*, *121*, 1631–1648. https://doi.org/10.1016/j.lingua.2011.05.008

Antoniou, M., & Chin, J. L. L. (2018). What Can Lexical Tone Training Studies in Adults Tell Us about Tone Processing in Children? *Frontiers in Psychology*, *9*, 1–11. https://doi.org/10.3389/fpsyg.2018.00001

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). https://doi.org/10.3758/s13428-019-01237-x

Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics*, *19*(4), 537–552. https://doi.org/10.1017/S0142716400010353

Baddeley, A. D. (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

Baddeley, A. D. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. In *Cortex* (Vol. 112, pp. 91–106). Masson SpA. https://doi.org/10.1016/j.cortex.2018.05.015

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*(4), 981–1005. https://doi.org/10.3758/s13414-019-01725-4

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

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*(1), 33–58. https://doi.org/10.1017/S0272263118000074

Barcroft, J., & Sommers, M. S. (2014). Effects of variability in fundamental frequency on L2 vocabulary learning : A comparison between learners who do and do not speak a tone language. *Studies in Second Language Acquisition*, *36*(3), 423–449. https://doi.org/10.1017/S0272263113000582

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

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3, Pt.2), 1–46. https://doi.org/10.1037/h0027577

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97–103. https://doi.org/10.1037/0096-1523.32.1.97

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*, 167–200. https://doi.org/10.1016/0378-4266(91)90103-S

Best, C. T. (2019). The Diversity of Tone Languages and the Roles of Pitch Variation in Non-tone Languages: Considerations for Tone Perception Research. *Frontiers in Psychology*, *10*(February), 1–7. https://doi.org/10.3389/fpsyg.2019.00364

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In M. J. Munro & O.-S. Bohn (Eds.), *Second Language Speech Learning: the role of language experience in speech and production* (Issue January, pp. 13–34). John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0060676

Bock, K., & Levelt, W. (1994). Language Production: Grammatical Encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). Academic Press.

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer* (6.0.48). http://www.praat.org/

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch Ability As an Aptitude for Tone Learning. *Language Learning*, *66*(4), 774–808. https://doi.org/10.1111/lang.12159

Braun, B., Galts, T., & Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research*, *30*(3), 323–350. https://doi.org/10.1177/0267658313510926

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*(4), 585–594. https://doi.org/10.1016/j.wocn.2011.06.002

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). {glmmTMB} Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective.* Liber.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1). https://doi.org/10.32614/RJ-2018-017

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. X., Schwarz, I.-C., & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, *36*(6), 1459–1491. https://doi.org/10.1017/S0142716414000496

Burnham, D., Reynolds, J., Vatikiotis-Bateson, E., Yehia, H., & Ciocca, V. (2006). The perception and production of phones and tones: The role of rigid and non-rigid face and head motion. *Proceedings of the 7th International Seminar on Speech Production*, 1–8.

Burnham, D., Singh, L., Mattock, K., Woo, P. J., & Kalashnikova, M. (2018). Constraints on Tone Sensitivity in Novel Word Learning by Monolingual and Bilingual Infants: Tone Properties Are More Influential than Tone Familiarity. *Frontiers in Psychology*, *8*(JAN). https://doi.org/10.3389/fpsyg.2017.02190

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. . In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude.* (pp. 83–118). Newbury House.

Chambers, E. W., Colin De Verdière, É., Erickson, J., Lazard, S., Lazarus, F., & Thite, S. (2010). Homotopic Fréchet distance between curves or, walking your dog in the woods in polynomial time. *Computational Geometry: Theory and Applications*, *43*(3), 295–311. https://doi.org/10.1016/j.comgeo.2009.02.008

Chan, I. L., & Chang, C. B. (2019). Perception of nonnative tonal contrasts by Mandarin-English and English-Mandarin sequential bilinguals. *The Journal of the Acoustical Society of America*, *146*(2), 956–972. https://doi.org/10.1121/1.5120522

Chan, R. K. W., & Leung, J. H. C. (2020). Why are Lexical Tones Difficult to Learn? *Studies in Second Language Acquisition*, *42*(1), 33–59. https://doi.org/10.1017/S0272263119000482

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

Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126. https://doi.org/10.1121/1.4976037

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. University of California Press.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, *31*(6). https://doi.org/10.1080/23273798.2016.1156715

Chen, A., Peter, V., Wijnen, F., Schnack, H., & Burnham, D. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language*, *180–182*. https://doi.org/10.1016/j.bandl.2018.04.006

Chen, F., & Peng, G. (2018). Lower-level acoustics underlie higher-level phonological categories in lexical tone perception. *The Journal of the Acoustical Society of America*, *144*(3), EL158–EL164. https://doi.org/10.1121/1.5052205

Chen, J., Best, C. T., & Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics*, *83*, 101013. https://doi.org/10.1016/j.wocn.2020.101013

Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PLoS ONE*, *12*(7). https://doi.org/10.1371/journal.pone.0180656

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

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary language. *Developmental Psychology*, *32*(5), 867–873. https://doi.org/10.1037/0012-1649.32.5.867

Chiao, W.-H., Kabak, B., & Braun, B. (2011). When more is less: Non-native perception of level tone contrasts. *Proceedings of the Psycholinguistic Representation of Tone Conference*, 42–45. https://pdfs.semanticscholar.org/86d2/63cf6d486dca683f77ae0ef4d0ea4c3bb2c5.pdf

Choi, W. (2021). Musicianship Influences Language Effect on Musical Pitch Perception. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.712753

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory*, *14*(7), 804–813. https://doi.org/10.1080/09658210600680020

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., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, *59*(2), 165–179. https://doi.org/10.3758/BF03211886

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Ding, H., Hoffmann, R., & Jokisch, O. (2011). An Investigation of Tone Perception and Production in German Learners of Mandarin. *Archives of Acoustics*, *36*(3), 509–518. https://doi.org/10.2478/v10168-011-0036-6

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*(8), e7191. https://doi.org/10.7717/peerj.7191

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words: Research report. *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*(3), 199–210. https://doi.org/10.1016/j.jml.2010.12.004

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

Engstrand, O. (1997). Phonetic Interpretation of the Word Accent Contrast in Swedish: Evidence from Spontaneous Speech. *Phonetica*, *54*(2), 61–75. https://doi.org/10.1159/000262211

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*. https://doi.org/10.1016/j.wocn.2012.02.007

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Pub. Co.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, *24*(3), 369–383. https://doi.org/10.1017/S0142716403000195

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. In *Journal of Phonetics* (Vol. 15).

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). York Press. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Francis, A. L., Ciocca, V., & Chit Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029–1044. https://doi.org/10.3758/BF03194832

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. https://www.r-project.org

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage Differences in Tone Perception: a Multidimensional Scaling Investigation. *Language and Speech*, *21*(1), 1–33. https://doi.org/10.1177/002383097802100101

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83–94. https://doi.org/10.3758/BF03210559

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., & Tamaoka, K. (2015). Predicting lexical accent perception in native Japanese speakers: An investigation of acoustic pitch sensitivity and working memory. *Japanese Psychological Research*, *57*(2), 143–154. https://doi.org/10.1111/jpr.12076

Goss, S., & 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., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *The Journal of the Acoustical Society of America*, *115*(5), 2545–2545. https://doi.org/10.1121/1.4783674

Grabe, E., Kochanski, G., & Coleman, J. (2004). The intonation of native accent varieties in the British Isles: potential for miscommunication? *Regional Variations in Intonation*, 9–31. https://doi.org/10.1007/s12603-014-0559-4

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375–401. https://doi.org/10.1177/00238309030460040201

Grazia Busà, M., & Urbani, M. (2011). a Cross Linguistic Analysis of Pitch Range in English L1 and L2. *Proc. 17th International Congress of Phonetic Sciences (ICPhS XVII)*, *August*, 380–383.

Green, P., & MacLeod, C. J. (2016). SIMR : an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Guevara-Rukoz, A., Lin, I., Morii, M., Minagawa, Y., Dupoux, E., & Peperkamp, S. (2017). Which epenthetic vowel? Phonetic categories versus acoustic detail in perceptual vowel epenthesis. *The Journal of the Acoustical Society of America*, *142*(2), EL211–EL217. https://doi.org/10.1121/1.4998138

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56 A*(7), 1213–1236. https://doi.org/10.1080/02724980343000071

Gussenhoven, C. (2005). Transcription of Dutch Intonation. In *Prosodic Typology* (pp. 118–145). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.003.0005

Gut, U. (2009). Research methodology in L2 phonological acquisition and the corpus-linguistic approach. In *Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic* (pp. 39–62). Lang, Peter.

Haendler, Y., Lassotta, R., Adelt, A., Stadie, N., Burchert, F., & Adani, F. (2020). *Bayesian Analysis as Alternative to Frequentist Methods: A Demonstration with Data from Language-Impaired Children's Relative Clause Processing*.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

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

Hao, Y.-C. (2018). Contextual effect in second language perception and production of Mandarin tones. *Speech Communication*, *97*. https://doi.org/10.1016/j.specom.2017.12.015

Hao, Y.-C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, *54*, 151–168. https://doi.org/10.1016/j.wocn.2015.10.003

Hartig, F. (2020). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. http://florianhartig.github.io/DHARMa/

Hayes-Harb, R., & Barrios, S. (2021). Native English speakers and Hindi consonants: From cross-language perception patterns to pronunciation teaching. *Foreign Language Annals*. https://doi.org/10.1111/flan.12566

Hoenig, J. M., & Heisey, D. M. (2001). The Abuse of Power. *The American Statistician*, *55*(1), 19–24. https://doi.org/10.1198/000313001300339897

Huang, B. H., & Jun, S. A. (2011). The effect of age on the acquisition of second language prosody. *Language and Speech*, *54*(3), 387–414. https://doi.org/10.1177/0023830911402599

Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, *71*, 52–63. https://doi.org/10.1016/j.neuropsychologia.2015.03.019

Ingvalson, E. M., Nowicki, C., Zong, A., & Wong, P. C. M. (2017). Non-native speech learning in older adults. *Frontiers in Psychology*, *8:148*, 1–10. https://doi.org/10.3389/fpsyg.2017.00148

Jongman, A., Qin, Z., Zhang, J., & Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America*, *142*(EL163). https://doi.org/10.1121/1.4995526

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neuroscience*, *9*, 53. https://doi.org/10.1186/1471-2202-9-53

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, *192*, 15–24. https://doi.org/10.1016/j.bandl.2019.02.004

Kaland, C. (2021). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*, 1–30. https://doi.org/10.1017/S0025100321000049

Kan, R. T. Y., & Schmid, M. S. (2019). Development of tonal discrimination in young heritage speakers of Cantonese. *Journal of Phonetics*, *73*, 40–54. https://doi.org/10.1016/j.wocn.2018.12.004

Katsura, M. (1969). Notes on Some Phonological Aspects of Northern Thai. *Tonan Ajia Kenkyu (The Southeast Asian Studies)*, *7*(2).

Kawahara, S. (2015). The phonology of Japanese Accent. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 445–492). De Gruyter Mouton.

Kirby, J., & Giang, Đ. L. (2021). Relating Production and Perception of L2 Tone. In *Second Language Speech Learning* (pp. 249–272). Cambridge University Press. https://doi.org/10.1017/9781108886901.010

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, *13*(4), 646–653. https://doi.org/10.1006/nimg.2000.0738

Köhnlein, B. (2020). Tone Accent in North and West Germanic. In *The Cambridge Handbook of Germanic Linguistics* (pp. 143–166). Cambridge University Press. https://doi.org/10.1017/9781108378291.008

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism*, *11*(2), 261–271. https://doi.org/10.1017/S1366728908003416

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*(1), 81–95. https://doi.org/10.1016/j.jneuroling.2009.09.001

Ladd, D. R. (2012). Analysis and transcription of intonation. In *Intonational Phonology* (pp. 87–130). Cambridge University Press. https://doi.org/10.1017/cbo9780511808814.004

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Laméris, T. J., & Graham, C. (2020). L2 Perception and Production of Japanese Lexical Pitch. *Journal of Monolingual and Bilingual Speech*, *2*(1), 106–136. https://doi.org/https://doi.org/10.1558/jmbs.14948

Laméris, T. J., & Post, B. (2022). 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*, 026765832210900. https://doi.org/10.1177/02676583221090068

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

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. https://doi.org/10.1121/1.2990713

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. In *Perception & Psychophysics* (Vol. 53, Issue 2).

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. https://cran.r-project.org/package=emmeans

Lewis, P. (1994). The Acquisition of Clicks by Non-Mother-Tongue Speakers. *Paper Presented at the Conference on Linguistics for the Language Professions (Stellenbosch, South Africa)*, 18.

Li, A., & Post, B. (2014). L2 Acquisition of Prosodic Properties of Rhythm. *Studies in Second Language Acquisition*, *36*(2), 223–255. https://doi.org/10.1017/S0272263113000752

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

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Ling, W., & Grüter, T. (2020). From sounds to words: The relation between phonological and lexical processing of tone in L2 Mandarin. *Second Language Research*. https://doi.org/10.1177/0267658320941546

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech*, *47*(2), 109–138. https://doi.org/10.1177/00238309040470020101

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. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2018*(September), 3791–3794. https://doi.org/10.21437/Interspeech.2018-2104

Mandel, J. (2009). *Adaptive pitch test, http://jakemandell.com/adaptivepitch/.*

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. In *Behavior Research Methods*. https://doi.org/10.3758/s13428-011-0168-7

Mattys, S. L., & Baddeley, A. D. (2019). Working memory and second language accent acquisition. *Applied Cognitive Psychology*, *33*(6), 1113–1123. https://doi.org/10.1002/acp.3554

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speaker of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258. https://doi.org/10.1006/jpho.2002.0174

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and Language in Contact : L2 Acquisition, Attrition and Languages in Multilingual Situations* (pp. 171–188). Springer. https://doi.org/10.1007/978-3-662-45168-7_9

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Moen, I., & Sundet, K. (1996). Production and perception of word tones (pitch accents) in patients with left and right hemisphere damage. *Brain and Language*. https://doi.org/10.1006/brln.1996.0048

Morett, L. M. (2020). The Influence of Tonal and Atonal Bilingualism on Children's Lexical and Non-Lexical Tone Perception. *Language and Speech*, *63*(2), 221–241. https://doi.org/10.1177/0023830919834679

Moyer, A. (Ed.). (2013). The scope and relevance of accent. In *Foreign Accent: The Phenomenon of Non-native Speech* (pp. 9–20). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511794407.002

Müller, M. (2007). Information retrieval for music and motion. *Information Retrieval for Music and Motion*, *January 2007*, 1–313. https://doi.org/10.1007/978-3-540-74048-3

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J. D. E., & Desimone, R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences*, *115*(28), E6630–E6639. https://doi.org/10.1073/pnas.1808412115

Nation, P. (2000). Learning Vocabulary in Lexical Sets: Dangers and Guidelines. *TESOL Journal*. https://doi.org/10.1002/j.1949-3533.2000.tb00239.x

Neri, A., Cucchiarini, C., & Strik, H. (2006). Selecting segmental errors in non-native Dutch for optimal pronunciation training. *IRAL - International Review of Applied Linguistics in Language Teaching*, *44*(4), 357–404. https://doi.org/10.1515/IRAL.2006.016

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2018). Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German. *Cognitive Science*, *42*, 1075–1100. https://doi.org/10.1111/cogs.12589

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

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. https://doi.org/10.1016/S0191-8869(99)00251-2

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Olsen, M. K. (2012). The L2 Acquisition of Spanish Rhotics by L1 English Speakers: The Effect of L1 Articulatory Routines and Phonetic Context for Allophonic Variation. *Source: Hispania*, *9517425415*(1), 65–82. https://doi.org/10.1353/hpn.2012.0008

Ota, M. (2003). The Development of Lexical Pitch Accent Systems: An Autosegmental Analysis. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *48*(3–4), 357–383. https://doi.org/10.1017/S0008413100000700

Ota, M. (2006). Children's production of word accents in Swedish revisited. *Phonetica*, *63*(4), 230–246. https://doi.org/10.1159/000097307

Ota, M. (2016). *Prosodic Phenomena* (J. L. Lidz, W. Snyder, & J. Pater, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199601264.013.5

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*(JUN). https://doi.org/10.3389/fpsyg.2011.00142

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced Second Language Learners' Perception of Lexical Tone Contrasts. *Studies in Second Language Acquisition*, *41*(1), 59–86. https://doi.org/10.1017/S0272263117000444

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2020). Even in the Best-Case Scenario L2 Learners Have Persistent Difficulty Perceiving and Utilizing Tones in Mandarin. *Studies in Second Language Acquisition*, 1–29. https://doi.org/10.1017/s027226312000039x

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of Musical Disorders. *Annals of the New York Academy of Sciences*, *999*(1), 58–75. https://doi.org/10.1196/annals.1284.006

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367. https://doi.org/10.1016/S1364-6613(03)00150-5

Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for Shared Cognitive Processing of Pitch in Music and Language. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0073372

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

Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult Learning of Novel Words in a Non-native Language: Consonants, Vowels, and Tones. *Frontiers in Psychology*, *9*(JUL). https://doi.org/10.3389/fpsyg.2018.01211

Post, B., Stamatakis, E. A., Bohr, I., Nolan, F., & Cummins, C. (2015). Categories and gradience in intonation A functional Magnetic Resonance Imaging study. In *The Phonetics/Phonology Interface: Sounds, representations, methodologies*. https://doi.org/10.1075/cilt.335.13pos

Qin, Z., & Jongman, A. (2016). Does Second Language Experience Modulate Perception of Tones in a Third Language? *Language and Speech*, *59*(3), 318–338. https://doi.org/10.1177/0023830915590191

Qin, Z., Tremblay, A., & Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics*, *73*. https://doi.org/10.1016/j.wocn.2019.01.002

Qin, Z., Zhang, C., & Wang, W. S. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America*, *149*(1), 435–446. https://doi.org/10.1121/10.0003330

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Ramachers, S., Brouwer, S., & Fikkert, P. (2017). How Native Prosody Affects Pitch Processing during Word Learning in Limburgian and Dutch Toddlers and Adults. *Frontiers in Psychology*, *8*(SEP). https://doi.org/10.3389/fpsyg.2017.01652

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, *77*(2), 571–591. https://doi.org/10.3758/s13414-014-0791-3

Roll, M., Söderström, P., & Horne, M. (2011). The marked status of Accent 2 in Central Swedish. *ICPhS XVII Regular Session Hong Kong*, 17–21.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. https://doi.org/10.1068/p5117

Sadakata, M., Weidema, J. L., & Honing, H. (2020). Parallel pitch processing in speech and melody: A study of the interference of musical melody on lexical pitch perception in speakers of Mandarin. *PLOS ONE*, *15*(3), e0229109. https://doi.org/10.1371/journal.pone.0229109

Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology*, *5*(4), 489–522. https://doi.org/10.1515/lp-2014-0016

Schmidt, E., Pérez, A., Cilibrasi, L., & Tsimpli, I. (2020). Prosody facilitates memory recall in L1 but not in L2 in highly proficient listeners. *Studies in Second Language Acquisition*, *42*(1), 223–238. https://doi.org/10.1017/S0272263119000433

Schmitz, J., Díaz, B., Fernández Rubio, K., & Sebastian-Galles, N. (2018). Exploring the relationship between speech perception and production across phonological processes, language familiarity, and sensory modalities. *Language, Cognition and Neuroscience*, *33*(5), 527–546. https://doi.org/10.1080/23273798.2017.1390142

Segerup, M., & Nolan, F. (2004). Gothenburg Swedish word accents: a case of cue trading? *Nordic Prosody - Proceedings of the IXth Conference*, 225–233.

Shih, C., & Lu, H. Y. D. (2010). Prosody transfer and suppression: Stages of tone acquisition. *Proceedings of the International Conference on Speech Prosody*, *December*.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293. https://doi.org/10.1177/0023830909357156

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. https://doi.org/10.1017/S0272263114000047

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, *22*(2), 286–296. https://doi.org/10.1080/09541440902771299

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019). The acquisition of phonological alternations: The case of the Mandarin tone sandhi process. *Applied Psycholinguistics*, *40*(6), 1495–1526. https://doi.org/10.1017/S0142716419000353

Tang, W., Xiong, W., Zhang, Y. xuan, Dong, Q., & Nan, Y. (2016). Musical experience facilitates lexical tone processing among Mandarin speakers: Behavioral and neural evidence. *Neuropsychologia*, *91*, 247–253. https://doi.org/10.1016/j.neuropsychologia.2016.08.003

The Importance of Music Education (DFE-00086-2011). (2011).

Tsukada, K., & Kondo, M. (2019). The Perception of Mandarin Lexical Tones by Native Speakers of Burmese. *Language and Speech*, *62*(4), 625–640. https://doi.org/10.1177/0023830918806550

van Dommelen, W. A., & Husby, O. (2009). The perception of Norwegian word tones by second language speakers. *The Journal of the Acoustical Society of America*, *125*(4), 2773–2773. https://doi.org/10.1121/1.4784743

van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335. https://doi.org/10.1016/j.jml.2003.10.003

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. https://doi.org/10.1016/j.wocn.2018.07.008

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188–196. https://doi.org/10.1016/j.lindif.2010.02.004

Wang, X. (2013). Perception of mandarin tones: The effect of L1 background and training. *Modern Language Journal*, *97*(1), 144–160. https://doi.org/10.1111/j.1540-4781.2013.01386.x

Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, *25*(03), 2766. https://doi.org/10.1017/S0142716404001213

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*(February), 1033–1043. https://doi.org/https://doi.org/10.1121/1.1531176

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. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*. https://doi.org/10.1016/j.wocn.2010.10.001

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

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

Wiener, S., Ito, K., & Speer, S. R. (2020). Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinations. *Studies in Second Language Acquisition*, 1–26. https://doi.org/10.1017/S0272263120000418

Wiener, S., & Lee, C. Y. (2020). Multi-Talker Speech Promotes Greater Knowledge-Based Spoken Mandarin Word Recognition in First and Second Language Listeners. *Frontiers in Psychology*, *11*(February), 1–14. https://doi.org/10.3389/fpsyg.2020.00214

Wong, P. C. M., Kang, X., Wong, K. H. Y., So, H.-C., Choy, K. W., & Geng, X. (2020). ASPM -lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances*, *6*(22), eaba5090. https://doi.org/10.1126/sciadv.aba5090

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

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7). https://doi.org/10.3758/s13414-017-1361-2

Wu, H., Ma, X., Zhang, L., Liu, Y., Zhang, Y., & Shu, H. (2015). Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Frontiers in Psychology*, *06*(APR). https://doi.org/10.3389/fpsyg.2015.00436

Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, *46*(1), 86–100. https://doi.org/10.1016/j.wocn.2014.06.005

Xu, Y. (2013). ProsodyPro - A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody*, 7–10.

Yang, C. (2019). The effect of L1 tonal status on the acquisition of L2 Mandarin tones. *International Journal of Applied Linguistics*, *29*(1), 3–16. https://doi.org/10.1111/ijal.12223

Yang, R. (2015). The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics*, *43*(1B), 453–472. https://doi.org/10.1353/jcl.2015.0035

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Yu, A. C. L., Lee, C. W. T., Lan, C., & Mok, P. P. K. (2021). A New System of Cantonese Tones? Tone Perception and Production in Hong Kong South Asian Cantonese. *Language and Speech*, 1–25. https://doi.org/10.1177/00238309211046030

Yu, K., Li, L., Chen, Y., Zhou, Y., Wang, R., Zhang, Y., & Li, P. (2019). Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology*, *56*(11). https://doi.org/10.1111/psyp.13448

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, *136*(3), 1320–1333. https://doi.org/10.1121/1.4887462

Yu, K., Zhou, Y., Li, L., Su, J., Wang, R., & Li, P. (2017). The interaction between phonological information and pitch type at pre-attentive stage: an ERP study of lexical tones. *Language, Cognition and Neuroscience*, *32*(9), 1164–1175. https://doi.org/10.1080/23273798.2017.1310909

Zhang, H. (2018). *Second Language Acquisition of Mandarin Chinese Tones*. Brill Rodopi. https://doi.org/10.1163/9789004364790

Zhang, K., & Peng, G. (2017). The Relationship Between the Perception and Production of Non-Native Tones. *Interspeech 2017*, 1799–1803. https://doi.org/10.21437/Interspeech.2017-714

Zhang, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S. Y. (2018). The effect of speech variability on tonal language speakers' second language lexical tone learning. *Frontiers in Psychology*, *9*(OCT). https://doi.org/10.3389/fpsyg.2018.01982

Zhang, Y., & Kirby, J. (2020). The role of F 0 and phonation cues in Cantonese low tone perception . *The Journal of the Acoustical Society of America*, *148*(1), EL40–EL45. https://doi.org/10.1121/10.0001523

Zhu, M., Chen, X., & Yang, Y. (2021). The effects of native prosodic system and segmental context on Cantonese tone perception by Mandarin and Japanese listeners. *The Journal of the Acoustical Society of America*, *149*(6), 4214–4227. https://doi.org/10.1121/10.0005274

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. *7th*

International Conference on Speech Prosody 2014, *08-12-Sept*(October), 1037–1041. https://doi.org/10.21437/SpeechProsody.2014-196