



Research Article

Cite this article: Laméris, T. J., Llompart, M., & Post, B. (2023). Non-native tone categorization and word learning across a spectrum of L1 tonal statuses. *Bilingualism: Language and Cognition*, 1–15. <https://doi.org/10.1017/S1366728923000871>

Received: 23 November 2022
Revised: 11 November 2023
Accepted: 21 November 2023

Keywords:

Individual variability; lexical tone; word learning; working memory; musical experience

Corresponding author:

Tim Joris Laméris;
Email: t.j.lameris@hum.leidenuniv.nl

Abstract

Adults differ in the ease with which they acquire lexical tones in a non-native language. Individual differences have been attributed to several factors, such as the role that pitch plays in a learner’s L1 to signal lexical meaning (L1 tonal status), the shape of the tones to be acquired (tone types), as well as extralinguistic factors (such as musical experience and working memory). Here, we ask whether learners from a spectrum of L1 tonal statuses (Dutch, Swedish and Japanese, and Thai) differ in their tone word learning facility, whilst we simultaneously investigate the effects of tone type, and musical experience and working memory. Our findings suggest that above and beyond L1 tonal status, the strongest predictor of tone word learning was pre-lexical tone processing (measured by a tone categorization task), although the strength of the link between pre-lexical and lexical processing may be modulated by L1 tonal status.

Introduction

Adults find it relatively difficult to learn lexical tone – the primary use of pitch to determine word meaning – in a non-native language (Pelzl et al., 2020, 2021), but some individuals appear to learn tones more easily than others do. Previous studies have identified several factors that could explain this individual variability in learning facility. Among these are i) the role that pitch plays in the native language (L1) to signal lexical meaning (‘L1 tonal status’); ii) the specific shape of the non-native (L2) tones and the potential interaction with tonal or intonational patterns from the L1 (‘tone types’); and iii) individual extralinguistic factors such as musical experience and working memory.

The effect of L1 tonal status

Studies that examine the first factor (L1 tonal status) are rooted in theoretical accounts that propose that a lexically contrastive speech feature in a non-native language may be relatively easy to learn for learners whose L1 also uses that feature contrastively, compared to learners whose L1 does not. Examples of such accounts are the “Feature Hypothesis” (McAllister et al., 2002) for the non-native acquisition of duration and the “Levels of Representation Account” (Francis et al., 2008, p. 269) for acquisition of non-native tones. Crucially, languages cannot usually be characterized in a binary way as either using a specific speech feature or not, but they are usually on a spectrum. In the case of pitch, some languages do not use it as a primary cue for lexical purposes (non-tonal languages like Dutch, which typically use pitch as a primary cue at a phrasal level in intonation), whereas some use it to differentiate a limited set of words (pitch-accent languages like Swedish and Japanese), and yet others use it at the syllable level to distinguish lexical items (tonal languages like Thai). We will henceforth use the term ‘L1 tonal status’ to describe such typological differences. If lexical tones are indeed relatively easier to learn for individuals whose L1 also uses pitch contrastively, a question that follows is if this relative ease is incremental according to different degrees of L1 tonal status. A study by Schaefer and Darcy (2014) sought to answer this very question. They recruited a group of speakers on a spectrum of different L1 tonal statuses: Mandarin (high), Japanese (high-intermediate), English (low-intermediate), and Standard Korean (low), and tested these speakers on their ability to discriminate Thai tonal contrasts. Their results support the intuition that L1 tonal status facilitates non-native tone processing, and that it does so incrementally. That is, speakers with the highest L1 tonal status (Mandarin) were faster and more accurate in Thai tone discrimination than speakers with an intermediate L1 tonal status

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

(Japanese), who in turn outperformed speakers with a low L1 tonal status (English and Korean).

Based on these findings, Schaefer and Darcy proposed the “Functional Pitch Hypothesis” (henceforth: “FPH”), which posits that L1 tonal status shapes perception of a non-native tone system. The FPH predicts that it may be particularly challenging to transfer sensitivity to pitch variations from a given domain in the L1 to a smaller domain in the L2. For instance, non-tonal L1ers, who are familiar with pitch variations at the phrasal level for non-lexical purposes (intonation), may find it relatively difficult to apply this sensitivity at a syllable level for lexical purposes (lexical tone) in an L2.

Although the overall predictions of the FPH are supported by a number of studies that show that learners with a high L1 tonal status outperform learners with a lower L1 tonal status in non-native tone processing (R. K. W. Chan & Leung, 2020; Peng *et al.*, 2010; Wayland & Guion, 2004), many studies do not find such clear differences (Cooper & Wang, 2012; Francis *et al.*, 2008; Gandour & Harshman, 1978; So & Best, 2010), and some only find a selective advantage for speakers of a higher L1 tonal status for specific tone types (Burnham *et al.*, 2015; Francis *et al.*, 2008; Hao, 2012; Qin & Jongman, 2016; Zhu *et al.*, 2021). For instance, Burnham *et al.* (2015) found that non-tonal L1ers (English) were less accurate than the aggregate of pitch-accent (Swedish) and tonal (Thai, Cantonese, Mandarin) L1ers to discriminate Thai tones. However, when looking at discrimination accuracy per tone type, they found differences between English speakers and the aggregate of pitch-accent and tonal speakers’ discrimination accuracy for only four out of ten possible tonal contrasts (Burnham *et al.*, 2015, p. 1475). In other words, a perceptual advantage for speakers from higher L1 tonal statuses was not found across the board, but only for specific tone types.

The effect of tone type

Some tone types appear to be inherently easy or difficult to process according to their acoustic properties. For instance, neurological evidence suggests that adults are better at registering F0 rises than falls in the brainstem (Krishnan *et al.*, 2010). Further, dynamic-dynamic tone contrasts (i.e., between two contour tones) appear to be more difficult to process than contrasts between a dynamic and a static (level) tone (Burnham *et al.*, 2018).

In addition, certain tone types may be differentially easy or difficult based on a learner’s L1. For instance, recent versions of the Perceptual Assimilation Model (PAM) propose that listeners perceptually assimilate non-native tone categories to native tone categories (Best, 2019; J. Chen *et al.*, 2020). In a nutshell, when non-native tones clearly assimilate to separate native tone types in a one-to-one manner (‘two-category assimilation’), they are perceptually distinct and easy to perceive, but when they assimilate to a single tone category in a many-to-one manner (‘single-category assimilation’), they may be confusing and difficult to perceive. For instance, Burnham *et al.* (2015) showed that Mandarin-L1ers outperformed Cantonese-L1ers in the discrimination of Thai static-static tone contrasts. They suggest that this is due to the greater number of static tones in Cantonese, which may increase the likelihood of perceptual confusion (Burnham *et al.*, 2015, p. 1479).

Because the ease of non-native tone processing appears to be affected by the potential interference with native tone categories (in addition to inherent difficulties of certain tone types based

on their acoustic properties), the very question of whether a higher L1 tonal status facilitates non-native tone processing needs to factor in the possibility that certain tone types may be differentially difficult depending on a learner’s L1. This appears to be particularly important for learners from a tonal L1 background, who exhibit clear L1-specific performance for specific tone types in pre-lexical perception (J. Chen *et al.*, 2020; Hao, 2012; Tsukada & Kondo, 2019), and in tone word learning (Laméris *et al.*, 2023; Laméris & Post, 2022). It is less clear whether non-tonal and pitch-accent L1ers also assimilate non-native tone categories to native pitch-accent or intonational patterns, and would therefore exhibit similarly clear L1-specific performance per tone type. Although some studies suggest that this may be the case (Braun *et al.*, 2014; Braun & Johnson, 2011), it appears that non-tonal L1ers may be less prone to strong interference from native prosodic categories (pitch-accent or intonational) on non-native tone processing in comparison to tonal L1ers (Laméris & Post, 2022; Reid *et al.*, 2015; So & Best, 2010; A. Yu *et al.*, 2021). Instead, non-tonal L1ers may process non-native tones more psychoacoustically (Best, 2019, p. 5; K. Yu *et al.*, 2019).

The effect of extralinguistic factors

A third potential source of individual variability in tone learning facility is extralinguistic in nature. Here, we focus on two factors that we deemed relevant for the present study: musical experience and working memory (WM). Musical experience refers to the years of musical practice an individual may have had, often through formal training. According to theoretical models like OPERA, musical experience is assumed to enhance pitch acuity in the domain of music, which may be transferred to the domain of speech (Choi, 2021; Patel, 2011). Various studies on non-native tone processing have found facilitative effects of musical experience (Bowles *et al.*, 2016; Götz *et al.*, 2023; Wong *et al.*, 2020; Wong & Perrachione, 2007), although such effects may be task-dependent (Chang *et al.*, 2016).

WM refers to the capacity to recall and process strings of information, as can be operationalized by a digit span task (Baddeley, 2003; Mattys & Baddeley, 2019). Evidence on the effects of WM on non-native tone processing is mixed: some studies showing (some) facilitative effects (Bowles *et al.*, 2016; Goss, 2020; Ingvalson *et al.*, 2017; Laméris & Post, 2022) whereas others do not (Goss & Tamaoka, 2019; Perrachione *et al.*, 2011). Despite these mixed findings from the tone-learning literature, we deemed WM to be a relevant factor of interest in the present study because our study involved novel word learning, for which WM appears to be facilitative (Atkins & Baddeley, 1998; Gupta, 2003; Kormos & Sáfár, 2008; Linck *et al.*, 2014).

Motivation for present study

The present study is inspired by two previous studies (Burnham *et al.*, 2015; Schaefer & Darcy, 2014) that investigated non-native tone processing by L1ers on a spectrum of L1 tonal statuses. We expand on the work from these studies in two ways. First, we investigated tone processing not only at a pre-lexical level (by means of a tone categorization task), but also at a lexical level (by means of a tone word identification task), and probe the link between the two. 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 processing at the lexical level will provide a more complete account of tone acquisition. This is especially relevant in the light of the recurrent finding that while learners may do relatively well at the pre-lexical perception of challenging L2 features – such as vowels (Díaz et al., 2012; Llompert, 2021), consonants (Darcy et al., 2013; Hayes-Harb & Masuda, 2008), stress (Dupoux et al., 2008), or tones (Pelzl et al., 2019) – they may not effectively link these to word meaning at the lexical level. Yet, the individual ability to perceive tones pre-lexically is often a strong predictor of individual ability to use those tones at a lexical level (Bowles et al., 2016; Cooper & Wang, 2012; Dong et al., 2019; Ling & Grüter, 2020; Wong & Perrachione, 2007)¹.

Interestingly, the few studies that assessed the link between pre-lexical and lexical processing cross-linguistically suggest that predictive strength of tone categorization on tone word learning may be modulated by a learners' L1 tonal status. For instance, tone categorization performance predicts Cantonese tone word learning for non-tonal English speakers, but not for tonal Thai speakers (Cooper & Wang, 2012). This may be because the beneficial effect from pitch processing skills – as can be operationalized by tone categorization but also by musicianship – and from L1 experience with a tonal language may not be additive (cf. Choi, 2021; Laméris & Post, 2022; Maggu et al., 2018; S. Chen et al., 2020). Indeed, extralinguistic pitch processing skills may only facilitate non-native tone learning if the learner lacks “relevant experience” such as prior L1 knowledge of a tonal language (Choi, 2021). Based on these findings, we also examine whether tone categorization is differentially (and incrementally) facilitative to tone word learning according to different L1 tonal statuses.

As a second expansion from previous work, we investigate the effect of L1 tonal status whilst simultaneously addressing the effects of tone type and the relative difficulty of the tones in question – which may be inherent and/or L1-specific – and the effects of individual differences in musical experience and working memory, as the orthogonal effects of all these factors together were not addressed in the studies by Schaefer and Darcy (2014) and Burnham et al. (2015).

To this end, we explore the effect of different L1 tonal statuses – low (Dutch), intermediate (Swedish and Japanese), and high (Thai) – on non-native tone learning viewed at a pre-lexical and at a lexical level, as well as the relationship between the two levels. We formulate the following research questions:

RQ1: Does L1 tonal status facilitate non-native tone perception and word learning, when the effects of tone type, musical experience, and working memory are simultaneously taken into account?

RQ2: Does tone categorization predict tone word learning, and if so, is this similar for speakers from different L1 tonal statuses?

The present study

A spectrum of L1 Tonal Statuses

We investigated non-native tone perception and word learning by native speakers of Dutch, Swedish, Japanese, and Thai, which represent a spectrum of different L1 tonal statuses, as summarized in Table 1.

Standard Dutch is a non-tonal language in which pitch alone is typically not used for lexical distinctions (Ramachers et al., 2017, p. 2). Its L1 tonal status is therefore low.

Central Swedish is a pitch-accent language in which words can be distinguished in meaning by an “acute” Accent I and a “grave”

Table 1. Respective L1 tone statuses according to language type and domain, adapted from Schaefer & Darcy (2014).

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

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 a non-tonal language like Dutch.

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, and different pitch patterns on otherwise segmentally identical words can be used for lexical distinctions.

Central Thai is a tone language with a high tonal status. It has 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 (Chao, 1968) of these tones are 315 (rising), 51/241² (falling), 45 (high), 33 (mid) and 21 (low) (Burnham et al., 2015, p. 1460; X. Wu et al., 2014, p. 90).

Previous studies involving Dutch, Swedish, Japanese or Thai speakers show inconsistent evidence of an incremental effect of L1 tonal status on non-native tone processing. For instance, Dutch speakers perform as accurately as tone language speakers in some tone perception tasks (A. Chen et al., 2016; Cutler & Chen, 1997) but they perform less accurately in lexical tasks (Zou et al., 2022). Perceptual studies with native speakers of Swedish (Burnham et al., 2015) and Japanese (Schaefer & Darcy, 2014; Zhu et al., 2021) suggest that their non-native tone perception accuracy sits somewhere between that of non-tone and tone language speakers. However, Braun et al. (2014) found that Japanese speakers performed worse than non-tonal (German) speakers in a lexical tone task. In what appears to be the only study that tested Thai and non-tonal speakers' perception and word learning of a language that was non-native to both groups, Cooper and Wang (2012) showed that Thai speakers without musical experience performed similarly to their English counterparts in Cantonese tone perception, but outperformed them in word learning.

Pseudoword tone system

We examined perception and word learning in a pseudoword tone system consisting of a low-level (11), a falling (51), and a peaking (141) tone. Figure 1 visualizes the pseudoword and respective L1 tone types. The static low-level tone contrasts with the dynamic falling and peaking tones in both height and direction (static-dynamic contrast). The fall-peak contrast constitutes a dynamic-dynamic contrast. Our motivation for choosing this

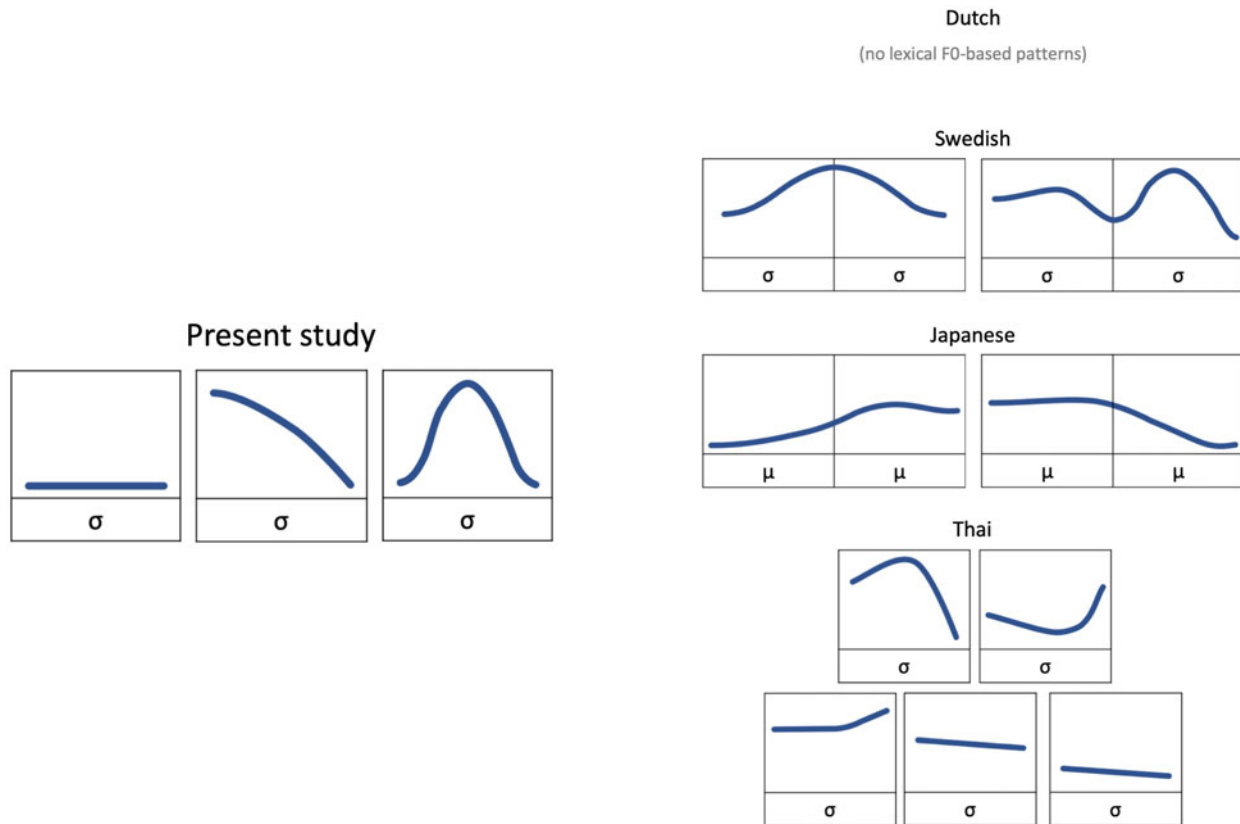


Figure 1. Overview of pseudoword tone system and lexical tone types in L1s.

Note: The tone type visualisations were adapted from Köhnlein (2020 pp. 154–155) for Swedish, Shport (2016, p. 744) for Japanese, and Burnham *et al.* (2014, p. 1461) for Thai.

tone system is that we hypothesize, as we will outline below, that the relative difficulty of the tones (*vis-à-vis* one another) should be similar for all L1s involved. Specifically, we hypothesize that the static level tone will be relatively easy, whereas the dynamic falling and peaking tones may be relatively difficult to categorize and to process at the word level.

We propose the abovementioned hierarchy of tone difficulty for two reasons. In the first place, our hypothesis is motivated by earlier observations that suggest that dynamic-dynamic tone contrasts are inherently more challenging than static-dynamic contrasts (Burnham *et al.*, 2018; Schaefer & Darcy, 2014). Secondly, an examination of any potential interference from L1 tone types, and how these may determine the relative ease/difficulty of specific tones, also leads to the same conclusion. That is, in the first place we assume that there will be no strong interference from Dutch intonational, and Japanese and Swedish pitch-accent types on the processing of our pseudoword tones, based on earlier findings involving non-native tone processing by non-tonal L1ers (Best, 2019, p. 5; Francis *et al.*, 2008, p. 2691 A. Yu *et al.*, 2021; but cf. Braun *et al.*, 2014). Although Thai speakers are likely to assimilate the pseudoword tones to their L1 tone types, we hypothesize that the relative hierarchy of difficulty of the pseudoword tones *vis-à-vis* one another (i.e., fall and peak > level) still holds. This is because we expect that none of the tones should clearly map in a two-category manner to Thai tone types, which could make the tones all relatively easy to distinguish. For instance, the falling 51 and peaking 141 pseudoword tones resemble the Thai 241 tone (J. Chen *et al.*, 2020, p. 6) and do not clearly map in a two-category manner to separate Thai tone types³.

With this tone system, we aim to investigate more directly whether a higher L1 tonal status in and of itself facilitates non-native tone perception and word learning. We note however, that we cannot exclude the possibility that some tones were in fact differentially easier or more difficult than one another due to interference with L1 intonational/pitch accent/tone types, and will consider this in the discussion.

Predictions

We make the following predictions regarding our research questions. **P1)** L1 tonal status incrementally facilitates non-native tone categorization and word learning of a tonal pseudoword system consisting of a low-level, a falling, and a peaking tone, such that Thai-L1 participants will outperform Swedish-L1 and Japanese-L1 participants, who in turn will outperform Dutch-L1 participants. In addition, musical and experience and working memory predict individual performance in tone categorization and word learning.

P2) Tone categorization predicts tone word learning performance; however, it may be more predictive for speakers of a low tonal status in comparison to speakers of a high tonal status.

Methods

Participants

The study was approved by the ethics board of the University of Cambridge. One hundred and fifteen participants took part. Participants were recruited through university networks, social

media, and *Prolific* (www.prolific.co) 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 professionals⁴. Following the exclusion criteria of Burnham et al. (2015) and Schaefer and Darcy (2014), native speakers of a non-tonal or pitch-accent language with knowledge of a tone language (4 Dutch, 4 Swedish, 6 Japanese) were not included. Five Thai speakers reported L2 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 (Maggu et al., 2018), and given that knowledge of the Mandarin tone system should not contribute to better perception of the pseudoword tones, since these do not all clearly map in a two-category fashion to Mandarin tone types. However, 4 Thai speakers who reported knowledge of Northern Thai were excluded because Northern Thai contains a fall-peak contrast (Katsura, 1969) to which the pseudoword tones could assimilate in a two-category fashion, which could facilitate their perception.

Further, because of an imbalance in the number of musicians and non-musicians across groups in the remaining participant pool, we here present data of 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 practice, and working memory (WM), measured by a backwards digit span task. Equivalence tests (Lakens et al., 2018) revealed that the groups were similar in terms of their musical experience and WM. An overview of the participant groups is provided in Table 2. Detailed participant demographics are in the Supplementary Material.

Stimuli

The audio stimuli consisted of a set of meaningless vowels ([a] [ε] [i]) for the tone categorization task and a set of pseudowords (/lala/ /lεlε/ /lili/; Table 3.) for the word identification task. Each of the stimuli carried either a low-level, a falling, or a peaking tone, resulting in nine vowel stimuli and nine pseudoword

stimuli for each task. The choice for disyllabic word stimuli was motivated by observations by Pelzl et al. (2020, p. 4) that monosyllabic word stimuli may have limited generalizability to real-world tone learning. Tone contrasts only occurred 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 on the tone contrast on one syllable, similar to the tone categorization task.

Visual stimuli in the tone categorization task consisted of drawings of the level, falling, and peak trajectories (Figure 1). In the word identification task, each aurally presented pseudoword was linked to an image representing its meaning. The images were gathered from a database by Rossion and Pourtois (2004) and represent nine high-frequency nouns (Battig & Montague, 1969; Van Overschelde et al., 2004).

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 pitch trajectories 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, whilst staying as closely as possible to the natural production in terms of the shape of the pitch trajectory. 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) (see Supplementary Materials for sound files and visualizations).

Procedures

The study was carried out online on the *Gorilla Experiment Builder* (Anwyl-Irvine et al., 2020). A copy of the online study is available via the Supplementary Materials. A battery of eight tasks (including word training sessions, see Table 4) was carried out in two sessions over two days. This was to limit the amount

Table 2. Participant demographics. Values are means with standard deviations in brackets.

	Dutch (n = 22)		Swedish (n = 15)		Japanese (n = 23)		Thai (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)	
Musical experience (years)	MU (n = 12) NM (n = 10)		MU (n = 7) NM (n = 8)		MU (n = 12) NM (n = 11)		MU (n = 10) NM (n = 10)	
MU = musicians, NM = non-musicians	10.20 (4.63)	1.10 (3.48)	11.10 (3.85)	0.38 (1.06)	8.92 (5.21)	0.81 (1.60)	8.10 (5.07)	0.60 (0.96)

Table 3. Pseudowords.

	Tone 1 (Low-level 11)	Tone 2 (Falling 51)	Tone 3 (Peak 141)
Segment 1 meaning	/la11.la11/ leaf	/la51.la11/ fork	/la141.la11/ television
Segment 2 meaning	/lε11.lε11/ chair	/lε51.lε11/ apple	/lε141.lε11/ hair
Segment 3 meaning	/li11.li11/ book	/li51.li11/ shirt	/li141.li11/ cat

Table 4. 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 identification (generalization)	10
Debriefing	5

of time spent in each session (approximately 25 minutes each), and to allow for overnight memory consolidation to facilitate for the word identification task on day 2 (Earle & Myers, 2015; Qin & Zhang, 2019). 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 filling out a questionnaire on their linguistic and musical background and 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.

Tone categorization task

In the tone categorization task, 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.

A 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 ('correct') or a red cross ('incorrect'), followed by the correct sound-contour 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 with 54 trials (6 presentations per stimulus) in a randomized order and without feedback.

Word training

The tone categorization task was followed by tone word training, which consisted of imitation and a feedback word identification task. These were expected to be efficient ways to stimulate retention of novel words in a relatively short period of time, based on previous studies (Baills *et al.*, 2019; M. Li & Dekeyser, 2017).

In the imitation block, participants were presented with the individual pseudolanguage words (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'). The presentation order was such that no segmental or tonal minimal pair followed one another. 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 pseudolanguage word and were prompted to identify the meaning of that word by clicking the corresponding tile from a 9-way choice answer board. 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 ('correct') or a red cross ('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.

Word identification task (Day 1)

The feedbacked block was followed by a word identification task without feedback. The set-up was the same as the feedbacked block, but there were 6 trials per unique word 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.

Working memory task

On day 2, participants started with a backwards digit span task, as a measure of WM capacity that is relevant to speech perception and word learning (Baddeley, 2003; Kormos & Sáfár, 2008). 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). 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. Working memory score was defined by the highest attained digit span.

Word identification (generalization)

After another round of word training (mimicry and feedbacked identification), participants completed a second word

identification task. This 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. This was to ensure that participants could generalize their acquired phono-lexical knowledge to novel stimuli, cf. Bowles et al. (2016); Finley (2013). After this 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 summary of the debriefing questionnaire is provided in the Supplementary Materials.

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 Bayesian inference to assess the effects of L1 tonal status, tone type, and extralinguistic factors on performance in the tone categorization and word identification task (day 2, generalization), operationalized by accuracy scores. Null responses and responses with unnaturally fast reaction times (< 250 ms) were removed, excluding 0.62% and 0.45% from each task, respectively.

We fitted Bayesian logistic regression models 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 a probability of direction (pd), which describes the probability that a parameter is positive or negative. We report findings for which i) the 95% CrI of the effect estimates in the posterior distribution did not contain zero, and ii) for 95% CrIs that did contain zero but that had a relatively high pd. We take both such findings to be ‘suggestions’ of an effect.

Following common practice (Haendler et al., 2020; Vasisht et al., 2018), models were constructed using weakly informative priors, with prior specification in *brms* set as (0, 3) for ‘Intercept’; (0, 1) for ‘b’; (0, 0.1) for ‘sd’ priors and LKJ(2) for correlation priors (Coretta et al., 2022). Four sampling chains with 3000 iterations each were run, with 1500 warm-up iterations. Model diagnosis was carried out by observing *R*hat values (ensuring these were close to 1), and by inspecting posterior draws using the *pp_draws()* command of the *brms* package. Multicollinearity between continuous variables (musical experience, working memory, and tone categorization) was checked using the *performance* package (Lüdtke et al., 2021) and revealed low levels of correlation (all Variance Inflation Factors < 5).

We built models with fixed effects and interactions as appropriate to our research questions. The model for tone categorization (dependent variable: correct/incorrect; 4250 observations; fitted with a Bernoulli distribution) contained fixed effects for *L1* (Dutch, Swedish, Japanese, Thai; sum contrast-coded), *Tone* (Level, Fall, Peak; sum contrast-coded), *Musical Experience* (Years of 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 (dependent variable: correct/incorrect; 4276 observations) was the same as for

the tone categorization task, but additionally contained a fixed effect of *Tone Categorization*, (Mean accuracy scores in the tone categorization task; centered and scaled) and an *L1:Tone Categorization* interaction to assess the effect of tone perception on tone word learning. We only report the findings from the word generalization task on day 2, as this was at the end of the word training (see Supplementary Materials for the day 1 findings).

To investigate interactions, we conducted planned comparisons between tone per L1, between L1s per tone, and for the estimates of musical experience, WM, and tone categorization per L1 using the *emmeans* package (Lenth, 2020). We report planned comparison results for which zero was not included in the 95% Highest Posterior Distribution (HPD), which we take to be suggestions for meaningful differences or estimates. Full statistical details (posterior distributions and planned comparisons) are in the Supplementary Materials.

Results

Tone categorization

Descriptive statistics

Table 5 shows accuracies for the tone categorization task per L1. An initial inspection reveals no stark difference between L1s in terms of accuracy.

Model results

Figure 2 plots predicted tone categorization accuracy per tone and L1. The model suggested an effect of *L1*, ($b = 0.42$ [0.01, 0.89]); *Tone* ($b = 1.57$ [1.24, 1.93]) and an *L1:Tone* interaction ($b = -0.55$ [-0.92, -0.20]).

Comparisons between tones per L1 suggested that level tones were more likely to be accurately categorized than falling and peaking tones in all groups (i.e., zero not included in any of the 95% HPDs). In addition, peaking tones were more likely to be accurately categorized than falling tones in all groups, except for the Japanese group, for which zero was included in the 95% HPD ($b = 0.59$ [-0.05, 1.24]).

Comparisons between L1s per tone suggested that Dutch speakers were less likely than Japanese ($b = -1.83$ [-2.97, -0.78]) or Swedish speakers ($b = -1.37$ [-2.58, -0.30]) to categorize level tones. In turn, Japanese ($b = 1.64$ [0.56, 2.83]) and Swedish speakers ($b = 1.20$ [0.09, 2.41]) were more likely than Thai speakers to categorize level tones. These comparisons should be viewed with caution because performance for level tone categorization appeared to be at ceiling. The comparisons further suggested that Thai speakers were more likely than Swedish speakers to accurately categorize peaking tones ($b = 0.78$ [0.06, 1.56]).

Figure 3 plots predicted tone categorization accuracy against musical experience and WM. The posterior distribution suggested an effect of *Musical Experience* ($b = 0.60$ [0.37, 0.83]) and an *L1: Musical Experience* interaction ($b = -0.52$ [-0.92, -0.15]). The estimates per L1 suggest that musical experience was particularly facilitative for Dutch ($b = 0.94$ [0.54, 1.36]) and Thai ($b = 1.16$

Table 5. Descriptive statistics for tone categorization task. Values are means with standard deviations in brackets.

	Dutch	Swedish	Japanese	Thai
Accuracy (%)	81.3 (13.3)	79.5 (12.0)	88.0 (10.1)	79.2 (17.3)

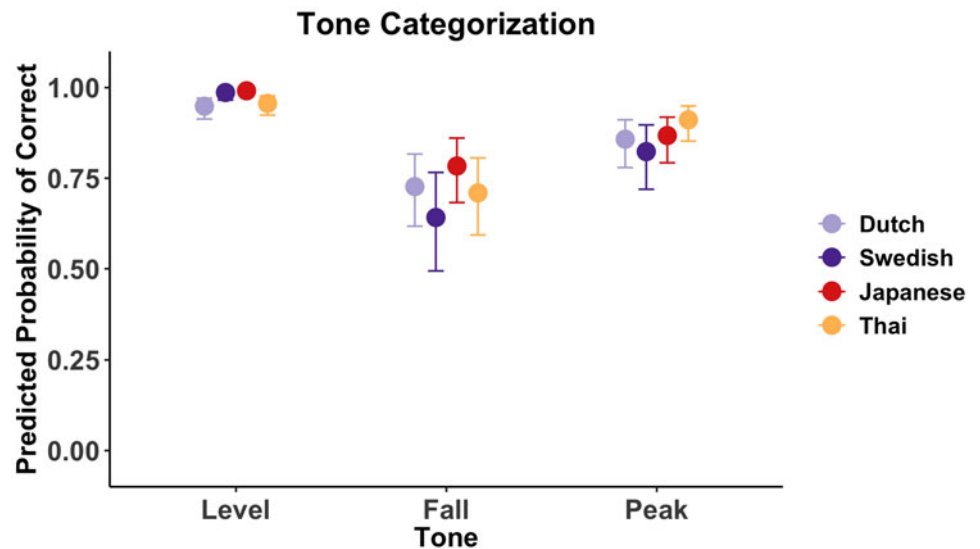


Figure 2. Predicted probability of tone categorization accuracy per tone and L1. Bars represent 95% Crls.

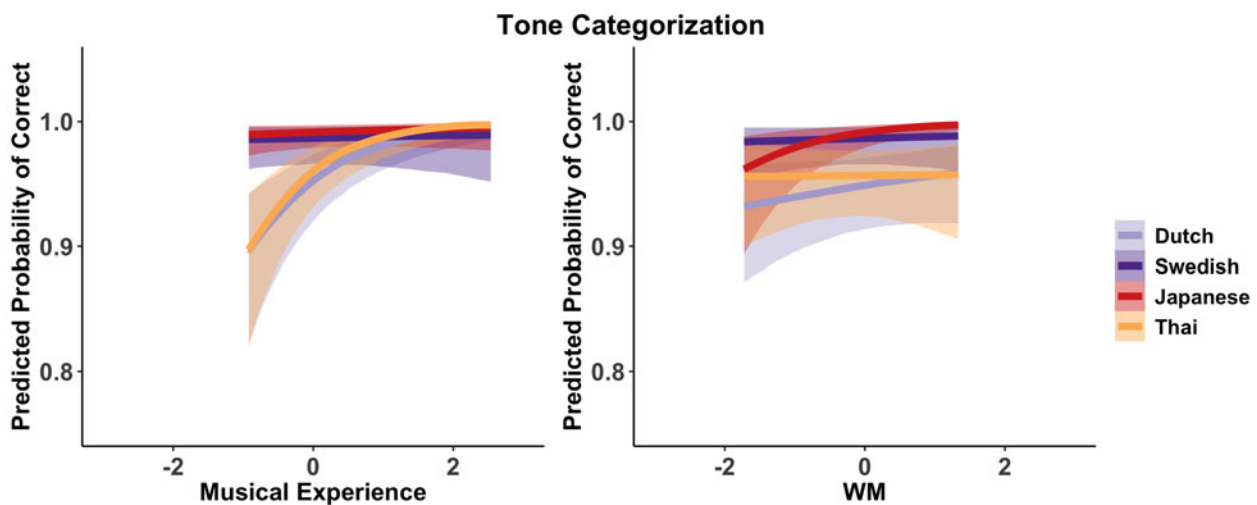


Figure 3. Predicted tone categorization accuracy against musical experience and WM (centered and scaled). Shading ribbons represent 95% Crls.

[0.66, 1.69]) participants, but not for Swedish ($b = 0.08$ [−0.35, 0.52]) and Japanese participants ($b = 0.21$ [−0.22, 0.68]).

The model also suggested an effect of WM ($b = 0.60$ [0.28, 0.92]) and an $L1:WM$ interaction ($b = 0.60$ [0.28, 0.92]). Estimates per L1 suggested that WM facilitated tone categorization for Japanese participants ($b = 0.90$ [0.53, 1.25]). This was less clear for Dutch ($b = 0.18$ [−0.14, 0.47]), Swedish ($b = 0.10$ [−0.34, 0.58]), or Thai participants ($b = 0.01$ [−0.41, 0.42]).

Word identification (generalization)

Descriptive statistics

Table 6 shows the accuracies in the word identification task (day 2, generalization). There appears to be no stark difference in accuracy between L1s. Table 6 also shows the proportion of “tone-only errors” (Wong & Perrachione, 2007), i.e., an error in which a participant misidentifies the meaning of a word purely because of its tonal properties, e.g., misidentifying /la11.la11/ ‘leaf’ as /la51.la11/ ‘fork’. The high proportion of tone-only errors

in each group suggests that most participants had retained the segmental, but not the tonal properties of the pseudolanguage words⁵.

Model results

Figure 4 plots predicted word identification accuracy per tone by L1. The posterior distribution suggested an effect of $L1$ ($b = -0.40$ [−0.73, −0.02]), $Tone$ ($b = 0.82$ [0.63, 1.02]) and an $L1:Tone$ interaction ($b = -0.73$ [−1.05, −0.44]).

Comparisons between tones within groups suggested that words with level tones were more likely to be correctly identified than words with falling and peaking tones (zero not included in any of the 95% HPDs). An exception to this was the Swedish group, for which there was a less clear suggestion that level tone words were more likely to be identified than falling tone words ($b = 0.46$ [−0.07, 0.99]) or peaking tone words ($b = -0.21$ [−0.87, 0.47]). Comparisons further suggested that in all groups, words with peaking tones were more likely to be correctly identified than words with falling tones.

Table 6. Descriptive statistics for the word identification task (day 2 generalization). Values are means with standard deviations in brackets.

	Dutch	Swedish	Japanese	Thai
Accuracy (%)	61.4 (27.0)	54.9 (24.3)	60.5 (22.6)	64.7 (22.2)
% of tone-only errors	81.1 (21.6)	75.4 (22.5)	71.0 (0.22)	80.6 (0.25)

Comparisons between L1s per tone suggested that Swedish speakers were less likely than Dutch ($b = -1.17 [-1.99, -0.28]$), Japanese ($b = -0.97 [-1.86, -0.12]$), and Thai speakers ($b = -1.85 [-2.70, -0.93]$) to identify words with level tones. Japanese speakers were in turn less likely than Thai speakers to identify words with level tones ($b = -0.86 [-1.69, -0.02]$). The comparisons further suggested that Japanese speakers were less likely than Dutch ($b = -1.04 [-1.64, -0.39]$), Swedish ($b = -0.77 [-1.46, -0.09]$), and Thai speakers ($b = -0.84 [-1.48, -0.25]$) to identify words with falling tones. Finally, Japanese speakers were less likely than Dutch ($b = -0.75 [-1.38, -0.19]$) and Thai speakers ($b = -0.67 [-1.27, -0.09]$) to identify words with peaking tones.

Figure 5 plots predicted word identification accuracy against musical experience, WM, and tone categorization. The model output suggested that there was an $L1:Musical\ Experience$ interaction ($b = 0.61 [0.23, 1.28]$), as well as an $L1:WM$ interaction ($b = -0.35 [-0.72, 0.03]$), with zero included in the 95% CrI and a probability of direction of 96.10%. Finally, the model suggested an effect of *Tone Categorization* ($b = 0.61 [0.23, 0.99]$), and an $L1:Tone\ Categorization$ interaction ($b = 0.63 [0.15, 1.10]$).

An observation of Figure 5 and the estimates and HPDs for the effect of musical experience per L1 suggest that musical experience facilitated word identification for the Swedish group ($b = 0.62 [0.08, 1.44]$), but not for the Dutch ($b = -0.23 [-0.67, 0.22]$), Japanese ($b = -0.08 [-0.48, 0.28]$), and Thai groups ($b = -0.22 [-0.72, 0.29]$).

Based on the estimates of the effect of WM per L1, there was no suggestion that WM facilitated word identification for Dutch ($b = 0.09 [-0.23, 0.41]$) or Swedish ($b = -0.17 [-0.64, 0.26]$) speakers, although it may have slightly facilitated performance

for Japanese ($b = 0.38 [-0.17, 0.87]$), and Thai speakers ($b = 0.39 [-0.02, 0.78]$).

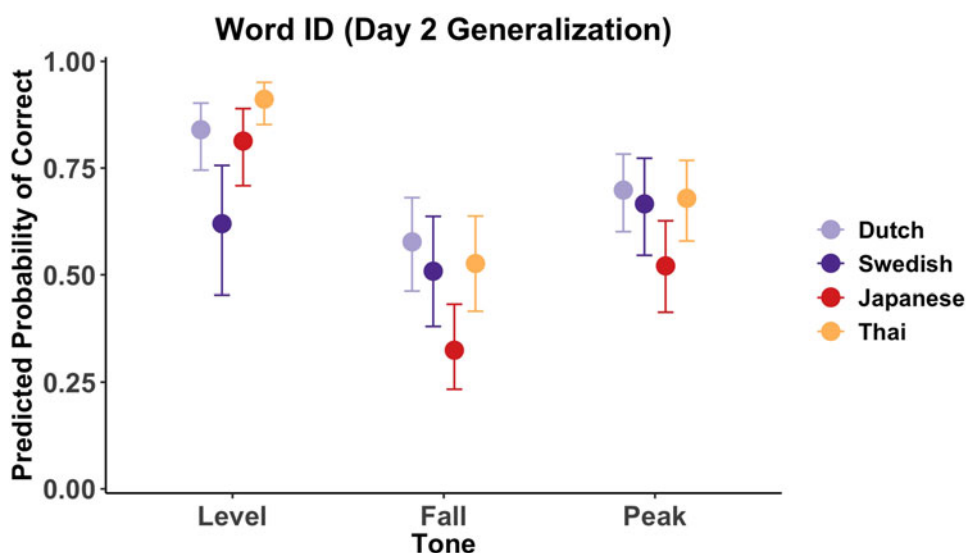
Figure 5 and the estimates of the effect of tone categorization per L1 suggest that tone categorization led to a higher likelihood of correct word identification in the Dutch ($b = 1.49 [0.95, 2.01]$), Swedish ($b = 0.76 [0.21, 1.32]$), Japanese ($b = 0.59 [-0.16, 1.43]$), and Thai group ($b = 0.57 [0.19, 0.92]$). The difference in estimate sizes suggests that tone categorization may have been particularly facilitative for Dutch speakers, but less so for Swedish and even slightly less so for Japanese and Thai speakers.

Discussion

Tone categorization

In this study, we examined the effect of different degrees of L1 tonal status on the perception and word learning of non-native tones, as well as the link between perception and word learning, whilst we simultaneously investigated the effects of tone type, musical experience and working memory. We built on from previous studies that similarly investigated the effect of different L1 tonal statuses on non-native tone processing, but that only investigated tone perception and that did not address these other factors together (Burnham et al., 2015; Schaefer & Darcy, 2014). Against our prediction informed by the Functional Pitch Hypothesis, we found no support for an incrementally facilitative effect of L1 tonal status on non-native tone categorization across the board. Apart from a number of between-L1 differences for level tone categorization (which need to be interpreted with caution because performance for level tones appeared to be at ceiling), and the observation that Thai speakers were more likely than Swedish speakers to categorize peaking tones, participants' tone categorization performance was relatively uniform and did not appear to be strongly modulated by their L1. They showed global trends in tone categorization accuracy, and categorized static low-level tones easily, whereas they categorized dynamic falling and peaking tones with greater difficulty.

The lack of a clear facilitative effect of L1 tonal status on tone perception (operationalized by tone categorization, here) replicates findings from other studies that similarly suggest that a higher degree of pitch functionality in the L1 does not necessarily

**Figure 4.** Predicted word identification accuracy per tone by L1. Bars represent 95% CrIs.

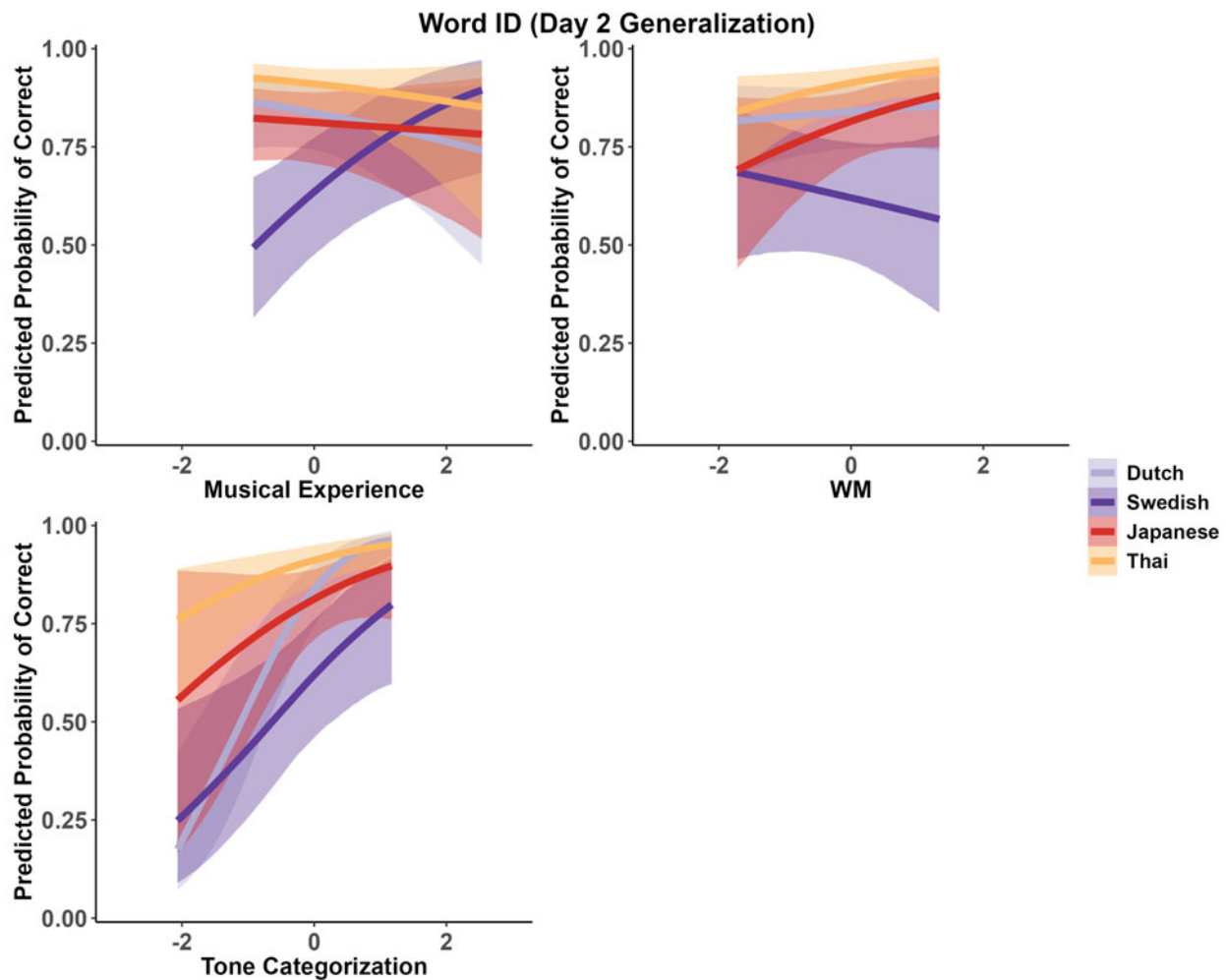


Figure 5. Predicted word identification accuracy against musical experience, WM, and tone categorization (centered and scaled). Shading ribbons represent 95% CrIs.

facilitate tone perception (A. Chen *et al.*, 2016; Cooper & Wang, 2012; Cutler & Chen, 1997; Francis *et al.*, 2008; Gandour & Harshman, 1978). Yet, it contrasts with Schaefer and Darcy (2014) and Burnham *et al.* (2015) that did find an incrementally facilitative effect of L1 tonal status on tone perception (operationalized by tone discrimination). The discrepancy between our and their findings is difficult to interpret given the methodological differences between the studies, but they could potentially be related to the type of target tone to be perceived, as our findings also suggest. Indeed, Burnham *et al.* (2015, pp. 1478-1489), who report differences between L1s per tone type, only found evidence for a facilitative effect of L1 tonal status in four out of the ten tone type contrasts tested.

In the present study, rather than finding robust between-group differences, we found robust between-tone differences applying cross-linguistically instead. It appears that L1 tonal status does not incrementally facilitate non-native tone categorization when the target tonal system is designed such that tones present the same relative difficulties (*vis-à-vis* one another) for all L1s involved. This draws a very interesting parallel with other fairly global L1-facilitating effects for L2 learning proposed in other domains, such as that of the L1 vowel inventory on L2 vowel perception. Based on the finding that listeners with large L1 inventories often outperform those with smaller L1 inventories in L2

vowel perception, it has been postulated that having a large L1 vowel inventory facilitates the learning of non-native vowels (Iverson & Evans, 2007, 2009; Kivistö-de Souza *et al.*, 2017). However, other studies have strongly challenged this idea by showing that said advantage is done away with, and that even the opposite pattern (smaller inventory > larger inventory) can be elicited, when the acoustic properties of the particular L2 target contrasts and the way in which they relate to the categories in the L1 inventory are taken into account (Alispahic *et al.*, 2017; Elvin *et al.*, 2014).

Above and beyond L1 tonal status, the present study showed that it was primarily musical experience that facilitated tone perception. This replicates previous findings regarding the effect of long-term musical training on pre-lexical tone perception, at least in certain perceptual tasks (Wong *et al.*, 2020; H. Wu *et al.*, 2015). The results, however, also suggested that the effect of musical experience was clearly facilitative for Dutch and Thai, but relatively weak or absent for Swedish and Japanese speakers. This goes against our expectation, informed by the 'lack of relevant experience' addition to the OPERA model (Choi, 2021), that the effect of any pitch-related skills would be particularly facilitative for non-tonal (Dutch) speakers, and less so for pitch-accent and least for tone language speakers. One possible explanation for this finding could be that musical experience

as expressed by the years of practice is not the most accurate proxy of music-derived pitch acuity in comparison to measures that directly gauge musical ability or musicality (Peretz & Hyde, 2003; Wallentin et al., 2010). In any case, the fact that we did observe an overall facilitative effect of musical experience (though not equally facilitative in all groups) highlights the relevance of including it as an extralinguistic factor to explain individual differences in tone categorization performance.

The estimates of the effect of working memory per L1 suggested that generally, participants with a higher backwards digit span did not perform better in 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, the estimate for the effect of WM for Japanese speakers suggested that digit span did lead to higher tone categorization accuracy. 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.

Word identification (generalization)

Results from the word generalization task showed that participants were able to identify the meaning of tonal pseudoword words when those were spoken by a new speaker. However, they appeared to struggle with identifying the meaning of words that formed tonal minimal pairs, as most errors that participants made were tone-only errors. This suggests that participants had acquired the words' segmental, but not the tonal properties, and highlights that encoding tone information into the lexical representations of newly-learned (pseudo)words is indeed rather challenging, as suggested by earlier tone word learning studies (Laméris & Post, 2022; Y. Li et al., 2022; Pelzl et al., 2021; Wong & Perrachione, 2007).

In addition, a comparison between the tone categorization and word identification results indicates that the difficulties experienced in the latter go above and beyond those involved in sheer pre-lexical tone perception. Although inherent task differences (e.g., a three-way versus a nine-way choice) may partly explain this discrepancy, this finding on tone processing extends to previous research on non-native segmental processing that shows that there is a substantial gap between L2 learners' perceptual acuity at the pre-lexical level and their actual ability to draw upon the critical L2 categories in situations requiring the establishment of form-meaning associations, i.e., word learning (cf. Hayes-Harb & Masuda, 2008; Llompart & Reinisch, 2020, 2021).

The central finding of the word identification task was, however, that accuracy levels across L1s were very similar. This suggests that L1 tonal status did not facilitate word identification performance. Parallel to the findings of the tone categorization task, it again appeared that dynamic-dynamic tone contrasts (falling and peaking tones) were the most difficult, whereas the low-level tones were easiest to process, and this was observed in all L1 groups. Nevertheless, unlike in the tone categorization task, there were more L1-specific patterns in accuracy of word identification (namely, that Swedish speakers were less likely than all other groups to identify level tone words, and Japanese speakers less likely than all other groups to identify falling tone words). These could be indicative of interactions between the L1 pitch accent systems and the target tone types. Although we can only speculate, it could be that Swedish speakers were inattentive

(or "deaf", cf. Dupoux et al. (1997, 2008)) to level tone words because these lack any pitch-based syllable prominence, unlike Swedish pitch-accented words which do have a pitch movement on the first syllable that may be phonetically and cognitively marked (Roll et al., 2011; but cf. Köhnlein, 2020, p. 458). As for Japanese speakers, the falling and peaking pitch patterns on the first syllables in the present pseudoword may have assimilated to canonical pitch patterns of Japanese first-mora accented words (Ishihara, 2006; Laméris & Graham, 2020) which would constitute a two-to-one perceptual assimilation pattern and increase the difficulty of distinguishing the falling and peaking patterns. A similarity judgment task (cf. J. Chen et al., 2020) would have to ascertain whether participants did indeed assimilate the pseudoword tone patterns, but a post-hoc questionnaire (see Supplementary Materials) suggests that participants did not find the tones to strongly resemble native prosodic categories.

Despite these two very specific observations, performance in the final word identification task was generally uniform across participant groups. Results revealed no clear evidence that L1 tonal status incrementally facilitates lexical encoding of tones, at least when operationalized by a word identification task of tonal pseudowords with tones following the same difficulty hierarchy for all L1s involved.

Similar to the tone categorization task, extralinguistic factors also predicted individual performance in word identification, to some extent. Musical experience predicted word identification in the Swedish group. In other groups, there was no indication that years of musical practice was associated with more accurate word identification. 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 this would require further investigation.

Although an effect of WM was predicted based on previous studies that show a link between WM capacity and the facility to learn non-native words (Baddeley, 2003; Kormos & Sáfár, 2008), it is possible that WM was relatively irrelevant to performance in the present study because individuals were learning *tone* words. 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 words (Bowles et al., 2016; Perrachione et al., 2011).

Link between tone perception and word learning

The strongest predictor of performance in word identification was not musical experience or working memory, but individual tone categorization performance. Across the board, it emerged that participants who scored higher in tone categorization were more likely to correctly identify tonal pseudowords. This adds to a growing body of research that assesses the relationship between pre-lexical perception and lexical representation and processing in second language learning (Darcy & Holliday, 2019; Melnik & Peperkamp, 2021; Simonchik & Darcy, 2017) and replicates findings by previous studies that show a strong link between pre-lexical and lexical tone processing, specifically when the former is assessed by 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 on the effect of musical experience, working memory, and tone categorization with those from Bowles et al. (2016),

who investigated Mandarin tone word learning in English-L1 participants. They found that measures of pitch processing, in particular tone categorization 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. They suggest that “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).

Yet, in the present study, which involved learners from a spectrum of L1 tonal statuses, the estimates of the effect of tone categorization per L1 suggested that L1 tonal status in fact modulated the strength with which tone categorization facilitated word identification. Tone categorization appeared to be most facilitative for Dutch speakers, less so for Swedish and Japanese speakers, and least for Thai speakers. This chimes in with studies that suggest that pre-lexical tone processing abilities may be more beneficial for non-tone language speakers than for tone language speakers in tone word learning (Cooper & Wang, 2012; Ling & Grüter, 2020). It may be that there is a trade-off between individual linguistic and extralinguistic ‘skills’, in which prior L1 experience with lexical tones reduces the need to rely on extralinguistic pitch processing skills to facilitate the task of tone word learning. Indeed, such a trade-off between individual resources is suggested by studies that show that the effects of L1 tone experience and musical experience are not additive (Choi, 2021; Maggu et al., 2018). If a higher degree of L1 tonal status indeed reduces the need for individual pre-lexical pitch processing skills – and if it does so in an incremental way as tentatively shown in the present study – this might be evidence of qualitative, but not quantitative differences across speakers of different of L1 tonal statuses in non-native tone word learning.

Conclusion

The findings from this study reveal that, when factoring in the simultaneous effects of tone type, musical experience, and working memory, an individual’s L1 tonal status is not strongly indicative of tone perception or tone word learning facility. Instead, the ease with which learners perceive non-native tonal contrasts appears to be largely guided by the target tone type, as learners exhibit global patterns in tone perception accuracy based on the inherent difficulty of the tone types at hand. The 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. Individual musical experience and working memory capacity may facilitate tone perception and word learning, but they appear to play a secondary role overall. In addition, although tone perception predicts tone word learning, it may do so most strongly in learners whose L1 tonal status is low, and less strongly in learners whose L1 tonal status is intermediate or high. In sum, the present study’s findings suggest that individual variability in tone word learning facility across a spectrum of L1 tonal statuses can be best captured by an L1-modulated feature-specific account of pitch processing.

Supplementary Material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S1366728923000871>

The separate tasks used in the online experiment are available at <https://app.gorilla.sc/openmaterials/446865>

Other Supplementary materials are available on <https://osf.io/kaczu/> and consist of the following:

1. Data and codes
2. Detailed participant demographics
3. Stimuli information (including sound files)
4. Model output
5. Summary of debriefing

Acknowledgements. We thank the three anonymous reviewers for their feedback on earlier versions of this manuscript. This study was conducted as part of T.J. Laméris’ doctoral research at the University of Cambridge, funded by the Economic and Social Research Council (Grant 2117864) and a St John’s College Learning and Research Fund. We would like to thank the volunteer participants for their participation, Alif Silpachai for his help with the Thai version of the experiment, and Yair Haendler and Stefano Coretta for their advice on the statistical analyses. All errors remain our own.

Notes

1. Note that unlike Schaefer and Darcy (2014) and Burnham et al. (2015), tone perception in these studies was operationalized not by a tone discrimination, but a tone categorization task in which a listener needs to identify a specific tone pattern on a sound devoid of lexical meaning. In line with these studies, we also chose a tone categorization task to operationalize pre-lexical tone processing.
2. This tone has been described both as rising-falling 241 (J. Chen et al., 2020, p. 6) and high-falling 51 (X. Wu et al., 2014). To avoid confusion, we will henceforth refer to this tone as ‘falling’ given that this is appears to be the most typical phonological description (I. L. Chan & Chang, 2019; J. Chen et al., 2020, p. 4).
3. An alternative scenario would be that the falling and peaking tone assimilate in a single-category or ‘category goodness’ manner to the Thai falling tone, which could make this contrast difficult for Thai speakers. This appears not to have been the case, since Thai speakers never underperformed in contour tone categorization or word identification in comparison to other speakers. It remains possible however, that a potential negative interference due to category goodness or single-category assimilation dampened any L1 perceptual advantage that Thai speakers could have had.
4. As in the previous studies that served as points of reference for the present study (Burnham et al., 2015; Schaefer & Darcy, 2014), our participants were resident in an English-speaking country, which may have dampened some of the L1 tonal status effects.
5. It might be argued that including ‘non-tone-only error’ responses, e.g., a misidentification of the target word /la11.la11/ as /le11.le11/ in the analysis, might not be relevant to examine tone word learning. Therefore, a separate model was fitted on the data that excluded such non-tone-only errors (see Supplementary Materials). However, this model revealed similar effects and interactions as the model fitted on the data that included all responses. Therefore we chose to run the analysis on the dataset including all response types, in line with previous tone word learning studies (Wong & Perrachione, 2007).

References

- Alispahic, S., Mulak, K. E., & Escudero, P. (2017). Acoustic Properties Predict Perception of Unfamiliar Dutch Vowels by Adult Australian English and Peruvian Spanish Listeners. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00052>
- 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](https://doi.org/10.1016/S0021-9924(03)00019-4)
- 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>
- 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>
- 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>
- 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>
- 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., & Johnson, E. K. (2011). Question or tone? 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>
- 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>
- 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., Luksaneyanawin, 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., 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>
- 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>
- 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, 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., Yang Y. & Yasin I. (2020). How musical experience affects tone perception efficiency by musicians of tonal and non-tonal speakers?. *PLOS ONE* 15(5), e0232514. <http://dx.doi.org/10.1371/journal.pone.0232514>
- Choi, W. (2021). Musicianship Influences Language Effect on Musical Pitch Perception. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.712753>
- 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>
- Coretta, S., Casillas, J. v., & Roettger, T. (2022). *Learn Bayesian Analysis for the Speech Sciences*. <https://learnb4ss.github.io/learnB4SS/index.html>
- 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>
- Darcy, I., & Holliday, J. (2019). Teaching an old word new tricks: Phonological updates in the L2 mental lexicon. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 10–26). Iowa State University.
- Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, 8(3), 372–420. <https://doi.org/10.1075/ml.8.3.06dar>
- 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>
- 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>
- Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A Destressing “Deafness” in French? *Journal of Memory and Language*, 36(3), 406–421. <https://doi.org/10.1006/jmla.1996.2500>
- 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>
- Earle, F. S., & Myers, E. B. (2015). “Sleep and native language interference affect non-native speech sound learning”: Correction to Earle and Myers (2015). *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1708–1708. <https://doi.org/10.1037/xhp0000160>
- Elvin, J., Escudero, P., & Vasiliev, P. (2014). Spanish is better than English for discriminating Portuguese vowels: acoustic similarity versus vowel inventory size. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01188>
- 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>
- Finley, S. (2013). Generalization to unfamiliar talkers in artificial language learning. *Psychonomic Bulletin & Review*, 20(4), 780–789. <https://doi.org/10.3758/s13423-013-0402-7>
- 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>
- 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>
- 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>
- Götz, A., Liu, L., Nash, B., & Burnham, D. (2023). Does Musicality Assist Foreign Language Learning? Perception and Production of Thai Vowels, Consonants and Lexical Tones by Musicians and Non-Musicians. *Brain Sciences*, 13(5), 810. <https://doi.org/10.3390/brainsci13050810>
- 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>
- 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*.

- 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>
- Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 24(1), 5–33. <https://doi.org/10.1177/0267658307082980>
- 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>
- Ishihara, T. (2006). *Tonal alignment in Tokyo Japanese*. University of Edinburgh.
- Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, 122(5), 2842. <https://doi.org/10.1121/1.2783198>
- Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126(2), 866–877. <https://doi.org/10.1121/1.3148196>
- Katsura, M. (1969). Notes on Some Phonological Aspects of Northern Thai. *Tonnan 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.
- Kivistö-de Souza, H., Carlet, A., Julkowska, I., & Rato, A. (2017). Vowel inventory size matters: assessing cue-weighting in L2 vowel perception. *Ilha Do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, 70(3), 33–46. <https://doi.org/10.5007/2175-8026.2017v70n3p33>
- 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>
- 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 extra-linguistic 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>
- Laméris, T. J., Li, K. K., & Post, B. (2023). Phonetic and Phono-Lexical Accuracy of Non-Native Tone Production by English-L1 and Mandarin-L1 Speakers. *Language and Speech*. <https://doi.org/10.1177/00238309221143719>
- Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://cran.r-project.org/package=emmeans>
- Li, M., & DeKeyser, R. (2017). Perception Practice, Production Practice, and Musical Ability in L2 Mandarin Tone-Word Learning. *Studies in Second Language Acquisition*, 39(4), 593–620. <https://doi.org/10.1017/S0272263116000358>
- Li, Y., Wang, M., Li, C., & Li, M. (2022). Phonological Training and Word Learning in a Novel Language. *Frontiers in Communication*, 7. <https://doi.org/10.3389/fcomm.2022.629349>
- 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>
- Llompарт, M. (2021). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition*, 24(3), 481–496. <https://doi.org/10.1017/S1366728920000656>
- Llompарт, M., & Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(8), 1590–1610. <https://doi.org/10.1037/xlm0000832>
- Llompарт, M., & Reinisch, E. (2021). Lexical representations can rapidly be updated in the early stages of second-language word learning. *Journal of Phonetics*, 88, 101080. <https://doi.org/10.1016/j.wocn.2021.101080>
- Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- 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>
- 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>
- Melnik, G. A., & Peperkamp, S. (2021). High-Variability Phonetic Training enhances second language lexical processing: evidence from online training of French learners of English. *Bilingualism: Language and Cognition*, 24(3), 497–506. <https://doi.org/10.1017/S1366728920000644>
- Ota, M. (2006). Children's production of word accents in Swedish revisited. *Phonetica*, 63(4), 230–246. <https://doi.org/10.1159/000097307>
- Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology* 2. <http://dx.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>
- Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2021). Advanced Second Language Learners of Mandarin Show Persistent Deficits for Lexical Tone Encoding in Picture-to-Word Form Matching. *Frontiers in Communication*, 6. <https://doi.org/10.3389/fcomm.2021.689423>
- 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., & 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](https://doi.org/10.1016/S1364-6613(03)00150-5)
- 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>
- 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., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLOS ONE*, 14(12), e0221498. <https://doi.org/10.1371/journal.pone.0221498>
- 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>

- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>
- 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>
- 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>
- Shport, I. A. (2016). TRAINING ENGLISH LISTENERS TO IDENTIFY PITCH-ACCENT PATTERNS IN TOKYO JAPANESE. *Studies in Second Language Acquisition*, 38(4), 739–769. <https://doi.org/10.1017/S027226311500039X>
- Simonchyk, A., & Darcy, I. (2017). Lexical encoding and perception of palatalized consonants in L2 Russian. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference* (pp. 121–131). Iowa State University.
- 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>
- 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 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>
- 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>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>
- 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>
- 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>
- 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, 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>
- 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>
- 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, A., Lee, C., Lan, C., & Mok, P. (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>
- 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>
- Zou, T., Caspers, J., & Chen, Y. (2022). Perception of Different Tone Contrasts at Sub-Lexical and Lexical Levels by Dutch Learners of Mandarin Chinese. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.891756>