DALARNA UNIVERSITY

Degree Thesis

Bachelor’s level (First cycle)

Mandarin lexical tone perception by native speakers of Greek

Author: Dimitrios Apostolou
School: Dalarna University
Supervisor: Man Gao
Examiner: Tao Yang
Subject/main field of study: Chinese Linguistics
Course code: GKI2QA
Credits: 15 hp
Date of examination:

Abstract:
This study examines the perception of Mandarin lexical tones by native speakers of Greek who are naïve to Mandarin. Being able to discriminate between Mandarin lexical tones is essential for effective and accurate communication in the language. While there is an increasing number of research studying the perception of Mandarin tone by speakers from various backgrounds, the Greek language has not received much attention. By employing an AXB discrimination task, this study tested the perception of Mandarin tone pairs T1-T2, T1-T4, and T2-T3 by native speakers of Greek. Due to their acoustic similarity, these three pairs are often found to be the most confusing tone contrasts for non-native listeners of Mandarin. Greek speakers (NG) had the highest accuracy for T1-T2 (0.88), and similar accuracy rates for T1-T4 (0.83) and T2-T3 (0.82). Subsequently, the Greek speakers were divided into two groups based on their experience with learning or playing a musical instrument, as this is a factor that has been shown to affect tone perception. Compared to non-musicians (NG1), the group of musicians (NG2) had higher accuracy rates for T1-T2 and T1-T4, and a similar accuracy rate for T2-T3. For NG1, from easiest to hardest, the accuracy rates were: T1-T2(0.93)>T1-T4(0.92)>T2-T3(0.82). For NG2, the accuracy rates were: T2-T3(0.82)⩾T1-T2(0.82)>T1-T4(0.74). These findings could improve our understanding in regard to the factors that can affect the perception of Mandarin lexical tone by native Greek speakers, which might contribute to language pedagogy.

Keywords: Chinese, Mandarin, Greek, lexical tone, perception study, discrimination task, AXB

摘要：
本论文旨在研究以零基础普通话希腊语母语人士对普通话声调范畴的感知能力。为了进行有效和准确的研究，本文实施了普通话发音辨别实验。虽然声调范畴感知能力的研究越来越多，但以希腊语母语者为研究对象的研究并不多见。通过 AXB 发音辨别测试，本论文调查了希腊语为母语的人士对普通话 T1-T2、T1-T4、T2-T3 三组声调组合的辨识感知。由于声音的相似性，这三对声调组合对非普通话母语的听众经常被发现是最容易被混淆的。结果发现，讲希腊语的人士（NG）对这三组声调组合辨识的准确度为：T1-T2（0.88）是最高的，T1-T4（0.83）和 T2-T3（0.82）是相似的。有研究认为音乐训练可以影响对声调的感知，所以接着实验参与者根据学习或演奏乐器的经验又被分成两组。与非音乐背景人群（NG1）相比，有音乐背景人群（NG2）T1-T2 和 T1-T4 的准确率更高，但对 T2-T3 组合的结果是一样的。若从最简单到最难排序，有音乐背景人群（NG1）的准确度是：T1-T2（0.93）>T1-T4（0.92）>T2-T3（0.82）。而无音乐背景人群（NG2），准确度是：T2-T3（0.82）⩾T1-T2（0.82）>T1-T4（0.74）。这些发现可以帮助理解以希腊语为母语的人对普通话声调感知困难的根源，所以对语言教学法的相关讨论会有所贡献。

关键词：汉语普通话声调、希腊语、感知能力、声调辨识、AXB 辨识测试

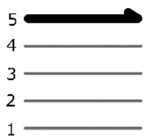
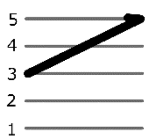
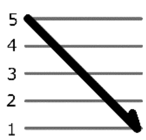
Contents

1 Introduction

Chinese is a language that is increasingly gaining in popularity and attention. However, it often proves challenging to the many individuals of various native language backgrounds that attempt to learn it as a second language. One evident reason is the complexity of the written language, which comprises thousands of logographic characters. Another reason is that Chinese is a tone language that makes use of pitch, or fundamental frequency (F0) when considered from an acoustic perspective, to change the core meaning of words (Yip, 2002). This feature is known as lexical tone and is mostly absent from non-tone languages such as those typically found in Western countries.

Mandarin, a designation commonly used to refer to the standardized form of Chinese spoken in the country and beyond (else known in English as Standard Chinese), has four lexical tones and one neutral tone in its inventory. As can be seen in Table 1, the four lexical tones are conventionally known as: Tone 1 (T1), Tone 2 (T2), Tone 3 (T3), and Tone 4 (T4). T1 is a level tone, while the rest are contour tones, as their pitch is dynamic. Syllables made of identical segments, for instance segmental syllable /ma/, vary in meaning depending on which tone they carry. Since tone is used to contrast meaning in Mandarin, being able to perceive the pitch patterns of all tones contrastively as distinct categories is essential to communicate efficiently and accurately.

Table 1
Mandarin lexical tone

Tone	Tone 1	Tone 2	Tone 3	Tone 4
Word (Pinyin)	妈 (mā)	麻 (má)	马 (mǎ)	骂 (mà)
Meaning	mother	hemp	horse	to scold
Description	high-level (55)	high-rising (35)	low-dipping (214)	high-falling (51)
Graphical representation				

Note. The numbers in parentheses refer to a scale system by Chao (1948), where ‘1’ indicates the lowest level in the pitch range of the speaker, and ‘5’ the highest level. The first number is the starting pitch, and the last is the ending pitch. A graphical representation of the system is included.

Greek, or more specifically Standard Modern Greek, is a stress accent language that does not utilize lexical tone. However, the pitch patterns over some syllables can be affected by a prosodic feature known as lexical stress, which makes syllables more prominent in order to mark linguistic information, including, in some cases, lexical meaning (Yip, 2002, p. 256). One basic difference with lexical tone though, is that there are many pitch patterns that a stressed syllable can carry without any change on the meaning of a word. The other difference is that the relationship of stress and pitch is indirect, as it is determined by intonation, which is another feature of prosody, that modulates pitch patterns on entire utterances (Arvaniti, 2020). In Greek, the position of stress is generally unpredictable, while, unlike other languages that also feature lexical stress such as English, its functional load is very high, as the meaning of many sets of words are distinguished by stress location (Arvaniti, 2007, pp. 130-131).

Other factors that can determine pitch variation in a language are the interaction of words when chunked together, the syntax, as well as pragmatics (Arvaniti, 2020). In a non-tone language, there are specific pitch patterns realized at syntactic boundaries associated with some specific meaning. In Greek particularly, there are eight different configurations of tonal melodies, and most of them can signify multiple meanings (Arvaniti & Baltazani, 2005, p. 95). Such tonal melodies, and their association with some specific meaning, can vary with every language.

Perception studies have shown that listeners of various language backgrounds with no experience in Mandarin, i.e., Mandarin-naïve listeners, do not perceive Mandarin lexical tones as accurately as native speakers of Mandarin do, while some tone pairs tend to be more confusable. For instance, T1 is often confused with T2, and vice versa; T1 is often confused with T4, and vice versa; and T2 is often confused with T3, and vice versa (e.g., So & Best, 2014; Tsukada, 2019). The reason why these three tone pairs, i.e., tone contrasts, are more challenging, has been attributed to their acoustical similarity; as tones T2 and T3 both have a dip followed by a rising contour, tones T1 and T4 both start at similar height, and tones T1 and T2 both end at a similar height (So & Best, 2010). But not all Mandarin-naïve listeners confuse these three Mandarin tone contrasts in the exact same manner. Their performance can vary, while their confusion patterns can be different. For instance, in Tsukada (2019), native speakers of Vietnamese, a tone language, found T2-T3 easier to discriminate compared to the other two pairs; native speakers of Thai, also a tone language, found T2-T3 as the hardest; while for native speakers of English, a non-tone language, the hardest pair was T1-T2.

The reason for this, assuming all things equal, is because individuals unfamiliar with another language are prone to perceive its features in accordance with the organization of their own linguistic system, and that includes the perception of non-native lexical tone (Arvaniti, 2020; Best, 2019). Thus, it is possible for two sounds that are distinct categories in an unfamiliar language to be confused, especially when they are very similar in acoustic terms but at the same time do not mark any meaningful distinction in one's own native language. This is in line with theoretical frameworks of speech perception, such as the Perceptual Assimilation Model (Best, 1995), or the Speech Learning System (Flege, 1995), as they both make similar assumptions.

In this study, the perception of the three Mandarin lexical tone contrasts (T1-T2, T2-T3, T1-T4) that have been consistently found in the existing literature to be more difficult to perceive for speakers of various language backgrounds, is going to be tested on native Greek speakers that have no experience with Mandarin (Mandarin-naïve). Greek is a language that has not received any attention from perception studies of Mandarin lexical tone, and it would be interesting to see how it compares with other languages. The role of music is also going to be examined by comparing the performance of Greek musicians and non-musicians, as it has been found to be a factor that can affect the perception of Mandarin tone (e.g., Alexander et al., 2005; Lee & Hung, 2008).

To summarize, this study will attempt to answer the following questions:

a. How well do Modern Greek native speakers discriminate Mandarin lexical tone pairs T1-T2, T1-T4, and T2-T3?
b. How do Modern Greek native speakers with considerable music training differ in their perception of the tone pairs T1-T2, T1-T4, and T2-T3, compared to Greek speakers without or with minimal training?
c. How does the pattern of Greek speakers in discriminating pairs T1-T2, T1-T4, and T2-T3 compare with the patterns by speakers of other languages as found in the literature?

2 Literature Review

There are several studies that tried to assess how speakers of diverse native language backgrounds perceive Mandarin lexical tone (e.g., Hao, 2012; So & Best, 2010, 2014; Tsukada, 2019). These studies have investigated Mandarin tone perception by native speakers of other tone languages (e.g., Cantonese: Hao, 2012; So & Best, 2010; Thai and Vietnamese: Tsukada, 2019), native speakers of pitch-accent languages (e.g., Japanese: So & Best, 2010; Tsukada & Idemaru, 2022; Swedish: Gao, 2016), native speakers of non-tone languages with lexical stress (e.g., English: Hao, 2012; So & Best, 2010, 2014; Tsukada: 2019), and native speakers of non-tone languages without lexical stress (e.g., French: So & Best, 2014; Korean: Tsukada & Han, 2019). Many of these studies have investigated the perception of speakers that are Mandarin-naïve (e.g., So & Best, 2010, 2014; Tsukada, 2019), some have solely included learners of Mandarin (e.g., Gao, 2016; Hao, 2012), while a few included both (e.g., Tsukada & Han, 2019; Tsukada & Idemaru, 2022). Finally, there is a number of studies that have focused on how Mandarin lexical tone perception is affected by music training (e.g., Alexander et al., 2005; Lee & Hung, 2008).

2.1 Native language influence on the perception of Mandarin lexical tone

While all languages employ pitch variation for linguistic purposes, they can differ in how they use it to convey lexical meaning. So and Best (2010) were interested to explore the effect of linguistic experience with native tonality on the perception of Mandarin lexical tone categories. To that end, they employed an identification task[1] to gauge the performance of three listener groups from various language backgrounds that had no experience with Mandarin. Participants in the first group were native speakers of Hong Kong Cantonese, a tone language that uses pitch variation in all lexical items; the second group consisted of native speakers of Japanese, a pitch-accent language with limited pitch variation in differentiating lexical items; the third group comprised native speakers of Canadian English, a

[1] An identification task is an experimental design common in perception studies where subjects are presented with stimuli that they need to give an explicit label regarding which category they judge it belongs to. In the case of Mandarin tone, for instance, listeners would be presented with a tone carrying syllable and then would have to decide which tone its carries.

non-tone language with very limited use of pitch variation to distinguish lexical meaning.

The results suggested that while experience with native tonality does have an influence on the perception of non-native lexical tone, with both Cantonese and Japanese outperforming English speakers, its effect is more related to constrains introduced by the way pitch is used in the native language. Cantonese speakers misidentified Mandarin T1 and T4 because the pitch patterns of these tones closely approximate two phonetic variations of Cantonese Tone 1. In a similar manner, Cantonese speakers confused pair T2-T3 because their pitch contour patterns resembled Cantonese Tone 2. While no similar effects were observed for the Japanese and English speakers, they noted that some tone contrasts are more difficult than others irrespective of the listeners' native language. More specifically T1-T2, T1-T4, and T2-T3 were more difficult for all language groups compared to T1-T3, T2-T4, and T3-T4, suggesting that the perception of tone contrasts with distinctive phonetic properties is less affected by native language.

Tsukada (2019) also investigated the effect of native tonality on the perception of Mandarin lexical tones but from a slightly different angle. In her study, she included two groups of native tone language speakers, Thai and Vietnamese; one group of native speakers of a non-tone language, i.e., Australian English; and a group of native speakers of Mandarin as a control group. The aim of the study was to assess discrimination accuracy rates for Mandarin lexical tones by three groups of non-native listeners that were naïve to Mandarin. Her intention was to explore how, and in which way, native and non-native listeners differ in Mandarin tone processing, as well as to find out whether there are differences between those listeners coming from a tonal and those coming from a non-tonal language background. While tone pairs T1-T2, T1-T4, and T2-T3 were more difficult for all non-native groups, the confusion patterns for these pairs differed between groups, which suggests that native language background still casts an influence on tone perception. For the English group, pair T1-T2 was the most difficult overall; for the Thai group, it was pair T2-T3; for the Vietnamese group, it was pair T1-T4.

Another prosodic feature that can determine how pitch variation is used in a language, and thus possibly affect the perception of non-native sounds, is lexical stress. So and Best (2014) examined the perception of Mandarin tone in a sentence environment by two groups of Mandarin naïve listeners. The first group consisted of native speakers of Australian English, a language with lexical stress, while the latter consisted of native speakers of French, a language without lexical stress. In their study, they carried out a discrimination task[2] in order to quantify the performance of participants in discriminating the six Mandarin tone pairs. The results of the discrimination task, with the target syllables being words embedded in a sentence at the position of the penultimate syllable, showed an identical confusion pattern for both groups. Moreover, the most phonetically similar tone pairs T1-T2, T1-T4, and T2-T3 were again found the hardest to perceive. Nevertheless, native language also played a role, as French speakers had better performance for all pairs, which was attributed to the lack of lexical stress in their native system, its absence simplifying the task of identifying pitch contour.

It should be noted that the acoustic properties of tone carrying syllables can vary depending on their relative position in a sentence, as has been reported in relevant studies (Ho, 1976; Howie, 1974; Nordenhake & Svantesson, 1983). For instance, according to Ho's research (1976), syllables uttered in isolation and in sentence final position exhibit more variation in their duration depending on which tone they carry; with T3 being the longest, followed by T2, and then by T1 and T4. On the other hand, his research also showed that when in other positions within a sentence, the difference in duration between syllables carrying different tones tends to be much smaller, and T3 even becomes shorter than T2. This can be important for tone perception, as duration can often be an additional cue for tone judgement. Regarding T2 and T3 in particular, both native speakers of Mandarin and Mandarin-naïve native speakers of English have been found to perceive elongated tone contours, which are intermediate between these two tones, as instances of T3 (Blicher et al., 1990). It has also been suggested that while native Mandarin speakers use duration as a secondary cue that does not affect their accuracy, non-native speakers use it as a cue that is similar in importance with F0

[2] A discrimination task is an experimental design that is also common in perception studies. In its most basic form, subjects are presented with two stimuli, and then have to judge whether they belong to the same category or not. For instance, in a Mandarin lexical tone perception study, listeners would be presented with two tone carrying syllables, and then should decide whether they carry the same tone or not.

(Chang, 2011). This can explain the discrepancy between the studies of So and Best (2014) and Tsukada (2019), where each found a different confusion pattern for native speakers of Australian English. In the former study, where target syllables were in the penultimate position of a sentence, participants could mainly rely on pitch patterns to discriminate tone, while in the latter study that used syllables in citation form, they could also leverage duration. It is possible that this is an effect of lexical stress, which is a feature of the prosody of English, as duration is a primary cue of syllable prominence (Arvaniti, 2020). Perhaps it is no coincidence that in Tsukada and Han (2019), Mandarin-naïve native speakers of Korean, an intonation language lacking lexical stress, discriminated pair T2-T3 least accurately in a task involving stimuli in citation form; stimuli that were identical to those used for Australian English in Tsukada (2019).

However, duration tends to become a less important cue as non-native listeners of Mandarin gain experience with the language or, alternatively, its effect on tone perception has certain limits. Tsukada & Idemaru (2022) explored how learners of Mandarin coming from two different native language backgrounds differ in their perception of Mandarin tone pairs. More specifically, in a study that used stimuli in citation form, they tested four groups of listeners: two groups of native Japanese speakers that differed in their experience with Mandarin, and two groups of native English speakers that also differed in their experience with Mandarin. All four groups confused pairs T1-T2, T1-T4, and T2-T3 the most. However, while both Japanese groups displayed the same confusion pattern, with T2-T3 being the most confusable overall, as in regard to the two groups of native English speakers, the Mandarin-naïve group found T2-T3 as the easiest of the three, while the experienced group found T2-T3 as the hardest. This is because, even though the experienced groups performed better for all pairs, the extent of the improvement was the lowest for T2-T3.

The reason why learners show the least improvement with pair T2-T3 is probably related to the similarity of the pitch patterns in these two tones. Chandrasekaran et al. (2010) reported that training with Mandarin lexical tone has been shown to enhance the ability of native speakers of English to follow pitch direction. In fact, according to their findings, good learners placed even more emphasis on this dimension of tone compared to poor learners. Therefore, it is reasonable to assume that the more distinct differences in pitch direction between tones in pairs T1-T2 and T1-T4 enable learners to perceive them better compared to Mandarin-naïve listeners, in contrast to pair T2-T3, where the overlap in pitch direction is very high. Furthermore, in addition to sharing T2-T3 as the most confusable pair, learners of Mandarin from various language backgrounds also displayed less variation in their accuracy rates. For instance, in Tsukada and Idemaru (2022), while the Mandarin-naïve Japanese group outperformed the Mandarin-naïve English group for some pairs, the two experienced groups showed a very similar performance for all pairs. Hao (2012) also came to a similar conclusion, in a study that investigated how learners of Mandarin coming from a tonal and a non-tonal native language background differ in the perception and production of non-native tone. By utilizing a task involving the tone identification of Mandarin syllables, she found no significant difference between two groups of native English and native Cantonese speakers, while both groups displayed the highest error rates for pair T2-T3.

Nevertheless, the above findings should not be interpreted as definite evidence that experience with Mandarin makes the listeners' native language less relevant as a factor affecting perception of Mandarin tone perception in general, and T2-T3 as the most difficult pair in particular. There is limited research on the subject to arrive at a general conclusion, while there is evidence which shows that native language can still have a strong effect on the confusion patterns of listeners that are also learners of Mandarin. To investigate whether the pitch accent feature of the Swedish language has any influence on the acquisition of Mandarin tone, Gao (2016) conducted an experiment with native Swedish speakers, that were also high school students and learners of Mandarin, by employing an identification task and stimuli that were monosyllabic words. According to the bidirectional error rates, the most confusing pair was T1-T2, rather than being T2-T3. This was attributed to the prosody of Swedish, where the contours of the two pitch accents of the language resemble the contours of T3 and T4. While this might be a fringe case, it still shows that native language can have a significant effect on tone perception if there is a close match between native and non-native categories, irrespective of the acoustic properties of the contrasts.

2.2 Music training influence on the perception of Mandarin lexical tone

Music and speech share fundamental frequency, or pitch, as a cue. There is a number of studies that have focused explicitly on how Mandarin lexical tone perception is affected by music training (e.g., Alexander et al., 2005; Lee & Hung, 2008). Music training has been found to be a factor that can affect the perception of lexical tone by making listeners more sensitive to pitch variation (e.g., Wong et al., 2007). Alexander et al. (2005) investigated the perception of two groups of native American English speakers with no previous experience of Mandarin; one group was made of musicians that had received continuous music training for eight or more years, while the other group was made of individuals with a maximum of three years of music training. They were interested to find out whether advanced processing of fundamental frequency, which is associated with music training, can carry over to the speech domain. They used stimuli produced in citation form but only after normalizing it for duration, in order to deny participants with an additional cue to base their judgements on. The difference in accuracy between musicians and non-musicians was significant in both the identification and discrimination tasks, with the former group outperforming the latter.

Gottfried et al. (2004) compared the discrimination of Mandarin tone by two groups of native speakers of American English, the former made of conservatory students, the latter made of college students studying an unrelated subject. The syllables that they used as stimuli were first recorded in a carrier sentence and then isolated in order to be used in the task. While both groups found pairs T1-T4 and T2-T3 as the most difficult, the conservatory students outperformed the non-conservatory students in discriminating all pairs. However, not all studies found musicians to have an absolute advantage in the perception of all six tones. Lee & Hung (2008) investigated Mandarin tone identification by two groups of native English speakers, one with musicians that had on average 15 years of experience, the other with non-musicians. They found that musicians outperformed non-musicians in the identification of all tones presented in citation form. More specifically the results showed that accuracy in the identification of T1, T2, and T4 greatly benefitted from musical training, while the difference for T3 was much smaller, with both groups being the least accurate with this tone. They attributed this difficulty to the presence of glottalization in some of the T3 tokens in their study.

Glottalization is a type of phonation that is also known as creaky voice. According to a definition by Keating et al. (2015), creaky voice in its prototypical form is characterized by a low rate of F0, irregular F0, and constricted glottis. In Mandarin, this phonation is largely induced by low pitch targets in T2, T3, and T4, with T3 reaching the lowest target of all, and thus being perceived as the creakiest (Huang, 2020). In contrast to the native English speakers in Lee and Hung (2008), it has been found that native Mandarin speakers identify T3 stimuli with creaky voice more easily compared to non-creaky stimuli (Huang, 2020). This suggests that the English speakers in Lee and Hung (2008) did not necessarily confuse creaky T3 stimuli because they failed to detect the low pitch level, but they might have lacked the experience to place them in the correct category.

2.3 Summary

To summarize, pairs T1-T2, T1-T4, and T2-T3 are typically found to be the most difficult to perceive for speakers from various native language backgrounds (e.g., So & Best, 2010; Tsukada, 2019). In regard to Mandarin-naïve speakers, while T2-T3 is often found to be the most confusable of the three (e.g., So & Best, 2010, 2014; Tsukada & Han, 2019), the manner in which pitch variation is used in the categories of one's own native language can have a direct effect on the confusion pattern, with T2-T3 even becoming the easiest of the three in some cases (Vietnamese: Tsukada, 2019). There is also evidence that native prosody can have an indirect effect on the discrimination of contrasts through features such as lexical stress. This might explain why English speakers appear to be able to take advantage of duration as an additional cue when it is salient enough in the stimuli they are presented with, in contrast to when this cue is not as salient, as manifested by a different confusion pattern for these three pairs, (So & Best, 2014; Tsukada, 2019). Furthermore, experience with Mandarin and experience with music have been found to affect perception in a similar manner, by making listeners from a non-tone background more sensitive to pitch direction (Chandrasekaran et al., 2010; Wong et al., 2007). Through this effect, the degree of similarity between the pitch patterns of contrasts seems to become a more important determiner of relative confusion compared to secondary cues such as duration, as the different results between native speakers of English that are learners of Mandarin and native speakers of English that are Mandarin-naive might suggest (Tsukada & Idemaru, 2022). This study is going to focus on Mandarin-naïve native speakers of Greek, a non-tone language with lexical stress that has received less attention from perception studies of Mandarin lexical tone. Music is also a factor that will be examined, as it can affect tone perception.

3 Methods

The present study aims to measure how accurately native speakers of Greek perceive Mandarin lexical tone, as well as the effect of music experience on tone perception. To fulfill this, this project follows a quantitative approach by employing an experimental design that allows the quantification of the perception of the three Mandarin tone contrasts that are often found in the literature to be the most confusable for non-native speakers. To this end, a discrimination task was used as the research instrument, which, in contrast to an identification task, does not require familiarization with Mandarin tone. In addition to the Greek speakers, a control group of native Mandarin speakers was recruited in order to set the benchmark of native-like accuracy. The sampling technique used for this study was social network or snowball sampling, which was chosen because, in addition to convenience, it also involves elements of random sampling (Buchstaller & Khattab, 2013). In the following subsections the methods are further analyzed.

3.1 Participants

A total of 21 listeners completed the discrimination task of this study, including 15 native Greek speakers and 6 native Chinese speakers. Native Greek speakers were subsequently divided into two groups, based on their amount of experience with learning and playing a musical instrument. The first group consisted of 8 (6 males and 2 females) native Greek speakers without any music experience, ranging in age from 31 to 52 (*mean age* = 38, *sd* = 6.14). The second group consisted of 7 (6 males and 1 female) native Greek speakers that had at least 4 years of music experience (*range* = 4 – 15, *mean* = 7.58, *sd* = 4.20), ranging in age from 29 to 39 (*mean age* = 33.29, *sd* = 3.99). The native Chinese speakers (5 females, 1 male) formed the control group. While there is a gender imbalance in participants of all groups, this researcher is not aware of any previous reports in the literature about gender being a factor that affects tone perception.

Informants were also asked to indicate whether they speak or learn other languages. All 15 native Greek participants were second-language speakers of English, their self-reported level in English ranging from intermediate (5 listeners) to advanced (9 listeners). A few of them also indicated having knowledge of additional languages; including French (3 listeners), German (2 listeners), Italian (2 listeners), Spanish (2 listeners), Indonesian (1 listener), and Russian (1 listener). None of the listeners reported having any experience with a tone language or a pitch-accent language.

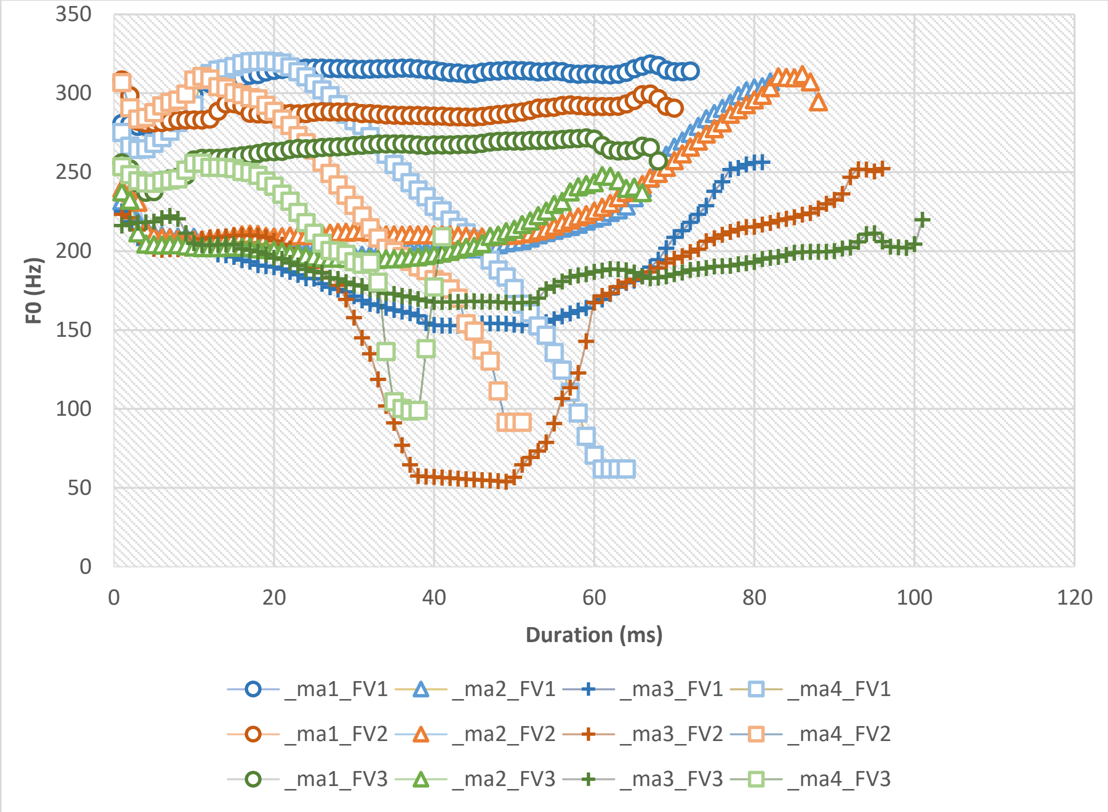
3.2 Materials

The speech stimuli for this study were derived from the Tone Perfect database (Ryu et al., n.d.). There are six utterances for every Chinese syllable in this database, with each syllable produced in isolation by three male and three female speakers. Only productions of the three female speakers were used as stimuli tokens. This decision was made in order to keep the duration of the experiment as short as possible, in an effort to minimize participant fatigue and the chance of abandonment. The reason for choosing the female speakers over the male speakers to be the talkers of this study was to retain a degree of comparability with relevant research in the literature. Among Mandarin lexical tone perception studies reviewed, those that did not involve talkers of both genders (e.g., So & Best, 2010; Tsukada & Han, 2019) utilized female talkers exclusively (Gao, 2016; Hao, 2012).

The target syllables in the discrimination task were Mandarin syllables /ma/ and /ni/. These two syllables were used because they consist of consonant and vowel combinations that are commonly found in other languages, including Greek, and as such were considered more familiar for participants, which would allow them to focus only on the tones. These two Mandarin segmental syllables were combined with the four Mandarin tones (Tone 1, Tone 2, Tone 3, Tone 4) to form a total of 8 syllables (ma1, ma2, ma3, m4, ni1, ni2, ni3, and ni4) that also exist as real words in Mandarin. Furthermore, the inclusion of tokens produced by three individual female talkers brought the total number of syllables used as stimuli to 24 (2 syllables x 4 tones x 3 talkers = 24 stimulus tokens). Tokens of syllable /ma/ were used in the main task to measure the performance of participants in discriminating tone contrasts, while tokens of syllable /ni/ were reserved for practice and presented before the main task. Practice trials were not considered in the calculation of the accuracy scores of the participants. The pitch contours of the twelve /ma/ syllables uttered by the three female speakers can be seen in Figure 1. The materials were originally in the MP3 file audio format but were all converted to the WAV format in order to be compatible with the software used for the experiment.

Figure 1

Pitch contours of the stimuli used in the discrimination task



Note. Contours traced with circular points indicate T1, triangular points indicate T2, cross points indicate T3, and rectangular points indicate T4. Shades of blue represent citations by talker FV1, shades of orange by talker FV2, and shades of green by talker FV3.

3.3 Procedure

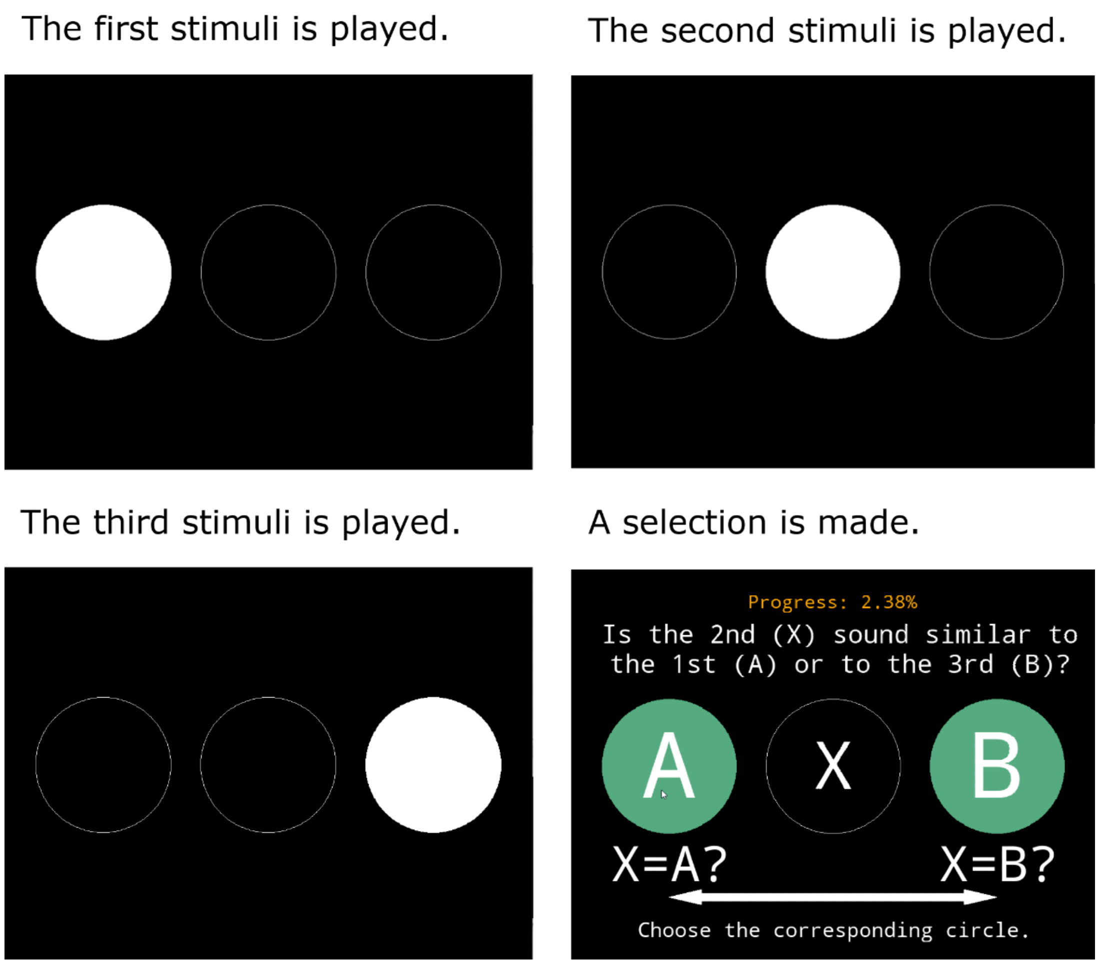
To assess how well participants can perceive Mandarin lexical tone contrasts, an AXB discrimination task was used. In this type of task, stimuli are presented in triads. Letters “A” and “B” represent tokens belonging to the first and the second contrasting sound categories respectively, while letter “X” represents a token that can belong to any of the two contrasting sound categories. The objective is to judge whether the second stimulus (X) belongs to the same category as the first or the third stimulus. This process, which is called a trial, is typically repeated for a number of times. In this study, participants were tested on their ability to discriminate between tokens of Mandarin lexical tone in a total of 86 trials. These trials were presented in three blocks corresponding to three tone contrasts (T1-T2, T2-T3, T1-T4). The first 4 trials in each block were for practice and were not analyzed. Accuracy rates were calculated by the remaining 72 trials, which were equally distributed into the three blocks (24 trials x 3 blocks).

In each trial, all three stimuli were tokens of a segmentally identical syllable (/ma/ in the main task, and /ni/ in the practice session), but varied in the tone they carried. In the AXB paradigm, there are four possible orders of stimuli presentation: AAB, ABB, BBA, BAA. For example, a trial testing the T1-T4 pair might consist of any of the following 4 sequences of syllables: [ma1] (A), [ma1] (A), [ma4] (B); [ma1] (A), [ma4] (B), [ma4] (B); [ma4] (B), [ma4] (B), [ma1] (A); [ma4] (B), [ma4] (B), [ma1] (A). This 4-trial format, which can reduce bias by involving all possible sequences of stimuli and distributing the correct responses equally, formed the basic unit of this task. In other words, trials were organized in sets of 4 trials. To ensure that listeners would not use the specific characteristics of a speaker's voice as a point of reference, all three individual syllables in each trial were always uttered by different talkers. For example, a trial testing pair T1-T4 might consist of: [ma1]-FV1 (A), [ma1]-FV2 (A), [ma4]-FV3 (B) (where FV1, FV2, and FV3 represent the first, the second, and the third female talker of this study respectively). Within each block, there were 6 sets of the 4-trial format. The order of talkers varied with every set but remained constant for all 4 trials within each set.

The experiment was implemented using the OpenSesame experiment builder software (Mathot et al., 2012) and then exported in the OSWeb extension for use with the Jatos server software (Lange et al., 2015), a tool specifically intended for running online studies. The experiment was uploaded on MindProbe (https://mindprobe.eu/), a website that provides free access to researchers for running online studies on Jatos. Links to participate in the experiment were then generated and provided to the participants, who could access it by using a browser on a personal computer, a smartphone, or any other device with similar capability. Once opened, the informant was first asked to provide their consent, and then to fill a small questionnaire asking some basic information. After this section was completed, the experiment component was loaded automatically. Before it began, the participants were presented with brief instructions about how to complete the task. They were told that the first and the third sound in every trial are always different, and that they should decide whether the second sound they hear is similar to the first or to the third. The discrimination task would begin immediately afterwards.

Figure 2

The four steps of a trial



The task required participants to respond to 72 trials. Each trial was a 4-step process which is illustrated in Figure 2. Three circles on black background, representing the three sounds of the AXB task, were always present for the duration of each trial. Depending on which sound was played, the corresponding circle would at the same time appear filled with white color, while the other two circles would simply appear with a white border. This was intended to provide listeners a visual cue to assist them with keeping track of which stimuli they were presented with at any point during the task. The interstimulus interval (ISI) was one second (1000ms). At the end of a trial, instructions would appear at the top of the experiment canvas asking participants to choose the circle that corresponds to the right answer, while additional visual cues, such as labeling and color, were utilized to minimize confusion. There was no time limit in choosing the answer, but it was not possible to repeat the playback of the stimuli. After a choice was made, the next trial started automatically.

The order by which the three main blocks were presented was randomized for every participant. Within each block, the 24 trials were randomized as well. There was no information about the contrasts involved at any point in the process. At the end of each block, there was a short mandatory break of about one minute. To keep listeners occupied during the waiting time, a humorous story would play for the participants running the Greek version of the experiment, and a recording of natural sounds for the participants running the English or the Chinese version. At the end of the third block, the experiment would finish. The total time needed for completion typically did not exceed 15 minutes.

3.4 Method of analysis

As the first step of the analysis process, the responses of the participants were coded as correct or incorrect depending on whether they chose the correct or the wrong answer in each trial. Based on this coding, three accuracy rates, one for each tone pair, were obtained for every participant, by calculating the ratio of correct responses to the total number of trials. As a final step, mean accuracy rates were derived from the individual accuracy rates of the participants. The mean accuracy rates represented the performance of each group in discriminating the three tone pairs and were used as the main measure for the analysis.

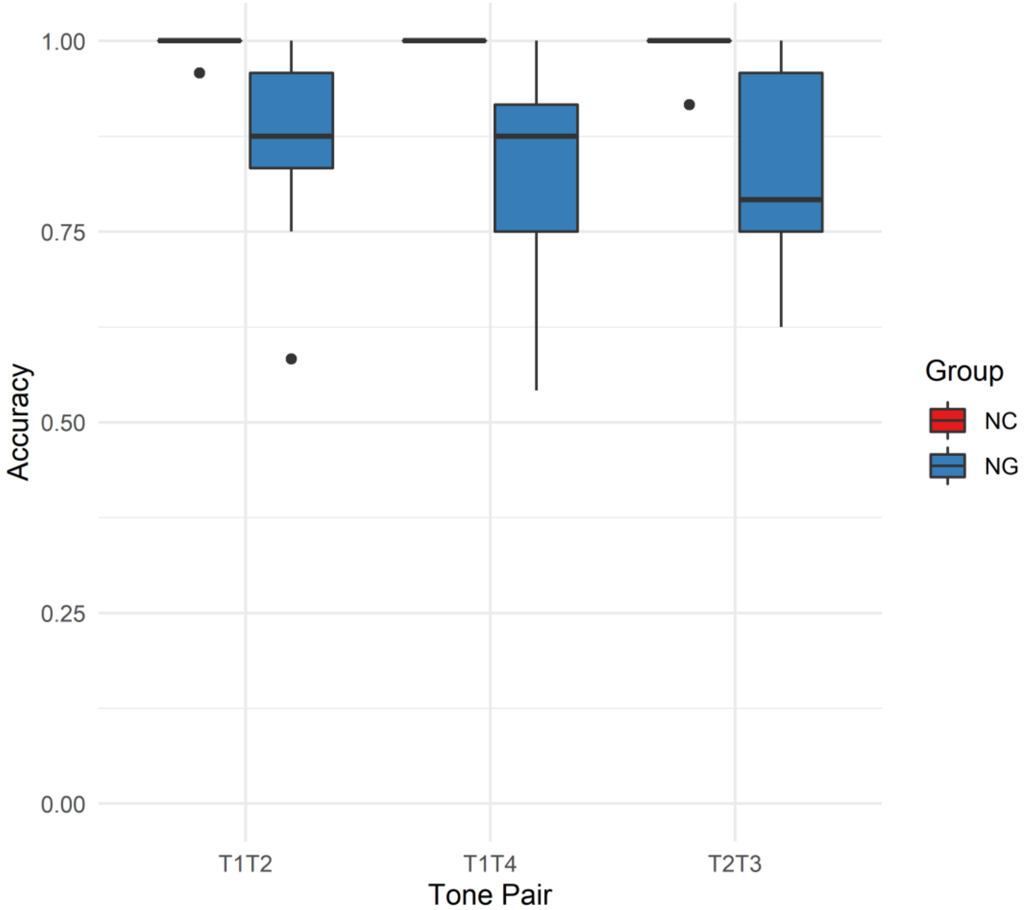
4 Results

Even though a total of 15 native Greek speakers participated in the experiment, not all were included in the analysis, because the data obtained from 2 of the respondents were evaluated as questionable. The performance of the remaining 13 native Greek listeners (NG), along with that of the 6 native Chinese listeners (NC), is graphically represented in Figure 3, where the distributions of the accuracy rates for the participants of the two groups can be seen as a function of tone pair.

The mean accuracy rates (with standard deviation in parentheses) of the NG group are 0.88 (0.12), 0.83 (0.14), and 0.82 (0.12) for tone pairs T1-T2, T1-T4, and T2-T3 respectively. The 13 native Greek speakers that took part in this study were better at discriminating T1-T2, while their performance was similar for T1-T4 and T2-T3. The corresponding mean accuracy rates of the NC group are 0.99 (0.02), 1.00, and 0.99 (0.03) respectively.

Figure 3

Tone pair accuracy rates distribution for native Greek and native Chinese speakers

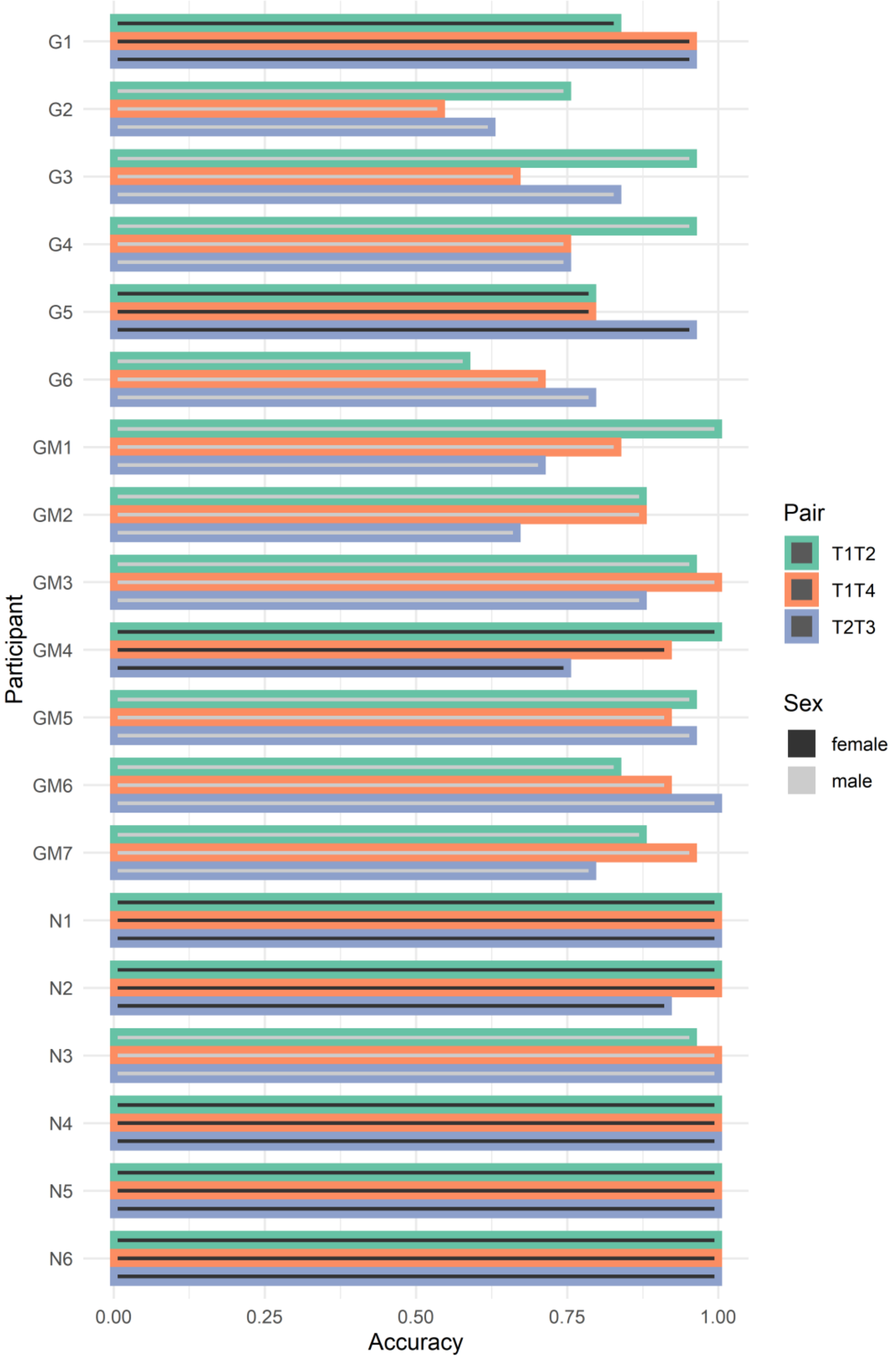


Note. The thick black line in each box represents the median, while the edges of the box mark 25% of the data above and below the median respectively. The extent of each box represents the middle 50% of the distribution and is known as the interquartile range (IQR). The lines extending from the boxes show the remaining upper and lower ends of the distribution, but only for measurements that fall within a distance of 1.5 times the IQR. Data points beyond that range are indicated as dots and are often termed as outliers or extreme values.

The discrimination accuracy rates of individual Greek participants are broken down in Figure 4. Regarding the individual scores of the NG group, 5 had the highest accuracy rates for T1-T2, 2 for T1-T4, and 2 for T2-T3; 1 had the lowest rates for T1-T2, 3 for T1-T4, and 5 for T2-T3. A few of the participants had the same accuracy rates for two of the three contrasts. More specifically, 1 had the highest accuracy rates for both T1-T4 and T2-T3, 1 for T1-T2 and T1-T4, and 1 for T1-T2 and T2-T3. On the other hand, 1 had the lowest accuracy rates for both T1-T2 and T1-T4, and 1 for T1-T4 and T2-T3. In the case of the NC group most participants had perfect scores for all pairs, with the exception of 2 participants. One of the two had a slightly lower rate for T1-T2, while the other for T2-T3.

Figure 4

Accuracy rates of native Greek speakers for all tone pairs.

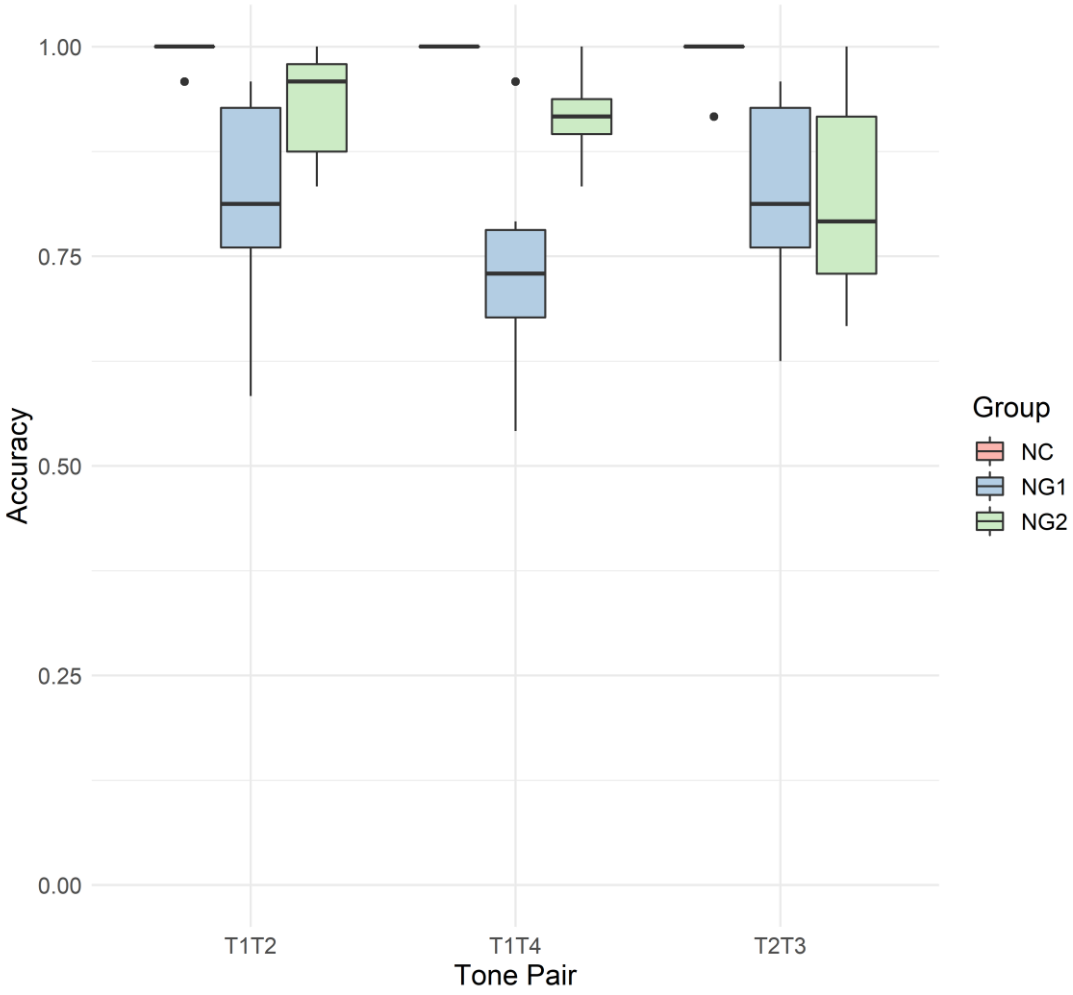


Note. Alphanumeric values on the y axis are used to code participants. Letter 'G' represents native Greek speakers, letter 'N' is for native Chinese speakers, while letter 'M' indicates music experience. The length of each bar indicates the accuracy rate for the specific tone pair as determined by the border color. A black fill color is used for females, and a light grey color for males.

To investigate the effect of music experience, the 13 native Greek speakers of the NG group were further divided into two groups. NG1 was composed of speakers without any music experience, while NG2 was consisted of speakers with music experience that ranged from 4 to 15 years. The 6 native Chinese speakers used as control formed group NC as before. The accuracy rates of all three groups for the three tone pairs can be seen in Figure 5.

Figure 5

Tone pair accuracy rates distribution for Greek listeners with music training (NG2) and without music training (NG1)



Note. NC represents native Chinese speakers, NG1 native Greek speakers without music experience, and NG2 native Greek speakers with music experience. The thick black line inside each box is the median or Q2, the upper and the lower edges of the box represent the Q3 and the Q1 respectively, the extent of the box is the interquartile range (IQR), while the whiskers above and below the edges reach up to a maximum of Q3+1.5*IQR and a minimum of Q1-1.5*IQR respectively to denote observations that fall within that range. Data points beyond that range are illustrated as dots.

As already indicated, the NC group was the most accurate, with mean accuracy rates that were perfect or near perfect. The NG1 group had the most difficulty with discriminating T1-T4 (0.74), while rates for T1-T2 (0.81) and T2-T3 (0.82) were similar. The NG2 had the most difficulty in discriminating T2-T3 (0.82), while achieving similar mean rates for T1-T2 (0.93) and T1-T4 (0.92). Compared to NG1, the accuracy rates of NG2 were higher for pair T1-T2 and T1-T4, but the performance of the two groups in discriminating T2-T3 was identical. The mean accuracy rates of all three groups are summarized in Table 2. Another thing to note is that while the NG2 had standard deviations for T1-T2 and T1-T4 that were half as high as those of NG1, for T2-T3, the standard deviations of the two groups were almost equal.

Table 2
Tone pair mean accuracy rates of groups NC, NG1, and NG2.

Tone pairs	NG1 (n=6)	NG2 (n=7)	NC (n=6)
T1-T2	0.81 (0.141)	0.93 (0.067)	0.99 (0.02)
T1-T4	0.74 (0.139)	0.92 (0.054)	1.00
T2-T3	0.82 (0.128)	0.82 (0.127)	0.99 (0.03)

5 Discussion

5.1 How well do Modern Greek native speakers discriminate Mandarin lexical tone pairs T1-T2, T1-T4, and T2-T3?

This study investigated the discrimination of Mandarin lexical tone pairs T1-T2, T1-T4, and T2-T3 by native speakers of Greek who have no knowledge of Mandarin. The Greek participants did not reach native-like performance, which is consistent with findings in the literature regarding Mandarin-naïve listeners from various language backgrounds (e.g., So & Best, 2010, 2014; Tsukada, 2019).

The Greek speakers (NG) that participated in this study discriminated most accurately pair T1-T2 (0.88), and less accurately T1-T4 (0.83) and T2-T3 (0.82), achieving similar accuracy rates for the latter two. Due to the task involved, their performance was to a certain degree affected by the acoustic properties of the tones, but the intonation categories in their own native language are likely to have played a role as well, either in aiding discrimination or constraining it. Although the Greek language does not use lexical tone, nevertheless, like any other language, it employs pitch variation to convey linguistic meaning as determined by intonation. Pitch patterns that are variations of a single category in Greek, while at the same time they resemble two distinct categories in Mandarin, could have affected discrimination rates negatively. As the study was not designed to detect interference from specific intonational categories that exist in native Greek prosody, and given the generally high accuracy rates they reached, it is not possible to confirm any hypothesis on that matter, other than note a small advantage in discriminating pair T1-T2, compared to T1-T4 and T2-T3.

5.2 How do Modern Greek native speakers with considerable music training differ in their perception of the tone pairs T1-T2, T1-T4, and T2-T3, compared to Greek speakers without or with minimal training?

When considering music experience as a factor, the non-musicians of the NG1 found pair T2-T3 (0.82) the easiest to discriminate, closely followed by T1-T2 (0.81), and T1-T4 (0.74) the hardest. The results for the NG2, on the other hand, revealed a positive effect of music experience on the discrimination of the tone pairs but not in a uniform manner. For NG2, pair T1-T2 (0.93) was the easiest to discriminate, T1-T4 (0.92) closely followed, while T2-T3 (0.82) was the hardest. Music experience not only aided native Greek speakers in discriminating pairs T1-T2 and T1-T4 better, but also to discriminate them equally well. In addition to that, the standard deviation for these two pairs was almost half of that of the NG1 group, indicating that the within group performance was now more homogenous as well. However, regarding pair T2-T3, no difference in accuracy rates between the NG1 and NG2 was observed (0.82 for both), including in standard deviation, suggesting that music training had a minimal to no effect on the discrimination of this pair.

Findings from the current study, related to the effect of music experience on non-native lexical tone perception, are, for the most part, consistent with the literature. Speakers of non-tone languages, such as English, have been found to rely more on pitch height rather than pitch direction (e.g., Gandour, 1983), while it has been reported that experience with music can enhance the ability to follow pitch direction (e.g., Wong et al., 2007). Attendance to pitch direction, rather than just height, can be one reason why NG2 outperformed NG1 in discriminating T1-T2 and T1-T4. This is because each of these two pairs involves tone contrasts with pitch patterns that are relatively distinct in terms of direction; in pair T1-T2, T1 is level and T2 is rising; in pair T1-T4, T1 is level and T4 is falling. On the other hand, for T2-T3, no advantage was noted for participants with music experience. One reason for this might be the high overlap in pitch direction between the tones of this contrast. As their tone contours resemble each other in terms of direction, to discriminate tones T2 and T3, listeners should be able to perceive the small differences in height and the timing of change in pitch direction. However, native English musicians have been shown to outperform non-musicians in all tone pairs, including T2-T3 (Gottfried et al., 2004). While syllable duration might have played a role in Gottfried et al. (2004), as it is a cue that musicians have been found to leverage more efficiently (Chen et al., 2020), in the current study it did not seem to provide Greek musicians with an advantage in discriminating pair T2-T3.

This effect of music experience is not enough to draw any concrete conclusions about the sources that complicate Mandarin tone perception for native Greek speakers, but it can still offer some clues in investigating them. For instance, in regard to T1-T4, the reason for the much lower performance of the NG1 in discriminating this pair relative to the other two pairs, is that the speakers are more affected by Greek intonation. One possibility is that the tone patterns of T1 and T4 form a single category in the Greek intonation system, or simply put they are both associated with the same linguistic meaning. In terms of intonational pitch patterns at the end of the sentence that indicate pragmatic meaning, Greek does not seem to have an intonational category resembling the sharp fall of Mandarin T4 (for a graphic representation, Arvaniti & Baltazani, 2005, p. 95, table 4.1). However, this does not mean that there are no sharp falls at the edge of Greek intonation tunes. A flat pitch pattern associated with a statement can actually involve a sharp fall over the last syllable when the ultimate syllable of the last word is stressed. Thus, a flat pitch and a falling pitch at the very last section of the tune can have the same pragmatic meaning, which might cause listeners that process it as speech, as speakers with no musical training are likely to do, to unconsciously compensate for these differences (Yip, 2002). By that account, music experience made the participants of the NG2 more sensitive to acoustic properties, allowing them to fully bypass a higher linguistic processing and perceive the contrasts in this pair as acoustic signals. However, another plausible, and much simpler, explanation is that non-musicians were confused by the creaky voice induced by very low pitch targets at the end of the syllable, as creakiness in T4, which can be often present in instances of this tone (e.g., Belotel-Grenie & Grenie, 1994; Kuang, 2017), has been shown to inhibit identification (Huang, 2020). On the other hand, musicians were able to base their judgement on pitch direction, in addition to syllable duration, and thus perform extremely well in the discrimination of pair T1-T4.

In the case of T2-T3 though, when considering the performance of the two groups, it is likely that native Greek speakers were aided by their native language. While those two tones have similar contours, i.e, first falling and then rising, a fact that is often cited as the reason why this pair is found as difficult to discriminate even for speakers of tone languages (e.g., So & Best, 2010), in theory musicians still have an edge, because they can better follow pitch direction. In addition to pitch, there were also some secondary cues available due to the stimuli being in citation form, such as duration, that, again in theory, the NG2 should have been in a better position to leverage. Regardless of the above point, the performance of the NG1 and NG2 was identical for this pair. One possibility is that both groups found it hard, while the cues they could use were not salient enough. However, the NG1 group displayed a higher overall accuracy rate for pair T2-T3 in relation to the other two pairs. When considering that pairs T1-T2 and T2-T3 share T2 as a contrast, along with the relatively high accuracy rates for these pairs that also happen to be almost equal, it is plausible that Greek speakers are more sensitive to the finer details in rising pitch patterns. By looking at the prosody of the Greek language, there seem to be many pitch patterns in Greek intonation with rising patterns (for details see, Arvaniti & Baltazani, 2005, pp. 86-102), that while not matching exactly those of Mandarin for T2 and T3, they could, nevertheless, have aided Greek speakers in discriminating between pairs involving this type of patterns. The fact that the NG2 group displayed the same performance for this contrast might be another indication that the discrimination of pair T2-T3, which involves two tones with similar pitch patterns, was aided by their native language, as the greater potential for tone processing that listeners with music experience have been reported to possess (Wong et al., 2007) did not seem to give them any advantage compared to NG1.

5.3 How does the pattern of Greek speakers in discriminating pairs T1-T2, T1-T4, and T2-T3 compare with the patterns by speakers of other languages as found in the literature?

The comparison of discrimination patterns for Mandarin tone pairs among speakers from other language backgrounds can be an indicator of native language interference in the perception of tones. Table 3 summarizes the relative difficulty in discriminating tone pairs T1-T2, T1-T4, and T2-T3 across speakers from various native language backgrounds in previous studies.

Table 3
Cross-study comparison of confusion patterns for pairs T1-T2, T1-T4, and T2-T3

Language	Relative difficulty (easiest to hardest)	Type of task	Stimuli used
Greek (non-musicians) ***(this study)***	*T2-T3 ≥ T1-T2 > T1-T4*	*Discrimination (AXB)*	*Monosyllabic words in citation form*
Cantonese **(So & Best, 2010)**	T1-T2 > T1-T4 > T2-T3	Identification	Monosyllabic words in a carrier sentence
English **(Tsukada & Idemaru, 2022)**	T2-T3 > T1-T4 > T1-T2	Discrimination (oddball)	Monosyllabic words in citation form
French **(So & Best, 2014)**	T1-T2 > T1-T4 > T2-T3	Discrimination (AXB)	Monosyllabic words in a carrier sentence

Japanese **(So & Best, 2010)**	T1-T4 > T1-T2 > T2-T3	Identification	Monosyllabic words in a carrier sentence
Korean (Tsukada & Han, 2019)	T1-T2 > T1-T4 > T2-T3	Discrimination (oddball)	Monosyllabic words in citation form
Swedish (learners) **(Gao, 2016)**	T1-T4 > T2-T3 > T1-T2	Identification	Monosyllabic words in citation form
Thai **(Tsukada, 2019)**	T1-T4 > T1-T2 > T2-T3	Discrimination (oddball)	Monosyllabic words in citation form
Vietnamese **(Tsukada, 2019)**	T2-T3 > T1-T2 > T1-T4	Discrimination (oddball)	Monosyllabic words in citation form

Note. All groups are naïve to Mandarin except when stated otherwise. The Greek speakers are non-musicians. The results for French, Cantonese, and Japanese are from studies that controlled for experience with music. For those studies that used an identification task, confusion patterns were calculated by aggregating bidirectional error rates. The larger the total value, the harder a pair was considered. For all languages, except for Greek that lack data, and Swedish where T2-T4 is actually the second most confusable pair after T1-T2, the tone pairs in this table represent the three most confusable pairs ranked from easiest to hardest.

More specifically, table 3 above lists the confusion patterns of listeners that are native speakers of 4 non-tone languages, including Greek, English, French, Korean; 3 tone languages, including Cantonese, Thai, and Vietnamese; and 2 pitch-accent languages, including Japanese, and Swedish. Of the remaining three non-tone languages, the most comparable to Greek is English, as they are both languages that employ lexical stress. Lexical stress might be an explanation why the English speakers that are listed on the table had higher discrimination rates for T2-T3 and T1-T4 than they had for T1-T2, while the Koreans, like the French, found T2-T3 as the hardest to discriminate. Since an extended duration is a cue of prominence that marks stressed syllables, the stimuli in citation form that were used in Tsukada and Idemaru (2022), a form in which instances of every tone tend to vary more in duration, could have potentially allowed English speakers to leverage the duration differences between T2 and T3, as well as between T1 and T4. This can offer an explanation regarding why English speakers in So and Best (2014), where stimuli were embedded in a carrier sentence and the target syllables varied much less in duration irrespective of the tone they carried, found T2-T3 as the hardest and T1-T2 as the easiest. On the other hand, it can also explain why speakers of Korean, a language which lacks lexical stress, when tested on a discrimination task that involved monosyllabic words in citation form, showed a confusion pattern for these three tone pairs that was identical to the pattern of the English speakers in So and Best (2014), rather than to the pattern of the English speakers in Tsukada and Idemaru (2022) that used a similar type of stimuli.

In regard to the Greek speakers that were also presented with monosyllabic words in citation form, having a lexical stress feature in their native language can partly explain the results for T2-T3, but not for T1-T4, which, unlike English speakers, was discriminated the least accurately. Another reason why this difference exists between Greek and English, which is also relevant to comparisons with other non-tone languages, is related to the intonation patterns that determine the pragmatic meaning of sentences. For instance, in regard to T1-T4, English most often associates a falling pitch pattern with statements (So & Best, 2010), while for Greek, statements are mainly marked by a flat pattern (Arvaniti & Baltazani, 2005). As for T1-T2, there is evidence from Greek prosody that native speakers of the language might be sensitive to the difference between flat pitch and rising pitch when distinguishing fine details in linguistic meaning. More particularly, in Greek *wh*-questions there could be two variations of the tune; the first is marked by a flat pitch at the beginning, followed at the last section by a rising pitch, which can indicate politeness or the seeking of information; the second is marked by a sustained flat pitch, which mainly indicates a statement (Baltazani et al., 2020).

For Swedish speakers though, bidirectional error rates showed that T1-T2 was the hardest for them, probably because their identification score for T2 was the lowest among all four tones. Another element that sets apart the Swedish speakers from other speakers is that their confusion pattern for the three most confusable pairs included T2-T4. This is to be expected though, as they identified T3 most accurately, and identified T2 least accurately (Gao, 2016). The Japanese speakers, on the other hand, displayed a pattern that matched the one of the Thai speakers, where T2-T3 was the most confusable. Interestingly enough, the results for Japanese speakers have been consistently the same across studies, regardless of whether the task involved the identification of tones and stimuli in the form of target syllables in connected speech (So & Best, 2010), or the discrimination between tone contrasts and stimuli in the form monosyllabic words in citation form (Tsukada & Idemaru, 2022). Ultimately, the confusion pattern of Greek speakers more closely matches the one of Vietnamese speakers; even the relation between their actual accuracy rates is comparable, since Vietnamese had similar rates for T1-T2 and T2-T3, while they scored much lower for T1-T4. Perhaps this is no accident as, per an explanation by Tsukada (2019), there are more than enough tone patterns in Vietnamese for Mandarin T2 and T3 to be assimilated in, while there is an absence of sharp falls resembling T4. As already discussed, there are many intonation categories in Greek involving rising patterns, and none that is distinguished by a sharp fall. Of course, this similarity between the two languages might only be superficial and only relevant to monosyllabic stimuli, as in non-tone languages the pitch patterns can vary depending on many factors, such as the length of the utterance, or the way words are chunked together.

To conclude this section, it is worth noting that while this cross-study comparison can offer some insights regarding specific features in a language that can affect non-native lexical tone perception, it should be interpreted cautiously. There is no single methodology in conducting a perception study, while the characteristics of the participants can vary a lot. In regard to the experimental design, some researchers opt for a discrimination task that involves contrastive stimuli, while others use an identification task that directly measures the categorical perception of a single stimulus. The type of the stimuli used can also have an impact, as determined by the phonetic properties of the syllables used, as well as by the gender, the number and other unique characteristic of the talkers. Tones in tokens uttered isolated in citation form can be different in terms of duration or contour shape, compared to when they are uttered in a sentential environment. Furthermore, not all studies have controlled for music experience, which is a factor that can affect perception of lexical tone, as also showcased in this study.

5.4 Limitations of this study and recommendations for future research

There are certain limitations specific to this study. One is concerning the number of the tone contrasts that was limited to only three. Comparing the accuracy rates for all six tone pairs would provide a more definite answer regarding whether these three more “acoustically similar” tone pairs are also the hardest to discriminate for native speakers of Greek. At present, it is impossible to know whether T2-T3 is the easiest over all six tone pairs available in Mandarin, or only over the three examined in this study. Another limitation is related to the number of participants and their generalizability to a larger population. In addition to that, although there are scant reports for gender being a factor in tone perception, a more balanced sample would be preferable. The stimuli used is another limitation, as it only consisted of one monosyllabic word.

Taking the above into consideration, it is recommended that future studies include all six tone contrasts, increase the number of participants, opt for a more balanced sample, and use more monosyllabic words as stimuli. The addition of native speakers of English should also be considered, as this would allow for more direct comparisons between Greek and English and provide with a basis for more meaningful indirect comparisons with other studies that also involve English speakers. Furthermore, comparing the perception of stimuli in citation form to that of stimuli in sentential form might offer insights regarding the effect of lexical stress on tone perception. This is also true for creaky voice, as its presence has been found to both aid and hinder the identification of tones. Finally, it would be interesting to investigate the difference between listeners that are Mandarin-naïve, and listeners that are learners of Mandarin, in order to find out if it is comparable to that between the Mandarin-naïve musicians and non-musicians of this study, as both musicians and learners of Mandarin have been found to attend more to pitch direction rather than pitch height.

6 Conclusion

This study investigated the perception of Mandarin lexical tone pairs T1-T2, T1-T4, and T2-T3 by native speakers of Greek. The 13 native Greek speakers that participated in the study discriminated most accurately T1-T2 and T1-T4, and least accurately pair T2-T3. While their accuracy rates were relatively high, they fell short of the perfect, or near perfect, rates of native Chinese speakers that were used as the control group of this study.

When music experience was considered, a factor that has been found to enhance the ability to follow pitch direction, Greek speakers who were taught and played a musical instrument for a minimum of 4 years clearly outperformed those Greek speakers that had no music experience. From easiest to hardest, the confusion pattern of musicians was T1-T2, T1-T4, and T2-T3; while of non-musicians it was T2-T3, T1-T2, and T1-T4. The biggest improvement between the two groups was for T1-T4, possibly hinting that these two pitch patterns do not form an important contrast in the Greek language, or perhaps indicating that the listeners of this study experienced some confusion due to creaky voice, a type of phonation that is caused by very low pitch targets, which might have been present in some of the tokens of T4. Interestingly, the accuracy rates for T2-T3 were exactly the same for both groups, which, taken together with the relatively high accuracy rates for this pair, might suggest that the reason why Greek speakers did not benefit from music experience is because they were already more sensitive to rising pitch patterns.

Finally, in comparison to other studies, the confusion pattern of Greek speakers with no music experience did not match to that of English speakers, which is another language with lexical stress, neither to that of speakers of any other non-tone language. Instead, it matched the confusion pattern of Vietnamese speakers, perhaps for similar reasons. More specifically, while there are enough lexical tones with rising patterns in the Vietnamese tone system for T2 and T3 to be assimilated in, there are not any with sharp falls for the assimilation of T4. To a similar effect, while there are many intonation categories marking pragmatic functions that involve rising patterns in Greek, there is hardly any that is distinguished by a sharp fall.

References

Alexander, J., Wong, P., & Bradlow, A. (2005). Lexical tone perception in musicians and non-musicians. *Proceeding of Interspeech*, Lisbon, Portugal, 397-400. https://doi.org/10.21437/Interspeech.2005-271

Arvaniti, A. (2007). Greek Phonetics The State of the Art. *Journal of Greek Linguistics*, *8*(1), 97-208. https://doi.org/https://doi.org/10.1075/jgl.8.08arv

Arvaniti, A. (2020). The Phonetics of Prosody. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.411

Arvaniti, A., & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 84-117). Oxford University Press. https://doi.org/DOI:10.1093/acprof:oso/9780199249633.003.0016

Belotel-Grenié, A., & Grenié, M. (1994). Phonation types analysis in standard Chinese. *Proceedings of ICSLP'94*, Yokohama, Japan, 343-346. https://doi.org/ 10.21437/ICSLP.1994-89

Best, C. T. (1995). A Direct Realist View of Cross-Language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-204). York Press.

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

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37-49. https://doi.org/10.1016/S0095-4470(19)30357-2

Buchstaller, I., & Khattab, G. (2013). Population samples. In R. J. Podesva & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 74-95). University Printing House.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456-465. https://doi.org/10.1121/1.3445785

Chang, Y.-h. S. (2011). Distinction between mandarin tones 2 and 3 for L1 and L2 Listeners. In J.-S. Zhuo [Ed.], *Proceedings of the 23rd North American conference on Chinese linguistics (NACCL-23): Vol. 1. Phonetics and Phonology (pp. 84-96)*. https://naccl.osu.edu/sites/naccl.osu.edu/files/NACCL-23_1_06.pdf

Chao, Y. R. (1948). *Mandarin Primer: An Intensive Course in Spoken Chinese*. Harvard University Press. http://books.google.com/books?id=31UGAQAAIAAJ

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

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 229-273). York Press.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149-175. https://doi.org/10.1016/s0095-4470(19)30813-7

Gao, M. (2016). Perception of lexical tones by Swedish learners of Mandarin. *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition,* Umeå, Sweden, 33-40. https://aclanthology.org/W16-6505

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

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

Howie, J. M. (1974). On the Domain of Tone in Mandarin. *Phonetica*, *30*(3), 129-148. https://doi.org/doi:10.1159/000259484

Huang, Y. (2020). Different attributes of creaky voice distinctly affect Mandarin tonal perception. *J Acoust Soc Am*, *147*(3), 1441. https://doi.org/10.1121/10.0000721

Lange, K., Kuhn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLoS One*, *10*(6), e0130834. https://doi.org/10.1371/journal.pone.0130834

Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.*, *124*(5), 3235-3248. https://doi.org/10.1121/1.2990713

Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav Res Methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

Nordenhake, M., & Svantesson, J.-O. (1983). *Duration of standard Chinese word tones in different sentence environments.* (Working paper No. 25). https://journals.lub.lu.se/LWPL/article/download/16909/15288/43162

Ryu, C., Mandarin Tone Perception & Production Team, & Michigan State University Libraries. (n.d.). *Tone Perfect: Multimodal Database for Mandarin Chinese.* Retrieved March 5, 2022, from https://tone.lib.msu.edu/

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences. *Lang Speech*, *53*(Pt 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

Tsukada, K. (2019). Are Asian Language Speakers Similar or Different? The Perception of Mandarin Lexical Tones by Naïve Listeners from Tonal Language Backgrounds: A Preliminary Comparison of Thai and Vietnamese Listeners. *Australian Journal of Linguistics*, *39*(3), 329-346. https://doi.org/10.1080/07268602.2019.1620681

Tsukada, K., & Han, J.-I. (2019). The perception of Mandarin lexical tones by native Korean speakers differing in their experience with Mandarin. *Second Language Research*, *35*(3), 305-318. https://doi.org/10.1177/0267658318775155

Tsukada, K., & Idemaru, K. (2022). Effects of First Language Background and Learning Experience in Perceiving Mandarin Lexical Tones: Learners and Nonlearners From English- and Japanese-Speaking Backgrounds. *Journal of Speech, Language, and Hearing Research*, *65(2),* 829-842. https://doi.org/doi:10.1044/2021_JSLHR-21-00146

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10(4),* 420-422. https://doi.org/10.1038/nn1872

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/cbo9781139164559