What makes second language perception of Mandarin tones hard?

Eric Pelzl

Article published in:
Chinese as a Second Language 漢語教學研究—美國中文教師學會學報
Vol. 54:1 (2019), pp. 51–78. DOI: 10.1075/csl.18009.pel

This article is under copyright. The publisher (John Benjamins) should be contacted for permission to re-use or reprint the material in any form.

WHAT MAKES SECOND LANGUAGE PERCEPTION OF MANDARIN TONES HARD? A NON-TECHNICAL REVIEW OF EVIDENCE FROM PSYCHOLINGUISTIC RESEARCH

Author: Eric Pelzl, The Pennsylvania State University

ABSTRACT

Mandarin Chinese tones are known to be difficult for second language learners. A large body of research has examined non-native perception of tones, and may provide useful and interesting insights about the sources of tone learning difficulty for Chinese teachers and learners. However, much of the literature is in journals that may be difficult to access or written in technical language that may be hard for non-specialists to understand. This review article aims to summarize key findings from this research in an accessible fashion. I will draw on the research to answer five broad questions: 1) Why are tones more difficult for some learners than others? 2) Why are some tones more difficult than others? 3) Why are tones in words more difficult than in meaningless syllables? 4) Why are tones in context more difficult than in isolation? 5) What can we do about tone learning difficulties?

摘要

众所周知，普通话的声调对二语者来说非常困难。当前声调感知方面的大量研究为汉语声调教学提供了丰富的营养。然而，由于电子期刊的版权限制和学术论文的专业性，这些研究成果很难为非专业人士（对外汉语教师和汉语学习者）直接服务。本文用浅显易懂的语言综述了汉语声调感知方面的一些重要成果，并借此希望回答以下五个问题：1）为什么声调的感知难度因人而异？2）为什么某些声调比其他声调更难感知？3）为什么成词声调比不成词声调更难感知？4）为什么句子中的声调比词语中的声调更难感知？5）降低声调感知难度的教学策略有哪些？

INTRODUCTION

Mandarin Chinese has a reputation as a hard language for native English speakers to learn. The US Foreign Service Institute classifies it among a select group of "super-hard languages."[1] Among the challenges often noted by learners and teachers, tones and characters stand out. However, unlike characters—which most people can learn given enough time and effort—tones can remain a challenge despite the long and dedicated labors of learners. Learners have lamented the difficulty of Chinese tones since at least the time of Mateo Ricci, and modern researchers have taken up that lament beginning with the first published experiment of second language (L2) Chinese tone perception by Kiriloff (1969), who noted, "To speakers of atonal languages the very concept of tone

[1] https://www.state.gov/m/fsi/sls/c78549.htm

differentiated syllables (tonemes) is puzzling." Moser (1991), tongue-in-cheek, puts it more succinctly: "tonal languages are weird."[2]

However, even if everyone can more or less agree that tones are difficult, the concept of 'difficulty' is not so straightforward. Who are tones difficult for? Everyone, or only learners with no tones in their native language? What is it that is difficult about tones? Hearing them, remembering them, noticing them in meaningful speech, or all of the above? Are tones always difficult, or are they only difficult in long, multi-syllable words or in sentences?

Many of these questions have been the focus of previous tone research. There are now well over a hundred tone perception studies, using techniques ranging from paper and pencil tests to neuroimaging. The goal of the present review is to distill some of the major insights from tone perception research in a way that is easy to understand for non-specialists. I will focus primarily on studies that deal with Chinese (mostly Mandarin), which are summarized in Tables A1 and A2, in the Appendix. On the basis of this body of research, I will try to provide answers for the following questions about L2 tone difficulty.

Why are tones more difficult for some learners than others?
Why are some tones more difficult than others?
Why are tones in words more difficult than in meaningless syllables?
Why are tones in context more difficult than in isolation?
What can we do about tone learning difficulties?

Before I address these questions, I will first provide a general description of Mandarin tones to help set the scene.

A DESCRIPTION OF MANDARIN TONES

Modern standard Mandarin Chinese (*Putonghua*) has four lexically contrastive tones that are differentiated by their pitch height (low, high) and contour (rising, falling, or dipping). By convention, the four tones are labeled with numbers, and most learners and teachers talk about the tones using these numbers. The first tone (T1) is a *high-level* tone. The second tone (T2) is a *rising* tone. The third tone (T3) is realized as a *low-falling* tone in most contextualized speech, but also can occur as a *low-dipping* tone in isolation or at the end of a phrase. The fourth tone (T4) is a *falling* tone. Figure 1 depicts these tones visually, in isolated and contextualized forms.

Along with their distinctive pitch patterns, tones can also differ from one another with respect to characteristics such as duration, voice quality (clear or creaky), and loudness. But among these, pitch appears to be the most noticeable characteristic for listeners, and is generally the main focus of tone research, as well as pedagogical descriptions.

In addition to the four citation tones, Mandarin has a so-called 'neutral tone' that occurs on unstressed syllables. The neutral tone gets its pitch height and shape largely from the tone of the preceding syllable (W.-S. Lee & Zee, 2014). While neutral tones are a topic of increasing interest in teaching discussions (e.g., Třísková, 2017, Sparvoli, 2017), I know of no extant experimental research examining the neutral tone in L2

---

[2] As Moser himself no doubt knows, among the totality of the world's languages, tone languages are not weird. They make up a large portion of the world's documented languages (cf. Maddieson, 2013; Yip, 2002), and are spoken by a tremendous portion of the world's population.

Mandarin, so it will not be a major part of discussion below. This should not be taken as evidence that it is unimportant, simply that researchers have yet to examine it experimentally.

While the above introduction to tones is fairly standard, it is insufficient in the context of the current review. When non-native speakers—especially those from non-tonal native languages—learn Mandarin, they not only need to learn the differences between the four tones, they also have to be able to use the tones linguistically to *recognize words*. In other words, to do full justice to the challenge of L2 Mandarin tone learning, we need to distinguish between learning *tone categories* and learning *tone words.* While these two objects of learning are clearly related, they are also separable. It is possible to learn tone categories well enough to accurately identify them without ever learning a single word of Mandarin—something that often happens in the laboratory experiments reviewed below. Similarly, it is possible to learn Mandarin words without accurately learning tone categories—a phenomenon observable in many beginning learners who can write the consonant and vowel letters for the Pinyin (romanization) of Chinese words, but are unable to provide the tone diacritics (¯ ˊ ˇ ˋ). We might consider this latter case to be *incomplete* learning, but it is nevertheless a real phenomenon that captures a possible stage in the L2 learning of any given Chinese word. Given that the outcome most learners and teachers really care about is word recognition, it is not enough to focus just on tone category learning.

It will be important to keep the distinction between tone categories and tone words in mind as we look to answer the questions below. In many cases, research has largely stopped at the stage of tone category learning, so we must be careful in drawing conclusions about what results mean for tone word learning.

With these thoughts in mind, we now turn our attention to questions of L2 tone difficulty.

WHY ARE TONES MORE DIFFICULT FOR SOME LEARNERS THAN FOR OTHERS?

Hearing linguistic pitch is not necessarily something we would expect to be difficult for humans. All languages use pitch to some extent, whether to express stress, intonation, or emotion, so it makes sense that people with normal, healthy hearing would be able to perceive pitch changes in speech. Nevertheless, the experience of both teachers and learners makes it clear that some people are better at this than others. Understanding the source or sources of the differences in people's pitch perception abilities has been a major goal of tone research. Among the factors explored are *linguistic experience*, *musical experience*, *pitch perception aptitude*, and *L2 proficiency*. We will consider each

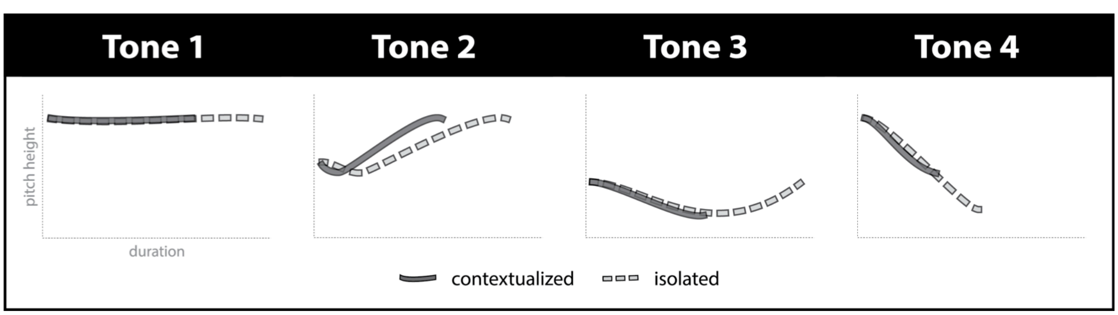

Figure 1. The four tones of Mandarin Chinese.

of these in turn, keeping in mind that, as I suggested above, most of these questions have been asked primarily for tone category learning, with only limited exploration of tone word learning.

Linguistic Experience

A person's *native language* experience is known to have a major impact on how they perceive sounds in another language. Because speakers of lexical tone languages have had intensive exposure to tones, we might suspect that they would have an advantage over non-tonal speakers when perceiving tone in a *new* tonal language. Extensive research generally bears this out (e.g., Bent, Bradlow, & Wright, 2006; Burnham et al., 2015; Chang, Yao, & Huang, 2017; Gandour, 1983; Hallé, Chang, & Best, 2004; So & Best, 2014 and *many* more). However, the tonal speaker's advantage is not always straightforward. Depending on the nature of the tone categories in a person's first language, there may be unique and stubborn patterns of confusion due to similarities or differences between those tones and the tones of the new language (e.g., Hao, 2012; Peng et al., 2010).

Even between non-tonal language speakers, there can be differences in how tones are initially perceived. Languages use pitch in a variety of ways, not just for lexical tones. Schaefer & Darcy (2014) examined completely inexperienced listeners' perception of Thai tones in people from languages thought to have varying degrees of "lexically contrastive pitch usage." This included Mandarin (lexical tone), Japanese (lexical pitch accent), English (lexical stress), and Korean (no lexically contrastive pitch). Their results strongly suggest that the role of pitch in a person's native language impacts the accuracy of their lexical tone perception, with Mandarin listeners having the highest accuracy and Korean listeners the lowest (for studies asking similar questions, but targeting Mandarin, see Braun, Galts, & Kabak, 2014; Braun & Johnson, 2011).

People who have previously learned a *second language* also appear to have an advantage when learning tone categories in a new language (Potter, Wang, & Saffran, 2016; Qin & Jongman, 2015; Wiener, Ito, & Speer, 2016). While we would expect that this advantage is specific to previous experience with a *tonal* L2, this has not been clearly established yet. At the same time, a number of studies have demonstrated rather convincingly that even rather limited experience with a tone language can cause changes in the way the brain processes linguistic tones (R. Lee, Hsu, Lin, Wu, & Tzeng, 2017; P. Li, Legault, & Litcofsky, 2014; Y. Wang, Sereno, Jongman, & Hirsch, 2003; J. Yang, Gates, Molenaar, & Li, 2015). It makes sense that these changes would then give the L2 learner an advantage for learning new tone categories.

In a similar vein, it also seems reasonable to expect that childhood experience with a tonal language—as in the case of so-called *heritage learners*—would give individuals an advantage in perceiving tones as adults. However, research along these lines is so far rather limited and results are a bit hard to interpret (cf. Tsukada, Xu, & Rattanasone, 2015).

Musical experience

Another major line of research examines the relationship between *musical experience* and linguistic perception. Many studies have shown a link between musical expertise and sensitivity to linguistic tones, with musicians outperforming non-musicians in tone identification accuracy (e.g., Alexander, Wong, & Bradlow, 2005; Cooper & Wang, 2012; Gottfried, 2007; C.-Y. Lee & Hung, 2008; M. Li & DeKeyser, 2017), as well as in neural measures of sensitivity to linguistic pitch (e.g., Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Wong, Skoe, Russo, Dees, & Kraus, 2007). This seems to confirm the impression Chinese teachers have had for many years. However, one caution in interpreting the results is that a connection between previous musical experience and heightened sensitivity to linguistic tone does not necessarily mean musical experience *causes* better tone perception. It may also be the case that people with heightened sensitivity to pitch naturally gravitate to music. In other words, the underlying cause for the relationship between music and sensitivity to tones might not be musical experience itself, but rather a biological aptitude for pitch (cf. Bowles, Chang, & Karuzis, 2016).

Tone aptitude

This brings us to another major theme of L2 tone research, namely, *tone aptitude*. Tone aptitude is typically thought to benefit people in two ways. First, people with higher levels of tone aptitude should ultimately be able to have more successful tone learning outcomes than those with less aptitude. Second, they should learn tones faster (Bowles et al., 2016). Several studies have revealed exactly these patterns, that is, both better and faster learning on the part of people with higher aptitude (Bowles et al., 2016; M. Li & DeKeyser, 2017; Perrachione, Lee, Ha, & Wong, 2011; Wong & Perrachione, 2007). In these studies, tone aptitude is understood as an individual ability to accurately perceive changes in pitch contour (Chandrasekaran, Sampath, & Wong, 2010; Gandour, 1983; Wong & Perrachione, 2007). In other words, learners who can pay attention to the shape of tones, rather than just differences in pitch height, appear to learn faster and obtain better overall accuracy. Interestingly, several aptitude studies have had measures of both musical ability and pitch aptitude. In each case, results suggested that positive tone learning outcomes were more strongly related to the pitch aptitude measures than to musical experience (for a thorough review of these issues, see Bowles et al., 2016).

L2 Proficiency

Though it may seem quite obvious, L2 proficiency is also consistently related to differences in the difficulties learners have in tone perception (e.g., C.-Y. Lee, Tao, & Bond, 2009, 2010a; L. Zhang, 2011; Zou, Chen, & Caspers, 2016). In most cases, proficiency is estimated simply by considering how long learners have been studying. Reassuringly for teachers, this tends to show the expected relationship, with learners in later years of study showing superior performance to those in earlier years—though occasionally, much like in actual classrooms, this does not quite work out (C.-Y. Lee, Tao, & Bond, 2010b). Importantly, results suggest not only that more proficient learners are more accurate in tone identification, but also that they are faster, and may undergo some qualitative changes in the way they attend to tones (cf. Zou et al., 2016). Critically, however, proficiency results so far have almost always been related to measures of tone category learning. This may give us an incomplete picture of learning outcomes. For example, in Pelzl, Lau, Guo, & DeKeyser (2018), we found advanced L2 learners to have rather extreme difficulties in a word recognition task that forced them to rely on tones. We will come back to this a bit later when we consider the challenges presented by tone word recognition in more depth.

Although the findings reviewed so far point to several factors that may make tones more difficult for some learners than others, there are still some important outstanding questions. For instance, do these results—typically found in completely inexperienced listeners or after only a brief period of training—predict *long-term* tone learning outcomes? Are those with lower aptitude doomed to tone purgatory, or can they catch up given enough time and/or effort? Some of the other studies reviewed below will suggest answers.

3. WHY ARE SOME TONES HARDER THAN OTHERS?

If research has shown that there are differences between learners, it has also shown that there are differences between tones. That is, some tones appear to be harder to perceive than others.[3] In particular, a consistent finding across a wide variety of studies is that, at least in isolated syllables, T2 is the most difficult tone for learners to identify, typically followed closely by T3 (Hao, 2012; Kiriloff, 1969; C.-Y. Lee, Tao, & Bond, 2013; Pelzl et al., 2018; So & Best, 2010, 2014; Sun, 1998; B. Yang, 2012). At the same time, with respect to T3, studies have occasionally reported dramatically different results in which it appears to be the *easiest* tone (C. Chang & Bowles, 2015; Maddox & Chandrasekaran, 2014).

It is likely that some of the wide swings in results for T3 come from variations in the specific way it sounds in the different recorded stimuli used between studies. For one, there does not seem to be consistent control—at least it is not consistently reported—for so-called creakiness in T3. Creakiness can be described as a mildly noisy interruption in the flow of a speaker's voice, and tends to occur when the pitch of a voice is very low. If some T3 stimuli in a given study are creaky, this might make T3 easier to differentiate from other tones. Another possible difference might come from the duration of T3 in different studies. When produced in isolation, T3 is often noticeably longer than the other tones. For studies that do not account for the duration of tones, it may be that this aspect makes T3 easy to identify. The inconsistent qualities of T3 put researchers in a bit of pickle. Do we want to control duration and creakiness? If we do, we run the risk of creating *artificial* difficulties for tone identification that do not clearly relate to the actual tone categories. But if we do not, it may be hard to know how pitch itself impacts the ease or difficulty of different tones. There is no right answer, but the issues do need to be considered when evaluating tone perception outcomes, as they might make us come to very different conclusions about the difficulty of T3.

Most often tone difficulty is assumed to be caused by similarities between tones.[4] However, determining what is similar between tones is not necessarily straightforward. If we focus on pitch, possible points of similarity might include overall shape (e.g., high-level, rising), but could also be limited to more isolated features such as pitch onset height, offset height, or some perceived 'average' pitch height. If perceived similarity was mainly based on pitch onset, this might explain why T1 and T4, which both have high pitch onsets, are often found to be confusable (e.g., Hao, 2012; Y. Wang, Spence,

[3] For related discussions of difficulty in tone *production*, see e.g., B. Yang (2012), C. Yang (2016), and H. Zhang (2016).

[4] But see H. Zhang (2016) for interesting arguments about possibly *universal* properties of tones that might cause difficulties.

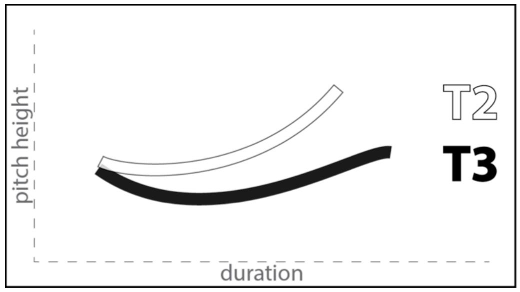

Figure 2. Similarity of shape of tones 2 and 3.

Jongman, & Sereno, 1999), whereas T1 and T2 appear less confusable. However, this is not an entirely satisfactory explanation, since in many cases we find that T1 and T2 are also confused, just less often than T1 and T4. In fact, several studies suggest that the answer lies not in the tones, but in the learners (see especially Chandrasekaran et al., 2010; Maddox & Chandrasekaran, 2014). Those who attend to pitch height will experience different confusions than those who attend to pitch contour.

The confusability of T2 and T3 requires a bit more discussion. They are without a doubt the most commonly confused tones, for native speakers, as well as L2 learners. Still, the source of this confusion is not entirely obvious. It is certainly true that they can have a large degree of similarity in their contours—at least in isolated single syllables. In isolation, acoustic analysis often shows T2, just like T3, has a slight dip at its onset (Figure 2). Another important factor that might create the impression of similarity is T3 sandhi, which can change the pitch of T3 to a rising tone—just like T2—when it occurs before another T3. For native speakers, it has been argued that T3 sandhi creates a strong mental relation between T2 and T3 and thus causes greater difficulty in distinguishing them (cf. Huang & Johnson, 2010). If L2 learners achieve high proficiency in Mandarin tones, they might experience the same sandhi-induced difficulties that native listeners experience. Yet, another possible reason for some of the difficulty learners experience in accurately identifying T2 and T3 may come from the influence of common teaching practices. This issue has been discussed repeatedly by teachers and scholars (Shi, 2007; Sparvoli, 2017; H. Zhang, 2014). In this case, the concern is that learners will tend to think the main factor that defines T3 is its dipping contour. Since this feature is also often present in T2, learners will tend to misidentify T2 as T3 when T2 dips, or perhaps T3 as T2 if the tone's dip is less apparent than its rise. However, so far no experimental research has explicitly set out to tease apart the potential causes for L2 *perceptual* confusion of T2 and T3 (for results with production experiments see H. Zhang, 2014, and He, Wang, & Wayland, 2016).[5]

4. WHY ARE TONES HARDER IN WORDS?

Compared to perception of tone categories, tone word recognition adds additional layers of potential challenge—particularly for non-tonal language speakers. Success in tone word recognition obviously requires the ability to perceive pitch differences and to

---

[5] One possible exception is a small study by Lin (1985), but results should be interpreted with caution due to many shortcomings by modern standards for empirical research (e.g., small number of participants, small number of stimuli, no control group, and no inferential statistics to support interpretations of quantitative results).

associate these differences with tone categories, as just reviewed. However, tone word recognition has a second important component. The listener needs to be able to encode the tone with a word's mental representation in long-term memory so that the tone can contribute to successfully recognizing that word when it is heard again in the future. This ability cannot be taken for granted. Even if a person successfully learns tone categories, they might well ignore them when learning new words, or they might attend to them in the moment, but be unable to store them in long-term memory for future use. Alternatively, a person may successfully memorize tone-to-word relationships when studying or in class, but still ignore the tones in real-life situations when listening to someone speak Mandarin. It is important to point out that in context it is often possible for someone to recognize a word *even though they do not recognize or know the tones*. But while the listener can 'cheat' and successfully recognize some—or even most—words without attending to tones, this will always be a disadvantage when the tone becomes essential for word recognition, such as when words are spoken in isolation, or when similar sounding words can only be differentiated by tones.

There are many possible states for tone word knowledge in a learner's head. We can take a simple example like the word gǒu (狗) 'dog'. A learner's knowledge of this word could have any of the following states: it may encode the *correct tone category* (gǒu), an *incorrect tone category* (gòu), a *non-existent tone category* (gôu), or *no tone category* at all (gou). The learner may be *aware* or *unaware* of their tone knowledge for this word, and may be *certain* of the knowledge (perhaps mistakenly), or very *uncertain*. Nevertheless, even if a learner's mental version of this word is in some sense incomplete, it is unlikely to cause regular difficulties in comprehension (though production might be another story).

All of this illustrates that tone word learning could conceivably be a very different sort of problem than tone category learning. This idea—that learning of tones is more than simply learning pitch categories—has been assumed in a large number of training studies carried out in the past decade (e.g., Alexander et al., 2005; Chandrasekaran et al., 2010; Cooper & Wang, 2013; Ingvalson et al., 2013; Perrachione et al., 2011; Wong & Perrachione, 2007). An illustrative example is a study by Wong and Perrachione (2007). They created a small artificial language with words that imitated three of the four Mandarin tones (T1, T2, and T4). Each word contrasted with two others such that only the tones could be used to differentiate them (e.g., the imaginary words: pēsh 'glass'**,** pésh 'pencil', and pèsh 'table'). People with no previous tone learning experience were given training to associate the words with pictures and then tested to see whether they could recognize the tone words, or if they would confuse one tone word with another. In other words, people in this study were not trained to recognize *tone categories* (T1, T2, T4), but to recognize *tone words*. This study, and others like it, show that some of the individual learner characteristics that predict tone category learning (musical ability, pitch aptitude) also predict tone word learning (Alexander et al., 2005; Bowles et al., 2016; Perrachione et al., 2011; Wong & Perrachione, 2007). Interestingly, a couple studies suggest that when non-musicians or learners with low pitch perception aptitude are first trained on tone category identification and then trained to learn tone words, they may be able to 'catch up' to musicians and high aptitude learners (Cooper & Wang, 2013; Ingvalson et al., 2013).

The findings from tone word training studies have been largely positive, suggesting that most people will be able to learn to recognize tone words given enough time and if appropriate steps are taken to address their individual aptitudes (on this latter point, see especially Perrachione et al., 2011). However, there are some limitations in the extent to which these results can be directly applied to real world tone learning. As noted above, these studies used small, artificial vocabularies—most often learners are trained on only 20-30 words, all nouns, and all with tones contrasting neatly with other words in the vocabulary. This inflates the information value of tones in the word learning task, and also simplifies the way tones occur in actual Mandarin.

In the real Mandarin language, words present people with a wide variety of learning challenges. For one, while the set of tone categories that have to be learned is small, the number of words to learn has no obvious limit. Words are also diverse in the way they sound. They vary in their length, both in terms of how long a single syllable is (xi vs. xiong), and in terms of how many syllables are in a word. While the most frequent words in Mandarin tend to be one-syllable long (i.e., monosyllabic), overall Mandarin words tend to be two-syllables long (i.e., disyllabic) (Duanmu, 2007). The number of syllables in a word may have dramatic effects on the ease or difficulty of learning it. While there are only four possible configurations for tones on a monosyllable, there are four-times-four configurations for disyllabic words (T1T1, T1T2, T1T3, T1T4, T2T1, T2T2, etc.)—and another four when neutral tones are included. Depending on how we believe tones are encoded in a learner's memory for a word, these different statistical properties could have major impacts on learning. In short, a full understanding of tone word learning has to be intimately connected to an understanding of the properties of Mandarin words.

Some recent work has begun to address these characteristics more fully. Chang & Bowles (2015) conducted a training study including both monosyllabic and disyllabic tone words. They found that the disyllabic words were much more challenging for their inexperienced learners. In a recent study (Pelzl et al., 2018), we tested advanced L2 learners who had achieved general fluency in Mandarin. These learners heard 120 items, 60 were real words, 30 were nonwords that differed from real words in the rhyme (what we often call the 'final' in Mandarin) of the first syllable (fángzi 'house' vs. féngzi), and 30 were nonwords that had an incorrect tone (fàngzi). We found that while these learners were able to reject nonwords using the rhyme cues (féngzi), they appeared to struggle greatly to reject nonwords on the basis of tone cues (fàngzi). This was rather surprising because these same learners were near-native in a challenging tone identification task for monosyllables, and they generally provided correct tones and definitions on a vocabulary test we administered after the experiment.

Another recent study (Wiener, Ito, & Speer, 2018) used an artificial language that closely imitated the distributional characteristics of Mandarin syllables and tones. Although all the words in the experiment were monosyllabic, the ways in which they contrasted with one another were not as simple as in previous studies. Instead, the probability of a given tone occurring with a given syllable varied for each syllable. So, for example, some words had many tone neighbors, just like the syllable *yi* in Mandarin (yī, yí, yǐ, yì all exist), while others hand few or none, like the syllable *gei* in Mandarin (in isolation, only gĕi exists, not gēi, géi, or gèi). Also, just like Mandarin, some words had many homophones (e.g., in imitation of syllables like *yi*). This allowed Wiener and his colleagues to test the different effects of syllable frequencies and the probability of a given tone occurring with a given syllable. They found that learners with previous Mandarin experience were able to learn the probabilistic relationships between syllables and tones, so that when low frequency syllable+tone combinations occurred, learners would expect the most common word with that combination. In other words, the results of their training study suggest that learning of tones is intricately linked to the probabilities of syllable+tone combinations occurring in Mandarin words overall. This type of learning is typically called 'statistical learning' and is known to be an important part of language learning both for children and adults.

To summarize, while there are still many gaps in our understanding of tone word learning, especially when the influence of a complex lexicon is considered, research in this area is expanding quickly. This is good news for teachers and learners, as few would think that simply identifying tones on single syllables is sufficient for making real progress in learning Mandarin. On the other hand, as we will consider next, even the complexities discussed so far fall dramatically short of the full range of issues that contribute to the difficulty of L2 tone learning.

5. WHY ARE TONES HARDER IN CONTEXT?

In this last section I wish to address a variety of complications that natural spoken language presents to L2 learners. To be useful, L2 tone word recognition needs to happen for speech at its natural speed in fully complex natural contexts. While the simple citation forms of tones may be a useful first step, ultimately learners need to deal with *contextual tone changes*, *speaker variation*, and all the *noise* that comes with natural speech. Research in these areas is overall rather limited, and we will have to be cautious in drawing conclusions based on just a few studies. Nevertheless, these issues are important in pushing L2 tone research towards a fuller understanding of the natural complexity of Mandarin speech, and such research is likely to shed light on why good performance on tone word recognition in laboratories and classrooms may not always scale up to real life situations.

Contextual tone changes

Outside of experiments, tones rarely occur in complete isolation. They occur in words, which are typically longer than a single syllable, and those words occur in sentences, which occur in discourse, spoken by people with emotions who also sometimes make speech errors—including, of course, tone errors (cf. Wan & Jaeger, 1998). So, in practice, multiple layers of contextual effects can influence the realization of any tone. If we consider the simplest contextual case of two syllables, even here we find that tone contours can undergo considerable change. The clearest examples are when a tone with a high or low offset precedes a tone with the opposite onset (e.g., a falling T4 ends low, and a level T1 starts high), in this case, the shape of both tones could be strongly influenced. Research with native speakers (Xu, 1997) and some work with L2 learners (C. Yang, 2016) has begun to investigate this type of circumstance. An important point to note is that even native listeners have a drop in accuracy in such cases.

Though not always examining contextual tone changes in detail, a number of studies have consistently found that disyllabic tone identification is more difficult than monosyllabic tone identification, and that initial syllables cause more difficulty overall than final syllables (Broselow, Hurtig, & Ringen, 1987; C. B. Chang & Bowles, 2015; Hao, 2012, 2018; Sun, 1998). These studies suggest some interesting interactions involved in perceptual confusions that seem to relate to the use of pitch in a listener's native language. For example, English speakers seem to naturally expect a falling tone on a final syllable, and consequently are happy to identify T4 in word final position. I will not test the reader's patience by recounting all of the specific patterns of individual positional tone errors here. Interested readers can check out the studies above.

Speaker Variation

Yet another potential challenge for L2 listeners is the great variability found in speech. We often think of the great variety of accents—including tones—found among native Chinese whose Mandarin may be strongly influenced by a local dialect. But even among speakers from a fairly homogeneous language region, variability is the norm. No two speakers' tones are the same, and even a single speaker's tones will vary across contexts. How well L2 listeners are able to deal with this natural variability is an important question and, to date, some of the most practical takeaways from research are related to such questions.

In training studies it has long been the assumption that having multiple speakers will help learners form more accurate and flexible mental tone categories by forcing them to focus on what is generalizable about tones across speakers, rather than the idiosyncratic features of any one speaker's tones (Y. Wang et al., 1999). For this reason most tone training studies use multiple speakers (typically two male, two female) and these voices are mixed together across the training. Some research strongly suggests that there is a kind of trade off involved in learning outcomes for high vs. low variability training (cf. Perrachione, Lee, Ha, & Wong, 2011). If learners are exposed to just a single speaker, they seem to make faster initial progress; however, their ability to recognize tones produced by an unfamiliar speaker may be rather weak. In contrast, when training includes multiple speakers, overall progress may be slower, but learners will be stronger in their abilities to recognize tone words spoken by unfamiliar voices.[6]

Importantly, the variability of training materials appears to have different impacts on learners according to their pitch perception aptitudes. Whereas high aptitude learners might excel in high variability training and achieve better overall learning outcomes, low aptitude learners are likely to struggle and may make significantly less progress than if they were trained with a single speaker (C. Chang & Bowles, 2015; Perrachione et al., 2011; Sadakata & McQueen, 2014). Perrachione et al. (2011) seemed to find a nice compromise. They used training sessions where any block of material was spoken by a single speaker, but across blocks, new speakers were introduced. This produced better outcomes for low aptitude learners, without significantly slowing down the high aptitude learners.

All of the above results occurred in laboratory conditions with computer training, so it is not yet clear how well they generalize to classroom settings. However, it makes sense for teachers to be aware of both the benefits of exposing learners to wider speaker variability, along with cautions to not bombard learners with a great mix of voices all at

---

[6] For a somewhat different type of trade-off in the context of statistical learning, see Wiener, Ito, & Speer (2018).

once. It seems likely that most classrooms are guilty of having too little speaker variability, rather than too much.

In addition to the sort of general speaker variability described above, social contexts can also lead to variability. Speakers in formal situations tend to speak more carefully and clearly—but at the same time also use more formal vocabulary and structures. In the case of Mandarin, this means the use of words that often come from classical Chinese. In such circumstances, it is possible that tones could play a more important role for listeners than when speech is more routine and predictable. On the opposite extreme, very casual speech may have an accelerated speech rate, less precise pronunciation, and more colloquial vocabulary and grammar (once again, likely to be less predictable for learners). In casual speech, speakers may also tend to be more emotional, with large shifts in overall intonation also impacting tones. While many researchers are aware of such factors, at present I am unaware of experimental research that tries to understand how well or poorly L2 Mandarin listeners do under such circumstances and what role tones play in their success or failure—though C. Yang (2016) and C. Yang & Chan (2010) do address some of the contextual influences that may impact tones in sentences (namely, question and statement intonation).

Noise

A final type of context to consider is *noise*. Unlike the gentle quiet that characterizes many research settings, Mandarin speech perception typically occurs with some level of noise competing for listeners' auditory attention. This, too, is an area that has received scant attention, despite its high relevance for L2 listeners. Lee and colleagues carried out two studies (C.-Y. Lee et al., 2010b, 2013) examining the effects of noise on native Chinese and L2 learners' abilities to identify tones on single syllables (i.e., without concern for word meanings). While they found that L2 learners were less accurate than native Chinese listeners overall, the impact of noise on identification was somewhat inconsistent in the two studies. Still, overall results seem to suggest that noise is in fact more problematic for L2 listeners than native speakers (for detailed discussion, see C.-Y. Lee et al., 2013). This area is certainly worthy of more attention, especially for meaningful speech comprehension, as some of the most common and important Chinese social occasions (i.e., meals) often occur in noisy settings such as restaurants or cafeterias.

6. WHAT CAN WE DO ABOUT TONE LEARNING DIFFICULTIES?

While scientific study does not always result in practical applications, in many cases, researchers engaged in tone perception work are interested in improving teaching and training outcomes. In this final section, then, I would like to address some of the more practical implications that have come out of tone perception research.

One clear finding from a large number of studies is that targeted training with tone identification can improve a listener's ability to learn tones—at least in the short term. This is most often done with computers, or other automated methods, as a supplement to more typical classroom activities (Liu et al., 2011; X. Wang, 2013; Y. Wang et al., 1999).[7] Computer assisted tone training programs have been available for

[7] Perhaps surprisingly, few studies have explored tone training methods designed for teacher-student or student-student interactions in a classroom (cf. Saito & Wu, 2014 for one exception, with Cantonese learners of Mandarin).

many years, but not always easy to find. With modern smart phones, many apps of this sort exist, and interested teachers and learners will be able to quickly find such resources. As suggested above, one mild caution is that long-term outcomes for this type of tone training remain uncertain.

To date, perhaps the studies of most interest to teachers are those showing the wide variety of techniques that appear to be helpful in tone training. First, several studies have shown that something as simple as the diacritic tone marks of Pinyin can help students recognize tones more accurately (Godfroid, Lin, & Ryu, 2017; Showalter & Hayes-Harb, 2013). More generally speaking, any iconic graphical cues for the tones appear to be helpful (Liu et al., 2011; X. Wang, 2013). Similarly, gestures imitating tone contours can also be helpful (Eng, Hannah, Leong, & Wang, 2013; Morett & Chang, 2015). Most of these results were obtained in carefully controlled laboratory settings with beginning level learners, and it is not always the case that they will work out as well in classrooms, or with different groups of learners. Still, for teachers who have not used them, diacritics, images, and gestures would be easy to test out in their own classrooms. For the many instructors who have long made use of such things, these results can be modest encouragement that they are likely on the right track.

Perhaps a bit more intriguing are methods that attempt to encode tones in writing without the use of diacritics, that is, in ways that force learners to know the tone if they want to write a word. Perhaps the most well-known example is Y.R. Chao's tonal spelling system, Gwoyeu Romatzyh (GR) (Chao, 1968). In GR, if you do not know a word's tone, you cannot properly spell the word. However, compared to Pinyin, GR is a much more complex system and requires considerable initial effort on the part of learners. More recently, it has been suggested that a similar effect might be obtained by using colors to encode the tones (Dummitt, 2008). Again, the idea is that if you do not know the tone, you cannot choose the appropriate color. Despite their ingenuity, so far neither of these approaches have found much support from experimental research. McGinnis (1997) conducted a classroom-based longitudinal study comparing the effects of learning with Pinyin and GR. He found that Pinyin was related to slightly better tone outcomes. A more recent study examined the use of colors in a training study, contrasting them with the use of numbers and iconic tone symbols (Godfroid et al., 2017). Despite some enthusiasm on the part of learners, who liked the use of colors, results showed that the best outcome was for tone symbols, and colors did not even outperform simple tone numbers. Future research may yet show the benefit of these clever encoding systems, but the limited effects found so far suggest it may not be worth the extra effort involved for learners or teachers.

One type of practical tone perception study that is striking for its absence is research examining corrective feedback for tones. Considering how much effort teachers put into thinking about how best to correct tone errors, and how often they do correct them, studies examining whether feedback has an impact on tone perception (and production) would appear to have major practical applications for Chinese teachers. We can hope that such studies will start to appear in the near future.

7. CONCLUSION

As I hope this review has made clear, there is now a sizable body of research on tone perception by non-native listeners. This research demonstrates that there are many difficulties involved with tone category and tone word learning. Some of these challenges may be short-lived, while others may persist for all but the most accomplished of L2 learners. On the other hand, all of these results have to be placed in the wider context of Mandarin speech learning. Generations of L2 learners have had deeply meaningful conversations with native Chinese speakers, despite some weaknesses in tone perception. Tones are important, but they are not the only, or even the most important, aspect of learning to comprehend Mandarin speech. It is my hope that future research will be able to help us pinpoint areas where tones are essential for L2 learners, and will suggest useful techniques to help learners overcome tone learning challenges in the most efficient and effective way possible.

REFERENCES

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Interspeech* (pp. 397–400). Retrieved from http://groups.linguistics.northwestern.edu/speech_comm_group/publications/2005/Alexander-Wong-Bradlow-2005.pdf

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

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

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, *66*(4), 774–808. https://doi.org/10.1111/lang.12159

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

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*, 585–594.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage Phonology: The Acquisition of a Second Language Sound System* (pp. 350–364). New York: Newbury House Publishers.

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

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influence by language experience. *Brain Research*, *1128*, 148–156.

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.

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *Journal of the Acoustical Society of America*, *136*(6), 3703–3716.

Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126. https://doi.org/10.1121/1.4976037

Chao, Y. R. (1968). *Grammar of spoken Chinese*. Berkeley, CA: Unversity of California Press.

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences*, *48*, 62–69. https://doi.org/10.1016/j.langsci.2014.12.002

Crinion, J. T., Green, D. W., Chung, R., Ali, N., Grogan, A., Price, G. R., … Price, C. J. (2009). Neuroanatomical markers of speaking Chinese. *Human Brain Mapping*, *30*(12), 4108–4115.

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.

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *The Journal of the Acoustical Society of America*, *134*(2), EL133–EL139.

Duanmu, S. (2007). *The Phonology of standard Chinese* (2nd edition). New York, New York: Oxford University Press.

Dummitt, N. (2008). *Chinese through tone & color.* New York, New York: Hippocrene Books.

Eng, K., Hannah, B., Leong, L., & Wang, Y. (2013). Can co-speech hand gestures facilitate learning of non-native tones? In *Proceedings of Meetings on Acoustics ICA2013* (Vol. 19, p. 060225). ASA. Retrieved from http://asa.scitation.org/doi/abs/10.1121/1.4799746

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*, 268–294.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Godfroid, A., Lin, C.-H., & Ryu, C. (2017). Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training. *Language Learning*. https://doi.org/10.1111/lang.12246

Gottfried, T. L. (2007). Music and language learning: Effects of musical training on learning L2 speech contrasts. In O.-S. Bohn & M. J. Munro (Eds.), *Language*

Experience in Second Language Speech Learning (pp. 221–237). Amsterdam: John Benjamins.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*(2), 207–231.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 57-77). Amsterdam: John Benjamins.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*, 395–421.

Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279. https://doi.org/10.1016/j.wocn.2011.11.001

Hao, Y.-C. (2017). Second Language Perception of Mandarin Vowels and Tones. *Language and Speech*, 0023830917717759.

Hao, Y.-C. (2018). Contextual effect in second language perception and production of Mandarin tones. *Speech Communication*, *97*, 32–42. https://doi.org/10.1016/j.specom.2017.12.015

Hao, Y.-C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, *54*, 151–168. https://doi.org/10.1016/j.wocn.2015.10.003

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. *Phonetica*, *67*(4), 243–267. https://doi.org/10.1159/000327392

He, Y., Wang, Q., & Wayland, R. (2016). Effects of different teaching methods on the production of Mandarin tone 3 by English speaking learners. *Chinese as a Second Language*, *51*(3), 252–265.

He, Y., & Wayland, R. (2013). Identification of Mandarin coarticulated tones by inexperienced and experienced learners of Mandarin. *Chinese as a Second Language Research*, *2*(1), 1–21.

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. *Phonetica*, *67*(4), 243–267. https://doi.org/10.1159/000327392

Ingvalson, E. M., Barr, A. M., & Wong, P. C. M. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech Language and Hearing Research*, *56*(3), 1045. https://doi.org/10.1044/1092-4388(2012/12-0024)

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*, 63–67.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A Cross-Linguistic PET Study of Tone Perception in Mandarin Chinese and English Speakers. *NeuroImage*, *13*, 646–653.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Research*, *1313*, 124–133. https://doi.org/10.1016/j.brainres.2009.11.061

Krishnan, A., Gandour, J. T., Smalt, C. J., & Bidelman, G. M. (2010). Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. *Brain and Language*, *114*(3), 193–198. https://doi.org/10.1016/j.bandl.2010.05.004

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In J. Leather & A. James (Eds.), *Sound Patterns in Second Langauge Acquisition* (pp. 59–80). Providence: Foris.

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235. https://doi.org/10.1121/1.2990713

Lee, C.-Y., Tao, L., & Bond, Z. S. (2009). Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics*, *37*(1), 1–15. https://doi.org/10.1016/j.wocn.2008.08.001

Lee, C.-Y., Tao, L., & Bond, Z. S. (2010a). Identification of acoustically modified Mandarin tones by non-native listeners. *Language and Speech*, *53*(2), 217–243. https://doi.org/10.1177/0023830909357160

Lee, C.-Y., Tao, L., & Bond, Z. S. (2010b). Identification of multi-speaker Mandarin tones in noise by native and non-native listeners. *Speech Communication*, *52*(11–12), 900–910. https://doi.org/10.1016/j.specom.2010.01.004

Lee, C.-Y., Tao, L., & Bond, Z. S. (2013). Effects of speaker variability and noise on Mandarin tone identification by native and non-native listeners. *Speech, Language and Hearing*, *16*(1), 46–54. https://doi.org/10.1179/2050571X12Z.0000000003

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, *53*(2), 157–165.

Lee, R. R.-W., Hsu, C.-H., Lin, S.-K., Wu, D. H., & Tzeng, O. J.-L. (2017). Learning transforms functional organization for Mandarin lexical tone discrimination in the brain: Evidence from a MEG experiment on second language learning. *Journal of Neurolinguistics*, *42*, 124–139. https://doi.org/10.1016/j.jneuroling.2016.12.004

Lee, W.-S., & Zee, E. (2014). Chinese phonetics. In C.-T. J. Huang, Y.-H. A. Li, & A. Simpson (Eds.), *The handbook of Chinese linguistics* (pp. 369–399). Malden, MA: Wiley Blackwell.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. *Journal of Pscyholinguistic Research*, *25*(5), 527–542.

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, 1–28. https://doi.org/10.1017/S0272263116000358

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, *58*, 301–324. https://doi.org/10.1016/j.cortex.2014.05.001

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

Liu, C. (2013). Just noticeable difference of tone pitch contour change for English-and Chinese-native listeners. *The Journal of the Acoustical Society of America*, *134*(4), 3011–3020.

Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., & MacWhinney, B. (2011). Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning*, *61*(4), 1119–1141. https://doi.org/10.1111/j.1467-9922.2011.00673.x

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research*, *1624*, 28–44. https://doi.org/10.1016/j.brainres.2015.07.014

Maddieson, I. (2013). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Plank Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/13

Maddox, W. T., & Chandrasekaran, B. (2014). Tests of a dual-system model of speech category learning. *Bilingualism: Language and Cognition*, *17*(04), 709–728. https://doi.org/10.1017/S1366728913000783

Maddox, W. T., Koslov, S., Yi, H.-G., & Chandrasekaran, B. (2015). Performance pressure enhances speech learning. *Applied Psycholinguistics*, 1–28. https://doi.org/10.1017/S0142716415000600

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715.

McGinnis, S. (1997). Tonal spelling versus diacritics for teaching pronunciation of Mandarin Chinese. *The Modern Language Journal*, *81*, 228–236.

Morett, L. M., & Chang, L.-Y. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*(3), 347–353. https://doi.org/10.1080/23273798.2014.923105

Moser, D. (1991). Why Chinese is so damn hard. *Sino-Platonic Papers*, *27*, 59–70.

Ning, L.-H., Loucks, T. M., & Shih, C. (2015). The effects of language learning and vocal training on sensorimotor control of lexical tone. *Journal of Phonetics*. https://doi.org/10.1016/j.wocn.2014.12.003

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

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

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. *Journal of the Acoustical Society of America*, *130*(1), 461–472.

Potter, C. E., Wang, T., & Saffran, J. R. (2016). Second language experience facilitates statistical learning of novel linguistic materials. *Cognitive Science*. https://doi.org/10.1111/cogs.12473

Qi, Z., Han, M., Garel, K., San Chen, E., & Gabrieli, J. D. E. (2015). White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics*, *33*, 14–28. https://doi.org/10.1016/j.jneuroling.2014.08.004

Qin, Z., & Jongman, A. (2015). Does second language experience modulate perception of tones in a third language? *Language and Speech*, 0023830915590191.

Repp, B. H., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics*, *18*, 481–495.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01318

Saito, K., & Wu, X. (2014). Communicative focus on form and second language suprasegmental learning. *Studies in Second Language Acquisition*, *36*(04), 647–680. https://doi.org/10.1017/S0272263114000114

Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology*, *5*(4). https://doi.org/10.1515/lp-2014-0016

Shi, J. (2007). On teaching tone three in Mandarin. *Journal of the Chinese Language Teachers Association*, *42*(2), 1–10.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, *29*(2), 185–200.

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.

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*(02), 195–221. https://doi.org/10.1017/S0272263114000047

Sparvoli, C. (2017). From phonological studies to teaching Mandarin tone: Some perspectives on the revision of the tonal inventory. In I. Kecskes & Chaofen Sun (Eds.), *Key issues in Chinese as a second language research* (pp. 81–100). New York, NY: Routledge.

Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. Honolulu, HI: Second Language Teaching & Curriculum Center.

Třísková, H. (2017). De-stress in Mandarin: clitics, cliticoids, and phonetic chunks. In I. Kecskes & Chaofen Sun (Eds.), *Key Issues in Chinese as a Second Language Research* (pp. 29–56). New York, NY: Routledge.

Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research*, *4*(2). https://doi.org/10.1515/caslar-2015-0009

Wan, I.-P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, *15*(3), 417–461. https://doi.org/10.1017/S0952675799003668

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, *97*(1), 144–160.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic Perception of Mandarin Tones by Chinese and American Listeners. *Brain and Language*, *78*(3), 332–348. https://doi.org/10.1006/brln.2001.2474

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI Evidence for cortical modification during learning of Mandarin lexical tones. *Journal of Cognitive Neuroscience*, *15*(7), 1019–1027.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*, 3649.

Wang, Y., Zhang, Y., Cooper, A., & Dovan, M. (2011). Effects of training on the processing of speech and non-speech tone: an event-related potential study. *Proceedings of the Psycholinguistic Representation of Tone Conference*. Presented at the Hong Kong. Retrieved from https://akcooper.files.wordpress.com/2010/08/plrt2011-wang-zhang-cooper-dovan_revision_submission.pdf

Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech*, 002383091876176. https://doi.org/10.1177/0023830918761762

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(04), 565–585.

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*, 420–422. https://doi.org/10.1038/nn1872

Xu, Y. (1997). Contextual tonal variation in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Yang, B. (2012). The gap between the perception and production of tones by American learners of Mandarin – An intralingual perspective. *Chinese as a Second Language Research*, *1*(1), 33–53. https://doi.org/10.1515/caslar-2012-0003

Yang, C. (2016). *The Acquisition of L2 Mandarin Prosody: From experimental studies to pedagogical practice*. Philadelphia, PA: John Benjamins Publishing Co.

Yang, C., & Chan, M. K. M. (2010). The perception of Mandarin Chinese tones and intonation. *Journal of the Chinese Language Teachers Association*, *45*(1), 7–36.

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*, *33*, 29–49. https://doi.org/10.1016/j.jneuroling.2014.09.004

Yip, M. (2002). *Tone*. New York, New York: Cambridge University Press.

Zhang, L. (2011). Meiguo liuxuesheng Hanyu shengdiaode yinwei he shengxue xinxi jiagong give Chinese character title here, followed by English translation in brackets, no italics, headline-style capitalization[*[Chinese Teaching in the World]*, *25*(2), 268–275.

Zhang, H. (2014). The third tone: Allophones, sandhi rules and pedagogy. *Journal of the Chinese Language Teachers Association*, *49*(1), 117–145.

Zhang, H. (2016). Dissimilation in the second language acquisition of Mandarin Chinese tones. *Second Language Research*, *32*(3), 427–451.

Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, *137*(3), 1452–1463.

Zou, T., Chen, Y., & Caspers, J. (2016). The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Bilingualism: Language and Cognition*, 1–13. https://doi.org/10.1017/S1366728916000791

APPENDIX

Table A1. Observational Studies. List of 46 observational studies with experiments targeting non-native perception of Mandarin tones.

Study	L1	L2 level	Measures	Syllables
Kiriloff (1969)	English	1st semester	Pinyin transcription; 4AFC	MS
Lin (1985)	English	beginner, intermediate, advanced	4AFC	MS
Broselow, Hurtig, & Ringen (1987)	English	naïve	4AFC	MS, DS, TS
Leather (1987)	Dutch, English	naïve	rating task	MS
Repp & Lin (1990)	English	naïve	speeded classification	MS
Lee & Nusbaum (1993)	English	naïve	speeded classification	MS
Lee, Vakoch, & Wurm (1996) *[Exp. 2]*	Canton., English	naïve	AX	MS
Gottfried & Suiter (1997)	English	8 with < 5 years 1 with > 20 years	4AFC	MS
Sun (1998)	English	1st, 2nd, 3rd, & 4th year	4AFC	MS, DS, TS
Klein, Zatorre, & Milner (2001)	English	naïve	AX w/ PET scan	MS
Wang, Jongman, & Sereno (2001)	English	naïve	4AFC (dichotic listening)	MS
Hallé, Chang, & Best (2004)	French	naïve	2AFC; AXB	MS
Alexander, Wong, & Bradlow (2005)	English: music/ non-music	naïve	2AFC	MS
Krishnan, Xu, Gandour, & Cariani (2005)	English	naïve	brainstem frequency following response	MS
Bent, Bradlow, & Wright (2006)	English	naïve	4AFC	MS
Chandrasekaran, Krishnan, & Gandour (2007)	English	naïve	passive discrimination (oddball) w/ ERPs (MMN)	MS
Gottfried (2007)	English: music/ non-music	naïve	4AFC (intact, silent center); AX	MS
Guion & Pederson (2007)	English, Japanese	naïve, ≥4 years	similarity ratings, multi-dimensional scaling	MS
Crinion et al. (2009)	'European'	naïve, 1-4 years experience	structural imaging (MRI)	*NA*
Lee, Tao, & Bond (2009)	English	1st, 2nd, & 3rd year	4AFC (intact, center-only, silent center, onset-only)	MS
Krishnan et al. (2010)	English	naïve	brainstem frequency following response	MS
Krishnan, Gandour, & Bidelman (2010)	English	naïve	brainstem frequency following response	MS
Lee, Tao, & Bond	English	1st, 2nd, & 3rd	4AFC (intact, center-	MS

(2010)		year	only, silent center, onset-only)	
Lee, Tao, & Bond (2012)	English	1st, 2nd, 3rd, & 4th year	4AFC (w/ multiple talkers & noise)	MS
Peng et al. (2010)	Canton., German	naïve	AX	MS
So & Best (2010)	Canton., English, Japanese	naïve	4AFC	MS
Yang & Chan (2010)	English	1st, 2nd, & 'advanced'	4AFC; intonation identification	MS, sentences
Huang & Johnson (2011)	English	naïve	difference rating; AX	MS
Braun & Johnson (2011)	Dutch	naïve	ABX	DS
Marie et al. (2011)	French: music/ non-music	naïve	discrimination of syllabic sequences	MS
Zhang (2011)	English	beginner, intermediate	4AFC; pseudoword recognition; multi-dimensional scaling	MS, DS
Hao (2012)	Canton., English	avg. 2.68 years	4AFC	MS, DS
Lee, Tao, & Bond (2013)	English	1st, 2nd, 3rd, & 4th year	4AFC	MS
He & Wayland (2013)	English	3 months, 12 months	4AFC	MS, DS
Liu (2013)	English	naïve	AX	MS
So & Best (2014)	English, French	naïve	AXB; categorization	MS
Chen, Liu, Kager (2015)	Dutch	naïve	AX	DS
Ning, Loucks, & Shih (2015)	English, Korean	naïve, mus., L2 learners	AX, pitch-shift task	MS
Tsukada, Xu, & Rattanasone (2015)	English, Canton. (heritage)	mixed exp., naïve	categorical discrimination test (oddball, 4AFC)	MS
Zou, Chen, Caspers (2016)	Dutch	naïve, beginner, advanced	ABX	MS
Hao & DeJong (2016)	English	intermediate	4AFC	MS
Shen & Froud (2016)	English	naïve, advanced	AX w/ tone contiuum	MS
Hao (2017)	English	naïve, 1st year, ≥ 4th year	AXB	MS
Hao (2018)	English	2nd year	4AFC	DS
Shen & Froud (2018)	English	naïve, advanced	passive and active oddball tasks w/ tone continuum & ERPs (MMN, P300)	MS
Pelzl et al. (2018)	English	advanced	4AFC; LDT; sentence judgment & ERPs	MS; DS; DS in sentences

KEY:
Canton.= Cantonese;

music= musicians,
n-music= non-musicians;
2AFC= two alternative forced choice identification;
4AFC= four alternative forced choice identification;
AX= sound discrimination task: *Is the second sound (X) the same as the first (A)?*;
ABX= sound discrimination task: *Is the last sound (X) the same as the first (A) or the second (B)?*
AXB= sound discrimination task: *Is the second sound (X) the same as the first (A) or the last (B)?*
oddball= sound discrimination task: *Is the new sound the same as the previous sound?*
MS= monosyllabic
DS= disyllabic
TS= trisyllabic

Table A2. Training Studies. List of 31 training studies targeting Chinese (or Chinese-like) tone languages.

Study	Target Language	L1 (proficiency)	Type	Syllable
Wang et al. (1999)	Mandarin	English (1st year)	pitch	MS
Wang et al. (2003) *fMRI*	Mandarin	English	pitch	MS
Wong & Perrachione (2007)	artificial (Mandarin)	English	lexical	MS
Francis et al. (2008)	Cantonese	English, Mandarin	pitch	MS
Song et al. (2008) *EEG*	artificial (Mandarin)	English	lexical	MS
Chandrasekaran et al. (2010)	artificial (Mandarin)	English	lexical	MS
Liu et al. (2011)	Mandarin	English, Korean	pitch pitch	MS DS
Perrachione et al. (2011)	artificial (Mandarin)	English	lexical	MS
Wang et al. (2011) *behavioral*	Mandarin	English	pitch	MS
EEG	Mandarin	English	pitch	MS
Wong et al. (2011)	artificial (Mandarin)	English	lexical	MS
Cooper & Wang (2012)	artificial (Cantonese)	English (musicians)	lexical	MS
Wang (2013)**	Mandarin	English, Hmong, Japanese	pitch	MS
Cooper & Wang (2013)	artificial (Cantonese)	English	pitch lexical	MS MS
Eng et al. (2013)	Mandarin	English	pitch	MS
Ingvalson et al. (2013)	artificial (Mandarin)	English	pitch lexical	MS MS
Showalter & Hayes-Harb (2013)	artificial (Mandarin)	English	lexical	MS
Braun, Galts, & Kabak (2014)	artificial (Mandarin)	French, German, Japanese	pitch, lexical	DS
Maddox & Chandrasekaran (2014)	Mandarin	English	pitch	MS
Sadakata & McQueen (2014)	artificial (Mandarin)	Dutch	pitch	DS
Saito & Wu (2014)**	Mandarin	Cantonese	lexical	MS

Chang & Bowles (2015)	artificial (Mandarin)	English	lexical	MS, DS
Lu et al. (2015) *EEG*	artificial (Mandarin)	English	pitch	MS
Maddox et al. (2015)	Mandarin	non-tonal L1	pitch	MS
Morett & Chang (2015)	Mandarin	English	lexical	MS
Qi et al. (2015) *MRI*	Mandarin	English	lexical	*NA*
Yang et al. (2015) *fMRI*	artificial (Mandarin)	English	lexical	MS
Zhao & Kuhl (2015)	Mandarin	English	pitch	MS
Bowles et al. (2016)	artificial (Mandarin)	English	lexical	MS, DS
Antoniou & Wong (2016)	artificial (Mandarin-Hindi)	English	lexical	MS
Lee et al. (2017) *MEG*	Mandarin	non-tonal L1 (beginners)	pitch	MS
Wiener et al. (2018)	artificial (Mandarin)	English (2nd year)	lexical	MS

Key:

MS = monosyllabic

DS = disyllabic

TS = trisyllabic

pitch = indicates outcomes were measures of phonetic pitch categorization

lexical = indicates outcomes were measures of lexical learning

Most studies in this table used behavioral methods to measure outcomes. Neurolinguistic measures are indicated in *italics* after the study name. Unless indicated in parentheses, participants were naïve, that is, prior to training they had no previous experience learning Mandarin (or another tonal language).