Chapter 3
Acoustic-Based and Knowledge-Based Processing of Mandarin Tones by Native and Non-native Speakers

Chao-Yang Lee and Seth Wiener

Abstract A fundamental issue in spoken language comprehension is how listeners process the acoustic signal to retrieve intended linguistic representations. This issue is discussed by reviewing selected studies on acoustic-based and knowledge-based processing of lexical tone in speech perception and spoken word recognition. Research on acoustic-based processing suggests that native listeners are able to use phonetic knowledge to compensate for compromised F0 information, whereas non-native listeners rely primarily on syllable-internal, canonical F0 information for tone identification. Research on knowledge-based processing shows that native listeners effectively track information, such as a syllable-tone's lexical status, the probability of syllable–tone co-occurrences, morpheme and word frequency, and the density of homophonous syllable–tone neighborhoods. Non-native listeners also show evidence of knowledge-based tone processing, although the difference between native and non-native listeners remains to be explored.

3.1 Introduction

Speech perception refers to the process in which listeners extract information from the acoustic signal and map that information onto some form of linguistic representation. Early work in speech perception focused on the discovery of acoustic correlates for consonant and vowel distinctions (Stevens & Hanson, 2010). Since the ultimate goal of speech perception is to map sound onto meaning, efforts have also been made to explicate the nature of lexical representation and process (Luce & McLennan, 2005). The relative ease of identifying speech sounds and spoken words in daily life often obscures the complexity of the sound-to-meaning mapping process. For example, the same sound or word spoken by different talkers can be acoustically

C.-Y. Lee (✉)
Division of Communication Sciences and Disorders, Ohio University, Athens, USA
e-mail: leec1@ohio.edu

S. Wiener
Department of Modern Languages, Carnegie Mellon University, Pittsburgh, USA

H.-M. Liu et al. (eds.), *Speech Perception, Production and Acquisition*,
Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-15-7606-5_3

different (Johnson, 2005). On the other hand, acoustically identical sounds or words can also be interpreted differently depending on the phonetic context (Ladefoged & Broadbent, 1957). These observations indicate that speech perception and spoken word recognition is shaped not only by auditory ability, but also by phonetic and linguistic knowledge of the listener. The overarching goal of research in speech perception and spoken word recognition, therefore, is to identify factors contributing to the mapping from the acoustic signal onto phonological and lexical representation. To that end, researchers have examined the nature of the acoustic signal, listener characteristics, and source of knowledge that contribute to the mapping process.

Whereas previous work has largely investigated these questions with respect to segments (e.g., Johnson & Mullennix, 1997), less work has explored how listeners process variability at the suprasegmental level. In this chapter, we address this issue with respect to the processing of lexical tones in speech perception and spoken word recognition by native and non-native listeners. In lexical tone languages, tones are functionally analogous to consonant and vowel phonemes. Lexical tones differ, however, from segmental phonemes in that they involve a distinct set of acoustic correlates in speech perception and consequently are processed in a different time course during spoken word recognition. These differences suggest that conclusions drawn from segmental processing do not necessarily apply to lexical tone processing in speech perception and spoken word recognition. Moreover, since tone languages constitute the majority of known languages in the world (Laver, 1994), examining lexical tone processing can advance our knowledge of cross-linguistic aspects of speech perception and spoken word recognition. Investigating native and non-native speech perception also has important theoretical and practical implications. Theoretically, since non-native listeners possess imperfect knowledge of the target language, performance by non-native listeners can reveal important insights about the relative contribution of auditory processing and linguistic knowledge in speech perception and spoken word recognition (Cutler, 2012). Practically, it is commonly reported that lexical tones are incredibly challenging for non-native speakers to acquire (e.g., Wang, Spence, Jongman & Sereno, 1999; see also Ingvalson & Wong, Chap. 4 of this volume). Identifying factors relevant to the processing of lexical tone in speech perception and spoken word recognition has the potential of informing pedagogical approaches to tone language instruction.

The above considerations motivate the following discussions in this chapter. In Sect. 3.2 we review basic facts about lexical tones, including their linguistic function, and acoustic and perceptual characteristics. In Sects. 3.3 and 3.4, we discuss how native and non-native listeners use two broad means of processing to overcome various sources of acoustic variability in speech perception. Within the experimental literature these forms are often discussed as 'bottom-up' and 'top-down' processing (e.g., Marslen-Wilson & Welsh, 1978). In the following sections, however, we organize our review as acoustic-based (Sect. 3.3) and knowledge-based processing of tone (Sect. 3.4).

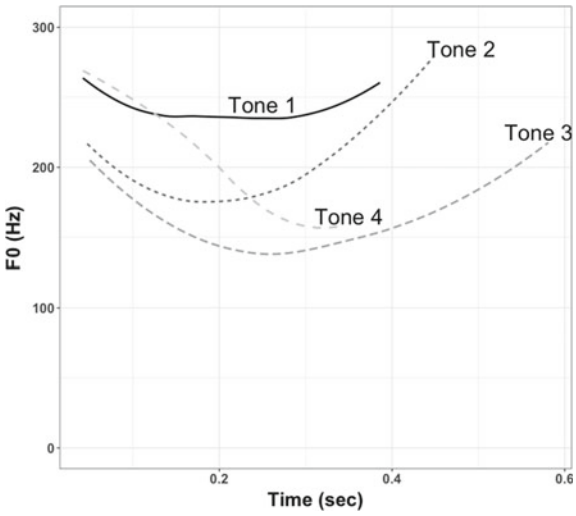

Fig. 3.1 F0 contours of the four Mandarin tones

3.2 Lexical Tones: Function, Acoustics, and Perception

In lexical tone languages, tones are functionally analogous to segments. That is, lexical tones can distinguish words just as segmental structure does. Ample research has established that fundamental frequency (F0) is the primary acoustic correlate of lexical tone (Howie, 1976). In Mandarin Chinese, for example, monosyllabic words can be distinguished by F0 variations over a syllable. As an example, the syllable *ma* with Tone 1 (high-level F0) means 'mother'; *ma* with Tone 2 (mid-rising F0) means 'hemp'; *ma* with Tone 3 (low-dipping F0) means 'horse'; and *ma* with Tone 4 (high-falling F0) means 'scorn.' In addition to F0, duration and amplitude contour serve as secondary cues to tone perception (Blicher, Diehl, & Cohen, 1990; Whalen & Xu, 1992). Nonetheless, F0 height and direction are the main acoustic cues used during tonal categorization and discrimination. The weight assigned to F0 height and direction is dependent upon a listener's language experience (Gandour, 1983). Figure 3.1 shows an example of the four tones, illustrating F0 change and duration differences.

3.3 Acoustic-Based Tone Processing

Previous research on lexical tones has primarily identified the acoustic correlates for specific tonal contrasts. However, relatively little is known about the effects of acoustic variability or adverse conditions on tone perception. As noted in the introduction, the primary puzzle in speech perception and spoken word recognition is how listeners accomplish perceptual constancy in the face of acoustic variability. In

particular, speech communication usually takes place in listening conditions that are less than optimal. Examining the effects of acoustic variability arising from adverse conditions can therefore inform the nature of speech perception (Guediche, Blumstein, Fiez, & Holt, 2013; Mattys, Davis, Bradlow & Scott, 2012). Similarly, investigating the impact of acoustic variability on non-native speech perception further elucidates how linguistic knowledge affects speech perception (Lecumberri, Cooke, & Cutler, 2010). The specific question being asked in this chapter is whether acoustic variability affects native and non-native tone perception in the same way or differently. Below we discuss selected studies on Mandarin tone perception to examine how native and non-native listeners with different levels of proficiency deal with various sources of acoustic variability in lexical tone perception.

3.3.1 Fragmented F0 Input

As noted earlier, F0 is the primary acoustic correlate of lexical tones. That means eliminating or reducing the amount of F0 information is likely to compromise lexical tone identification. However, there is evidence showing that native listeners are quite good at identifying tones from stimuli devoid of F0 information. Whalen and Xu (1992) manipulated Mandarin syllables such that F0 was removed but amplitude contour and duration were retained. Listeners were able to identify the majority of the tones, suggesting the use of duration and amplitude contours for tone perception. Liu and Samuel (2004) similarly showed that tone identification remained robust when F0 was neutralized with signal processing or whispered speech. Studies using the gating paradigm (Grosjean, 1980), where a stimulus is truncated systematically to manipulate the amount of acoustic information available to listeners, also showed that isolated Mandarin tones could be identified with less than half of a syllable (Lee, 2000). These findings demonstrate that native listeners are capable of using secondary acoustic cues to identify tones when F0 information is not available or substantially reduced.

A series of studies employing the silent-center syllable paradigm (Strange, Jenkins, & Johnson, 1983) further demonstrated that listeners are able to use fragmented F0 information to retrieve lexical tones. Gottfried and Suiter (1997) constructed four types of Mandarin consonant–vowel syllables with varying amounts of F0 information. Native and non-native listeners were able to identify the tones of the stimuli that included intact syllables, center-only syllables (with the first six and final eight glottal pulses removed), silent-center syllables (with all but the first six and final eight glottal pulses removed), and onset-only syllables (with all but the first six glottal pulses removed). Figure 3.2 shows waveforms of the four types of stimuli used. Native listeners were highly accurate in identifying tones in all but the initial-only syllables. For example, despite the absence of the majority of the tonal contour, the silent-center tones were identified as accurately as intact and center-only tones. Non-native listeners, on the other hand, were not able to identify the silent-center tones as accurately. These results indicate that native listeners were able to integrate

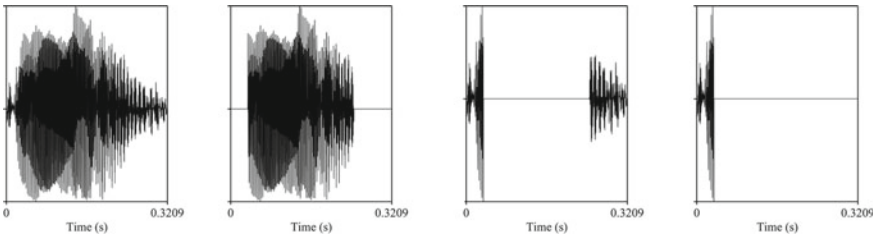

Fig. 3.2 From left: intact, center-only, silent-center, and onset-only stimuli

tonal information from the initial and final portions of the silent-center syllable to reconstruct the intended tones. Non-native listeners, however, did not take advantage of the dynamic tonal information in the remaining fragments of the syllable.

Lee, Tao, and Bond (2008) and Lee, Tao, and Bond (2010a) extended Gottfried and Suiter (1997) by using the same types of stimuli, but with a larger number of native listeners and non-native listeners with Mandarin experience ranging from one to three years of classroom instruction. Lee and colleagues also used reaction time as an additional response measure since it is usually considered a more sensitive measure of processing differences. The accuracy results replicated Gottfried and Suiter (1997); native listeners identified silent-center tones as accurately as the intact and center-only syllables. The reaction time results, however, revealed subtle differences between the modified syllables and the intact syllables, indicating a processing cost associated with the limited F0 information available from the fragmented syllables.

The ability of native listeners to recover missing tonal information is consistent with recent behavioral and neuroimaging evidence on Chinese sentence processing. Xu, Zhang, Shu, Wang, and Li (2013) showed that pitch-flattened sentences are as intelligible as normal sentences, indicating native listeners can use contextual information to access lexical meaning in sentences even if F0 information is altered substantially. Taken together, these findings suggest native listeners are quite capable of reconstructing lexical tones from reduced or altered F0 information at various levels of language comprehension.

3.3.2 Contextual Variation

The aforementioned studies on the perception of fragmented tones also investigated the role of context on tone perception. The effect of context on the acoustics and perception of lexical tones is well documented. In particular, the canonical F0 contour of a tone can be substantially altered by preceding and following tones due to tonal coarticulation (Xu, 1997). Nonetheless, native listeners are able to use their knowledge of the consequence of tonal coarticulation to compensate for contextual variations (Xu, 1994). That is, when asked to identify tones in context, native listener

do not simply rely on the canonical F0 contour of the target tone to judge tone identity. Rather, they interpret surface F0 contours with consideration of the acoustic consequence of tonal coarticulation.

Do non-native listeners also use contextual variations to facilitate tone perception from fragmented syllables? Gottfried and Suiter (1997) presented the fragmented tones with and without a following syllable *zi* ('word'), which had a high-falling F0 contour. The results showed that tone perception accuracy by native listeners was substantially higher when the fragmented tone stimuli were presented in context, but non-native tone identification performance remained the same irrespective of the presence or absence of context. Confusion patterns also differed between the native and non-native listeners. For example, native listeners misidentified onset-only Tone 4 as Tone 1 in isolation, presumably because the high F0 onset (but not the low F0 offset) of Tone 4 was present in the stimuli, which resembles the high onset of Tone 1 (see Fig. 3.1). However, the Tone 4-Tone 1 confusion disappeared when the context was present, presumably because the low offset of Tone 4 carried over to the following syllable, resulting in a lowered onset in the following Tone 1. In other words, native listeners managed to infer from the lowered onset of Tone 1 that the preceding tone had a low offset, which is consistent with Tone 4. Non-native listeners, on the other hand, did not show such a change in confusion pattern as a result of context.

Lee and colleagues (2008, 2010a) evaluated the contribution of context to fragmented tone identification by recording target tones in two carrier phrases. As a result, the offset F0 of the carrier tone and the onset F0 of the stimulus tone resulted in either a match or mismatch. The stimuli were presented in the original carrier phrases (matching contextual F0), excised from the carrier phrases (no contextual F0), or excised and cross-spliced with another carrier phrase (mismatching contextual F0). For the native listeners, there was no effect of splicing, indicating comparable accuracy and speed of processing across the three contexts. However, in the cross-spliced condition, syllables that were originally produced with a matching carrier tone were identified faster and more accurately, demonstrating native listeners' sensitivity to contextual tonal variations. Non-native listeners, on the other hand, did not show such sensitivity to the original tonal context. That is, non-native tone identification was not modulated by contextual tonal variations as native tone identification was.

Taken together, findings from the studies discussed so far suggest that native listeners are sensitive to dynamic tonal information from within a syllable (as indicated by the high accuracy in identifying silent-center tones) and from across syllables (as indicated by improved performance in context, and sensitivity to F0 mismatch between adjacent tones). Non-native listeners, on the other hand, appear to concentrate on syllable-internal, canonical F0 information (as indicated by accurate performance in center-only syllables but not silent-center syllables, and lack of sensitivity to contextual tonal variation). In summary, these findings support the idea that native and non-native listeners use different strategies in dealing with acoustic variability in tone perception. However, it remains to be seen whether this observation can generalize to the processing of other sources of acoustic variability. Next, we turn to two common challenges in speech perception: speaker variability and noise.

3.3.3 Speaker Variability

Listening to different speakers is the norm in speech communication. As noted in the introduction, the same sound or word spoken by different speakers can be quite different acoustically, yet listeners rarely have trouble understanding or adapting to an unfamiliar speaker as long as they speak the same language. How speaker variability is handled is particularly relevant to lexical tone perception. Since lexical tone perception relies primarily on F0, and since F0 range varies across speakers, listener most likely will have to interpret F0 in the acoustic signal relative to a speaker's specific range (Lee, 2009). Does speaker variability affect tone perception more than it does segmental perception? Are non-native listeners affected by speaker variability to a larger degree than native listeners?

Building on earlier studies using fragmented tone stimuli, Lee, Tao, and Bond (2009) examined the effects of speaker variability and context on Mandarin tone identification from intact, silent-center, center-only, and onset-only syllables. The stimuli were presented in isolation or with a precursor carrier phrase. Native and non-native listeners were put under time pressure to identify the tones of the syllables. The literature on segmental processing indicates that adapting to different speakers demands cognitive resources, and the demand usually results in less accurate and more time-consuming responses to multiple-speaker stimuli than single-speaker stimuli (Creelman, 1957). This observation is supported by Lee and colleagues (2009). As shown in Fig. 3.3, both native and non-native listeners had lower identification accuracy in multi-speaker presentation than in single-speaker presentation. However, there was no evidence that non-native listeners were affected to a greater extent by speaker variability than native listeners were. That is, unlike fragmented F0 and absence of tonal context, speaker variability does not appear to pose a disproportionate challenge to non-native listeners.

Lee, Tao, and Bond (2010b) further investigated the effect of speaker variability on native and non-native tone perception by using stimulus sets blocked by speaker and mixed across speakers. Previous studies showed that a mixed-speaker set is more challenging than a blocked-speaker set for tone identification (Wong & Diehl, 2003; Zhou, Zhang, Lee, & Xu 2008). Lee and colleagues presented monosyllabic Mandarin words produced by three male and three female speakers in two presentation formats (blocked by speaker and mixed across speakers) with five levels of signal-to-noise ratios (quiet, 0, −5, −10, and −15 dB) to native listeners and non-native listeners with Mandarin experience ranging from one to four years. Figure 3.4 shows the accuracy results. For both native and non-native listeners, responses to stimuli blocked by speaker were faster and more accurate than responses to stimuli mixed across speakers. Native listeners outperformed non-native listeners, but the additional demand of processing mixed-speaker stimuli did not compromise non-native performance disproportionately.

A possible explanation for the lack of difference between native and non-native tone perception in the speaker variability effect is the presence of tonal context in the stimuli. In particular, the stimuli in Lee et al. (2010b) were embedded in a

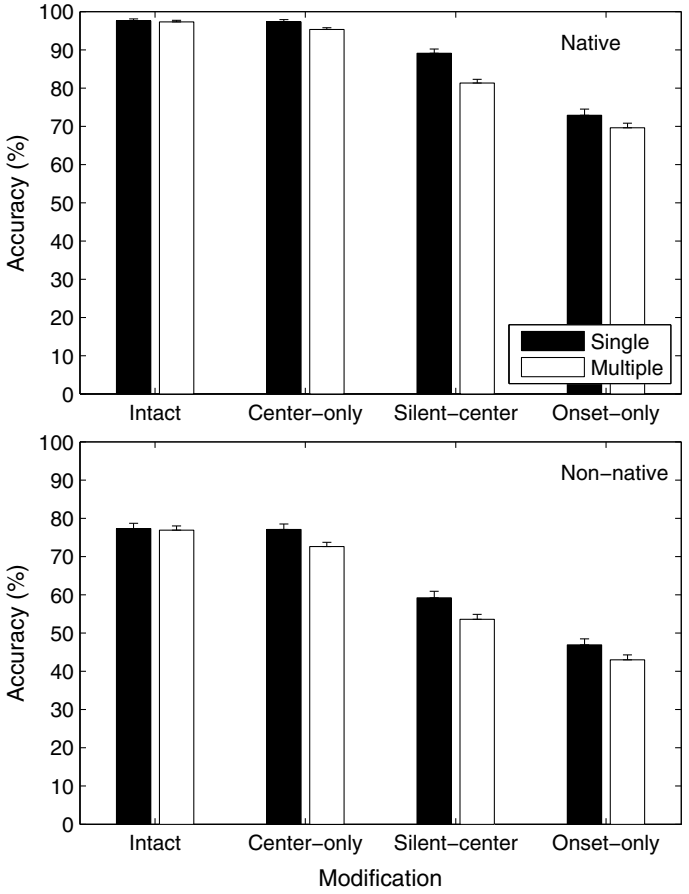

Fig. 3.3 Accuracy of identification of single- and multiple-speaker tones by native listeners (Lee et al., 2009) and non-native listeners (Lee, Tao, & Bond, 2010a). Error bar indicates standard error

carrier phrase *Qing3shuo1__* ('Please say__') to make sure listeners could actually hear the target tones against heavy background noise. Although the carrier phrase was relatively short, listeners could have obtained sufficient information about the speakers, which could have neutralized the processing difference between native and non-native listeners. To evaluate the potential impact of context, Lee, Tao, and Bond (2013) attempted to replicate Lee et al. (2010b) by presenting the same set of stimuli in isolation instead of with the carrier phrase. Lee et al. (2013) also analyzed the data by baseline tone identification proficiency (obtained from accuracy in the blocked presentation without noise) in addition to years of Mandarin instruction with the consideration that the number of years of Mandarin instruction does not necessarily reflect the actual proficiency. Both analyses yielded the same conclusion, and Fig. 3.5 shows the results of the analysis by baseline proficiency. There was no evidence

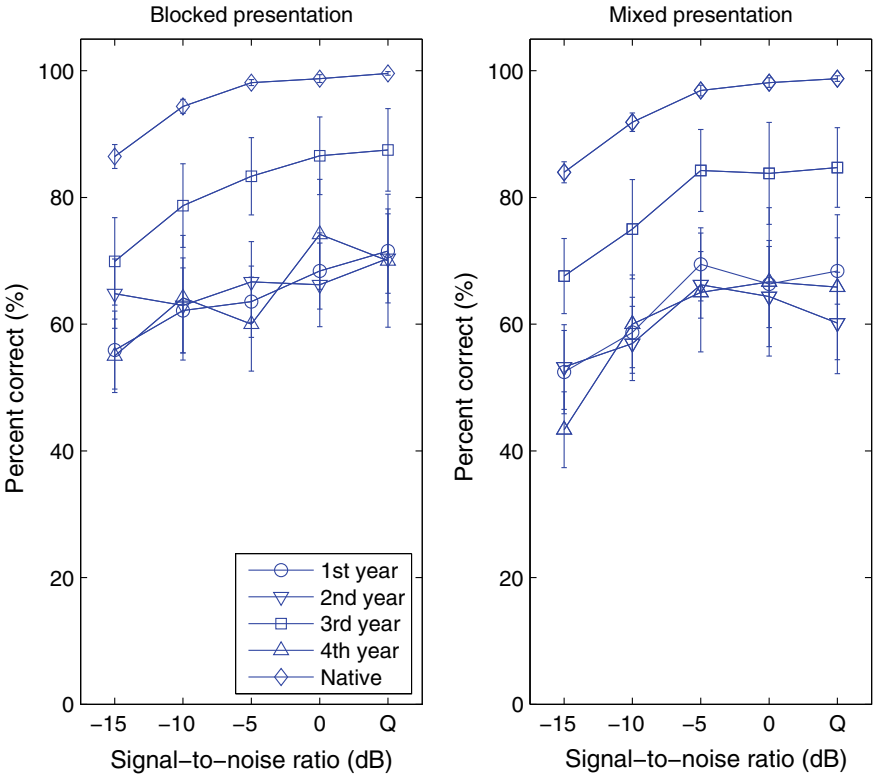

Fig. 3.4 Accuracy of Mandarin tone identification as a function of speaker variability (blocked/mixed presentation), noise level (quiet to −15 dB SNR), and listener background (native and 1–4 years of instruction) in Lee et al. (2010b), reproduced with permission from Elsevier. Error bar indicates standard error

that the mixed-speaker presentation affected non-native listeners disproportionately, suggesting that speaker variability did not pose a special challenge to non-native listeners.

3.3.4 Noise in Tone Perception

Lee and colleagues (2010b, 2013) also examined the effect of noise on tone perception. Noise is arguably the most common adverse condition in speech perception. There is evidence that the perception of segmental phonemes in isolated syllables is usually not compromised disproportionately for non-native listeners, leading to the proposal that the source of disproportionate non-native difficulty with noise is not at a relatively low level of processing such as identifying consonants and vowels in isolated syllables (Bradlow & Alexander, 2007; Cutler, Weber, Smits, & Cooper,

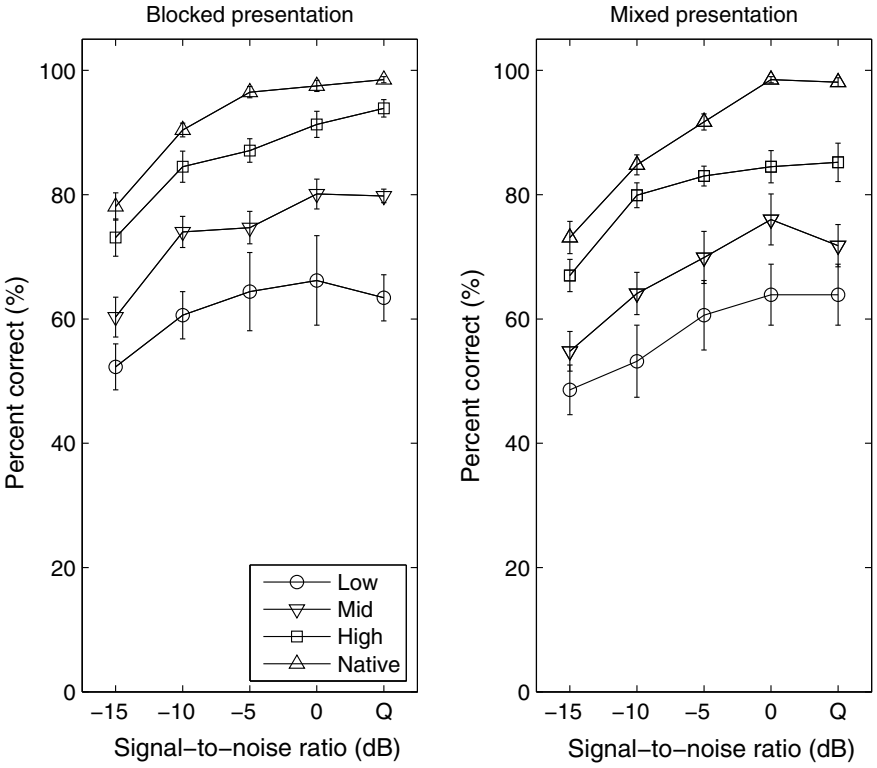

Fig. 3.5 Accuracy of Mandarin tone identification as a function of speaker variability (blocked/mixed presentation), noise level (quiet to −15 dB SNR), and listener background (native and level of proficiency) in Lee et al. (2013), reproduced with permission from Taylor & Francis Ltd. Error bar indicates standard error

2004). In contrast, non-native difficulties with speech perception in noise usually arise only when listeners are asked to process longer stretches of speech, which need more complex linguistic processing. In other words, disproportionate non-native difficulties with noise usually do not surface when the stimuli are relatively simple. Rather, non-native difficulties with noise accumulate across all levels of spoken language comprehension.

Does the conclusion drawn from the segmental literature regarding native and non-native speech perception generalize to tone perception in noise? Surprisingly, little evidence is available to address this question. As noted, Lee et al. (2010b) used multi-speaker tone stimuli embedded in speech-shaped noise with a precursor carrier phrase. The stimuli were presented to native listeners and non-native listeners with Mandarin instruction varying from one to four years. As Fig. 3.4 shows, noise compromised tone perception performance in all listener groups. There was, however, no evidence that noise compromised performance of listeners with less Mandarin experience disproportionately. Since the stimuli were simple syllables

that do not require complex linguistic processing, this result appears to be consistent with the proposal noted earlier that non-native difficulties with noise do not arise from processing simple consonants, vowels, and tones (Bradlow & Alexander, 2007; Cutler et al., 2004).

However, a follow-up analysis showed that when listeners were divided according to baseline performance instead of duration of Mandarin exposure, a significant noise level by baseline performance interaction emerged, suggesting disproportionate noise effect depending on Mandarin proficiency. As discussed earlier, Lee et al. (2013) replicated their earlier findings (2010b) with the same set of stimuli but without the carrier phrase. Their data were also analyzed in terms of baseline performance in addition to years of Mandarin instruction. As Fig. 3.5 shows, noise did affect some listener groups disproportionately. However, it was the listeners with higher proficiency that were affected disproportionately by noise. This result is rather counterintuitive because listeners with lower proficiency were expected to be affected more by adverse conditions because of their less robust knowledge of the target language. Lee et al. (2013) speculated that the less proficient listeners could not identify tones well in the easy, baseline conditions initially. Therefore, the extent to which their performance could be reduced became more constrained compared to the more proficient listeners.

3.4 Knowledge-Based Tone Processing

Prior experience with a language affects how speech is processed. During language acquisition, listeners amass knowledge about a language's phonological and morphological structure, as well as the statistical regularities of spoken sounds and words (Dahan, Magnuson, & Tanenhaus, 2001; Vitevitch & Luce, 1999). Listeners use this knowledge to assess the likelihood of the acoustic signal-to-representation match. Whereas models of spoken word recognition still debate how and at what stage this knowledge influences word recognition (e.g., Cutler, 2012; Norris, McQueen, & Cutler, 2000; Samuel, 2001), there is agreement that listeners are sensitive to a variety of information about the input and draw on this information at some stage of word recognition. A well-known example of knowledge-based processing of speech is the Ganong effect (1980), which demonstrates that listeners tend to identify ambiguous speech sounds in a manner that results in a word. For instance, an ambiguous sound between the two velar plosives /g/ and /k/ tends to be identified as /g/ when English listeners hear it before *–ift* as in 'gift.' In contrast, that same ambiguous sound is identified as /k/ when listeners hear it before *–iss* as in 'kiss.'

Can listeners use knowledge-based processing to overcome tonal variability and poor acoustic-based processing? In many ways, Mandarin serves as an ideal language to test this research question given the language's constrained syllable phonology, tendency for tone contours to align with a syllable (Xu, 1999), and the direct mapping of syllable–tone combinations to morphemes, words, and written characters (DeFrancis, 1986; Duanmu, 2007; Myers, 2010). For example, the first-person

pronoun 'I/me' is the syllable *wo* produced with the dipping third tone. This syllable–tone combination serves as an individual morpheme; this morpheme can stand alone as a word, which in turn can be written with the character 我. While most modern Mandarin words by type are multisyllabic/multimorphemic, spoken corpora like SUBTLEX-CH (Cai & Brysbaert, 2010) indicate that the majority of speech tokens uphold this 1:1:1:1 mapping from syllable–tone to morpheme to word to written character. This suggests that listeners could potentially draw on different sources of knowledge to overcome tonal variability. We first outline these potential sources of information and then identify what role they may play in knowledge-based processing of tone.

3.4.1 Sources of Linguistic Knowledge

Mandarin makes use of roughly 400 unique (C)V(C) syllables: a syllabary roughly one sixth the size of the English syllabary (Duanmu, 2009; see Fon, Chap. 2 of this volume). The syllable serves as the critical or 'proximate' unit in Mandarin perception and production (Chen, Chen & Dell, 2002; O'Seaghda, Chen, & Chen, 2010). Given the limited number of syllable types and their privileged role in speech, native speakers may track syllables' distributional properties in a way that is beneficial for word recognition. One way a listener may do that is by tracking the frequency at which a syllable token occurs in speech. Like other speech sounds, syllables occur with certain statistical regularities. The syllable *shi*, for example, occurs so frequently that the Chinese linguist Chao Yuen Ren famously wrote a poem consisting of 92 characters, all of which share the syllable *shi* but differ in tone: *Shi1 shi4 shi2 shi1 shi3* 'The story of Mr. Shi eating lions.' Native speakers, therefore, may be aware that a syllable like *shi* occurs frequently in speech while a syllable like *nuan* occurs infrequently in speech.

At the syllable–tone level, a listener may track whether a particular syllable co-occurs with all four tones. An often-overlooked feature of Mandarin is that the four lexical tones are not evenly distributed across all syllables. Due to the historic evolution of tone, some syllables appear with all four tones, whereas some syllables only appear with one tone. As examples, the syllable *gei* only appears with tone 3 as *gei3*. The syllable *cong* appears with only two of the four tones as *cong1* and *cong2*. The syllable *ban* appears with three of the four tones as *ban1*, *ban3* and *ban4*. The syllable *shi* appears with all four tones as *shi1*, *shi2*, *shi3*, and *shi4*.

Calculations based on spoken and written corpora (Cai & Brysbaert, 2010; McEnery & Xiao, 2008; Wang, 1986) as well as a modern usage dictionary (CC-CEDICT, 2013) reveal that of the roughly 400 syllables in use, less than a third appear with all four lexical tones. Due to such gaps, modern Mandarin makes use of only around 1400 unique syllable–tone combinations (Duanmu, 2007). Listeners may not need to process tonal information for words in which the segmental information is sufficient for word identification (e.g., *gei*), whereas listeners may need to carefully process tone when the syllable co-occurs with all four tones (e.g., *shi*).

Table 3.1 Number of unique morphemes given a Mandarin syllable and tone

Syllable	Tone 1	Tone 2	Tone 3	Tone 4
you	10	19	11	14
li	0	32	16	61
qiong	1	11	0	0
cou	0	0	0	3

Native listeners may be aware of these co-occurrences and tolerate different degrees of tonal variability given the syllable–tone nonword gaps present in the lexicon.

At the morpheme level, Mandarin possesses a relatively high degree of homophony. Wen (1980) reports that 11.6% of Mandarin words have homophones, compared to only 3.1% of English words. Similarly, Duanmu (2009) estimates Mandarin's homophone density at 9.0 and that of English at roughly 2.4. On average, a particular syllable–tone combination corresponds to 11 semantically and orthographically unrelated morphemes or characters (Perfetti & Tan, 1998), though this can fluctuate across different syllables. For instance, the syllable–tone combination *gei3* only corresponds to one morpheme, whereas the syllable–tone combination *yi4* corresponds to approximately 90 homophonous morphemes (CC-CEDICT, 2013). As a result, tone varies greatly in its informativeness for morpheme identification. In many cases, without additional disambiguating speech or context, listeners may still be faced with a dense tonal neighborhood of homophonous morphemes (Packard, 1999, 2000). To illustrate the asymmetric role of tone in morpheme identification, Table 3.1 shows four syllables and their number of unique morphemes per tone according to a modern usage dictionary (CC-CEDICT, 2013).

The syllables *you* and *li* both occur frequently in speech and both appear with relatively dense tonal neighborhoods. For the syllable *you,* each tone combination corresponds to at least 10 morphemes. For the syllable *li*, even though there is a syllable–tone gap, the remaining combinations all correspond to a relatively large number of morphemes, with *li4* corresponding to over 60 homophonous morphemes. Tone appears to be less informative for these high frequency syllables associated with dense tonal homophone neighborhoods simply because additional context may be required for accurate morpheme identification.

These two high token frequency syllables are juxtaposed with the two low token frequency syllables *qiong* and *cou.* The syllable *qiong* almost exclusively appears with the second tone as *qiong2—qiong1* only corresponds to one extremely rare morpheme. *Qiong3* and *qiong4* are syllable–tone nonword gaps. The syllable *cou* only occurs with the fourth tone, thereby rendering tone redundant since the syllable alone is sufficient to access these morphemes.

Given the relatively small number of syllable–tone combinations, and the varying size of these tonal neighborhoods, a listener may track the probability of a certain syllable co-occurring with each tone. Using the four syllables in Table 3.1 as examples, and assuming each morpheme occurs at the same rate (i.e., not accounting for any lexical frequency effects), a listener has a 100% probability of hearing the

syllable *cou* with the fourth tone, an extremely high probability (roughly 92%) of hearing *qiong* with the second tone, an over 50% probability of hearing *li* with the fourth tone, and roughly equal probabilities of hearing *you* with each tone. Thus, in addition to knowledge of whether a syllable co-occurs with all four tones, listeners may draw on knowledge regarding the relative size of a syllable's tonal neighborhood. That is, a listener may not only be aware that *qiong3* and *qiong4* are both syllable–tone nonword gaps but also that *qiong2* is much more probable than *qiong1* in speech given the size of each tonal neighborhood.

At the word/character level, syllable–tone combinations vary in their frequency of occurrence resulting in a Zipfian distribution (Zipf, 1935). Crucially, lexical frequency has implications for the activation of specific members within a tonal neighborhood. For instance, even though the syllable–tone combination *you3* only corresponds to 11 morphemes—*you2* and *you4* both correspond to more morphemes—one of those *you3* morphemes is the verb 'to have.' This verb 有 is the eleventh most frequently spoken word within the 33.5 million-word corpus SUBTLEX-CH (Cai & Brysbaert, 2010). Therefore, despite the syllable *you* appearing with all four tones and each tonal neighborhood resulting in a relatively similar number of homophonous morphemes, the lexical frequency of 'to have' causes listeners to experience more *you3* exemplars than *you1, you2* or *you4* exemplars. As a result, native listeners may expect and recognize the verb 'to have' more often and faster than other *you3* words, and faster than other *you1*, *you2* and *you4* words (e.g., Janssen, Bi, & Caramazza, 2008; Wiener & Turnbull, 2016; Zhou & Marslen-Wilson, 1994, 1995).

In sum, native Mandarin listeners may track a variety of information about the spoken language. Listeners may track syllable information such as how frequently a syllable token occurs, whether it co-occurs with all four tones, and what the probabilities of these co-occurrences are. Listeners may also track the density of homophonous syllable–tone neighborhoods and the degree to which tone informs morpheme or word identification. Additionally, listeners may track the lexical frequency of particular words within a tonal neighborhood and across the Mandarin lexicon. We next discuss whether this statistical knowledge could be used during speech processing, particularly as a means of overcoming variability in the speech signal.

3.4.2 Evidence for Knowledge-Based Tone Processing

Fox and Unkefer (1985) first tested whether native Mandarin listeners draw on their knowledge of syllable–tone co-occurrences during tone categorization. Participants listened to syllable–tone combinations created from a nine-step tone continuum and were forced to categorize the stimulus as either Tone 1 or Tone 2. In the baseline condition, participants heard a word at each end of the continuum, such as *fei1* 'to fly' and *fei2* 'fat.' Results demonstrated categorical perception of tones similar to categorical perception of phonemes (e.g., Francis, Ciocca, & Ng, 2003; Harnad, 1987). These results were then compared to conditions in which one end of the continuum

contained a non-word, such as *hei1* 'black' and *hei2* (non-word) or *shei1* (non-word) and *shei2* 'who.' Native listeners' tonal category boundaries shifted in the direction of the non-word endpoints for both word/non-word and non-word/word when compared to the word/word continuum baseline. When the study was run on native English speakers with no knowledge of Mandarin, no such tonal category boundary shift was observed. These results indicated that native listeners are aware of syllable–tone co-occurrences and non-word gaps and draw on this knowledge when they process ambiguous speech sounds. Moreover, non-native speakers lack this lexical knowledge and therefore do not draw on it during non-native tonal categorization.

Fox and Unkefer's study demonstrated that knowledge of syllable–tone combinations affects tone categorization, but to what degree does phonological, morphological, and lexical information affect spoken word recognition? In an eye tracking study, Wiener and Ito (2015) demonstrated that listeners are sensitive to syllable token frequencies, syllable–tone co-occurrences probabilities and the relative size of tonal neighborhoods. Moreover, the authors showed that this statistical information affects the earliest stages of spoken word recognition. Wiener and Ito recorded participants' eye movements while four frequency-controlled characters were displayed on a monitor. Participants were instructed to click on the character that matched the perceived spoken syllable–tone combination. Participants were auditorily presented with target words consisting of either a high or low token frequency syllable carrying either the most or least probable tone for that particular syllable. For instance, the high token frequency syllable *you* is most likely to appear in speech with the third tone as *you3* largely due to the high frequency word 'to have' (i.e., adjusting for lexical frequency, given the syllable *you,* listeners have well over a 50% probability of hearing *you3*). In contrast, *you1* is the least likely *you* combination to occur because there are fewer *you1* morphemes and no *you1* word occurs at a high frequency (i.e., listeners have a less than chance probability of hearing *you1*). In addition to the on-screen target (e.g., *you3*), the three other on-screen words included the tonal competitor, which shared the same syllable but the opposite tonal probability (e.g., *you1*), and two other distractors that carried different, non-target syllables.

Response time results from Wiener and Ito demonstrated that native Mandarin listeners mouse clicked fastest on low token frequency syllables carrying the most probable tone and slowest on low token frequency syllables carrying the least probable tone. Participants' eye fixations revealed a similar pattern; participants looked fastest to characters corresponding to a low token frequency syllable with a more probable tone and slowest for characters corresponding to a low token frequency syllable with a less probable tone. Slower fixations and mouse-click responses to low token frequency syllables with less probable tones reflected the competition from the more probable tonal competitor. For high token frequency syllables like *you,* no effect of tonal probability was observed in eye fixations. Mouse-click response times showed a trend in which high probability tones were identified faster than low probability tones, but not at a significantly different speed. The authors' results suggest that Mandarin listeners track and use syllable token frequencies, as well as syllable-specific tonal probabilities, though these probabilities are primarily used when listening to low token frequency syllables (i.e., syllables that occur less often

in speech and tend to carry fewer tonal homophones). Wiener and Ito argued that listeners form syllable-specific tonal hypotheses (cf. Ye & Connine, 1999) as soon as the speech signal begins to unfold, in part, because tone is more informative for word identification on these infrequent syllables.

In a follow-up gating study, Wiener and Ito (2016) used the same stimuli to test how much of the acoustic signal is required for listeners to begin forming a syllable-specific tonal hypothesis. In the first gate, participants heard only the syllable onset. In each successive gate, participants heard 40 ms increments of the vowel. After each stimulus was heard, participants were forced to respond with the perceived syllable–tone combination. Results indicated that minimal acoustic cues from the onset and 40 ms of the vowel triggered knowledge-based processing. Listeners more accurately identified high token frequency syllables (and their tones) than low token frequency syllables (and their tones). An analysis of listeners' correct-syllable–incorrect-tone responses revealed an effect of tonal probability for low token frequency syllables; participants reported the most probable tone for the perceived low token frequency syllable, even when that tone was acoustically dissimilar to the truncated stimuli. When the perceived syllable was a high token frequency syllable, participants did not demonstrate the same knowledge-based processing of tone; participants reported a tone that was acoustically similar to the truncated stimuli. This probability effect was short lived. After hearing the onset and 120 ms, participants began reporting more acoustically similar tones irrespective of the perceived syllable.

To summarize, a small yet growing body of work has shown that native Mandarin listeners track and use a variety of information regarding syllable token frequencies, syllable–tone co-occurrences probabilities, tonal neighborhood densities, and syllable–tone lexical frequencies. This knowledge may be used to overcome tonal uncertainty, improve tonal categorization, and improve spoken word recognition. Knowledge-based processing of tone appears to be most useful in the identification of a relatively unique set of lexical candidates given infrequent syllable tokens and/or fewer tonal homophones.

Can non-native listeners make use of a similar knowledge-based tone processing mechanism? This question was explored in a series of artificial language learning studies (Wiener, 2015). Monolingual English speakers and L2 learners with over a year of classroom Mandarin experience learned an artificial tonal language in which visual nonce symbols were associated with Mandarin-like monosyllables and tones. The stimuli were designed to mimic Mandarin's rich statistical regularities including high and low syllable token frequencies and varied syllable–tone co-occurrence probabilities. After four consecutive days of training on the artificial language, monolingual participants demonstrated evidence of knowledge-based processing of tone in their mouse-clicks (Wiener, Ito, & Speer, 2016). The L2 participants—like the native Mandarin speakers tested in Wiener and Ito (2015)—looked at a higher rate to symbols corresponding to low token frequency syllables with more probable tones than to symbols corresponding to low token frequency syllables with less probable tones. Interestingly, Wiener, Ito, and Speer (2018) found that when learners of the artificial language were trained on multi-speaker input, they relied more on their knowledge of syllable–tone co-occurrence probabilities and less on the incoming acoustic

signal. As a result, learners exposed to single speaker speech or low speaker variability recovered from incorrect probability-based predictions of tone more rapidly than participants exposed to multi-speaker speech or high speaker variability. These results suggest that L2 learners may, in part, rely on knowledge-based tone processing as means of overcoming speaker variability and initial tonal perceptual uncertainty (Wiener & Lee, 2020).

More recently, Wiener, Lee, and Tao (2019) expanded Wiener et al.'s (2016, 2018) findings by investigating knowledge-based tone processing in intermediate L2 Mandarin learners. The authors modified Wiener and Ito's (2016) gating stimuli and tested L2 learners before and after roughly 10–12 weeks of university classroom instruction. An L1 group also performed the gating task as a baseline. Tone-only and syllable–tone word accuracy results at the final gate (i.e., the full acoustic signal) revealed that the L1 group was statistically more accurate than the L2 group at both tests. Although the L2 group made modest tone-only and word accuracy improvements, these gains were not statistically significant. Analyses of the early gates, in which the acoustic information was truncated and listeners had to rely on their Mandarin syllable–tone knowledge, revealed that both L1 and L2 listeners identified high token frequency syllables (and their tones) more accurately than low token frequency syllables (and their tones). An analysis of correct-syllable–incorrect-tone responses revealed that L1 speakers reported more probable syllable–tone combinations when faced with the consonant and up to 80 ms of vowel information, thus corroborating Wiener and Ito (2016). L2 learners only showed a trend toward greater probability-based errors, suggesting that more experience with Mandarin syllable–tone combinations may be required to trigger native-like knowledge-based tone processing.

It is hoped that future work will continue to explore what additional acoustic factors promote knowledge-based processing of tone, how L1 and L2 speakers navigate between acoustic-based and knowledge-based processing of tone, and what are the neurocognitive correlates of these processes (e.g., Yu, Wang, & Li, Chap. 5 of this volume; see also Politzer-Ahles, Wiener, & Zhang, 2017).

3.5 Summary

In this chapter, we explored how native and non-native listeners make use of the acoustic signal and their linguistic knowledge to process tones in speech perception and spoken word recognition. The review of selected studies on processing acoustic variability in tone perception showed that not all sources of acoustic variability are equally disruptive to native and non-native tone perception. Whereas most adverse conditions compromised non-native tone perception disproportionately (fragmented F0, contextual variation, & noise), speaker variability appears to affect native and non-native tone perception similarly. It seems that non-native tone perception is compromised disproportionately only when syllable-internal, canonical F0 information is removed or altered. This observation is consistent with the

observation that non-native listeners rely primarily on syllable-internal, canonical F0 information for tone identification, whereas native listeners are able to use their knowledge of tonal coarticulation and contextual tonal variation to compensate for lost F0 information. Consequently, when syllable-internal, canonical F0 information is reduced (as in fragmented tones) or altered (as in tones excised from original tonal context), non-native tone perception tends to be disrupted disproportionately. In contrast, speaker variability does not affect non-native tone perception disproportionately because speaker variability does not remove or alter syllable-internal, canonical F0 information.

From a knowledge-based perspective, native listeners are able to process tone by drawing on their previous experience with speech. Because tone cannot occur devoid of segmental information, we identified multiple sources of information listeners may track and rely on during speech perception and spoken word recognition: syllable token frequencies, syllable–tone co-occurrence probabilities, syllable–tone homophone densities, and syllable–tone lexical frequencies. These various sources of information are the result of listeners generalizing over the phonological, morphological, and lexical patterns that emerge across the lexicon. Non-native learners are also able to process tone in a knowledge-based manner, though the timing and degree to which learners rely on this mechanism appears to depend upon learners' Mandarin experience and the degree of acoustic variability in the speech signal. In sum, native and non-native listeners process tones similar to that of segments, through acoustic-based and knowledge-based processing to achieve speech perception and accurate spoken word recognition.

References

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,* 37–49.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121,* 2339–2349.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5,* e10729.

CC-CEDICT. (2013). Online Chinese dictionary. https://www.mdbg.net.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46,* 751–781.

Cutler, A. (2012). *Native listening.* Cambridge, MA: MIT Press.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America, 116,* 3668–3678.

Creelman, C. D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America, 29,* 655.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367.

DeFrancis, J. (1986). *The Chinese language: Fact and fantasy.* University of Hawaii Press.

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

Duanmu, S. (2009). *Syllable structure: The limits of variation.* New York: Oxford University Press.

Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics, 13,* 69–89.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non) categorical perception of lexical tones. *Perception & Psychophysics, 65,* 1029–1044.

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

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

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

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28,* 267–283.

Guediche, S., Blumstein, S. E., Fiez, J. A., & Holt, L. L. (2013). Speech perception under adverse conditions: Insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience, 7,* 126.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In *Categorical perception: The groundwork of cognition* (pp. 1–52). Cambridge University Press.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones.* Cambridge: Cambridge University Press.

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes, 23,* 1191–1223.

Johnson, K. A. (2005). Speaker normalization in speech perception. In D. Pisoni, R. Remez (Eds.), *The handbook of speech perception* (pp. 363–389). Wiley-Blackwell.

Johnson, K., & Mullennix, J. W. (1997). *Talker variability in speech processing*. Morgan Kaufmann Publishers Inc.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication, 52,* 864–886.

Lee, C.-Y. (2000). *Lexical tone in spoken word recognition: A view from Mandarin Chinese.* Unpublished doctoral dissertation. Brown University.

Lee, C.-Y. (2009). Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study. *Journal of the Acoustical Society of America, 125,* 1125–1137.

Lee, C.-Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36,* 537–563.

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,* 1–15.

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

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,* 900–910.

Lee, C.-Y., Tao, L., & Bond, Z. S. (2013). Effects of speaker variability and noise on identifying isolated Mandarin tones by native and non-native listeners. *Speech, Language and Hearing, 16,* 46–54.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech, 47,* 109–138.

Luce, P. A., & McLennan, C. T. (2005). Spoken word recognition: The challenge of variation. In D. Pisoni & R. Remez (Eds), *The handbook of speech perception* (pp. 590–609). Wiley-Blackwell.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.
Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27,* 953–978.
McEnery, T., & Xiao, R. (2008). The Lancaster Corpus of Mandarin Chinese (LCMC). https://www.lancaster.ac.uk/fass/projects/corpus/LCMC/.
Myers, J. (2010). Chinese as a natural experiment. *The Mental Lexicon, 5,* 421–435.
Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299–325.
O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but segments in English. *Cognition, 115,* 282–302.
Packard, J. L. (1999). Lexical access in Chinese speech comprehension and production. *Brain and Language, 68,* 89–94.
Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach.* Cambridge: Cambridge University Press.
Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 101–118.
Politzer-Ahles, S., Wiener, S., & Zhang, C. (2017). Predictive tones facilitate Mandarin lexical identification: evidence from ERPs. In *International Conference on Theoretical East Asian Psycholinguistics.* Hong Kong, Hong Kong, 10–12 March.
Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12,* 348–351.
Stevens, K. N., & Hanson, H. M. (2010). Articulatory-acoustic relations as the basis of distinctive contrasts. In W. J. Hardcastle & J. Laver, F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (2nd ed., pp. 424–453). Wiley-Blackwell.
Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America, 74,* 695–705.
Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.
Wang, H. (1986). *Modern Chinese frequency dictionary.* Beijing: Beijing Language Institute Press.
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,* 3649–3658.
Wen, W. (1980). Cong Yingwen de tongxingci lai kan hanyu pinyin wenzi de tongyinci [A study of Chinese homophones from the view of English homographs]. *Yuwen xiandaihua [Modernizing Our Language], 2,* 120–124.
Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49,* 25–47.
Wiener, S. (2015). *The Representation, Organization and Access of Lexical Tone by Native and Non-Native Mandarin Speakers.* Unpublished doctoral dissertation. The Ohio State University.
Wiener, S., & Ito, K. (2015). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition & Neuroscience, 30,* 1048–1060.
Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics, 56,* 38–51.
Wiener, S., Ito, K., & Speer, S. R. (2016). Individual variability in the distributional learning of L2 lexical tone. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Speech Prosody* (pp. 538–542). Boston, MA: Speech Prosody 2016.
Wiener, S., Ito, K., & Speer, S. R. (2018). Early L2 spoken word recognition combines acoustic-based and probability-based processing. *Language and Speech, 61,* 632–656.

Wiener, S., & Lee, C.-Y. (2020). Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology, 11*, 214.

Wiener, S., Lee, C.-Y., & Tao, L. (2019). Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning, 69,* 527–558.

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech, 59,* 59–82.

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46,* 413–421.

Xu, G., Zhang, L., Shu, H., Wang, X., & Li, P. (2013). Access to lexical meaning in pitch-flattened Chinese sentences: An fMRI study. *Neuropsychologia, 51,* 550–556.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95,* 2240–2253.

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

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics, 27,* 55–105.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14,* 609–630.

Zhou, N., Zhang, W., Lee, C.-Y., & Xu, L. (2008). Lexical tone recognition with an artificial neural network. *Ear and Hearing, 29,* 326–335.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes, 9,* 393–422.

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes, 10,* 545–600.

Zipf, G. K. (1935). *The Psycho-biology of language*. Oxford, England: Houghton, Mifflin.