

Asymmetries in the perception of Mandarin tones: evidence from mismatch
negativity

Stephen Politzer-Ahles^{1,2*}, Kevin Schluter³, Kefei Wu³, Diogo Almeida³

¹*Faculty of Linguistics, Philology & Phonetics, University of Oxford, United Kingdom*

²*NYUAD Institute, New York University Abu Dhabi, United Arab Emirates*

³*Psychology Program, New York University Abu Dhabi, United Arab Emirates*

**Address correspondence to*

Faculty of Linguistics, Philology & Phonetics

University of Oxford

Clarendon Press Centre

Walton Street

Oxford OX1 2HG

United Kingdom

Tel.: +971-56-689-9497

E-mail: stephen.politzer-ahles@ling-phil.ox.ac.uk

Asymmetries in the perception of Mandarin tones: evidence from mismatch negativity

Most investigations of the representation and processing of speech sounds focus on their segmental representations, and considerably less is known about the representation of suprasegmental phenomena (e.g., Mandarin tones). Here we examine the mismatch negativity (MMN) response to the contrast between Mandarin Tone 3 (T3) and other tones using a passive oddball paradigm. Since the MMN response has been shown to be sensitive to the featural contents of speech sounds in a way that is compatible with *underspecification* theories of phonological representations, here we test the predictions of such theories regarding suprasegmental phenomena. Assuming T3 to be underspecified in Mandarin (because it has variable surface representations and low pitch), we predicted that an asymmetric MMN response would be elicited when T3 is contrasted with another tone. In two of our three experiments, this was observed, but in non-Mandarin-speaking participants as well as native speakers, suggesting that the locus of the effect was perceptual (acoustic or phonetic) rather than phonological. In a third experiment, the predicted asymmetry was limited to native speakers. These results highlight the importance of distinguishing phonological and perceptual contributions to MMN asymmetries, but ultimately support abstract phonological models in which certain information is underspecified in long-term memory.

Keywords: mismatch negativity; lexical tone; Mandarin; speech perception; underspecification

1 Introduction

Speech is a highly variable phenomenon, and any given word or speech sound may be realized in many different ways (both in terms of its articulation and its acoustic signal) depending on factors such as the context in which it is produced and the experience and vocal tract physiology of the speaker producing it. This poses a challenge to listeners, who must map this variable signal onto long-term memory representations of sounds or words if they are to successfully comprehend speech. To account for this, a variety of

theories have been proposed regarding the nature of long-term memory representations of sounds (a speaker's *phonology*) and of words (a speaker's *lexicon*). Regarding this topic, there is a well-known debate over whether memory representations of words and sounds are abstract, ignoring certain sources of variation or redundant features of sounds (e.g., Archangeli, 1988; Lahiri & Reetz, 2002), or whether these representations fully store as much information as possible (e.g., Goldinger, 1998; Mitterer, 2011). Most of this debate has focused on the representation of segmental aspects of sounds (e.g., consonants and vowels), whereas suprasegmental aspects (e.g., pitch) are less well understood, even though languages in which suprasegmental aspects make up an important part of phonology (such as tone languages) constitute a large portion of, and perhaps a majority of, the world's languages (Maddieson, 2005). The present paper, therefore, tests the predictions of an abstract model of phonological representation and processing (Lahiri & Reetz, 2002) for Mandarin tones. Namely, we test the hypothesis that certain lexical tones, just like certain consonants and vowels, have less specific memory representations than others (i.e., only containing a subset of the features present in the memory representations of other tones), using electrophysiological measures to examine the perception of different tonal contrasts at the brain level. In addition to providing an opportunity to extend such models to account for the representation of suprasegmental sound features, Mandarin tones also offer, as we demonstrate below, some of the strongest kind of evidence to date for a phonological locus (rather than a lower-level perceptual locus) of empirical effects which have been argued to support abstract models of phonology and the lexicon.

Much evidence supporting abstract representations of speech sounds comes from measurements of neural activity during speech comprehension using electrophysiology. Electrophysiological techniques, including electroencephalography (EEG) and

magnetoencephalography (MEG), provide a powerful means to examine the neural processing of speech sounds. EEG and MEG allow researchers to probe the early and automatic brain responses to the perception of sounds, without requiring the participants to perform metalinguistic tasks such as identification or discrimination. Furthermore, electrophysiology can help reveal the time course of, and qualitative differences between, the perception of different kinds of speech sounds and discrimination between them. Of particular relevance to the question of speech processing is the mismatch negativity (MMN) component of EEG (and its magnetoencephalographic counterpart, the mismatch field).

1.1 The MMN, and MMN asymmetries, in language research

The MMN is an electrophysiological response (event-related potential, or ERP) elicited by an infrequently repeated sound (called a *deviant*) embedded in a stream of more frequently repeated sounds (called *standards*), for instance, by each *s* in "*ffff sff sffffffffff sfff s...*". The MMN is argued to be generated not by the low-level acoustic analysis of the deviant (e.g., [s]), but by the contrast between the deviant and the memory representation of the standard (for review, see Näätänen, Jacobsen, & Winkler, 2005; Näätänen, Paavilainen, Rinne, & Alho, 2007). Mere difference detection is, of course, relatively easy, and many species of nonhuman animals can show MMN responses (see reviews in Harms et al., 2014; Näätänen et al., 2007; Walker, Carpenter, Downs, Cranford, Stuart, & Pravica, 2001). Interestingly, however, the MMN may also be elicited by the detection of differences between highly *abstract* memory representations. For instance, while the elicitation of the MMN relies on the presentation of stimuli in an oddball design with many standards from one discrete category and a few deviants from a different discrete category, the MMN can

nonetheless be elicited in a paradigm where there is substantial variation within both the standards and the deviants (in terms of acoustic features such as duration, pitch, location, etc.) as long as one feature (e.g., intensity) sets the varied standards and the varied deviants apart—in other words, when no one unique stimulus token acts as the standard, the "standard" can be a feature or a memory representation of a feature (Pakarinen, Houtilainen, & Näätänen, 2010). This is also the case for linguistic stimuli, in which case MMN can be elicited by features such as voicing (in a paradigm with [p, t, k] standards and [b, d, g] deviants; Phillips, Pellathy, & Marantz, 2000) or lexical tone (in a paradigm where both standards and deviants involve a variety of vowels at a variety of intensities, but are uniquely distinguished by tone; Wang, Gu, He, Chen, & Chen, 2012). Furthermore, the MMN based on linguistic features has been shown to depend on the standard and deviant tokens falling on opposite sides of a perceptual categorical boundary (Phillips, Pellathy, Marantz, Yellin, et al., 2000; Xi, Zhang, Shu, Zhang, & Li, 2010) and on the sound contrast being phonologically meaningful (i.e., such two words might differ only on that one sound) in the listener's language (Kazanina, Phillips, & Idsardi, 2006). In short, there is ample evidence that the MMN can index not just simple change detection, but also discrimination between abstract linguistic categories. Thus, the MMN can shed light on the organization of a speaker's phonological system.

Interestingly, many perceptual contrasts elicit *asymmetric* MMNs. That is to say, the MMN response may be greater when one category serves as the standard and the other as deviant than vice versa. For example, the *e* tokens in "*o o o o e o o e o o o o o o o e*" may elicit larger and/or earlier MMNs than the *o* tokens in "*e e e e o e e o e e e e e e e o*" (Eulitz & Lahiri, 2004), even when controlling for intrinsic physical differences (i.e., by subtracting the ERPs elicited by *e* standards from those elicited by *e*

deviants, and likewise for *o* standards and *o* deviants). The amplitude of the MMN is typically correlated with the degree of difference between tokens, such that a larger acoustic difference between standard and deviant yields a larger MMN (Sams, Paavilainen, Alho, & Näätänen, 1985); an asymmetry based solely on direction of presentation, then, is surprising, since the difference between the acoustics of the two sounds is presumably the same regardless of the direction in which they are presented. The fact that perceptual distance alone does not account for all the MMN effects, and that asymmetries occur, is potentially informative because it yields insights about the memory representations of sounds. In other words, since the MMN relies on the listener's comparing a just-heard deviant signal against the memory trace of one or more standard signals heard earlier, then the presence of an asymmetric response can allow one to make inferences about whether one standard has a different kind of memory representation than another.

While these asymmetries can be used to draw conclusions about the architecture of the speech processing system and the representations used in language processing, doing so can be challenging in practice. This is because there are many circumstances under which asymmetric MMN responses are observed, and these circumstances are not all based on phonological factors and are not always easily dissociable. Asymmetries can emerge at the acoustic level of processing (i.e., the physical sensation of sounds) or at the phonetic level of processing (i.e., the decoding of physical sounds into linguistic percepts, or *phones*), both of which are different than the phonological level of processing (involving, i.e., the mapping of phones onto abstract linguistic structure, and questions of how these linguistic units are represented in long-term memory).

Circumstances that yield asymmetric MMNs include the following:

Presence vs. absence of acoustic elements. Both auditory and visual MMNs show greater amplitudes when the deviant has an additional element or feature relative to the standard than when the deviant is missing a feature that was present in the standard (Czigler, Sulykos, & Kecskés-Kovács, 2014; Nordby, Hammerborg, Roth, & Hugdahl, 1994; Sabri & Campbell, 2000; Timm, Weise, Grimm, & Schröger, 2011; Winkler & Näätänen, 1993). Comparable effects for linguistic stimuli have also been observed in behavioral tasks, in which participants recognize increased presence of a feature faster and more accurately, and imitate increased presence more, compared to decreased presence of a feature (e.g., Nielsen & Scarborough, 2015).

Refractoriness of the N1 response. The MMN overlaps in time and topography with the N1 component, an exogenous ERP component related to physical properties of the stimulus. Thus, ERPs in the same time window as the MMN may reflect physical properties of the stimuli rather than the abstract change detection process in the mind—for instance, pure tones with rising or falling frequency elicit a larger N1 when they have a steeper slope (Yingling & Nethercut, 1983). In many linguistic MMN experiments, subtraction is used to attempt to rule out these factors: the MMN is calculated by subtracting the brain potentials elicited by standard tokens from the brain potentials elicited by the same physical stimulus when it is presented as a deviant in a separate block. This subtraction method putatively removes the influence of differences in exogenous components such as the N1 that are driven mainly by intrinsic low-level factors in the stimulus. However, the N1 component is attenuated as a stimulus is repeated (Jääskeläinen et al., 2004; McEvoy, Hari, Imada, & Sams, 1993), like a standard stimulus in an oddball paradigm would be. Thus, subtracting the standard-stimulus ERPs from the deviant-stimulus ERPs may not completely remove low-level N1 effects from the waveform. In other words, if one stimulus intrinsically elicits

greater N1 than the other, then a traditional MMN subtraction may fail to account for that, because this stimulus may show the heightened N1 when it appears as a deviant but not when it appears as a standard. The result, then, may appear to be an asymmetric MMN when the asymmetry is actually due to a lower-level N1 present in the deviants (possibly with different amplitudes for different deviants) and not the standards. Schröger and Wolff (1996) suggest an alternative MMN measurement technique that can better separate the MMN from the N1.

Perceptual bias. MMNs may be larger when the deviant is more articulatorily peripheral (pronounced with the blade of the tongue near the edges of the vowel space, as in [i, a, u]), than the standard (Polka & Bohn, 2011). This effect has been better studied in behavioral paradigms than in MMN; behaviourally, such perceptual biases for vowels are well-attested and appear to be language-universal. This asymmetry may occur because of these vowels' cross-linguistic status as references for the vowel space, or perhaps because of acoustic salience associated with these vowels, which exhibit a high degree of formant convergence (i.e., the first through third formants of the acoustic signal fall at similar frequencies).

Prototypicality. MMNs are larger when the standards are relatively prototypical members of their phonological category and the deviants are not, compared to vice versa (Ikeda, Hayashi, Hashimoto, Otomo, & Kanno, 2002); this may be because atypical standards activate a weaker representation of the corresponding category.

Lexical status. MMNs are larger when the deviant corresponds to an existing word than when it corresponds to a pseudoword (e.g., Shtyrov & Pulvermüller, 2002), and when the deviant includes a sequence with high phonotactic probability as opposed to one with low probability (Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005),

presumably reflecting the activation of stored cortical representations of words or phonotactic sequences.

Underspecification of phonological representations. Finally, and most important for the present study, MMNs are often larger when the standard is fully specified than when it is *underspecified* (Cornell, Lahiri, & Eulitz, 2011, 2013; de Jonge & Boersma, 2015; Durvasula, Trotter, & Beretta, 2015; Eulitz & Lahiri, 2004; Hestvik & Durvasula, 2016; Law, Fung, & Kung, 2013b; Scharinger, Bendixen, Trujillo-Barreto, & Obleser, 2012; Scharinger, Merickel, Riley, & Idsardi, 2011; Scharinger, Monahan, & Idsardi, 2012; Schluter, Politzer-Ahles, & Almeida, 2015, to appear; Shafer, Schwartz, & Kurtzberg, 2004; Walter & Hacquard, 2004).

The reasoning used to support the underspecification account of the asymmetric MMN results is as follows. In theories of underspecification (and many other phonological theories), the phonological representation of a speech sound is presumed to be a constellation of atomic features: for instance, the sound /t/ is made up of distinct features indicating that it is a stop consonant (made by creating and then releasing a complete closure in the vocal tract), is usually aspirated in English (the release of the closure is followed by a puff of air before the following vowel begins), and has a coronal place of articulation (it is produced by putting the tip of the tongue on the alveolar ridge behind the upper front teeth, or nearby areas), among other features. Some speech sounds are argued to have *underspecified* features, which are those that are not stored in the abstract representation of the sound, but which are included in the eventual pronunciation of a sound, presumably by being filled in with default articulatory codes (Archangeli, 1988). Underspecified features have some special properties that fully specified features do not, such as susceptibility to change in assimilation contexts. For instance, coronal place of articulation (a relatively easy and

neutral place to articulate a consonant [Ladefoged, 1975]), may be underspecified in the memory representations of coronal sounds like /t/ and /n/. One piece of evidence for this is that, in many languages, sounds that would normally be coronal are often instead pronounced at other places of articulation, but sounds that are non-coronal do not become coronal. For example, the <n> in "a quick run picks you up" or "rainbow" is often pronounced as [m] because of place-of-articulation assimilation (Gaskell, 2003), but the <m> in "a quick rum does you in" or "humdinger" is not pronounced as [n]. Default features like coronal place of articulation have been argued to not be stored in the lexicon (Archangeli, 1988; Lahiri & Reetz, 2002, 2010); such theories have been proposed both for economy of storage, and to account for how variant surface forms are mapped onto lexical representations (see, e.g., Lahiri & Reetz, 2010, for details).

In addition to assuming that memory representations in the lexicon do not include these underspecified features, these accounts link phonological representations to the MMN by predicting that memory representations created when hearing these sounds as standards in an oddball paradigm also do not include the underspecified feature. On the other hand, these features *are* present when initially perceiving the sound, because regardless of the memory representation of a sound, the acoustic signal associated with that sound when it reaches the ear contains certain physical cues: even if a speaker's mental representation of, e.g., /n/ does not specify coronal place of articulation, she still ultimately pronounces it with her tongue in such a place, and this has acoustic consequences which will reach the listener's ear. In other words, hearing a sound as a deviant in an oddball paradigm activates all its features, but when the sound is converted into a more abstract memory representation (as would be the case when it is repeated as a standard) the representation of *underspecified* features may be stripped away. This in turn explains the underspecification-related MMN asymmetry. For

example, when a sound with a non-coronal place of articulation, like [k], is used as a standard, its articulation feature remains fully specified in memory; thus, a coronal articulation feature extracted from an incoming deviant [t] sound will be compared against a fully-specified memory representation and strongly clash with it, yielding a large MMN. On the other hand, when the standard is [t], the coronal articulation feature may initially be extracted when the sound is heard, but when the standards are converted into long-term memory representations this feature is abstracted away. Thus, an incoming deviant sound like [k] will not strongly clash with the memory representation, since the feature on which [k] would normally be able to clash (place of articulation) is not actually present in the memory representation. While this theory, and the neurophysiological evidence in support of it, has been developed mainly on the basis of sounds with coronal versus non-coronal places of articulation (Cornell, Lahiri, & Eulitz, 2011, 2013; Eulitz & Lahiri, 2004; Scharinger, Bendixen, et al., 2012; Schafer et al., 2004; Scharinger et al., 2011; Schluter et al., to appear; Walter & Hacquard, 2004), MMN asymmetries putatively related to underspecification have also been observed for features such as vowel height (de Jonge & Boersma, 2015; Scharinger, Monahan, et al., 2012), voicing in stops (Hestvik & Durvasula, 2016) and fricatives (Durvasula et al., 2015; Schluter et al., 2015), and lexical tone (Law et al., 2013b).

This line of research has yielded important insights into both the architecture of the human language system and the neural mechanisms subserving speech processing. Importantly, this account makes a strong claim about the nature of mental representations of sounds—i.e., that not all aspects of a sound are stored to the same extent—which has major implications for a model of speech processing. Currently, however, it is still challenging to attribute the results of this line of research uniquely to phonological factors. As discussed above, many other non-phonological factors can

contribute to perceptual asymmetries. In many cases these perceptual biases are cross-linguistic and develop at very early ages (Polka & Bohn, 2011); for example, Tsuji, Mazuka, Cristia, & Fikkert (2015) observed a coronal-labial perceptual asymmetry (similar to that predicted by underspecification accounts described above) for both Dutch-learning and Japanese-learning infants as young as 4 months, which is likely earlier than underspecified lexical representations develop. Results such as these suggest that many of the abovementioned perceptual asymmetries attributed to phonological underspecification might instead be due to language-universal *phonetic* biases that are part of the language perception architecture even before phonological representations have developed.

Most previous studies reporting MMN asymmetries tested only one group of speakers (with the exception of Shafer et al., 2004), which makes it difficult to adjudicate between phonological accounts of the asymmetry, such as underspecification, and other potential accounts based on potentially language-universal effects such as perceptual biases. Because phonology involves abstract mental representations, one way to attempt to rule out such accounts is to show that an MMN asymmetry only occurs in a paradigm that encourages abstraction (when there are multiple tokens of both the standards and deviants, such that there is within-category variation) and not in a paradigm that does not (when there is only one repeated token of the deviant and one of the standard, allowing the brain to detect the difference just by focusing on low-level perceptual features). Hestvik and Durvasula (2016) indeed report a in which a voicing contrast on stops elicits an asymmetric MMN in a high token-variability paradigm but not in a single-token paradigm, suggesting that the asymmetry was likely based on phonological factors; Durvasula and colleagues (2015), testing a voicing contrast on fricatives, find that the high token-variability paradigm a long-

lasting asymmetry whereas the single-token paradigm elicits an early asymmetric MMN followed by a later symmetric effect (with MMNs for both directions of the contrast), suggesting that underspecified phonological representations were accessed earlier during deviance detection and fully specified phonetic or auditory representations were accessed later. We note, however, that several other studies reviewed above (de Jonge & Boersma, 2015; Scharinger, Bendixen, et al., 2012; Walter & Hacquard, 2004) *do* observe asymmetries when using one token each of standard and deviant. This suggests that either the manipulation of within-category token variability does not always modulate the presence or absence of MMN asymmetries, or MMN asymmetries observed in single-token paradigms were likely due to low-level acoustic factors rather than to phonological representation. The factors that lead some one-token-per-category experiments to show MMN asymmetries and some not to are currently unclear.¹ Below we describe additional way, which forms the basis of the present study, to rule out language-universal perceptual contributions to MMN asymmetries.

1.2 The present study

In addition to comparing single-token to multiple-token paradigms, it is also important to test for asymmetries in different groups of speakers with different

¹ One explanation for the inconsistency that we considered was that some of these studies used single *synthetic* tokens and others used single *natural* tokens; natural tokens may encourage more abstract phonological encoding. However, this does not seem to fully account for the different results, as some of the single-token studies reporting asymmetries also used synthetic stimuli (de Jonge & Boersma, 2015; Walter & Hacquard, 2004).

phonological systems. Observing an asymmetry in one group of speakers and not another would provide powerful evidence for a language-specific phonological basis, rather than a language-universal perceptual (phonetic or acoustic) basis, of the asymmetry. Unfortunately, some putatively underspecified features, such as coronal place of articulation, are cross-linguistically common (Paradis and Prunet, 1991; although see, e.g., Avery and Rice, 1989, for evidence that coronal underspecification is not universal), and thus it would be nontrivial to find a suitable control group for such a test. On the other hand, complex lexical contour tones vary substantially across languages and even across dialects within a language, making it easy to compare speakers who have a given feature in their phonological system against speakers who do not. Such a comparison is invaluable for making an argument that an observed MMN asymmetry is due specifically to the listener's phonological system. Indeed, as we argue in the discussion, we believe that combining cross-paradigm comparisons (per, e.g., Hestvik & Durvasula, 2016) with cross-language comparisons when possible should be seen as the gold standard for evaluating hypotheses about phonological representation using neurolinguistic and psycholinguistic data.

For these reasons, suprasegmental features found in tonal contrasts would be a fertile testing ground in which to investigate MMN asymmetries. In particular, the tonal system of Standard Mandarin is ideal for addressing both of the limitations described above. Chinese contour tones are complex, multidimensional suprasegmental features (Gandour & Harshman, 1978): while they are distinguished mainly by fundamental frequency (pitch), other dimensions such as duration (Blicher, Diehl, & Cohen, 1990; Lin, 1988; Liu & Samuel, 2004), intensity (Coster & Kratochvil, 1984; Whalen & Xu, 1992), and vowel quality (Rose, 2015; Yang, 2011) also play a role in distinguishing tones. This makes them an interesting test case for theories of underspecified features.

Furthermore, lexical contour tones are only part of the long-term memory representations of speech sounds for speakers of these languages, and not for people who do not speak the language. It is thus straightforward to predict that listeners who do not speak a language with lexical contour tones will not show phonological effects on the MMN, although they may still show lower-level effects based on cross-linguistic perceptual biases.²

An additional benefit of using lexical tones in the present study is that tone languages are currently underrepresented in research on this topic: almost all studies on asymmetries in the MMN, particularly underspecification-related asymmetries, have been conducted on segmental contrasts and on Indo-European languages. Little is known about the processing and representation of suprasegmental cues, particularly complex multidimensional ones like lexical tones, in the context of underspecification theory and perceptual asymmetries.

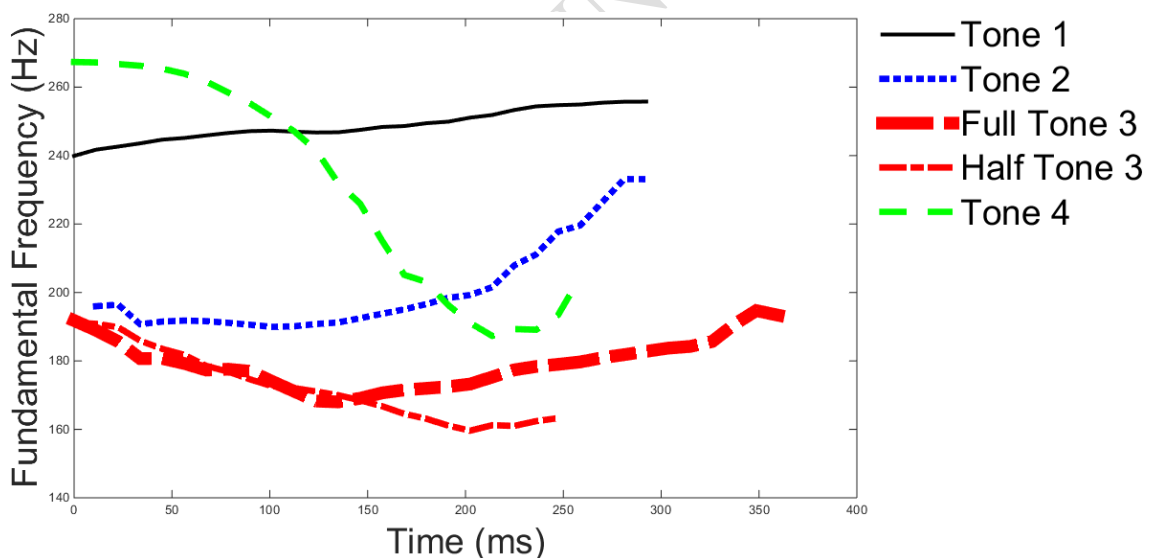
1.3 Mandarin tones and the MMN

Standard Mandarin has four lexical tones. For example, *shou*¹ 收 means "collect", whereas *shou*² 熟 means "ripe", *shou*³ 手 "hand", and *shou*⁴ 瘦 "skinny".³ In the standard dialect, tone one (T1) is realized as a High tone (with a pitch value of 55 in the

² Note that non-Mandarin-speaking participants may still have phonological representations of tone, because intonation can serve linguistic functions—for instance, in English, stress allows for a contrast between the noun *SUBject* and the verb *subJECT*. But such participants will presumably not have phonological representations of the specific tone contours that correspond to Mandarin tone categories.

³ Superscript numbers refer to tone categories.

Chao system [Chao, 1968], in which 5 indicates the highest pitch and 1 the lowest, and the two numbers indicate the pitch at the beginning and end of the tone), T2 as a Rising tone (35), T3 as Low (213), and T4 as Falling (51); see Figure 1. T3 may be realized as a simple low-falling contour or a complex falling-rising contour, depending on its prosodic context: in isolation, the tone dips to a very low pitch and then rises again at the end, whereas when followed by another syllable in the same intonational unit, the final rise is generally not pronounced. T3 and T2 are involved in an alternation relationship known as third tone sandhi: when two T3 syllables are adjacent, the first is pronounced as T2: /T3.T3/ → [T2.T3] (M. Chen, 2000; Yip, 2002; Zhang, 2010). For example, while the morphemes *wo*³ 我 ("I") and *xiang*³ 想 ("think") both canonically are pronounced with T3, when they are produced together in a phrase 我想... ("I think...") then the pronunciation is instead *wo*² *xiang*³.



T3 bears several properties of underspecified features. It undergoes asymmetric alternation—i.e., T3 may sometimes be realized as T2, but no other tone is realized as T3—which many putatively underspecified segments do (although alternation is neither necessary nor sufficient for a segment to be considered underspecified). There is also evidence that some children acquire the contrast between T3 and other tones earlier than contrasts between other tone pairs (Qu, 2013; see, however, Wong, 2012, and Wong,

Schwartz, & Jenkins, 2005, for a conflicting account). In Standard Mandarin, T3 is a low tone, which is considered typologically less marked (a common property of underspecified segments) than high tones and rises, on the basis of properties such as the avoidance of placing contrastive stress on a syllable with low tone (Kuo, Xu, & Yip, 2007) and the disproportionate use of high tones and rises, rather than low tones, for communicative purposes associated with getting listeners' attention (Hsu, Evans, & Lee, 2015).

Asymmetries in tone perception have been observed at very young ages in infants exposed to Mandarin. Yeung, Chen, and Werker (2013) found that both Mandarin-learning and Cantonese-learning four- and nine-month-olds distinguished two Cantonese tones better when they were familiarized to a high-rising tone than when they were familiarized to a mid-level tone. They argue that the high-rising tone may have allowed the infants to construct a better representation of the acoustic space, as this tone covers a larger range of pitch values; such an argument points to a non-phonologically-based perceptual bias driving tone perception (although it should be noted that English-learning infants in the study did not show this effect). In addition, Tsao (2008) found that one-year-old Mandarin-learning infants were better at distinguishing T3 in the context of T1 than vice versa (although no such asymmetry was observed in the contrast between T3 and T2). These studies provide some evidence for early-developing perceptual biases in tone perception; an open question is whether phonological organization, as opposed to perceptual biases, also plays a role in the perception and recognition of tones in the developed (adult) phonological system.

There is a large body of MMN research on adult tonal processing in both Mandarin (Bidelman & Chung, 2015; Chandrasekaran, Krishnan, & Gandour, 2007, 2009; Chandrasekaran, Gandour, & Krishnan, 2007; Gu, Li, Wang, Hou, Huang, &

Chen, 2012; Hsu, Lin, Hsu, & Lee, 2014; Lu, Wayland, & Kaan, 2015; Luo, Ni, Li, Li, Zhang, Zeng, & Chen, 2006; Ren, Yang, & Li, 2009; Wang, Gu, He, Chen, & Chen, 2012; Xi et al., 2010; Zheng, Peng, Chen, Zhang, Minett, & Wang, 2014; among others) and Cantonese (Gu, Zhang, Hu, & Zhao, 2013; Jia, Tsang, Huang, & Chen, 2015; Law, Fung, & Kung, 2013a, b; Tong, McBride, Lee, Zhang, Shuai, Maurer, & Chung, 2014; Tsang, Jia, Huang, & Chen, 2011; Zheng et al., 2014; among others). Most of this work, however, has focused on the localization of tonal processing in general, on what cues contribute to discrimination, and on which contrasts are perceived better or worse than which other contrasts, regardless of presentation order. Testing for asymmetries, on the other hand, by definition requires testing for differences in presentation order within the same contrast, and few studies have done this. Lu and colleagues (2015) report both directions of each contrast they test, but this experiment did not include any contrasts with T3 and did not test native speakers; Gu and colleagues (2013) also recorded both presentation orders of each contrast they test, but as their research question was not about perceptual asymmetries they collapse across presentation orders and contrasts in their analysis. One study to explicitly test for asymmetries related to underspecification is that of Law and colleagues (2013b), who examined Cantonese tones. In the Cantonese tonal system, low-register tones may also be underspecified, and Law and colleagues indeed found an MMN asymmetry such that the MMN was smaller when low level tones served as the standard than when high level tones did.⁴ However, this study did not test a group of naïve listeners with no Cantonese phonology; therefore, as was the case for many of the other MMN asymmetry studies discussed above, it is difficult to rule out potential low-level perceptual contributions to the asymmetry (i.e.,

⁴ The authors also tested a contrast between a low level tone and a low falling-contour tone, but this contrast did not yield a significant asymmetry.

factors discussed above such as the detection of feature presence or absence, N1 refractoriness, or universal phonetic biases).

In the present study, therefore, we systematically test for asymmetries in the processing of Mandarin contour tones, and we examine the same contrasts in both native Mandarin-speaking listeners and naïve listeners. Observing an asymmetry in the Mandarin-speaking listeners but not the naïve listeners would constitute strong evidence for a phonological locus of the effect. Several potential patterns of results may be predicted.

On the one hand, if there is an asymmetric MMN effect and it occurs because T3 is underspecified, then it should be observed on a contrast between T3 and any other tone, regardless of their phonological relationship. (This is assuming, however, that tones are represented with single features. If the representation of a tone comprises multiple features (for example, pitch onset and offset features, or pitch height and slope features), then this may be different. For example, even though coronal place of articulation is underspecified, [t] would not necessarily show an asymmetric MMN when contrasted with [a], because this is not a minimal pair differing only in place of articulation; the same might apply to tonal contrasts. For more discussion of the potential mental representation of tone features, see e.g. Xu and Wang (2001) and Zhang (2010). We return to this point in the Discussion; see section 5.2.)

On the other hand, if there is an asymmetry and it is due to knowledge of phonological alternations, then it may only occur in the T3~T2 contrast (because T2 is known to be a possible realization of T3, but not vice versa) and not in contrasts between T3 and other tones. Therefore, in Experiments 1 and 2 below, we test three contrasts: T3 vs. T2 (i.e., T3 and its phonologically related tone), T3 vs. T4 (T3 and a phonologically unrelated tone), and T2 vs. T4 (a contrast not involving T3 at all).

Whereas the T3 vs. T2 and possibly the T3 vs. T4 contrasts would be expected to show MMN asymmetries, we had no *a priori* phonological reason to predict asymmetries for the T2 vs. T4 contrast, and thus included it as a control. For the phonologically unrelated pair we chose to contrast T3 with T4, rather than with T1, because T3 and T4 are both contour tones, whereas T1 is a level tone. In Experiment 3, we test all six possible contrasts, between all four tones. Experiment 1 tested T3 with a full falling-rising contour, whereas Experiments 2 and 3 tested the half-T3 variant with a simpler low-falling contour, which is more similar in complexity to the other tones' contours.

2 Experiment 1

2.1 Methods

2.1.1 Participants

Sixteen native Mandarin speakers (9 women, mean age = 20, SD = 1.6) and sixteen naïve speakers with no knowledge of Mandarin (12 women, mean age = 22, SD = 8.7) were recruited from the New York University Abu Dhabi community to take part in the experiment. The naïve speakers came from a variety of language backgrounds, but none had experience with Mandarin or any other tone language. All were right-handed as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971; a Chinese version [Li, 1983] was used for the Mandarin speakers). All participants provided their informed consent and were paid for their participation, and experimental procedures were approved by the Institutional Review Board of New York University Abu Dhabi. Detailed demographic information about the participants is available in Supplementary File 1.

2.1.2 Materials

The stimuli for all experiments are available in Supplementary File 2. The materials for this experiment comprised the syllable *yi* ([i])⁵ in tones 2, 3, and 4. The log parts-per-million token frequencies (from Da, 2004) for these syllables are 3.04 (*yi*²), 3.73 (*yi*³), and 3.77 (*yi*⁴). This syllable was chosen because it is a single monophthong, and because, while it is an existing word in every tone, the lexical frequency of the T3 (low falling-rising) syllable is in between that for T2 and T4, meaning that if T3 behaves differently than the other tones then this behavior would probably not be due to lexical frequency alone. Even though these syllables correspond to existing words for Mandarin-speaking participants and not for non-Mandarin-speaking participants, this difference does not confound the interpretation of within-language asymmetries. While a lexicality effect might be predicted to cause an across-the-board larger MMNs for native speakers than naïve speakers, we had no *a priori* reason to predict that lexical status would cause a perceptual asymmetry for one group and not another, given that all the stimuli were existing words in Mandarin (whereas reports of within-language MMN asymmetries based on lexical status, such as Shtyrov & Pulvermüller, 2012, require the comparison of word deviants versus nonword deviants).

Stimuli were spoken by a female fluent native speaker of Mandarin from Shanghai in a quiet room using an H4n recorder (Zoom Corporation) and an HT5-P condenser headset microphone (Audix). The monophthong *yi* ([i]) was produced in all four tones, in a random order until ten repetitions per tone were collected, and once again with ten repetitions of a fixed order (T1, T2, T3, T4); *wu* ([u]) and *yu* ([y]) were also recorded in the same way, but for this experiment we only used *yi*. The continuous

⁵ Italicized forms are in Hanyu Pinyin, the Romanization used for Mandarin in mainland China.

A representation in the International Phonetic Alphabet is given in parentheses.

recording was noise-reduced offline in Audacity 2.0.6 (Audacity Team) using spectral noise gating. To create the stimuli, one of the *yi* tokens from the random-order block was selected to use as the carrier vowel, and one *yi* token (also from the random-order block) for each tone was selected to use as the pitch tier. For T3, a token with a full falling-rising contour and without glottalization was selected. Using Praat (Boersma & Weenink, 2014), these were all normalized to 300 ms by stretching (using the script `changeduration.praat` from the Praat Vocal Toolkit, <http://www.praatvocaltoolkit.com>). To produce tokens of different tones, the pitch tier from a token of that tone was imposed on the vowel base token, such that the tokens differed in nothing but F0. All tokens were normalized to 75 dB. Note that, while many studies on underspecification-related MMN asymmetries use a variety of tokens, here we used only one token per tone, in keeping with most MMN studies on lexical tone in Mandarin (see section 1.1); in Experiment 3 we test the processing of lexical tones in a paradigm with within-category token variability.

The tokens were played in isolation to several native Mandarin speakers, who verified that they were still identifiable as the expected syllables after manipulation. Only the T2, T3, and T4 tokens were used in this experiment. While T2 does alternate with T4 and T1 in certain lexical contexts (the words 一, 不, and [for some speakers] 七 and 八), these alternations are not productive to other lexical items, and thus we considered T2 and T4 to not be phonologically related in the same sense as T3 and T2 are.

2.1.3 Procedure

The experiment used a passive oddball design with six types of blocks: $yi^2 \rightarrow yi^3$, $yi^3 \rightarrow yi^2$, $yi^2 \rightarrow yi^4$, $yi^4 \rightarrow yi^2$, $yi^3 \rightarrow yi^4$, and $yi^4 \rightarrow yi^3$. (The syllable to the left of the arrow was presented as the standard in that block, and the syllable to the right as the deviant;

for example, in an $y_i^2 \rightarrow y_i^3$ block, y_i^2 served as the standard and y_i^3 as the deviant.) Each of these block types was divided into three sub-blocks, and each sub-block included 204 standards and 36 deviants mixed together (15.4% deviants), with an additional 20 standards presented before the first deviant in a sub-block. Thus, each condition had 672 standards and 108 deviants overall. The number of standards that intervened between deviants was between two and ten. The inter-trial interval was 500 ms.⁶ The stimuli were presented binaurally over tube earphones (Aearo Technologies), using Presentation stimulus delivery software (Neurobehavioral Systems) to control the experiment. Each participant selected a movie or television show with subtitles to watch silently, or a book to read, during the experiment. At the end of each sub-block in this experiment (but not Experiments 2 and 3) the participant was asked to list in Pinyin the sounds she heard during the preceding block (non-Mandarin-speaking participants, who

⁶ This inter-trial interval was chosen both to keep the experimental session relatively short, given the large number of conditions, and to mimic that used in several early studies on underspecification-related MMN asymmetries (Cornell, Lahiri, & Eutlitz, 2013, 2011; Eutlitz & Lahiri, 2004). It is worth noting, however, that short intervals may amplify the influence of phonetic discrimination and reduce the influence of phonological representation (Molnar, Polka, Baum, & Steinhauer, 2014); accordingly, many recent MMN asymmetry studies have used intervals between 500 and 1000 ms (Hestvik & Durvasula, 2016; Scharinger, Bendixen, et al., 2012; Scharinger et al., 2011; Scharinger, Monahan, & Idsardi, 2012).

Given that this aspect of the present design arguably makes it more difficult to observe language-specific phonological effects, it is not a confound where language-specific effects are indeed observed (i.e., Experiment 3 of the present study). Where language-specific effects are not observed (Experiments 1 and 2), however, this does raise for future research the question of whether such effects would have emerged in those experiments with a longer inter-trial interval. We thank an anonymous reviewer for drawing this to our attention.

do not know how to romanize Mandarin sounds, were asked to take their best guess),⁷ and was then allowed to take a self-paced break. The 18 sub-blocks were presented in a fully random order (because presenting six blocks in all possible orders would require a prohibitive number of lists—720—whereas dividing them into 18 blocks and presenting them in a random order would hopefully minimize effects of presentation order across participants). The experimental session lasted about two hours.

2.1.4 Data acquisition and analysis

EEG was continuously recorded from 66 Ag/AgCl electrodes (actiCAP, Brain Products) in a 10/20 layout. The sampling rate was 1000 Hz, and data were filtered online from 0.1–250 Hz. FCz served as the online reference and AFz as the ground. Offline analysis was conducted using EEGLAB functions (Delorme & Makeig, 2004) and MATLAB (The MathWorks, Inc.). Bad channels (no more than one per participant), if present, were interpolated offline, and the continuous data were then re-referenced to the average of both mastoids, segmented into epochs from -200ms to 500ms relative to the sound onset, demeaned using the mean of the whole epoch (Groppe, Makeig, & Kutas, 2009), and then decomposed into independent components using ICA (Makeig, Bell, Jung, & Sejnowski, 1996). For each participant, one to three independent components

⁷ These data were collected in order to assess the presence of verbal transformations (Warren, 1968) in the data; we do not discuss them further here. Also note that participants were instructed not to explicitly attend to the tones, and after the session participants usually reported that they were able to remember which tones they heard in a given block without having attended to the tones. Furthermore, several studies in which participants did explicitly attend to the stimuli nevertheless yielded similar MMN asymmetries as passive oddball experiments have (Hestvik & Durvasula, 2016, Experiment 1; Walter & Hacquard, 2004).

related to blinks or eye movements were identified based on visual inspection and removed from the data.⁸ The artifact-corrected epochs were then baseline-corrected based on the -100 to 0 ms time window, and subjected to a 30 Hz low-pass filter (Hamming windowed sinc FIR filter, 220 samples filter order). The first series of 20+ standards in each block, the first deviant in each block, the first standard after each deviant, and any epoch with voltage at any channel exceeding $\pm 75 \mu\text{V}$ in the -150 to 400 ms time window were excluded from further analysis. For each participant, more than 40 deviant trials per condition were retained; detailed information on the number of trials kept per condition and participant for each experiment is available in Supplementary Figure 1. Within each contrast (T2~T3, T2~T4, and T3~T4), the MMNs were calculated by subtracting the average ERP responses to each standard from the average ERP responses to the same token when it was used as a deviant (for example, standard yi^3 from the $yi^3 \rightarrow yi^2$ block was subtracted from deviant yi^3 from the $yi^2 \rightarrow yi^3$ block).

MMN amplitudes were compared using spatiotemporal clustering with non-parametric permutation statistics (Maris & Oostenveld, 2007), implemented in the

⁸ While artifact correction was not necessary to preserve sufficient trials for analysis in this experiment, it was necessary in Experiment 3, in which fewer trials were collected per participant. Thus, for the sake of comparisons across experiments, we report data for all experiment using this artifact correction technique. Where the pattern of results was different for data analyzed without artifact correction, this is noted in the text (for Experiments 1 and 2 only; for Experiment 3, which had fewer trials per condition, we did not attempt to conduct analysis without artifact correction).

FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011).⁹ Within each speaker group, spatiotemporal clusters between 100 and 300 ms were identified where there were significant interactions of the factor CONTRAST (3 levels: T2~T3, T2~T4, and T3~T4) and DIRECTION (2 levels).¹⁰ For example, within the T2~T3 blocks, $y_i^2 \rightarrow y_i^3$ was one direction and $y_i^3 \rightarrow y_i^2$ another; this factor did not have the same meaning across all contrasts because, for example, a contrast direction defined in terms of whether T3 was the standard or deviant would of course be meaningless for the T2~T4 contrast. Therefore, a main effect of DIRECTION is uninterpretable, but a simple effect of direction within any contrast would indicate that there was an asymmetric MMN for that contrast. The α level for clustering was .05, and at least three neighboring channels had to surpass this threshold to be included in a cluster. Cluster-level p -values were estimated based on 500 permutations of the data; if the p -value was between .04 and .06

⁹ Analysis with spatiotemporal clustering is ideal because it reduces researcher degrees of freedom (Simmons, Nelson, & Simonsohn, 2011) in choosing time windows and electrodes for analysis, and also addresses the multiple comparisons problem. However, for ease of interpretation and for comparison with previous studies, we also analyzed the data using parametric statistics on the mean amplitudes around the MMN peaks; the results from this analysis are described in Supplementary File 5.

¹⁰ Note that, while the experiment follows a $2(\text{GROUP}) \times 3(\text{CONTRAST}) \times 2(\text{DIRECTION})$ mixed design, the issue of how to use non-parametric permutation statistics on EEG data to test an interaction between a between-subjects factor and a polytomous (three levels) within-subjects factor is nontrivial, and is not implemented in FieldTrip (see, e.g., <http://mailman.science.ru.nl/pipermail/fieldtrip/2011-September/004244.html>). Thus, given our *a priori* interest in each group and each contrast, we directly test the within-subjects CONTRAST \times DIRECTION interaction within each group. The analysis with parametric statistics (Supplementary File 5), however, does use the full mixed design.

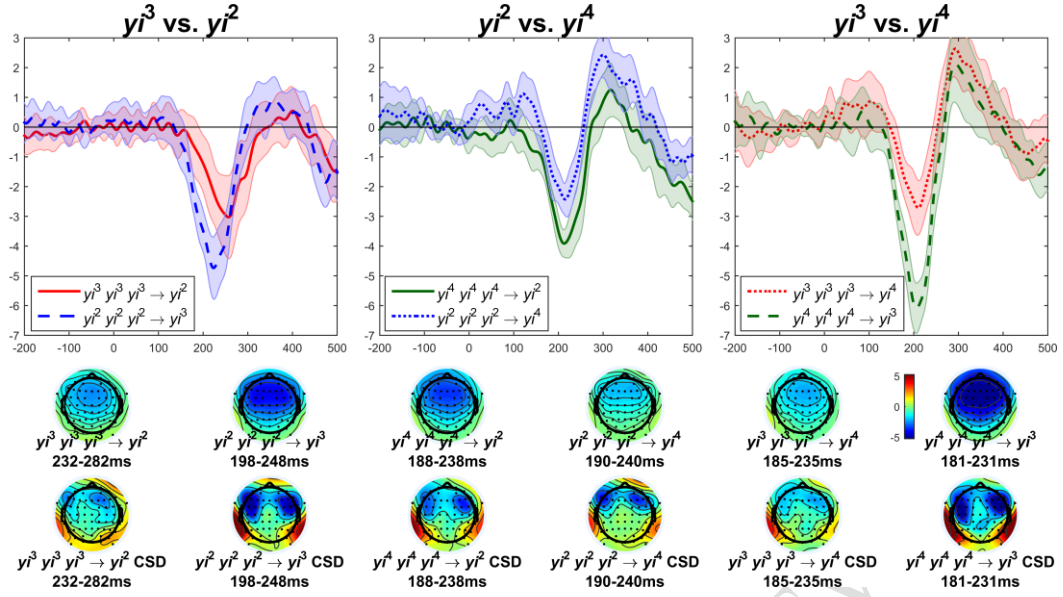
(i.e., very near the significance threshold) it was re-calculated using 5000 permutations. When there was a significant interaction, condition means within the cluster were calculated by averaging the amplitudes of all samples in the cluster for each participant, and compared using t tests with p -values Bonferroni-corrected based on the fact that six post-hoc comparisons were made (the simple effect of DIRECTION within each of the three contrasts, and then the three pairwise comparisons of the DIRECTION difference scores between the three contrasts).

2.2 Results

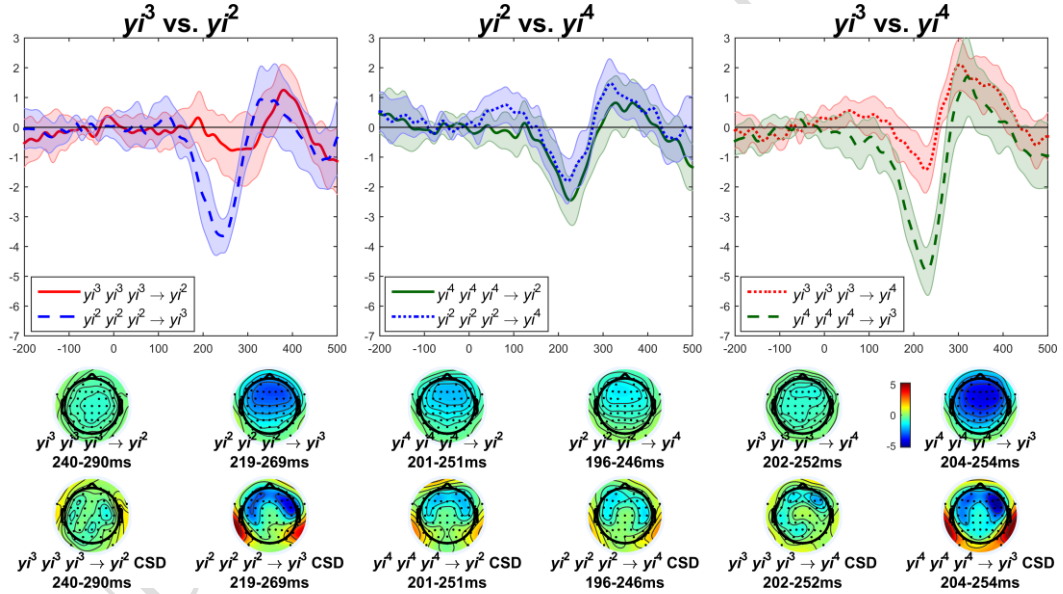
The MMNs for each direction of each contrast, at electrode Fz and time-locked to the sound onset, are shown in Figures 2 and 3, along with topographic maps of the average MMN amplitude for a 51-ms time window centered on each peak, and current source density maps (Kayser & Tenke, 2006a, b; maps were calculated using the CSD Toolbox [Kayser, 2009]) of the MMN amplitude in this time window.¹¹ Figure 2 shows the results for native Mandarin-speaking participants, and Figure 3 the results for naïve participants. The averaged ERP data for this and the following experiments are available in Supplementary File 3. Plots showing the standard and deviant raw waveforms as along with the MMN difference waves are available in Supplementary File 4.

¹¹ Nose-referenced MMN effects, as opposed to other effects in this time window such as N1, often show a polarity reversal over the scalp, with a negativity at midline frontal locations and a positivity at the mastoids (Näätänen et al., 2007). As our montage did not include a nose reference, and we used our mastoids as the reference, checking for polarity reversal at the mastoids was not possible. Transforming the data into current source density, a reference-free representation, allows us an alternative way to show this topographic pattern.

Mandarin-speaking participants



Non-Mandarin-speaking participants



From visual inspection of the data it is clear that native speakers showed robust MMNs for all conditions, and that the T3~T2 (phonologically related) and T3~T4 (phonologically unrelated) contrasts both showed asymmetries such that MMNs were smaller when T3 (i.e., low falling-rising tone) served as the standard and another tone as the deviant, than vice versa. This asymmetry also appears stronger than the asymmetry, if any, observed in the T2~T4 contrast. Finally, native speakers appear to show roughly the same pattern. These findings were confirmed in the statistical analysis.

For native speakers, the CONTRAST \times DIRECTION interaction yielded a marginal effect ($p = .069$), based on a cluster extending from 156-223 ms after the stimulus onset and including the 46 channels Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, P7, P3, Pz, O1, AF7, AF3, AF4, AF8, F5, F1, F2, F6, FT9, FT7, FC3, FC4, FT8, C5, C1, C2, C6, TP7, CP3, CPz, P5, P1, PO3, POz, and FCz. On the mean amplitude of all samples in this cluster, all three contrasts showed significant directional asymmetries ($|t/s(15)| > 3.28$, $ps < .03$), but the magnitude of the asymmetry in the T3~T4 contrast was greater than in the T2~T4 contrast ($t(15) = 4.35$, 95% CI = 0.89...2.61, $p = .003$).

For non-Mandarin-speaking listeners, there was a significant CONTRAST \times DIRECTION interaction ($p = .004$) due to a cluster from 154-273 ms after the stimulus onset encompassing the 43 channels Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, Pz, P4, AF7, AF3, AF4, AF8, F5, F1, F2, F6, FT7, FC3, FC4, FT8, C5, C1, C2, C6, CP3, CPz, CP4, P2, P6, and FCz. Within this cluster, there were significant asymmetries for the T3~T2 contrast ($t(15) = -5.30$, 95% CI = -2.50...-1.07, $p < .001$) and the T3~T4 contrast ($t(15) = -6.94$, 95% CI = -3.07...-1.63, $p < .001$), but not for the T2~T4 contrast ($t(15) = -1.04$, 95% CI = -1.09...0.38, $p = .317$). There was a second marginal interaction effect ($p = .092$), based on an earlier cluster from 112-148 ms and including 32 channels: F3, Fz, F4, F8, FC1, FC2, FC6, C3, Cz, C4, T8, CP1, CP2, CP6, Pz, P4, AF4, AF8, F1, F2, F6, FC3, FC4, FT8, FT10, C1, C2, C6, CPz, CP4, P2, and FCz. In this cluster, there was a significant asymmetry in the T3~T4 contrast ($t(15) = -4.02$, 95% CI = 1.4389...-0.4410, $p = .001$) but not in the other contrasts ($|ts|(15) < 1.10$, $ps > .290$).

2.3 Discussion

Robust effects were elicited for tonal contrasts by both listener groups; based on polarity and topography, the effects are consistent with MMN effects.¹² Crucially, the MMN effects revealed asymmetries in the brain's error detection system. For both listeners who are native speakers of Mandarin and listeners with no tone language experience, asymmetric MMN effects were elicited when T3 (low falling-rising tone) was contrasted with another tone. These effects were all in the same direction—the MMN was attenuated when T3 was the standard and another tone the deviant, compared to the reverse case. Asymmetries in a contrast that did not involve T3 (T2~T4) were generally weaker or absent. Given that T3 is argued to be in the low register whereas T2 and T4 are argued to be in the high register (Kuo et al., 2007), the results are consistent with those of Law and colleagues (2013b), who showed a similar asymmetry between a

¹² Note that, while some of the effects may appear to be late relative to typical MMNs, this is likely because "0" in our plots and analysis represents the onset of the sound, not the disambiguation point of the tone, which is likely much later (and which varies across different tonal contrasts—for instance, T2 and T3 become different much later than T3 and T4 do, see Figure 1). While it is possible that differences in the peak latency of the MMN (due potentially to differences in the disambiguation point of a given contrast) could cause apparent amplitude differences within a time window, we note that 1) the exploratory time-window statistics (Supplementary File 5) used peak-picking, which is unaffected by differences in peak latency as it tests a different time window for each condition; 2) visual inspection of the waveforms suggests that the amplitude differences observed were not due only to latency differences; and 3) the statistical analysis likewise did not show the patterns that would be expected if two MMNs had identical amplitudes at different latencies (in which case the test would reveal a negative cluster closely followed by a positive cluster, or vice versa).

low and a high tone. The present study, however, also found such an asymmetry for non-Mandarin-speaking participants, which casts doubt upon the assumption that this effect is due to the tone's phonological representation. Since the non-Mandarin-speaking participants have no phonological system for Mandarin tones, the effect could only be phonological in nature (as opposed to based on general perceptual mechanisms) if low tones like T3 are universally underspecified, even in speakers whose languages do not involve lexical tone. While there is indeed some evidence that high tones are typologically more marked than low tones (see, e.g., the review in Hsu et al., 2015), this is nonetheless a rather extreme claim to make. It could be tested, however, by examining tone-elicited MMN asymmetries in other Mandarin dialects, which have similar phonological tone systems with different phonetic realizations of the same tones. For example, the Tianjin, Jinan, and Taiyuan dialects all have the same four tonal categories as Standard Mandarin, and all of them exhibit the same sandhi relationship between T3 and T2 (as well as some additional sandhi relationships of their own; e.g., Zhang & Liu, 2011). But in Tianjin, T3 is a mid rising tone and T2 a high level tone; in Jinan, T3 is high level and T2 mid falling; and in Taiyuan, T3 high falling and T2 low level (Zhang, 2010). If the results of the present experiment were due to a cross-linguistic underspecification of low tones, then they should be observed on different tones in other dialects. On the other hand, if the results were due to other acoustic properties of the stimuli, such as the shape or height of the tonal contour (see below), then they might not be replicated in, for instance, Jinan, which has a high level tone as opposed to a low falling contour or low falling-rising contour. Thus, dialectal differences offer a promising testing ground for future research to examine the locus of tone-related perceptual asymmetries.

Assuming, however, that the asymmetric MMN effects elicited by T3 were not due to phonology, then what caused them? We hypothesized that they were due to the complexity of the T3 tonal contour. Recall that the MMN is larger when the deviant is acoustically more complex than the standard, by virtue of containing an additional cue, as opposed to vice versa (Czigler et al., 2014; Nordby et al., 1994; Sabri & Campbell, 2000; Timm et al., 2011; Winkler & Näätänen, 1993). This may indeed be the case for the present tones, as T3 has a complex falling-rising contour (see Figure 1), and thus has a salient turning point, which the other tones do not. T3 has another contextual variant, however, which is less acoustically complex. The full falling-rising contour of T3 is only spoken when the tone falls at the end of an intonational unit, before a pause; otherwise the final rise is not produced (Zhang & Lai, 2010). This contour reduction process (called "half-third sandhi", as opposed to third-tone sandhi discussed in 1.1, which neutralizes T3 and T2) yields a realization of T3 which is arguably more similar in acoustic complexity to the other tones (see Figure 1). In Experiment 2 we used a reduced T3, rather than a full T3, to see whether effects attributable to phonology (i.e., asymmetries limited to Mandarin-speaking participants) would emerge once acoustic complexity was roughly controlled.

3 Experiment 2

3.1 Methods

3.1.1 Participants

Sixteen right-handed native Mandarin speakers (10 women, mean age = 22.8, SD = 8.2) and sixteen right-handed naïve speakers with no knowledge of Mandarin (8 women, mean age = 20.7, SD = 2.3) took part in the experiment. Eight of the native

speakers had also participated in Experiment 1 (see Supplementary File 1 for details). Like in Experiment 1, the naïve speakers were native speakers of a variety of languages, but none had experience with tone languages. All participants provided their informed consent and were paid for their participation, and experimental procedures were approved by the Institutional Review Board of New York University Abu Dhabi. Detailed demographic information about the participants is available in Supplementary File 1.

3.1.2 Materials

The materials were *wu* ([u]) monosyllables extracted from the same recording as that used in Experiment 1, and processed in the same way. *wu* rather than *yi* was used in order to increase the generality of the results. Like *yi*, each tonal realization of *wu* corresponds to at least one real word; the log parts-per-million token frequencies (from Da, 2004) for the syllables are 3.24 (wu^2), 3.19 (wu^3), and 3.38 (wu^4).

3.1.3 Procedure

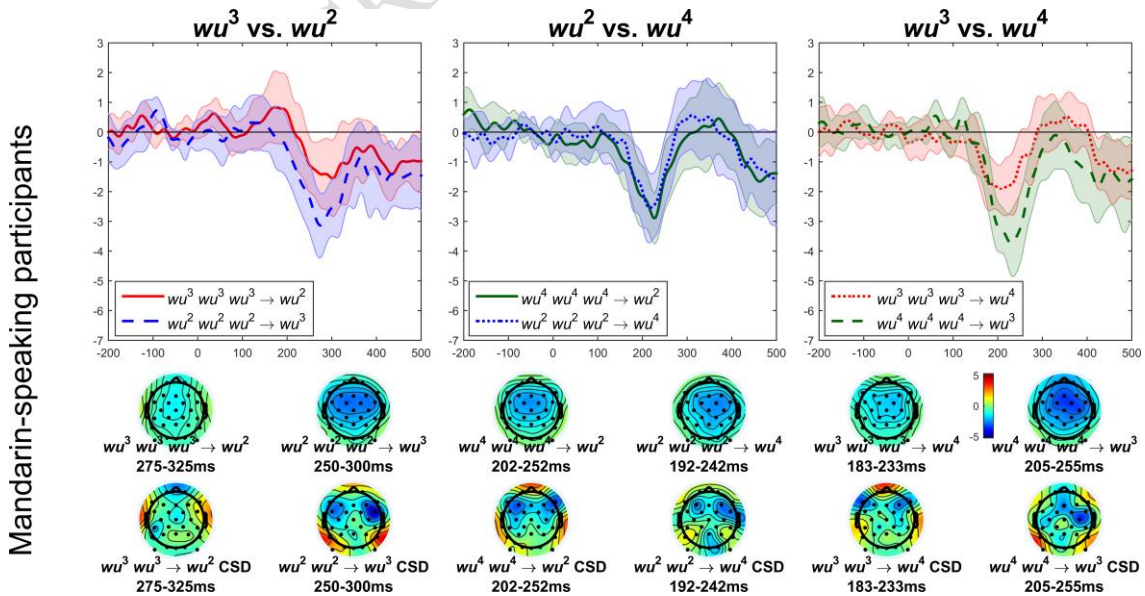
The procedure was identical to that in Experiment 1, except that participants were no longer asked to list the sounds they heard, and each sub-block had 265 standards and 40 deviants with an additional 20 standards at the beginning of each sub-block (13.1% deviants not counting the initial standards; for a total of 120 deviants and 855 standards per condition). This was the number of standards and deviants that had been intended for Experiment 1 as well, but a programming error in the script used to control stimulus presentation in Experiment 1 had caused a lower number of trials to be presented in that experiment.

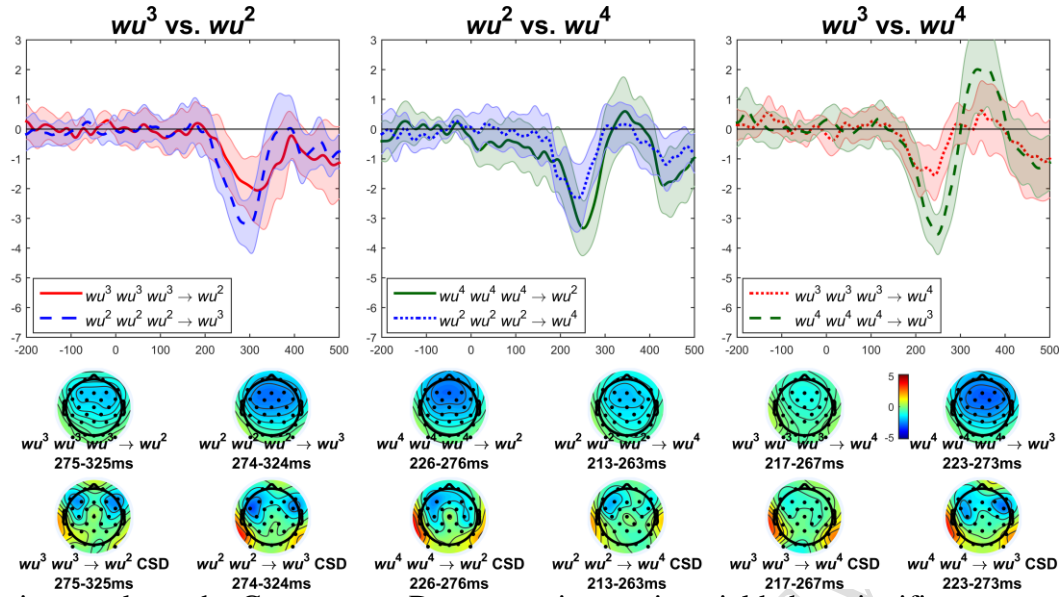
3.1.4 Data acquisition and analysis

Data acquisition and analysis were identical to that in Experiment 1, except that only 34 electrodes were used, and that for one participant we interpolated two bad channels. The lowest number of deviant trials remaining per condition for any participant was 35 (see Supplementary File 1 for detailed counts).

3.2 Results

The MMNs for each direction of each contrast, at electrode Fz, are shown in Figures 4 (Mandarin-speaking participants) and 5 (non-Mandarin-speaking participants), and plots of the standard and deviant waves are available in Supplementary File 4. Upon visual inspection the effects appear similar, albeit weaker, than those in Experiment 1: the T3~T2 (phonologically related) and T3~T4 (phonologically unrelated) contrasts yielded asymmetries for each group, whereas the asymmetry was less robust for the T2~T4 contrast. In other words, contrasts including T3 (low tone) elicited stronger asymmetries. These findings were confirmed in the statistical analysis.





For native speakers, the CONTRAST \times DIRECTION interaction yielded no significant

effects ($p > .343$), but the main effect of DIRECTION did ($p = .024$). But because of our *a priori* interest in the effect of DIRECTION within each contrast, our predictions based on Experiment 1, and the fact that DIRECTION is a dummy factor that is not meaningful on its own, we proceeded to run a spatiotemporal clustering test of DIRECTION within each contrast, in order to clarify the nature of the asymmetries. For the T3~T2 contrast, there was a significant negativity ($p = .018$, Bonferroni-corrected based on 3 comparisons), lasting from 185 to 294 ms and including the 21 channels F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P3, Pz, and FCz. Because the test was coded as reflecting the $wu^2 \rightarrow wu^3$ MMN minus the $wu^3 \rightarrow wu^2$ MMN, the fact that the cluster was negative indicates that the MMN was more negative (i.e., larger) when T3 was the deviant and T2 the standard, compared to when T3 was the standard and T2 the deviant. For the T3~T4 contrast there was a significant negativity ($p = .024$, Bonferroni-corrected based on 3 comparisons), lasting from 194 to 268 ms and including the 13 channels Fp2, Fz, FC1, FC2, FC6, C3, Cz, C4, T8, CP2, CP6, Pz, and FCz; again, this cluster was due to the MMN's being attenuated when T3 was the

standard, compared to when T4 was the standard. Finally, for the T2~T4 contrast, there was no significant effect (no clusters found).

For non-Mandarin-speaking listeners, again the CONTRAST \times DIRECTION interaction was not significant (no clusters) but the DIRECTION main effect was ($p = .022$), so we proceeded to run three more spatiotemporal clustering tests to examine the simple effect of DIRECTION within each contrast. There was a significant asymmetry for the T3~T2 contrast ($p = .030$, Bonferroni-corrected)¹³ based on a negative cluster from 219 to 300 ms and including the 20 channels F3, Fz, F4, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P3, Pz, P4, and FCz; there was also a significant asymmetry for the T3~T4 contrast (Bonferroni $p = .012$) due to a negative cluster from 209 to 299 ms and encompassing the same 20 channels. In both of these cases the asymmetry was due to smaller MMN amplitudes when the standard was T3 than when the standard was a different tone. For the T2~T4 contrast, the asymmetry was not significant ($p = .168$).

3.3 Discussion

Experiment 2 approximately replicated the results of Experiment 1. In short, even when using a variant of T3 (low tone) with a less complex contour, asymmetric MMN effects between T3 and other tones were still observed, and again the effects were such that MMN was weaker when T3 served as a standard and another tone as the deviant than vice versa. As in Experiment 1, the same effect was seen in naïve listeners and thus is likely not due to phonological knowledge. The MMN for the T2~T4 contrast did not show significant asymmetries for either group (whereas in Experiment 1 there was a

¹³ In the analysis of data that were not corrected with ICA, this effect was non-significant, although it was numerically in the same direction.

significant asymmetry for Mandarin speakers, but its magnitude was smaller than that for the contrasts involving T3), although it is clear from Figures 2-5 that it often showed trends in the same direction; as we did not have *a priori* predictions about MMN asymmetries for this contrast and because they were not reliably replicated across experiments and speaker groups, we have little to say about this effect here and instead leave it to future research.

Since the asymmetry in contrasts involving T3 again appeared in both speaker groups, it still cannot be attributed to phonological factors. At this point it is difficult to say what aspect of the stimulus caused the asymmetry. We note that, while we considered half T3 to be roughly comparable in contour complexity to the other tones, there are still numerous differences between the contours (see Figure 1), so there are many potential candidates for acoustic or phonetic features that could have driven the asymmetry. More systematic research on the role of tonal contour shape in the generation of MMN asymmetries (along the lines of, e.g., Gu, Wang, He, & Chen, 2012; Tsang et al., 2011) is needed to determine what frequency aspects of the signal have the greatest influence on the MMN generator during change detection in the context of lexical tones.

Another possibility is that the asymmetry is not due to greater acoustic salience of T3, but to a cross-linguistic perceptual bias towards this tone. Such cross-linguistic perceptual biases are well documented in the perception of vowels, where peripheral (as opposed to central) vowels have been argued to act as Natural Referent Vowels (Polka & Bohn, 2011); Natural Referent Consonants have also been proposed to exist in the perception of consonant place of articulation (Bundgaard-Nielsen, Baker, Kroos, Harvey, & Best, 2015). It is unclear, however, how such an account would translate into tone perception. One proposal for why corner vowels act as Natural Referent Vowels is

that they serve as "anchors" that mark the extent of the vowel space for a speaker and thus aid in talker normalization; Yeung and colleagues (2013) have suggested that Cantonese tones that cover most of the pitch space may have a similar effect, although this would not explain the present results since T3 covers less of the pitch space than T4 and yet clearly showed a perceptual bias over T4 in our experiments. Another proposal is that the vowels which are most distinguishable from each other serve as Natural Referent Vowels (Bundgaard-Nielsen et al., 2015), but this also does not appear to apply to the present results; for example, T3 and T4 are much more distinguishable from each other than T3 is from T2 (Howie, 1968), which would suggest that these tones might be perceptual anchors compared to T2 and thus would each show asymmetries with T2 and not with each other; however, T3 showed a strong asymmetry with T4. Finally, Natural Referent Vowels may have a special status because they happen to exhibit acoustic formant convergence (Schwartz, Abry, Boe, & Vallée, 2005), but it is not clear how this concept, would translate into the perception of lexical tones, which are cued mainly by fundamental frequency (pitch) rather than formant frequency (vowel quality). Nonetheless, the present dataset cannot conclusively distinguish between acoustic-salience and perceptual-bias accounts of the asymmetry; therefore, we leave this issue to future research, and here we only conclude that the present experiments show evidence for a non-language-dependent, non-phonological basis for the asymmetry, while remaining agnostic about what the specific locus of that asymmetry was.

It is possible that Experiments 1 and 2 did not show phonological effects because of the nature of the paradigm used. Each tonal category was only represented by one repeated token, meaning that the listener could detect change based only on low-level perceptual features, and had no need to abstract the features away and realize

phonological category representations. To minimize this sort of low-level processing and maximize the need for abstract phonological processing, in Experiment 3 we use a large variety of different tokens within each tone category. In a design like this, the stimuli only fall into an oddball arrangement at an abstract phonological level, and thus MMN effects are more likely to reflect phonological categorization (Eulitz and Lahiri, 2004; Hestvik & Durvasula, 2016; Kazanina et al., 2006; Phillips, Pellathy, Marantz, Yellin, et al., 2000; among others). In addition to using multiple different tokens of each tonal contour (under which situation Mandarin tones have been shown to yield categorical perception effects in MMN; Xi et al., 2010), we also superimposed these various tonal contours on several different carrier vowels to yield both phonetic and phonological variability within tone categories; this sort of design has yielded MMNs for abstract voicing categories (Phillips, Pellathy, & Marantz, 2000) and lexical tone categories (Wang et al., 2012). Thus, this is an ideal paradigm to test for phonologically-driven asymmetries in the MMN elicited by tonal contrasts.

4 Experiment 3

4.1 Methods

4.1.1 Participants

Twenty-three right-handed native Mandarin speakers (17 women, mean age = 24.4, SD = 9.6) and twenty-three right-handed naïve speakers who were native speakers of various non-tonal languages (12 women, mean age = 20.75, SD = 1.8; one of these speakers had studied Japanese as a second language) took part in the experiment; three of the Mandarin speakers were recruited via word-of-mouth from outside the university community. Eight of the native speakers and one of the naïve speakers had also

participated in Experiment 1 and/or 2 (see Supplementary File 1 for details). All participants provided their informed consent and were paid for their participation, and experimental procedures were approved by the Institutional Review Board of New York University Abu Dhabi. Detailed demographic information about the participants is available in Supplementary File 1. It was necessary to recruit more participants than were run in Experiments 1 and 2, because each participant contributed fewer trials (see section 4.1.3). An additional nine participants joined the experiment but were not included in the final analysis: five (three Mandarin-speaking and two non-Mandarin-speaking) because of excessive artifacts in their recording, two Mandarin-speaking left-handed participants, and two non-Mandarin-speaking participants who had to abort the experiment before completing the session.

4.1.2 Materials

Stimuli were spoken by a female native speaker of Mandarin from Beijing using the same setup as in Experiment 1. All four tones were produced on five carrier vowels, the monophthongs *yi* ([i]), *wu* ([u]), *yu* ([y]), *a* ([a]), and *e* ([ɤ]), repeated ten times in random order and ten times in a fixed order. *yi*, *wu*, *yu*, and *e* occur in all four tones, while *a* only occurs in tone 1 or without tone (in unstressed positions, e.g. as a particle). The type frequencies of tones 1 through 4 regardless of segmentals (expressed as the number of legal syllables with each tone) are 258, 224, 254, and 358, and the token frequencies (expressed as the percentage of tokens in the corpus with a given tone) are 16.7%, 18.4%, 14.8%, and 42.5%.

After noise reduction in Audacity (Audacity Team), further processing was carried out in Praat (Boersma & Weenink, 2014). We selected three clear tokens of each tone (for T3, tokens with the half T3 reduced contour and without glottalization were selected), and one clear token of each vowel. All tokens were normalized to 300 ms

duration, and the fundamental frequency of each tone was extracted and overlaid onto each vowel, yielding 60 tokens ($4 \text{ tones} \times 3 \text{ tokens of each tone} \times 5 \text{ vowels}$). These were normalized to 75 dB in intensity, and played in isolation to several native Mandarin speakers, who verified that they were still identifiable as the expected syllables.

Because the present study used the Schröger and Wolff (1996) method of calculating MMNs, we needed to record seven tones. Specifically, because deviants consisted of 1/7 of the trials in experimental blocks (where the remaining 6/7 of trials were standards), they had to be compared to trials in a control block where the same tokens made up 1/7 of the trials but where there was no standard (i.e., where six other tones each made up another 1/7 of the trials). In other words, the control blocks consisted of seven different tones, each presented at a 1/7 ratio relative to the other tones. As Mandarin only has four lexically distinctive tones, we created three artificial tones by extracting and modifying portions of English speech from the recording (portions where the experimenter was conversing with the speaker). No specific criteria were used to select portions, except that they had to be portions of speech with a pitch contour that corresponded to no existing tone category in Standard Mandarin. The pitch contours from these speech portions were shifted up, shifted down, or left unchanged to yield three contours for each artificial tone, and then overlaid onto the five vowel tokens, just like the real tones were, to create an additional 45 stimuli ($3 \text{ tones} \times 3 \text{ tokens of each tone} \times 5 \text{ vowels}$) for use in the control blocks (see 4.1.3, Procedure). Artificial tones were necessary to ensure that each tone took up 1/7 of the trials (rather than 1/4) in a control block, which then would be comparable to the proportion of trials a tone took up when presented as a deviant in any contrast. The artificial tone stimuli are available, along with the critical stimuli, in Supplementary File 2.

4.1.3 Procedure

Unlike Experiments 1 and 2, in Experiment 3 we also tested contrasts involving T1; thus, there were six contrasts overall, with two presentation directions per contrast, yielding twelve types of blocks: T1→T2, T2→T1, T1→T3, T3→T1, T1→T4, T4→T1, T2→T3, T3→T2, T2→T4, T4→T2, T3→T4, and T4→T3. Each of these block types was divided into two (rather than three) sub-blocks, and each sub-block included 238 standards and 40 deviants (14.4% deviants), with an additional 20 standards at the beginning of each sub-block, such that each condition overall had 516 standards and 80 deviants. Reducing the number of trials relative to Experiments 1 and 2 was necessary to keep the experiment relatively short, given the increased number of conditions.

Furthermore, an additional block type was added: a control block, in which tokens from each of the four tones, plus tokens from each of the three artificial tones (see 4.1.2, Materials) were presented with equal frequency—meaning that each tonal category in this block was presented at a 1/7 (14.3%) ratio compared to other tones, approximately equal to the ratio of deviants to standards in normal blocks. MMNs were calculated by subtracting ERPs of a given sound when presented in this control block from the ERPs of the same sound when presented as a deviant. Calculating MMNs in this way, rather than subtracting ERPs of the sound when presented as a standard, reduces the influence of N1 refractoriness on the MMN time window, thus giving a more conservative estimate of the MMN component itself rather than a combination of MMN and N1 (Schröger and Wolff, 1996; see the Introduction for a more detailed explanation of this point). The addition of this control block (divided into two sub-blocks like the others) meant that the experiment included 26 sub-blocks, from 13 block

types. The experiment took about 2.5 hours. The procedure was otherwise identical to that of Experiment 2.

4.1.4 Data acquisition and analysis

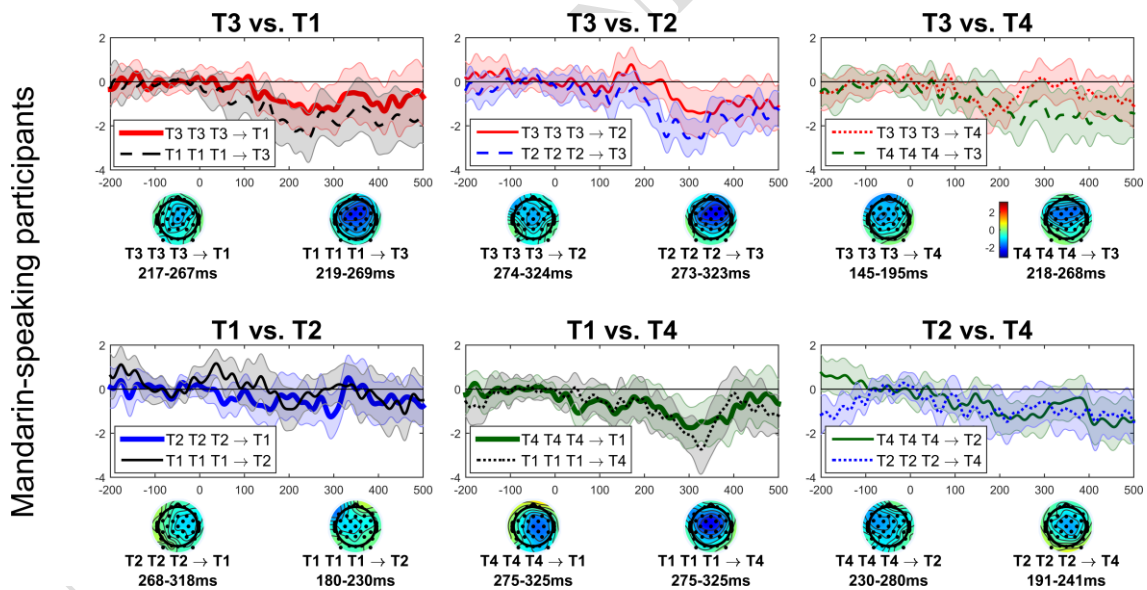
Data acquisition and analysis were the same as in Experiment 2, except for how the MMNs were calculated. Rather than subtracting a standard waveform from the deviant waveform, it was the waveform elicited by a given sound in the control block that was subtracted from the waveform elicited by the same sound as a deviant in each contrast. The lowest number of trials remaining per condition for any participant was 23 (see Supplementary File 1 for detailed counts). Furthermore, cluster analyses were carried out over the 0-400 ms time window, rather than 100-300 ms, because the effects were less focal in this experiment than in the previous ones and the narrower time window did not capture the full extent of all effects; analyses with the 100-300 ms time window yielded the same pattern of results but sometimes yielded clusters encompassing the entire time window.

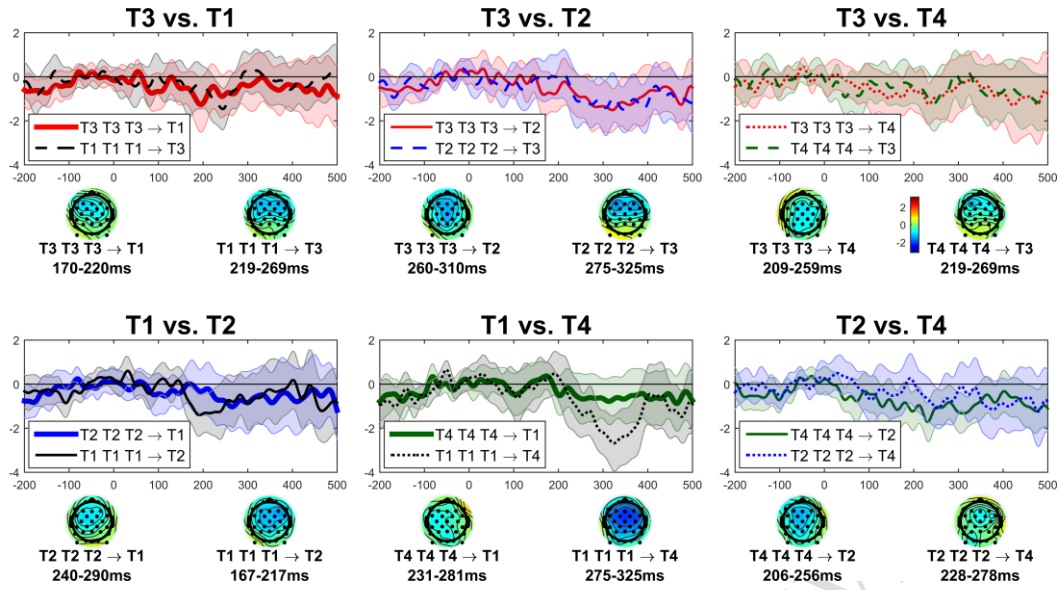
4.2 Results

The MMNs for each direction of each contrast, at electrode Fz, are shown in Figures 6 (Mandarin-speaking participants) and 7 (non-Mandarin-speaking participants), and plots of the standard and deviant waves are available in Supplementary File 4. The most striking aspect of the results is that the MMNs are far weaker, with less prominent peaks, than those in the previous experiments. This is to be expected, given that the contrast in this experiment was made much less obvious via the introduction of substantial within-category variability. While it is possible to observe clear MMN peaks under substantial stimulus variability, it requires a large number of stimuli (for instance, while Wang and colleagues, 2012, and Pakarinen and colleagues,

2010, both showed relatively clear MMN peaks, they presented participants with 360 and 180 deviant tokens per condition, respectively, whereas we were only able to present 80 per condition because of the large number of conditions). The use of a different subtraction procedure may also have contributed to the difference, although the plots in Supplementary File 4, using the same subtraction procedure as in the previous experiments, show a similar pattern as the pattern obtained with the new subtraction procedure; see discussion in section 4.3.

Upon close inspection, though, the results for Mandarin-speaking participants once again show a trend towards weaker MMNs when T3 (low tone) is the standard and another tone is the deviant. Furthermore, no other tone shows an across-the-board asymmetric trend like this. For non-Mandarin-speaking participants no clear pattern is evident. These findings were confirmed in the statistical analysis.





For native speakers, the CONTRAST \times DIRECTION interaction yielded a non-significant effect ($p = .140$), based on a cluster extending from 281-329 ms and including 13 channels: F7, F3, Fz, FC5, FC1, FC2, T7, C3, Cz, CP5, CP1, P3, and FCz. The main effect of DIRECTION was also nonsignificant ($ps > .123$). But given that this test is rather conservative and could not take advantage of the structure of the conditions (for example, the T3~T1 contrast is related to the T3~T2 contrast more closely than it is to the T2~T4 contrast), and the fact that we had clear predictions (based on the results of Experiments 1 and 2) about the nature of the interaction, we conducted planned comparisons to take advantage of our *a priori* predictions. For each tone, we constructed a measure of the asymmetry effect by subtracting the average of all MMNs in which that tone was a standard from the average of all MMNs in which that tone was a deviant (for instance, the T3 asymmetry effect was represented as $\{\text{mean}(\text{T1} \rightarrow \text{T3}, \text{T2} \rightarrow \text{T3}, \text{T4} \rightarrow \text{T3}) - \text{mean}(\text{T3} \rightarrow \text{T1}, \text{T3} \rightarrow \text{T2}, \text{T3} \rightarrow \text{T4})\}$). We then ran a two-tailed cluster test on each asymmetry effect (Bonferroni-corrected based on four comparisons) to test for across-the-board asymmetries for any given tone; note that, while a two-tailed test

is overly conservative for T3 (the case where we had an *a priori* prediction to observe negative asymmetries across all contrasts in which this tone is involved), we used it since we did not predict any particular direction of asymmetry for other tones. T3 yielded a significant asymmetry effect ($p = .034$) due to a cluster from 209-353 ms including the 25 channels Fp2, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, Pz, P4, P8, Oz, O2, PO10, and FCz.¹⁴ None of the other tones yielded significant asymmetries ($ps > .238$).

¹⁴ Follow-up tests for each tone pair within this cluster, Bonferroni-corrected based on three comparisons, showed a significant negativity for the T3~T2 contrast ($t(22) = -2.86$, 95% CI = $-2.07 \dots -0.33$, $p = .028$), a marginal negativity for the T3~T4 contrast ($t(22) = -2.23$, 95% CI = $-1.81 \dots -0.07$, $p = .108$), and a trend towards negativity for the T3~T1 contrast ($t(22) = -1.75$, 95% CI = $-1.79 \dots 0.15$, $p = .284$). The fact that not all comparisons were significant arguably challenges the interpretation of the data as reflecting an across-the-board asymmetry for T3. We believe, however, that the data still suggest an across-the-board asymmetry, because this was a conservative, Bonferroni-corrected test, and because, regardless of the statistical outcome, each contrast involving T3 showed a numerical asymmetry in the expected direction. Given that the present paradigm makes it very difficult to detect MMNs at all (for reasons described in the text above) and that the effects in the present study were not surprising differences but rather were replications of effects that had already been observed in the two preceding experiments and had been specifically predicted here, we argue that the asymmetries observed in this experiment can still be considered real even in the face of somewhat equivocal statistical tests. Treating the three tone pairs as separate studies and comparing them using meta-analytic techniques (Braver, Thoenes, & Rosenthal, 2014) also provides evidence for an across-the-board effect; specifically, the Q-test, which assesses the homogeneity of a set of effect sizes (Rosenthal & Rubin, 1982) was non-significant ($\chi^2(2) = 0.27$, $p = .128$), suggesting that all three tone pairs elicited similar

For non-Mandarin-speaking participants, the CONTRAST×DIRECTION interaction was not significant ($p = .868$), nor was the main effect of DIRECTION ($p = .403$). For the sake of comparison with the Mandarin-speaking participants, we tested for across-the-board asymmetries with any given tone (using the same method described above for the analysis of Mandarin-speaking participants); again, no contrasts were significant ($ps > .702$).

4.3 Discussion

For Mandarin-speaking listeners, the results approximately replicated the previous experiments, with a trend (in a conservative analysis) towards asymmetric MMN effects when T3 (low tone) was contrasted with other tones, and no asymmetries in other contrasts that did not involve T3.

The effects in this experiment were weaker than in Experiments 1 and 2, and only reached statistical significance in planned comparisons. Nonetheless, as these effects were essentially replicating effects that we had predicted and had already observed in two prior experiments, and because they appeared in the specific comparisons where we had expected them to appear, and not in the other comparisons, we consider them sufficient evidence for the presence of a real asymmetry effect. We also note that the effects in question (the CONTRAST×DIRECTION asymmetry, and the simple effects of Direction in the T3~T1 and T3~T2 contrasts, for native speakers and not for naïve speakers) were significant in an exploratory analysis of the MMNs computed via standard subtraction (data available in Supplementary File 3) rather than

MMN asymmetries. Likewise, the pooled z -score across the three tone pairs (Braver et al., 2014) was -3.16 ($p < .001$), suggesting that the three tone pairs provided strong evidence for a real asymmetry when analyzed as a whole.

the Schröger & Wolff (1996) method, although we consider the latter method superior for the reasons described above.

Crucially, in this experiment the non-Mandarin-speaking listeners did not show the same pattern as the Mandarin-speaking listeners, failing to show asymmetries for contrasts involving T3. Therefore, unlike the previous experiments, it is possible to argue that the asymmetries observed for Mandarin speakers in this experiment are due to phonological knowledge.

The main difference between Experiment 3 and the previous experiments was the inclusion of within-category token variability in Experiment 3. This is the most likely explanation for the difference in the results; within-category variability forced the language processing system to generate abstract representations of the standards, thus enhancing the presence of phonological effects on the MMN, while simultaneously decreasing the presence of low-level perceptual effects by ensuring that it was not the case that standards and deviants would differ on only one physical feature. While the use of within-category token variability does not always rule out perceptual biases (for example, infants in Tsuji et al., 2015, show putatively perceptual asymmetries in the face of variable tokens), it likely does for adults who have a phonological system to fall back on, particularly in the oddball MMN paradigm (see, e.g., Hestvik & Durvasula, 2016; Phillips, Pellathy, Marantz, Yellin, et al., 2000).

There are two other potential explanations for the difference between experiments that we can rule out. Firstly, the difference is not due to the use of a different subtraction method for calculating MMNs in the present experiment (see section 4.1.3, Procedure); as shown in the figures in Supplementary File 4, roughly the same pattern of results—with larger asymmetries, if anything—obtains when the data are analyzed using the same subtraction method that was used in Experiments 1 and 2. It is possible, however,

that the asymmetries observed in Experiments 1 and 2 reflected a summation of phonological effects and N1 refractoriness effects for Mandarin speakers, and only N1 refractoriness effects in the non-Mandarin-speaking participants, which would then explain why an asymmetry remained for Mandarin speakers but not non-speakers in Experiment 3 when the N1 refractoriness effect was removed. The fact that Mandarin-speaking participants showed smaller asymmetries than non-Mandarin-speaking participants in Experiment 1, however, seems inconsistent with this account since, if the effects of phonology and N1 refractoriness were additive, then Mandarin speakers should always have shown larger asymmetries. Thus, we believe the disappearance of the asymmetry for non-Mandarin-speaking participants in Experiment 3 is not due to the use of a different subtraction technique, but rather is probably due to the use of a high token-variability paradigm. Secondly, one might be concerned that the failure for non-Mandarin-speaking participants to show asymmetries in Experiment 3 may be due to a floor effect. In other words, because the design made it very difficult to detect categorical differences at all, putting non-Mandarin-speaking participants at the floor in terms of MMN generation, there may have been insufficient variability left to detect asymmetries. While this may be true, it does not change the interpretation of the difference as having its locus in phonology. Other than the two groups' differing phonological systems and experience with tones, there is no reason for the non-Mandarin-speaking group to show a floor effect when the Mandarin speakers do not. In other words, if the MMN asymmetries for Mandarin speakers were due to any non-phonological effect, they should also have been observed in the non-speakers. We also note that, while the MMN effects were small, there were still significant MMNs for each group (in one-tailed cluster tests comparing the average of all MMNs against zero, Bonferroni-corrected based on two comparisons; a $p = .004$ effect for Mandarin

speakers and a $p = .036$ effect from 222-334 ms for non-speakers). This suggests that non-Mandarin-speaking participants were still able to detect categorical differences (which is not surprising, given that abstract difference detection with a complex multidimensional stream of standards is known to be possible for contrasts that are not phonologically meaningful, as demonstrated by Pakarinen and colleagues, 2010), but that the abstract representations they were able to develop for tone were not phonologically complex (i.e., with some representations more specified than others) in the way that native speakers' representations were.

5 General discussion

In the present study we have presented three experiments demonstrating nearly across-the-board asymmetries in native Mandarin speakers' neural processing of the lexical tone 3 (T3), the low tone. From an empirical standpoint, the results demonstrate a special status of T3 in tone perception, which was not reported previously since most studies on lexical tone perception have not compared both presentation orders within a single contrast (see section 1.1). Furthermore, the results also contribute to our broader understanding of MMN asymmetries. While most research on asymmetries in the mismatch negativity (MMN) paradigm that is grounded in theoretical phonology has tested only one group of speakers, the present study took advantage of the language-specific status of T3, as a construct that does not exist in the phonological systems of listeners who do not know Mandarin, to test for phonological contributions to perceptual asymmetries in both native listeners and naïve listeners. This allowed us to identify both cases in which the asymmetry is driven by low-level, language-universal phonetic or acoustic factors (i.e., when both speaker groups show an asymmetry) and cases in which it is driven by phonological knowledge (when only the native Mandarin-

speaking participants show an asymmetry). Below we discuss the insights of the present study into the underlying causes of MMN asymmetries (with a particular focus on the factors contributing to asymmetries in lexical tone perception), and the consequences of the present findings for the phonological representation of lexical tones.

5.1 Different causes of MMN asymmetries

As reviewed in the Introduction, there are numerous factors that can cause a single contrast to yield a larger MMN in one presentation order than another. This is even possible when the MMN subtraction is based on identical tokens used respectively as standard and deviant—the asymmetries described above are all still observable even when using this subtraction method. Therefore, it is crucial to be able to rule out alternative explanations before concluding that an MMN asymmetry indicates something special about the abstract phonological status of one sound.

When a contrast shows an asymmetric MMN, the asymmetry could be explained in terms of the standards or in terms of the deviants. For instance, among the factors reviewed in the Introduction, lexicality effects on the MMN appear to be wholly based on the lexicality of the *deviant* (Shtyrov & Pulvermüller, 2002), whereas underspecification-related effects are argued to be based on the phonological representation of the *standard* (Eulitz & Lahiri, 2004, among others). This raises the question, then, of whether asymmetries related to T3 occur because the MMN is attenuated in blocks with T3 standards, or is amplified in blocks with T3 deviants. These are not necessarily empirically distinguishable on the basis of the ERP data alone in the present study. These accounts cannot be distinguished just on the basis of comparing standard and deviant waves specifically (as opposed to comparing MMN

difference waves, as was done in the present study) to see which was modulated.¹⁵ First of all, the different tones have different physical properties that could cause them to elicit different exogenous components of ERP regardless of how they are phonologically or perceptually encoded—indeed, subtracting out these low-level acoustic effects is one of the reasons the MMN is a valuable paradigm for comparing brain responses to different speech sounds. Secondly, accounts that place the locus of the MMN effect on differences in the deviant or differences in the standard do not necessarily predict that differences will be confined to the deviant ERP waveform or to the standard ERP waveform. For example, while the underspecification account of MMN asymmetries assumes that asymmetries arise because underspecified standards are encoded differently, this does not entail that underspecified standards will elicit different ERPs on their own. Rather, the effect is still based on the MMN that occurs only after the deviant is heard: specifically, the underspecified memory representation that is derived from the standards causes the mismatch between deviant and standard to be weaker, thus attenuating the MMN that is elicited when the deviant is perceived. Crucially, even though the asymmetry is based on how the standard is encoded, the point at which the asymmetry is reflected in the ERP is still when the deviant is perceived. The direct ERP correlates of making specified versus underspecified memory representations (i.e., what is happening when the standards are perceived) are not known, and indeed would be difficult to detect given that this would require comparing raw ERPs elicited by acoustically different sounds.

¹⁵ For the present dataset, inspection of the standard and deviant waves separately (Supplementary File 4) shows that the standard waveforms were relatively constant across all contrasts involving T3, and the differences arose in the deviant waveforms.

Even though the data alone cannot necessarily tell us whether the MMN was attenuated when T3 was the standard or amplified when T3 was the deviant, these accounts can be adjudicated between via of the predictions of the various asymmetry mechanisms outlined in the Introduction. For example, the non-language-specific effects observed in Experiments 1 and 2 could not be due either to the lexicality of the deviants or the category-prototypicality of the standards, since T3 and other tones do not differ in these respects for listeners who do not speak Mandarin; nor are they likely to be due to phonological underspecification (see section 2.3). Therefore, by process of elimination, the most likely accounts for the effect are those based on the acoustic salience or perceptual bias of the deviant compared to the standard. On the other hand, these explanations are not applicable in Experiment 3 because they predict asymmetries to occur in both speaker groups; for that experiment, underspecification accounts for the results better than any of the other accounts reviewed in the Introduction.

In the present study, when a single token of each sound was used (Experiments 1 and 2), asymmetries were seen in both Mandarin-speaking and naïve participants. If we had not tested naïve participants, it would have been tempting to attribute the asymmetry in native speakers to phonological knowledge, particularly since some other studies using single tokens to yield an MMN asymmetry have also made representational claims (e.g., de Jonge & Boersma, 2015; Scharinger, Bendixen, et al., 2012; Walter & Hacquard, 2004). The fact that it was observed in naïve participants as well, however, suggests that the asymmetry has a low-level perceptual basis. As noted in the Introduction, subtracting standard ERPs from the same tokens that were used as deviants does not necessarily remove phonetic/acoustic contributions or other non-underspecification-based contributions from the MMN. Furthermore, while it is possible for a given study to argue against a phonetic or acoustic explanation of the result—for

instance, if certain acoustic properties of the stimuli would not have predicted the same result that was observed—such an argument can probably never be comprehensive, as there might always be another acoustic factor which we have not thought of or measured and which explains the results. For the same reason, creating nonlinguistic control stimuli (such as harmonic noise) that are matched to various acoustic properties of the linguistic stimuli will not necessarily dissociate all possible perceptual sources of the effect (furthermore, when an acoustic cue encodes phonological contrast in a language, that acoustic cue may elicit linguistic MMN effects even in nonlinguistic stimuli; see Gu et al., 2013). Using a single-token experiment as a control experiment for comparison against a multiple-token experiment, as was done by Hestvik and Durvasula (2016), is an effective means to rule out acoustic explanations of MMN asymmetries in some cases, although in the present study such an approach would have yielded a false negative: in other words, if we only compared the multiple-token native speaker experiment (Experiment 3) to the single-token native speaker experiments (1 and 2) we would have concluded that the asymmetry was not phonologically based, whereas comparison to the multiple-token naïve speaker experiment revealed a phonological basis. In cases like the present one, where there are both low-level perceptual and higher-level phonological factors contributing to an asymmetry, the asymmetry may be observed in both single-token and multiple-token paradigms. Likewise, a cross-linguistic comparison alone would not necessarily support a language-specific phonological account of MMN asymmetries, as if we had run Experiment 3 alone then the lack of asymmetry for naïve listeners may have been because they were unable to auditorily encode the tones they heard. The combination of the cross-language and cross-method comparisons in the present study, though, suggests that both speaker groups are able to auditorily encode the tones and show lower-level, language-universal

perceptual effects in a single-token paradigm that does not strongly tap into phonological representations, whereas only the Mandarin-speaking group showed sparse phonological encoding in the multiple-token paradigm that examines phonological representations. Thus, we believe that, for testing abstract phonological accounts of MMN asymmetries, the gold standard should be combining both cross-linguistic comparisons and within-language, across-paradigm comparisons.

For many underspecification-related asymmetries, such a comparison may not be possible, if the same segment is underspecified universally. For example, coronal place of articulation, e.g., the place of /t/ as opposed to /k/ or /p/, is considered to be underspecified in most languages (Paradis & Prunet, 1991). There may be a physiological basis for the prevalence of coronal underspecification—the tip of the tongue is a faster and easier-to-move articulator than most others, and the coronal area (the back of the teeth and alveolar ridge) is easier to reach with this articulator than other places of articulation are, so there are strong articulatory reasons for such a place of articulation to be an unmarked default. Thus, when testing a perceptual asymmetry between coronal and non-coronal segments, for example, there might be no appropriate "na ïve" language group to use as a control. This is why Mandarin lexical tone was a useful test case in the present study: it is very unlikely that non-Mandarin-speaking listeners would have any reason to have a less specified representation of T3 than other tones. In fact, because of the inclusion of this group, we believe the present study offers some of the strongest support to date for the argument that asymmetries in MMN can reflect sparse phonological representations (Eulitz & Lahiri, 2004, among others). Other contrasts that would be useful test cases include those in which the underspecified-specified relationship differs across languages. Voicing or aspiration contrasts are an example of this (Keating, 1984). For instance, in languages like English and Mandarin,

/p/ is specified as aspirated, and /b/ underspecified (unaspirated). On the other hand, in languages like Spanish and Modern Standard Arabic, /b/ is specified (as voiced) and /p/ underspecified (voiceless). In a case like this, observing a reversal of the MMN asymmetry (i.e., an asymmetry in one direction for one language group and in the opposite direction for another) would be even stronger evidence for underspecification than what was shown in the present study.

Testing multiple speaker groups in the present study allowed us to reveal both cases in which the asymmetry appears to be based on language-universal perceptual biases, and cases in which it is likely based on language-specific phonological structure. This finding is consistent with research on infant perception showing that universal perceptual biases might account for some perceptual asymmetries, such as the commonly observed coronal-labial asymmetry (Tsuji et al., 2015). The present results, in which both language-specific and non-language-specific effects were observed in similar paradigms, suggest that phonological and phonetic/acoustic accounts of these asymmetries can be reconciled: specifically, they suggest that both factors play a role in perception, and that the single-token paradigm used in Experiments 1 and 2 drew on more cross-linguistic perceptual biases (thus yielding asymmetries for both language groups) whereas the high-variability token paradigm used in Experiment 3 instead tapped into language-specific sound organization (thus yielding an asymmetry only for the Chinese-speaking group). Such a result would be expected if, as suggested by Tsuji and colleagues (2015), acoustic or phonetic biases are in of themselves the reason why underspecified phonological representations develop. Under such an account, one could predict that speakers of both languages have perceptual biases for tones, which emerged very early in development or were innate, but that these biases had no effect on the development of non-tonal phonological systems (where such biases are not relevant to

the categorization of native speech sounds) whereas these biases became part of the architecture of tonal phonological systems. Thus, in the present study, Mandarin-speaking participants showed reflections of these biases both in paradigms that tap phonological representations (where such biases are explicitly built into the language architecture, perhaps by way of underspecification) and those that tap phonetic/acoustic discrimination (where such biases are perhaps an epiphenomenon of other aspects of the perceptual system). Non-Mandarin-speaking participants, on the other hand, only showed these biases in a paradigm that taps phonetic/acoustic discrimination and not in a paradigm that taps their phonological organization.¹⁶ In short, the present results link and reconcile other recent studies showing that perceptual asymmetries can be grounded in cross-linguistic perceptual biases (Tsuji et al., 2015, and the MMN literature reviewed in the Introduction) and that both phonological abstraction and phonetic/acoustic biases can coexist in the language perception system (e.g., Hestvik & Durvasula, 2016).

While the present study provided strong evidence that phonetic/acoustic perceptual biases contribute to asymmetries in tone perception (Experiments 1 and 2), it remains unclear just what set of acoustic features caused these asymmetries. There are numerous differences between the tone contours tested, and any one of these differences is a candidate; further research is necessary to shed light on the nature of this perceptual effect. While we are unable to explain the effect at present, it would be a valuable avenue for future study, as better understanding the cause of this effect can also inform our understanding of how different pitch cues are used during speech perception and what causes cross-linguistic perceptual biases in the perception of tones. As for

¹⁶ We would like to thank an anonymous reviewer for suggesting this characterization of the results.

phonological cues, the present study allowed us to observe that, even when removing the environment that supported perceptually-grounded asymmetries, asymmetries related to phonological knowledge were still observable. This suggests that both phonetic/acoustic and phonological factors play a role, possibly in parallel or possibly interactively, during speech perception, and that the respective roles of these cues are dissociable using experimental manipulations like those in the present study.

Another open question is how Mandarin-like a listener's tone system needs to be to show phonologically-based asymmetries like those in the present study. We found that naïve speakers with no tone language experience did not show phonological asymmetries in Mandarin tone perception, but we did not test speakers of other, non-Mandarin tone languages. Yeung and colleagues (2013) found that both Mandarin-learning and Cantonese-learning infants showed similar asymmetries in the perception of Cantonese tones, but English-learning infants did not. This suggests that there may be some early-emerging perceptual biases that are shared across tone languages but not across all languages. In the future, therefore, Chinese languages and other tone languages could prove a fruitful testing ground for distinguishing between language-specific phonological mechanisms on one extreme, language-universal perceptual mechanisms at another extreme, and tone-language-specific perceptual mechanisms in the middle.

5.2 The role of underspecification in the phonological representation of lexical tones

The results of Experiment 3 suggest that there is something special about the phonological representation of T3. Throughout the present paper we have situated our predictions about the representational status of T3 in terms of underspecification theory, in particular the Featurally Underspecified Lexicon model (Lahiri & Reetz, 2002,

2010). Under such an explanation, T3 standards yielded attenuated MMNs for the same reason that standards with a coronal place of articulation (for example) do: that other tones are fully specified, or marked, in the mental lexicon of the proficient Mandarin speaker, whereas T3 is underspecified or unmarked, and treated as a default.

In section 1.1 we reviewed several reasons why such a prediction might make sense—namely, that T3 is contextually limited and is changed into another tone (via third tone sandhi) in some contexts, that the contrast between T3 and other tones may be acquired early by children (Qu, 2013), and that T3 is a low tone in Standard Mandarin. There are also, however, some challenges to the assumption that T3 is underspecified. First of all, whereas many underspecified segments may be treated as a default because they are easy to articulate (as is the case for coronal segments, as well as for voiceless unaspirated consonants, which are contrasted with specified forms that have either longer aspiration periods or longer prevoicing periods), it is not clear that T3 is articulatorily any easier than other sounds. Indeed, the full contour of T3 (with a final rise) is probably the most extreme tone in Standard Mandarin, is difficult for children to perceive and produce (Wong, 2012; Wong et al., 2005), and in many speakers carries non-modal creaky voice (Chao, 1968) which may also add articulatory effort and perceptual salience. Secondly, less marked segments are often higher in lexical frequency than other segments (Paradis and Prunet, 1991), and T3 is not; in fact, T4 has the highest type and token frequency in Mandarin, whereas T3 has the second-lowest token frequency and only the second-highest type frequency (Zhang & Lai, 2010). Finally, while T3 is indeed a low register tone in Standard Mandarin, it is not low in many other Mandarin dialects, but otherwise shows the same phonological behavior as Standard Mandarin T3 (Zhang, 2010). Thus, any markedness-based analysis which uses the low-ness of T3 to account for its synchronic phonological behavior (e.g., Kuo et al.,

2007), or to justify the assumption that it is underspecified, will fail for many widely spoken dialects.

If the MMN asymmetry were not due to any underspecified status of T3, what alternative accounts would remain? One possibility is that tone height alone causes the asymmetry, with low-register standards yielding smaller MMNs than the reverse case. Such an account would be consistent with the results of Law and colleagues (2013b), who found asymmetric MMN responses for Cantonese tones. They attribute their effect to underspecification of low register tone; in fact, their effect could not be due to knowledge of phonological alternation or behavior, as Cantonese lexical tones do not undergo the same sort of alternations as Mandarin tones (i.e., paradigmatic alternations based solely on intonational context).¹⁷ On the other hand, in a behavioral study Francis and Ciocca (2003) found what seems to be the opposite pattern—discrimination was better when a low tone was presented first than when it was presented second—although they explain the finding in terms of knowledge of downdrift (a cross-linguistic phenomenon whereby words further and further into an utterance are pronounced with lower and lower pitch) rather than in terms of the underlying representations of the tones. While an account of the data based on pitch height cannot be ruled out on the basis of the present experiments, it could be confirmed or ruled out in the future by testing speakers of dialects in which T3 is not a low tone but still undergoes the same

¹⁷ Cantonese tones do undergo another type of alternation, called *bin yam*, which is a derivational morphological operation that expresses changes in meaning (Yu, 2007). Unlike Mandarin third-tone sandhi, however, *bin yam* is not limited to one tone category (all tones are susceptible to it) and is not a strictly phonological operation.

phonological behavior as it does in Standard Mandarin (see section 2.3).¹⁸ If the phonologically-based asymmetry in Experiment 3 is driven just by the tone's low-ness, the effect might disappear or reverse in a dialect such as Taiyuan, where T3 is in the high register and T2 in the low register. On the other hand, if the asymmetry is driven by other properties of the tone's phonology (such as the fact that T3 has a variable surface representation), then the effect should be replicated even in a dialect like this, because T3 still alternates with T2 in this dialect even though the specific pitch values of these tone categories are different than in Standard Mandarin. We also note that, if pitch height is the explanation for the asymmetry observed in the present experiment, its explanatory power for cross-dialectal phenomena is subject to the limitations mentioned above.

One explanation that we can rule out for the present dataset is the computation of tone sandhi in particular. As described in the introduction, T3 alternates asymmetrically with T2, such that a T2 syllable is possibly a surface realization of an underlying T3, whereas a T3 token cannot be a realization of an underlying T2. It would have been possible, then, to predict MMN asymmetries on this basis: when hearing a stream of T3s as the standard, a T2 deviant may have only weakly mismatched the

¹⁸ A related case that could be informative to study is tones with different voice qualities. While the major acoustic cue for tone is typically fundamental frequency (pitch), tone really is a complex phenomenon involving a constellation of cues including, for instance, duration, intensity, and phonation type (e.g., creaky voice or breathy voice) in addition to pitch (Chao, 1968; Gandour & Harshman, 1978; Karlgren, 1962; Norman, 1988; among others). Such cases may allow for testing whether perceptual asymmetries are due strictly to the pitch values of tones, or to phonological knowledge of the tones' distribution and alternation behavior.

memory representation of the standard, since T2 is a possible realization of T3 and may even activate memory traces of T3 sounds.¹⁹ The same could not be true of a T3 deviant in a stream of T2 standards. Under such an explanation, however, asymmetries would not be expected for T3~T1 or T3~T4 contrasts. We observed asymmetrical trends, however, for both of these contrasts in Experiment 3, which casts doubt upon this explanation for our data.

Another possibility is that it is not T3 itself that is underspecified, but some sub-component of it. A longstanding question in the phonology of lexical tone is whether contour tones are represented by unitary, possibly dynamic, features (e.g., "High", "Rise", "Low", and "Fall") or by concatenations of static features (e.g., representing a falling tone as a High-Low sequence); see Zhang (2010) and Wang (1967) for review. Our predictions and discussions above have treated tone as a unitary feature, and assumed that asymmetries between T3 and other tones occur because T3 itself is differently specified than other tones, rather than that some sub-component of T3 is differently specified. It may be possible that an analysis of tones as feature bundles consisting of multiple sub-features could also offer an account for the asymmetric MMN observed in the previous study—for instance, the analysis proposed by Yip (Kuo et al., 2007) argues that only T3 consists of a Low-Low sequence (this could also be seen as a Low-underspecified sequence, where the underspecified portion is realized as Low by default but as High in sandhi contexts), whereas no other Standard Mandarin

¹⁹ Whether T2 activates memory representations of T3 is an open question that has not, to our knowledge, been tested at the neural level, although it has been tested behaviorally with auditory priming (Chien, Sereno, & Zhang, 2016; Zhou & Marslen-Wilson, 1997), implicit priming (J.Y. Chen & Chen, in press; Y. Chen, Shen, & Schiller, 2011; Politzer-Ahles & Zhang, in press), and lexical decision paradigms (Weiner & Turnbull, 2013).

tones include any Low sub-feature. If such an analysis were the case, the present results may also shed light on the neural processing of these component features of tones. We note, however, that sequences of features might just be convenient shorthands used in formal analysis of tone, rather than actual reflections of the mental representation of tone; for example, phonetic study has suggested that the underlying pitch targets for contour tones are dynamic and unitary slopes, rather than feature sequences (Xu & Wang, 2001, among others). Nonetheless, there is evidence from phonological behavior that a tone may comprise several hierarchically organized features, such as height features and contour features (Wang, 1967); future cross-linguistic study would be useful to examine whether the asymmetries observed in the present experiment are due to underspecification of a whole tone, like T3, or to underspecification of one of its component features that sets it apart from other tones.

Finally, it is possible that results consistent with underspecification could also be accounted for by episodic theories with the additional assumption that categories which encompass a greater variety of exemplars would yield weaker memory representations. That is to say, since the category of Tone 3 would include exemplars with a variety of surface realizations (full T3, half T3, and sandhi T3, which is similar to T2), a T3 token may elicit a weaker memory representation than other tones (while Tones 2 and 4 also undergo major coarticulatory phonetic changes [Chandrasekaran et al., 2009; Kuo et al., 2007; Shen, 1990] these arguably are less diverse than the variation in T3). Thus, apparent underspecification (and consequent MMN asymmetries) may instead be an epiphenomenon of a having a more diverse set of exemplars associated with that category.²⁰ While this argument cannot be ruled out on the basis of the present dataset,

²⁰ We thank Karthik Durvasula and Caicai Zhang for suggesting this argument.

we note that underspecification-related asymmetries have also been observed in contrasts that do not undergo phonological alternation, such as English /s/ vs. /f/ (Schluter et al., accepted), although such an episodic effect may still be possible under acoustic rather than phonological variability (see Tsuji et al., 2015). Overall, then, while it is still unclear whether a traditional underspecification account can apply to Mandarin T3, it does so far seem to be the most consistent with the MMN findings of the present study. Converging evidence from several multiple suggests that tones are not all processed in the same ways but that some have privileged status; for example, in addition to the present neurophysiological data, Chen (2013) observes a similar special status of T3 based on child acquisition patterns. Underspecification theory may offer a principled explanation for such patterns. Similarly, assuming that one tone or pitch accent is not lexically specified has proven valuable in explaining phonological patterns in other tone languages (such as Norwegian; Wetterlin & Lahiri, 2012). Further research is likely necessary to uncover the specifics of the mental representation of lexical tones and to test how these theories of representation may apply to other tones in other dialects; and the present study offers an example of how neurophysiological data can shed light on these representational issues.

6 Conclusion

The present study showed, in three experiments using the mismatch negativity (MMN), a robust asymmetry in the processing of Standard Mandarin tones. When Mandarin low tone (tone 3, or T3) was contrasted with another tone, MMN amplitude was reduced when T3 served as the standard and another tone as the deviant. By comparing the results from native Mandarin-speaking listeners against the results of naïve listeners, we were able to show that this effect can be separated into both perceptually-based asymmetries and phonologically-based asymmetries. The present

study both sheds new light on how lexical tones are processed, and offers strong evidence in support of accounts (such as underspecification) that relate asymmetries in neural discrimination responses to the abstract phonological representations of sounds. This provides a crucial piece to a large body of research suggesting that the architecture of the speech processing system is based on abstract representations with different possible levels of specificity, as opposed to fully-stored representations. Finally, the results demonstrate that both abstract, language-specific phonological structures and more language-universal perceptual biases coexist in the same language perception system, and that cross-linguistic comparisons and experimental manipulations can dissociate them.

Acknowledgements

This research was supported in part by grant G1001 from the NYUAD Institute, New York University Abu Dhabi. We are indebted to Caicai Zhang, Yu-Fu Chien, and Yiya Chen for valuable feedback on previous versions of this work. We also thank Hongbo Yu, Mingzhu Hou, and João Menezes for assistance with stimulus recording. This research is based in part on pilot research carried out by the first author at the Center for Brain and Cognitive Sciences at Peking University. The authors are responsible for any errors in the manuscript.

References

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183-207.

- Avery, P., & Rice, K. (1989). Segment structure and coronal underspecification. *Phonology*, 6, 179-200.
- Bidelman, G., & Chung, W. (2015). Tone-language speakers show hemispheric specialization and differential cortical processing of contour and interval cues for pitch. *Neuroscience*, 305, 384-392.
- Blicher, D., Diehl, R., & Cohen, L. (1990). Effects of syllable duration on the perception of the Mandarin tone 2/tone 3 distinction: evidence of articulatory enhancement. *Journal of Phonetics*, 18, 37-49.
- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4., retrieved 2014 from <http://www.praat.org/>.
- Bonte, M., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology*, 116, 2765-2774.
- Braver, S., Thoemmes, F., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9, 333-342.
- Bundgaard-Nielsen, R., Baker, B., Kroos, C., Harvey, M., & Best, C. (2015). Discrimination of multiple coronal stop contrasts in Wubuy (Australia): a Natural Referent Consonant account. *PLoS ONE*, 10, e0142054.
- Chandrasekaran, B., Gandour, J., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: a multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, 25, 196-210.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128, 148-156.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. (2009). Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear & Hearing*, 30, 552-558.
- Chao, Y. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press
- Chen, J.-Y., & Chen, T.-M. (in press). Mechanism and locus of tone 3 sandhi during Mandarin Chinese spoken word production. *Journal of Chinese Linguistics*.
- Chen, M. (2000). *Tone sandhi: patterns across Chinese dialects*. (Cambridge Studies in Linguistics 92.) Cambridge: Cambridge University Press.
- Chen, Y., Shen, R., & Schiller, N. (2011). Representation of allophonic tone sandhi variants. *Proceedings of Psycholinguistics Representation of Tone* (satellite workshop to *International Congress of Phonetic Sciences*).
- Chien, Y., Sereno, J., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi. *Language, Cognition and Neuroscience*, 31, 179-189.

- Cornell, S., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain Research*, 1394, 79-89.
- Cornell, S., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 757-772.
- Coster, D., & Kratochvil, P. (1984). Tone and stress discrimination in normal Peking dialect speech. In: Hong, B. (Ed.), *New Papers in Chinese Linguistics*. Canberra, Australian National University Press, 119-132.
- Czigler, I., Sulykos, I., & Kecskés-Kovács, K. (2014). Asymmetry of automatic change detection shown by the visual mismatch negativity: an additional feature is identified faster than missing features. *Cognitive & Affective Behavioral Neuroscience*, 14, 278-285.
- Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In: Zhang, P., Xie, T., Xu, J. (eds.), *The studies on the theory and methodology of the digitized Chinese teaching to foreigners: Proceedings of the 4th International Conference on New Technologies in Teaching and Learning Chinese*. Beijing: The Tsinghua University Press, 501-511.
- de Jonge, M., & Boersma, P. (2015). French high-mid vowels are underspecified for height. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134, 9-21.
- Durvasula, K., Trotter, D., & Beretta, A. (2015). Voicing underspecification in English fricatives. *Poster presented at 7th Annual Meeting of the Society for the Neurobiology of Language*.
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16, 577-583.
- Francis, A., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of the Acoustical Society of America*, 114, 1611-1621.
- Gandour, J., & Harshman, R. (1978). Crosslanguage differences in tone perception: a multidimensional scaling investigation. *Language and Speech*, 21, 1-33.
- Gaskell, M. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31, 447-463.
- Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251.
- Groppe, D., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage*, 45, 1199-1211.

- Gu, F., Li, J., Wang, X., Hou, Q., Huang, Y., & Chen, L. (2012). Memory traces for tonal language words revealed by auditory event-related potentials. *Psychophysiology*, 49, 1353-1360.
- Gu, F., Wang, X., He, K., & Chen, L. (2012). Preattentive auditory processing of pitch glide. *Tonal Aspects of Languages 2012*.
- Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *NeuroImage*, 83, 637-645.
- Harms, L., Fulham, W., Todd, J., Budd, T., Hunter, M., Meehan, C., ... Michie, P. (2014). Mismatch negativity (MMN) in freely-moving rats with several experimental controls. *PLoS ONE*, 9, e110892.
- Hestvik, A., & Durvasula, K. (2016). Neurobiological evidence for voicing underspecification in English. *Brain and Language*, 152, 28-43.
- Howie, J. (1968). Experiments on the perception of Mandarin vowels and tones. *Haskins Laboratories Status Report: SR-13/14* (pp. 189-196).
- Hsu, C., Evans, J., & Lee, C. (2015). Brain responses to spoken F0 changes: is H special? *Journal of Phonetics*, 51, 82-92.
- Hsu, C., Lin, S., Hsu, Y., & Lee, C. (2014). The neural generators of the mismatch responses to Mandarin lexical tones: an MEG study. *Brain Research*, 1582, 154-166.
- Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., & Kanno, A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters*, 321, 133-136.
- Jääskeläinen, I., Ahveninen, J., Bonmassar, G., Dale, A., Ilmoniemi, R., Levänen, ... & Belliveau, J. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 6809-6814.
- Jia, S., Tsang, Y., Huang, J., & Chen, H. (2015). Processing Cantonese lexical tones: evidence from oddball paradigms. *Neuroscience*, 305, 351-360.
- Karlgren, G. (1962). *Sound and Symbol in Chinese*. Hong Kong: Hong Kong University Press.
- Kayser, J., & Tenke, C. (2006a). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clinical Neurophysiology*, 117, 348-368.
- Kayser, J., & Tenke, C. (2006b). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates. *Clinical Neurophysiology*, 117, 369-380.
- Kayser, J. (2009). Current source density (CSD) interpolation using spherical splines – CSD Toolbox (Version 1.1)

- [<http://psychophysiology.cpmc.columbia.edu/Software/CSDtoolbox>]. New York State Psychiatric Institute: Division of Cognitive Neuroscience.
- Kazanina, N., Phillips, C., and Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103, 11381–11386.
- Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286-319.
- Kuo, Y., Xu, Y., & Yip, M. (2007). The phonetics and phonology of apparent cases of iterative tone change in Standard Chinese. In: Gussenhoven, C., Rial, T. (eds.), *Tones and Tunes Vol 2: Experimental Studies in Word and Sentence Prosody*. Berlin: Mouton de Gruyter, 211-237.
- Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory Phonology*, 7, 637–675.
- Lahiri, A., & Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *Journal of Phonetics*, 38, 44-59.
- Law, S., Fung, R., & Kung, C. (2013a). An ERP study of good production vis-à-vis poor perception of tones in Cantonese: implications for top-down speech processing. *PLoS ONE*, 8, e54396.
- Law, S., Fung, R., & Kung, C. (2013b). Asymmetric mismatch negativity and underlying representations of lexical tones in Cantonese. *Proceedings of the 51st Academy of Aphasia*.
- Li, T. (1983). Distribution of left/right handedness among Chinese people. *Xinli Xuebao (Acta Psychologica Sinica)*, 15, 268-276.
- Lin, M. (1988). Pǔtōnghuà shēngdiào de shēngxué xìngzhì yú zhījué zhēngzhào [Acoustic characteristics and perceptual cues for tones in Standard Chinese]. *Zhōngguó Yǔwén [Chinese Linguistics]*, 3, 182-193.
- Liu, S., & Samuel, A. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47, 109-138.
- Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – a behavioural and ERP study. *Brain Research*, 1624, 28-44.
- Luo, H., Ni, J., Li, Z., Li, X., Zhang, D., Zeng, F., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 19558-19563.
- Maddieson, I. (2005). Tone. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press.

- Makeig, S., Bell, A., Jung, T., & Sejnowski, T. (1996). Independent component analysis of electroencephalographic data. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems* 8 (pp. 145-151). Cambridge: MIT Press.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177-190.
- McEvoy, L., Hari, R., Imada, T., & Sams, M. (1993). Human auditory cortical mechanisms of sound lateralization: II. Interaural time differences at sound onset. *Hearing Research*, 67, 98-109.
- Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 496-513.
- Molnar, M., Polka, L., Baum, S., & Steinhauer, K. (2014). Learning two languages from birth shapes pre-attentive processing of vowel categories: Electrophysiological correlates of vowel discrimination in monolinguals and simultaneous bilinguals. *Bilingualism: Language and Cognition*, 17, 526-541.
- Morey, R. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, 4, 61-64.
- Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology*, 42, 25-32.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, 118, 2544-2490.
- Nielsen, K., & Scarborough, R. (2015). Perceptual asymmetry between greater and lesser vowel nasality and VOT. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Nordby, H., Hammerborg, D., Roth, W., & Hugdahl, K. (1994). ERPs for infrequent stimulus omissions and inclusions of stimulus elements. *Psychophysiology*, 31, 544-552.
- Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869.
- Oldfield, R. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.
- Pakarinen, S., Houtilainen, M., & Näätänen, R. (2010). The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology*, 121, 1043-1050.

- Paradis, C., & Prunet, J. (1991). Introduction: asymmetry and visibility in consonant articulations. In C. Paradis & J. Prunet (Eds.), *The Special Status of Coronals, Internal and External Evidence*, 1-28. San Diego: Academic Press.
- Phillips, C., Pellathy, T., & Marantz, A. (2000). Phonological feature representations in auditory cortex. Unpublished manuscript.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., ... Roberts, T. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience*, 12, 1038-1055.
- Polka, L., & Bohn, O. (2011). Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. *Journal of Phonetics*, 39, 467-478.
- Politzer-Ahles, S., & Zhang, J. (in press). Evidence for the role of tone sandhi in Mandarin speech production. *Journal of Chinese Linguistics*.
- Qu, C. (2013). *Representation and acquisition of the tonal system of Mandarin Chinese*. Ph.D. thesis, McGill University.
- Ren, G., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience*, 162, 87-95.
- Rose, P. (2015). Tonation in three Chinese Wu dialects. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Rosenthal, R., & Rubin, D. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, 92, 500-504.
- Sabri, M., & Campbell, K. (2000). Mismatch negativity to inclusions and omissions of stimulus features. *NeuroReport*, 11, 1503-1507.
- Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography & Clinical Neurophysiology*, 62, 437-448.
- Scharinger, M., Bendixen, A., Trujillo-Barreto, N., & Obleser, J. (2012). A sparse neural code for some speech sounds but not for others. *PLoS ONE*, 7, e40953.
- Scharinger, M., Merickel, J., Riley, J., & Idsardi, W. (2011). Neuromagnetic evidence for a featural distinction of English consonants: sensor- and source-space data. *Brain and Language*, 116, 71-82.
- Scharinger, M., Monahan, P., & Idsardi, W. (2012). Asymmetries in the processing of vowel height. *Journal of Speech, Language, and Hearing Research*, 55, 903-918.
- Schluter, K., Politzer-Ahles, S., & Almeida, D. (2015). Underspecification of voice features in the mental larynx. *Talk presented at the 23rd Manchester Phonology Meeting*.
- Schluter, K., Politzer-Ahles, S. & Almeida, D. (to appear). No Place for /h/: an ERP investigation of English fricative place features. *Language, Cognition and Neuroscience*.

- Schwartz, J., Abry, C., Boe, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. *Speech Communication*, 45, 425-434.
- Shen, X. (1990). On Mandarin tone 4. *Australian Journal of Linguistics*, 10, 41-59.
- Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.
- Shafer, V., Schwartz, R., & Kurtzberg, D. (2004). Language-specific memory traces of consonants in the brain. *Cognitive Brain Research*, 18, 242-254.
- Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *NeuroReport*, 13, 521-525.
- Timm, J., Weise, A., Grimm, S., & Schröger, E. (2011). An asymmetry in the automatic detection of the presence or absence of a frequency modulation within a tone: a mismatch negativity study. *Frontiers in Psychology*, 2.
- Tong, X., McBride, C., Lee, C., Zhuang, J., Shuai, L., Maurer, U., & Chung, K. (2014). Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: an ERP study. *Psychophysiology*, 51, 1158-1168.
- Tsang, Y., Jia, S., Huang, J., & Chen, H. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: the effects of pitch contour and pitch height. *Neuroscience Letters*, 487, 268-272.
- Tsao, F. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Zhōnghuá Xīnlǐxué Kān [Chinese Journal of Psychology]*, 50, 111-124.
- Walker, L., Carpenter, M., Downs, C., Cranford, J., Stuart, A., & Pravica, D. (2001). Possible neuronal refractory or recovery artifacts associated with recording the mismatch negativity response. *Journal of the American Academy of Audiology*, 12, 348-356.
- Walter, M., & Hacquard, V. (2004). MEG evidence for phonological underspecification. *14th International Conference on Biomagnetism*, 292-293.
- Wang, W. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33, 93-105.
- Wang, X., Gu, F., He, K., Chen, L., & Chen, L. (2012). Preattentive extraction of abstract auditory rules in speech sound stream: a mismatch negativity study using lexical tones. *PLoS ONE*, 7, e30027.
- Warren, R. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin*, 70, 261-270.
- Weiner, S., & Turnbull, R. (2013). Mandarin word recognition: tone sandhi, accidental gaps and the lexicon. *Poster presented at the 87th Annual Meeting of the Linguistic Society of America*.

- Wetterlin, A., & Lahiri, A. (2012). Tonal alternations in Norwegian compounds. *The Linguistic Review*, 29, 279-320.
- Whalen, D., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25-47.
- Winkler, I., & Näätänen, R. (1993). Event-related brain potentials to infrequent partial omissions in series of auditory stimuli. In H. Heinze, T. Münte, & G. Mangun (Eds.), *New Developments in Event-Related Potentials*, 219-226. Boston: Birkhauser.
- Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, 40, 141-151.
- Wong, P., Schwartz, R., & Jenkins, J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, 48, 1065-1079.
- Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170, 223-231.
- Xu, Y., & Wang, Q. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication*, 33, 319-337.
- Yang, R. (2011). The phonation factor in the categorical perception of Mandarin tones. *Proceedings of the 17th International Congress of Phonetic Sciences*.
- Yingling, C., & Nethercut, G. (1983). Evoked responses to frequency shifted tones: tonotopic and contextual determinants. *International Journal of Neuroscience*, 22, 107-118.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yu, A. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology*, 24, 187-214.
- Zhang, J. (2010). Issues in the analysis of Chinese tone. *Language and Linguistics Compass*, 4, 1137-1153.
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27, 153-201.
- Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, 68, 161-191.
- Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., & Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: an investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry*, 53, 874-882.
- Zheng, H., Peng, G., Chen, J., Zhang, C., Minett, J., & Wang, W. (2014). The influence of tone inventory on ERP without focal attention: a cross-language study. *Computational and Mathematical Methods in Medicine*, 2014, 961563.

Zhou, X., & Marslen-Wilson, W. (1997). The abstractness of phonological representation in the Chinese mental lexicon. In Chen, H. (Ed.), *Cognitive Processing of Chinese and Related Asian Languages*, 3-27. Hong Kong: The Chinese University Press.

Figure captions

Figure 1. The tones of Standard Mandarin (values taken from recordings made for Experiments 1 and 2). The y-axis represents fundamental frequency (i.e., pitch).

Figure 2. MMNs for Mandarin-speaking participants (Experiment 1). Each box shows the MMNs at Fz for both directions of one contrast; ribbons represent a Cousineau-Morey 95% confidence interval around the MMN (Morey, 2008). In the color version of the figure (online), lines of the same color represent conditions with the same standard tone, and lines of the same line style represent conditions with the same deviant tone. Topographic maps in the first row represent the mean of the MMN for each condition over a 51-ms time window centered on the most negative-going point in the 100-300 ms time window; the second row of topographic maps represents current source density (Kayser & Tenke, 2006a, b) of these topographies.

Figure 3. MMNs for non-Mandarin-speaking participants (Experiment 1).

Figure 4. MMNs for Mandarin-speaking participants (Experiment 2).

Figure 5. MMNs for non-Mandarin-speaking participants (Experiment 2).

Figure 6. MMNs for Mandarin-speaking participants (Experiment 3). Current source density maps are not shown. Also note that the vertical scale for this and the next figure is different from that in the previous ERP figures.

Figure 7. MMNs for non-Mandarin-speaking participants (Experiment 3).

Supplementary file information

Supplementary file 1. Demographic information and number of trials retained per condition for participants in all experiments (.xls).

Supplementary file 2. Audio files used as stimuli in the experiments (.zip containing .wav files)

Supplementary file 3. ERP averages for all experiments (.mat). The file contains three MATLAB structures, one for each experiment. The fields of the MATLAB structure give the channel location information (a structure), time in milliseconds corresponding to each sample (a vector of integers), the identifiers for native and non-native participants (each a cell array of strings), and an 'averages' structure containing the ERPs. The 'averages' structure has a separate field for each condition, and within that a field for each participant; the ERP for a given condition/participant combination is a Channels \times Samples matrix. The condition names are strings delimited by underscores, where the first part indicates the standard of that block, the second part indicates the deviant, and the third part indicates whether the present ERP is the standard ERP or the deviant ERP; for instance, the ERP labelled 'yi2_yi3_std' is that elicited by standards in the $yi^2 \rightarrow yi^3$ block, and the ERP labelled 'yi2_yi3_dev' is that elicited by deviants in the same block. For Experiment 3 there are also ERPs labelled 'control1' through 'control7', referring to tones in the random control blocks (for instance, 't3_t1_dev' refers to T1 deviants in the T3 \rightarrow T1 block, and 'control1' refers to those same T1 syllables when presented in the random control block). These files are provided for researchers interested in verifying the claims of this paper through reanalysis of the data; researchers interested in using the data for other purposes (such as to test new research questions) should obtain prior agreement from the authors (per section 8.14 of the

American Psychological Association's *Ethical Principles of Psychologists and Code of Conduct* <http://www.apa.org/ethics/code/index.aspx?item=11#814>).

Supplementary file 4. Plots showing (at Fz) the MMN (colored line) for each condition in each experiment overlain with the standard and deviant waveforms on which it was based (.zip file containing six images). Note that for Experiment 3, the MMNs in these plots will not exactly match the MMNs in Figures 6 and 7 in the main paper, since the MMNs in those figures were calculated based on control items rather than standard items (see section 4.1.3 for details).

Supplementary file 5. Summary of statistical results based on peak-picked time windows on electrode Fz as opposed to spatiotemporal clustering (.pdf)