ASYMMETRIES IN THE PERCEPTION OF MANDARIN TONES: EVIDENCE FROM MISMATCH NEGATIVITY

Stephen Politzer-Ahles, Kevin Schluter, Kefei Wu, & Diogo Almeida

New York University Abu Dhabi
spa268@nyu.edu

ABSTRACT

While asymmetrical neural responses to segmental contrasts have been used to articulate theories of the featural representation of speech sounds, little is known about asymmetrical responses to suprasegmental phenomena. The present study tested the neural processing of Mandarin tones using a passive oddball paradigm. For both native Chinese speakers and naive speakers with no Chinese experience, Tone 3 (T3), which alternates with T2 in certain contexts, elicited asymmetrical mismatch negativity (MMN) effects. Specifically, when contrasting T3 and another tone (T2 or T4), a smaller MMN was elicited when T3 was the standard than when the other tone was the standard. On the other hand, no asymmetry was observed between T2 and T4, a pair that does not productively alternate. The results suggest that T3 standards have an underspecified lexical representation.

Keywords: Mandarin, tone, mismatch negativity, EEG, underspecification

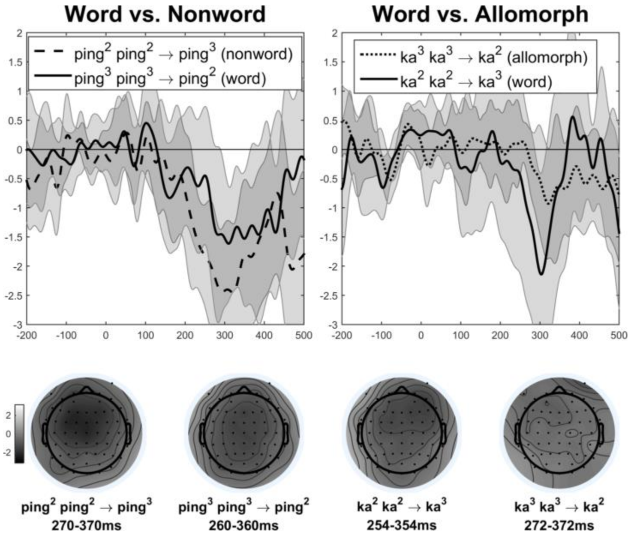
1. INTRODUCTION

Electroencephalography (EEG) provides a powerful means to examine the neural processing of speech sounds. The mismatch negativity (MMN) is an electrophysiological component elicited by an infrequently repeated sound embedded in a stream of more frequently repeated sounds (e.g., by ***s*** in "*f f f f* ***s*** *f f* ***s*** *f f f f f f f f* ***s*** *f f f* ***s***..."). While the MMN is a low-level effect that is not restricted to language processing, it is nevertheless sensitive to abstract linguistic structure [13]. Importantly, directional asymmetries in the amplitude of the MMN have been argued to reflect linguistic phenomena. For instance, larger MMNs are elicited when a phonetically more peripheral vowel like [i] is used as the infrequent deviant stimulus amongst a stream of less peripheral vowels like [e] as the frequent standard [14], or when a deviant segment is embedded in a stream of standards that are phonologically underspecified [5, 16]. Such asymmetries, however, have primarily been tested in Indo-European languages, and on segmental contrasts. While there is a growing body of MMN research on tonal contrasts (see [6] and references therein; see [2] for an MMN test of a tone-allotone contrast), this work has mainly focused on different contrasts, rather than testing for directional asymmetries within the same contrast (but see [10]). The present study uses MMN to test for asymmetric neural processing of suprasegmental tonal contrasts in Mandarin Chinese.

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] 受 "receive". In the standard dialect, tone one (T1) is realized as a High tone, T2 as a Rising tone, T3 as Low, and T4 as Falling. T3 may be realized as a simple low-falling contour or a complex falling-rising contour, depending on its prosodic context. 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 T3: /T3.T3/ → [T2.T3].

This alternation raises questions about how T3 and T2 are processed. In particular, since a T2 syllable is a valid potential realization of an underlyingly T3 syllable, whereas the converse is not the case, could there be asymmetries in the perception of these tones? Data from a pilot MMN experiment suggest such an asymmetry: while the study was meant to test a different comparison, it instead showed that T3 deviants embedded in a stream of T2 standards tended to elicit more negative MMN than T2 deviants with T3 standards (Figure 1). A similar asymmetry between low and high tones has been reported for Cantonese [10]. The present study was conducted to investigate the Mandarin asymmetry more systematically. In addition to testing whether or not this asymmetry would be replicated, we also tested whether T3 would show asymmetrical MMN effects when contrasted with a phonologically unrelated tone. The pilot data do not show whether the asymmetry is limited to the contrast between T3 and T2 (which are related to one another because of third tone sandhi), or whether T3 elicits asymmetrical MMN effects across the board. The present study, therefore, tested the contrasts between T3 and T2 (which alternate with one another), T3 and T4 (which do not alternate with each other), and T2 and T4. Furthermore, we

Figure 1: Pilot data for for the T2-T3 contrast. Each line represents a difference wave (e.g., event-related potential responses to *ping*^{3} as a deviant minus those to the same *ping*^{3} used as the standard in another block). Ribbons represent ±2 standard errors.



tested both full T3 (with a complex falling-rising contour) and half T3 (a variant with a simple falling contour, which is produced in non-XP-final positions where T3 sandhi does not apply).

2. METHODS

2.1. Participants

Data for the full T3 experiment were collected from 16 New York University Abu Dhabi students (aged 18-23, 7 men) who were native speakers of Mandarin from mainland China, and 16 control participants who had no experience speaking Mandarin (18-54, 4 men). The half T3 experiment had 16 Chinese-speaking participants (18-49, 6 men) and 16 non-Chinese-speaking participants (18-26, 8 men). Participants provided written informed consent and were paid for their participation. All methods were approved by the Institutional Review Board of New York University Abu Dhabi.

2.2. Materials and procedure

The stimuli for the full T3 experiment comprised the syllable *yi* in each tone. The log parts-per-million frequencies (from [4]) for these syllables are 4.22 (*yi*^{1}), 3.04 (*yi*^{2}), 3.73 (*yi*^{3}), and 3.77 (*yi*^{4}). This syllable was chosen because it is a single monopthong ([i]), and because the lexical frequency of the T3 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.

Stimuli were spoken by a native speaker of Mandarin from Shanghai. *yi*, *wu*, and *yu* were produced in all four tones, ten repetitions of each, in a random order. To create the stimuli, one of these tokens was selected to use as the vowel, and one token for each tone was selected to use as the pitch tier (for T3, a token with a full falling-rising contour and without vocal fry was selected). Using Praat [1], these were all normalized to the same duration. To produce tokens of different tones, the pitch tier from a token of that tone was imposed on the vowel base token. All tokens were normalized to 75 dB. Thus, the tokens differed in nothing but F0 contour. Only the T2, T3, and T4 tokens were used in this experiment. While T2 does alternate with T4 and T1 in certain contexts (the words 一, 不, and [for some speakers] 七 and 八), these alternations are not productive, and thus we considered T2 and T4 to be non-alternating.

The experiment used a passive oddball design with six types of blocks: *yi*^{2}→*yi*^{3}, *yi*^{3}→*yi*^{2}, *yi*^{2}→*yi*^{4}, *yi*^{4}→*yi*^{2}, *yi*^{3}→*yi*^{4}, and *yi*^{4}→*yi*^{3}. (The syllable to the left of the arrow refers to the syllable that was used as the standard in that block, and the syllable to the right refers to the deviant.) Each of these block types was divided into three short blocks, and each short block included 224 standards and 36 deviants, such that each condition overall had 672 standards and 108 deviants. Each trial began with at least 20 standards, and 2–10 standards intervened between each deviant. Each syllable was 300 ms long, and the inter-trial interval was 500 ms. The stimuli were presented binaurally over tube earphones. Each participant selected a movie or television show with subtitles to watch silently during the experiment. The experimental session lasted about 2 hours.

The stimuli and procedure for the half T3 experiment were the same, except that the carrier syllable was *wu* ([u]; log PPM frequencies 2.68 for *wu*^{1}, 3.25 *wu*^{2}, 3.19 *wu*^{3}, 3.38 *wu*^{4}), and the pitch contour for T3 was that of half-third tone (a simple falling contour) rather than full third-tone. Each block had 267 standards and 40 deviants, for a total of 120 deviants and 801 standards per condition.

2.3. Data acquisition and analysis

EEG was continuously recorded from 66 Ag/AgCl electrodes (actiCAP, Brain Products) for the full T3 experiment, and 34 positions for the half T3 experiment. The sampling rate was 1000 Hz, and data were filtered online from 0.1–1000 Hz. FCz served as the online reference and AFz as the ground. Offline data were re-referenced to the average of both mastoids, low-pass filtered at 30 Hz, segmented into epochs from -200ms to 500ms relative to the sound onset, and baseline-corrected (-100 to 0ms). The first series of 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 ±75 μV in the -150 to 400 ms time window were removed from further analysis. For each participant, at least 25 deviant trials per condition were retained. 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 (e.g. 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).

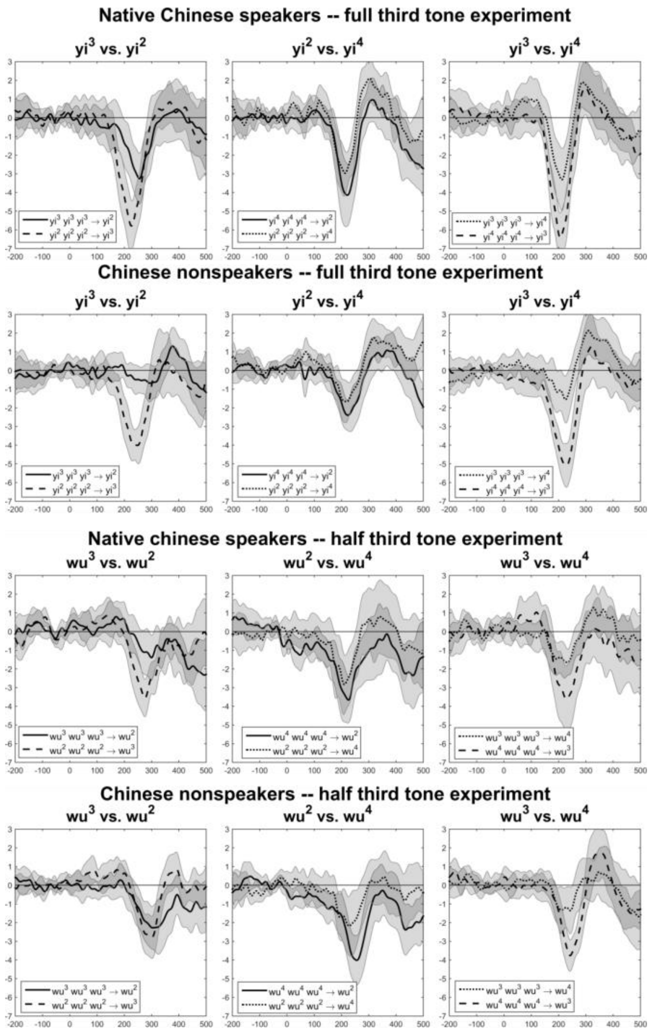
MMN amplitudes were compared using spatiotemporal clustering with non-parametric permutation statistics [11]. Spatiotemporal clusters between 100 and 300 ms were identified where there were significant interactions of the factor Contrast (3 levels: T2~T3, T2~T4, and T3~T4) and a dummy factor indicating the direction of the contrast (e.g., within the T2~T3 blocks, $yi^2 \rightarrow yi^3$ was coded as 'A' and $yi^3 \rightarrow yi^2$ as 'B'; the 'A' and 'B' conditions do not mean anything on their own, but a simple effect of direction within any contrast would indicate that there was an asymmetrical MMN for that contrast).

3. RESULTS

MMN waves for each contrast in the full T3 experiment are shown in the upper half of Figure 2. For both groups, MMNs were asymmetrical in the T2~T3 contrast and the T3~T4 contrast, but may not have been in the T2~T4 contrast. Significant Contrast×Direction interaction clusters were identified over front-central sites in both groups (Chinese speakers: 160-260 ms, p = .001; non-Chinese speakers: 155-289 ms, p = .002), indicating that the presence or absence of an asymmetry (represented by the Direction factor) varied as a function of the contrast. We resolved the interaction by using pairwise t-tests to compare the two MMNs in these clusters within each contrast (p-values were Bonferroni corrected by multiplying each by 3). The pairwise tests revealed that for the T2~T3 contrast (left), $yi^3 \rightarrow yi^2$ yielded a less negative (i.e., smaller) MMN than $yi^2 \rightarrow yi^3$ (Chinese: $t(15)$ = 3.28, 95% CI = 0.58 – 2.75, p = .015; non-Chinese: $t(15)$ = 5.80, 95% CI = 1.35 – 2.93, p < .001). For the T3~T4 contrast (right), $yi^3 \rightarrow yi^4$ yielded a less negative (i.e., smaller) MMN than $yi^4 \rightarrow yi^3$ (Chinese: $t(15)$ = 3.09, CI = 0.66 – 3.61, p = .022; non-Chinese: $t(15)$ = 5.33, CI = 1.37 – 3.19, p < .001); and for the T2~T4 contrast (center), the MMNs did not significantly differ after Bonferroni correction (Chinese: $t(15)$ = 1.96, CI = -0.11 – 2.64, p = .206; non-Chinese: $t(15)$ = 1.05, CI = -0.57 – 1.69, p = .926). This pattern of results was also observed in a traditional analysis using analysis of variance on mean amplitudes over pre-selected time windows.

For the half T3 experiment (bottom portion of Figure 2), the results were similar, except that there is no longer an apparent asymmetry in the T2~T3 contrast for non-speakers of Chinese. Chinese-speaking participants showed a significant, somewhat right-lateralized frontocentral cluster for the Contrast×Direction interaction, from 191-301 ms including 21 electrodes (p = .018), in which there were significant asymmetries for T2~T3 ($t(15)$ = 4.52, CI = 1.00 – 2.79, p = .001) and T3~T4 ($t(15)$ = 3.15, CI = 0.54 – 2.82, p = .020), but not for T2~T4 ($t(15)$ = 1.16, CI = -0.64 – 2.18, p = .791). Non-Chinese-speaking participants showed a significant frontocentral interaction cluster from 218-294 ms including 24 electrodes (p = .012), with significant asymmetries for T2~T4 ($t(15)$ = 2.55, CI = 0.21 – 2.33, p = .067) and T3~T4 ($t(15)$ = 6.11, CI = 1.22 – 2.52, p < .001), but not for T2~T3 ($t(15)$ = -1.51, CI

Figure 2: MMN waveforms at electrode Fz.



= -1.30 – 0.22, p = .456).

4. DISCUSSION

Chinese speakers and non-speakers showed asymmetrical processing of the contrast between T3 and other tones: specifically, MMN amplitude was reduced when T3 was the standard, compared to when another tone was the standard. This asymmetry was observed both in the contrast with T2, which T3 alternates with, and with T4, which it does not alternate with. No asymmetry was observed between T2 and T4. As for non-speakers of Chinese, they showed the same pattern as Chinese speakers in the experiment using full T3, but half T3 in the second experiment did not robustly elicit asymmetrical MMNs across all contrasts tested.

Several situations can elicit asymmetrical MMNs. A contrast between a typical member of a phonological category and a less typical member yields a larger MMN when the typical member is the standard [7]; this, however, is not likely to be the cause for the present results, since the tones tested belong to different categories, at least for the native speakers. A contrast between a more extreme and a less extreme segment (e.g., a corner vowel /i/ versus a more central vowel /e/) also yields asymmetrical MMN [14], but we had no *a priori* reason to consider T3 less extreme than the other tones. The fact that T3 undergoes phonological alternation could mean that it creates a weaker expectation (compared to a sound that does not undergo alternation) when used as the standard in an oddball paradigm, but this would not explain the asymmetry in the full T3 experiment for the control group, who had no knowledge of Mandarin tone sandhi. Two more viable explanations are the acoustic complexity of the stimuli, and underspecification.

Full T3 deviants may have elicited greater MMNs because they have more complex contours than the other tones. It has been shown that the MMN is sensitive to the number of features in the standards and deviants: a deviant with an added feature (e.g., deviants consisting of combined white noise and sine tones after standards consisting of only white noise or only sine) elicits a larger MMN than the reverse case [3, 12]. T3, which has a falling-rising contour when pronounced in isolation or phrase-finally, may also have been considered to have an additional phonetic feature by the listeners. If this were the case, we would expect that half T3 (the low-falling realization of T3 in non-phrase-final position, which does not have a complex contour) would not show asymmetrical MMN with other tones, and this is exactly what we observed for non-speakers of Chinese in the second experiment. The pattern in the full T3 experiment suggests that listeners without a tonal phonology are still sensitive to acoustic complexity. However, since asymmetries were observed in the native speakers in the half T3 experiment when acoustic complexity was roughly controlled, acoustic complexity alone cannot fully explain the asymmetrical MMN; rather, for listeners with a tonal phonology, there must be other factors contributing to the effect as well. An acoustic complexity account also does not explain why [10] found an asymmetrical MMN between Cantonese low and high level tones, but not between contour and non-contour low tones.

The asymmetrical MMN effect remaining after acoustic complexity was ruled out can likely be accounted for by underspecification. In many MMN studies, the contrast between an phonologically underspecified standard (usually a coronal segment) and a fully specified deviant yields a smaller MMN than when the standard is fully specified; this is argued to occur because a deviant cannot phonologically *mismatch* features that are not fully specified in the standard [5, 9]. If T3 were phonologically underspecified on some feature that separates it from the other tones, this would cause the observed results. Indeed, it has been argued that Low tones are typologically less marked than other tones, are acquired earlier, and that Mandarin T3 in particular exhibits special properties such as a tendency not to bear contrastive stress [8, 15, 17]; these properties could help motivate an analysis in which T3 (a Low tone in standard Mandarin) is underspecified. It bears mention, though, that T3 does not share some of the other common properties of underspecified coronals which have often shown directional asymmetries in MMN paradigms: T3 is not more frequent than other tones in either type or token frequency [18], nor is it articulatorily easier or more default-like. More importantly, an analysis predicating the underspecification of T3 on its Low tone feature would predict that T3 would *not* be underspecified in Mandarin dialects such as Tianjin, Jinan, and Taiyuan, where T3 corresponds to a mid-rising, high, or mid-falling tone, respectively [19], and thus that MMN effects in these dialects would look quite different. This is an empirical question; nonetheless, there are theory-internal reasons that motivate us to doubt an analysis in which third tone sandhi is based on a Low feature in standard Mandarin but on an entirely different mechanism in these other dialects.

Future work will test these tones, as well as T1, in other segmental contexts, and on speakers of other Mandarin dialects.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4., retrieved 2014 from http://www.praat.org/.

[2] Chandrasekaran, B., Krishnan, A., Gandour, J. 2009. Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear & Hearing* 30, 552-558.

[3] Czigler, I., Sulykos, I., Kecskés-Kovács. 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Lahiri, A., Reetz, H. 2010. Distinctive features: phonological underspecification in representation and processing. *Journal of Phonetics* 38, 44-59.

[10] Law, S., Fung, R., Kung, C. 2013. Asymmetric mismatch negativity and underlying representations of lexical tones in Cantonese. *51st Academy of Aphasia*.

[11] Maris, E., Oostenveld, R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* 164, 177-190.

[12] Nordby, H., Hammerborg, D., Roth, W., Hugdahl, K. 1994. ERPs for infrequent stimulus omissions and inclusions of stimulus elements. *Psychophysiology* 31, 544-552.

[13] Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., Roberts, T. 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience* 12, 1038-1055.

[14] Polka, L., Bohn, O. 2011. Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. *Journal of Phonetics* 39, 467-478.

[15] Qu, C. 2013. *Representation and acquisition of the tonal system of Mandarin Chinese*. Ph.D. thesis, McGill University.

[16] Schluter, K., Politzer-Ahles, S. Almeida, D. No Place for /h/: an ERP investigation of English fricative place features. Submitted.

[17] Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

[18] Zhang, J. 2010. Issues in the analysis of Chinese tone. *Language and Linguistics Compass* 4, 1137-1153.

[19] Zhang, J., Lai, Y. 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* 27, 153-201.