Processing neutral tone under the non-attentional condition: a mismatch negativity study

Wei-Jing Zhou[1], Zhi-Yan Wang[1,†], Su-Wan Wang[1,†], Lun Zhao[2,*]

[1] *School of International Studies, Yangzhou University, Yangzhou, 225127 Jiangsu Province, P. R. China*

[2] *School of Educational Sciences, Liaocheng University, Liaocheng, 252000 Shandong Province, P. R. China*

*Correspondence: lunzhao6@gmail.com (Lun Zhao)
[†]These authors contributed equally.

The neutral tone is a unique tone form in Mandarin as it distinguishes from four canonical tones or full tones on the one hand and integrates phonetic, morphological, syntactical and prosodic information on the other hand. Research to date has been focusing on its unique and variant acoustic features. However, little is known about how native Mandarin speakers process such a unique tone. In the present study, the mismatch negativity was used to explore the comparison-based pre-attentive change detection of Mandarin neutral tone. The mismatch negativity at the time window of 400-800 ms post-first-tone onset was obtained by subtracting event-related potentials to standard neutral tone from event-related potentials to a deviant natural tone. The source analysis of mismatch negativity showed the cortex generator was located at the left temporal lobe. The data suggest that Chinese native speakers process neutral tone automatically under non-attentional conditions, as revealed by the mismatch negativity data aligned with a neutral tone, and that neutral tone does exist as an automatically recognizable one in native Mandarin speakers' tone system.

1. Introduction

It is widely accepted that neutral tone, i.e., T0, is a unique tone form in Mandarin as it distinguishes from four canonical tones on the one hand and integrates phonetic, morphological, syntactical and prosodic information on the other hand [1–6]. Phonetically speaking, T0 has a much shorter duration with reduced pitch contour, yet prosodically speaking, it embodies both tonal and stress information at the same time. Such a special tonal category has intrigued several studies to explore its characteristic duration, specific contour patterns, categorization rules as well as phonetic and phonological motivations based on the interaction of universal default rules and language-specific phonetics principles, all of which cast light on the nature of T0 in Mandarin, i.e., T0 can be both lexically specified and rule-derived [1–22]. However, little is known about how native Mandarin speakers process such a unique tone.

As event-related potentials (ERPs) outperform behavioral procedures by providing a continuous measure of processing between a stimulus and a response, the present study adopted this extensively used technique in neuroscience, cognitive psychology, and cognitive science to probe into the fundamental issue of Mandarin natives' processing of peculiar T0, i.e., whether native Mandarin speakers can automatically process T0 and detect incongruent one under non-attentional condition.

All languages employ vowels and consonants as discrete contrastive subcomponents of syllables to differentiate lexical meaning. In addition to vowels and consonants, some languages also employ tones, i.e., pitch variations, to distinguish different lexical meanings, such as the syllable *ma* in Mandarin Chinese can convey 4 different lexical meanings when the syllable is aligned with 4 distinctive tones: mā (in Tone 1, in short T1), má (T2), mǎ (T3), mà (T4) meaning "mother", "hemp", "horse" and "scold" respectively. Many Languages like Mandarin are typical Tone languages, in which tone variations represent different lexical meanings. In contrast, those who do not use tone variation to differentiate lexical meanings are not tone language. They are thus termed as non-tone language, which on the other hand employ stress (a combination of pitch, duration and intensity) to distinguish the lexical or grammatical meanings of the same written form, such as the word *record* in English: when the first syllable is stressed, it is a noun *REcord*; when the second syllable is stressed, it is a verb *reCORD*. According to Yip [21], 60-70% of existing languages are tone language, including many Asian, African and indigenous American languages and a few European and South Pacific languages [22]. Recent studies on the perception of T0 by children from tone- and non-tone L1 backgrounds revealed that Dutch infants could discriminate continuously neutral and canonical tones, regardless of age [12].

In contrast, native Mandarin infants in the first year of life unexpectedly failed to achieve discrimination. Presumably, Dutch infants took the neutral-canonical tone contrast as

lexical stress contrast due to the stress prominence in their L1 Dutch. Unlike Dutch, Mandarin is not a stress-timed language. The Mandarin infants' failure to acquire T0 within the first year of their birth indicates that tonal categories' mastery is a protracted process longer than expected [12].

To date, however, not much is known about the fundamental issues of adult Mandarin native speakers' perception of T0, such as whether they can process T0 automatically under non-attentional conditions. The automatically adequate task-independent deviant events under non-attentional conditions are an important cognitive function for human survival. Besides, several crucial issues remain unsettled. First, if native infants and young children are not the optimal subjects to detect the perception of T0, what about adult native Mandarin speakers with mature tonal mastery? Second, monosyllabic carrier words adopted in the previous perception experiments are not ideal for generating the perception of T0 because the realization of T0 varies with the tonal category of its preceding syllable and linguistic structure of target T0 in the carrier words. Therefore, disyllabic words outperform monosyllabic words in eliciting the perception of T0, as disyllabic words are the predominant prosodic units in Mandarin and have more frequent usage than monosyllabic words in speech [12]. Third, the data collected from a forced-choice procedure and a visual fixation diagram may not be reliable enough to tract the direct and accurate on-line perceptual processing of T0 in pre-attentive conditions. To achieve higher research reliability, a more precise instrument is needed to probe into the fundamental issue of native Mandarin speakers' processing of T0. Based on previous studies' findings and the research gap in this field, the present study intends to investigate whether adult native Mandarin speakers can automatically process T0 in pre-attentive condition by recording and tracking the mismatch negativity (MMN) component in ERPs elicited by T0.

Generally, MMN is observed by submitting ERPs related to standard stimuli from ERPs elicited by deviants, reflecting the pre-attentive change detection under non-attentional conditions [23]. Auditory MMNs are composed of two subcomponents, the supra-temporal and the frontal subcomponents. The former reflecting pre-perceptional change detection is generated from the bilateral supra-temporal area. In contrast, the latter is mainly generated from the right frontal lobe, which is related to involuntary attention changes caused by input information change [20–22]. Recently, converging data indicates that the MMN indexes reliably the processing of automatic change detection is associated with the brain's detection of the temporal variation regularity [23]. Importantly, recent studies have established the relationship between MMN and speech and higher-order linguistic processing [23, 26, 27]. For instance, Pulvermüller *et al.* [26] exhibited that the MMN elicited by deviant spoken Finnish syllables had a greater amplitude at the end of a Finnish word than at the end of a pseudoword

and that this MMN related to word syllables was mainly generated from the left superior temporal lobe.

Moreover, Beauchemin *et al.* [27] found that the amplitude of MMN in the familiar voice deviants was higher than that in the unfamiliar voice deviants. This enhancement was not observed in another group of utterly unfamiliar subjects with neither strange voice. Thus, it is preliminarily believed that the specialized field of speech processing is especially suitable for familiar sounds rather than unfamiliar sounds. There is some degree of pre-attentive voice-familiarity assessment in the human brain moderating behavioral discrimination [27].

According to the Atkinson-Shiffrin [28] memory model, human memory capacity consists of two types: long-term memory, a stage where informative knowledge is held indefinitely, and short-term and working memory, which persist for only about 18 to 30 seconds. Long-term memory is further divided into two main types: explicit memory and procedural memory. The former is the conscious, intentional recollection of factual information, previous experiences and concepts, and the latter is unconscious memories of skills such as knowing how to get dressed, how to speak, how to ride a bicycle, etc., without having to re-relearn the skill each time. Thus procedural memory learns rule-like relations, and explicit memory learns arbitrary relations, which is achieved by long terms of learning and experience through multiple presentations of a stimulus and response. As T0 never exists as a self-sufficient, independent tone in Mandarin, it is always attached to one of the canonical tones in a multi-character prosodic word. Naturally, native Mandarin children are exposed and taught to learn T0 in the tone combination of the prosodic words since their infancy. They are then intensively instructed and explicitly assessed to use T0 and tone combination during speaking, reading and writing courses from an early age. Most likely, native Mandarin speakers' master of T0 results from both explicit memory and procedural memory, and their knowledge and skills of T0 belong to long term memory.

Due to the linguistic features of T0 and the potential cognitive mechanism in the native Mandarin speakers' processing of T0, it is assumed that T0 is related to the long-term memory traces and that the MMN can be produced by the input mismatch bias of the memory template formed by the standard stimuli related to various acoustic characteristics of speech. Therefore, if adult native Mandarin speakers can process T0 automatically in pre-attentive condition then the MMN related to T0 will be elicited.

2. Methods

2.1 Participants

Twenty-four healthy right-handed participants (11 female, mean age 25.44 ± 3.21 years) were recruited. All the participants were native Chinese speakers, with normal or corrected-to-normal vision, normal hearing, and no history of brain injuries or neurological problems during their lives.

Table 1. The tone patterns of stimuli in the present study			
Mandarin character	Pinyin	Congruent condition	Incongruent condition
妈妈	mā ma	mā (T1) ma (T0)	mā (T1) mā (T1)
爷爷	yé ye	yé (T2) ye (T0)	yé (T2) yé (T2)
婶婶	shěn shen	shěn (T3) shen (T0)	shěn (T3) shěn (T3)
弟弟	dì di	dì (T4) di (T0)	dì (T4) dì (T4)

They were paid for their participation. The local ethics committee approved the study, and the written informed consent was obtained from all the participants before the testing.

2.2 Stimuli and procedure

Reduplication is the most common word-formation in Mandarin and the most pervasive context for T0 usage in Mandarin. The present study's stimuli were 4 high-frequency two-character disyllabic reduplicative Chinese words combined with two kinds of tones: canonical tone of the first word and neutral tone of the second word. The two words were the same in the forms of Chinese characters and segments, but they differed in their tones, i.e., the second word's tone was a neutralized version of the first word according to rhythmic rules of disyllabic reduplicative words in Mandarin. Therefore, the congruent tone patterns of the 4 target disyllabic words were T1T0, T2T0, T3T0 and T4T0, respectively, i.e., T1/2/3/4 + T0, and the incongruent tone patterns of the 4 target disyllabic words were T1T1, T2T2, T3T3 and T4T4, respectively, i.e., T1/2/3/4 + T1/2/3/4 (see Table 1).

The probabilities of standard stimuli and deviant stimuli were 80% and 20%, respectively. Two words were stimuli binaurally presented 500 ms (about 250 ms for each) with the inter-stimulus interval of 1000 ms in two blocks. In one block, congruent neutral tones were used as standard stimuli (i.e., T1/2/3/4 + T0) and incongruent tones as deviant stimuli (i.e., T1/2/3/4 + T1/2/3/4). The reverse configuration was used in another block: incongruent tones as standards and congruent tones as deviants. It has been approved that this deviant-standard-reverse paradigm was possible to obtain the memory-comparison-based auditory MMN, which was obtained by subtracting the ERPs related to standard stimuli in one block from ERPs to the same stimuli as deviant in another block. A total of 500 stimuli in each block condition were presented in a pseudo-random order, with each deviant preceded by at least two standard stimuli. Two block conditions were counterbalanced across participants.

3. Results

Fig. 1 shows the grand-averaged different waveforms at selected electrode sites between ERPs elicited by standard and deviant stimuli. The negative deflection, auditory MMN elicited by neutral tone, is evident at the time window of 400-800 ms post stimuli onset.

To confirm the presence of MMN, the comparison between MMN amplitude and zero was conducted for T0 and T1/2/3/4, respectively. As shown by a one-sample *t*-test, at the time window of 400-600 ms post the first word (i.e., T1) onset, the mean amplitudes of MMN_T0 were significantly different from zero for each channel (ps < 0.02), but the amplitudes of the MMN_T1/2/3/4 were not significantly different from zero for each channel (ps > 0.1). At the time window of 600-800 ms post the first word (i.e., T1) onset, the mean amplitudes of MMN were significantly different from zero for each channel (ps < 0.01), regardless of MMN_T0 or MMN_T1/2/3/4 condition.

An ANOVA test showed that at the time window (400-600 ms), the more negative amplitudes were found for MMN_T0 (-1.9 uV) than for MMN_T1/2/3/4 [-0.02 uV; $F_{(1, 23)}$ = 15.65, $P < 0.001$, Partial η^2 = 0.336]. Within the time interval (600-800 ms), there was no significant difference between two MMNs [-2.7 uV and -2.6 uV for MMN_T0 and MMN_T1/2/3/4, respectively; F < 1]. The main effect of electrode site was also significant [$F_{(2, 46)}$ = 10.36, $P < 0.001$, Partial η^2 = 0.425 and $F_{(2, 46)}$ = 9.78, $P < 0.01$, Partial η^2 = 0.269 for 400-600 ms and 600-800 ms, respectively], showing the maximum of -1.0 uV and -2.6 uV at Cz site for 400-600 ms and 600-800 ms, respectively. The two-way interaction was not significant (F < 1).

As indicated in Fig. 2, the cortex current density analysis of MMN_T0 showed a maximum of 2.80/mm^3 at the left temporal lobe (Brodmann area 20, Inferior Temporal Gyrus, Temporal Lobe; MNI, -60, -55 and -20 mm for x, y, and z).

4. Discussion

T0 is a unique tone form in Mandarin as it distinguishes from four canonical tones or full tones and integrates phonetic, morphological, syntactical and prosodic information. In this study, we used the auditory MMN of ERPs related to the deviant-standard reverse oddball paradigm to explore the pre-attentive change detection of Mandarin neutral tone. The MMN was obtained by subtracting ERPs to standard neutral tone in one block from ERPs to deviant natural tone in another block. We found the MMN at the time interval of 400-800 ms post-first-tone onset, with the cortex generator of MMN located at the left temporal lobe. These results indicate that Chinese native speakers process T0 automatically under non-attentional conditions, as revealed by the MMN data aligned with a neutral tone, and that T0 does exist as an automatically recognizable one in native Mandarin speakers' tone system.
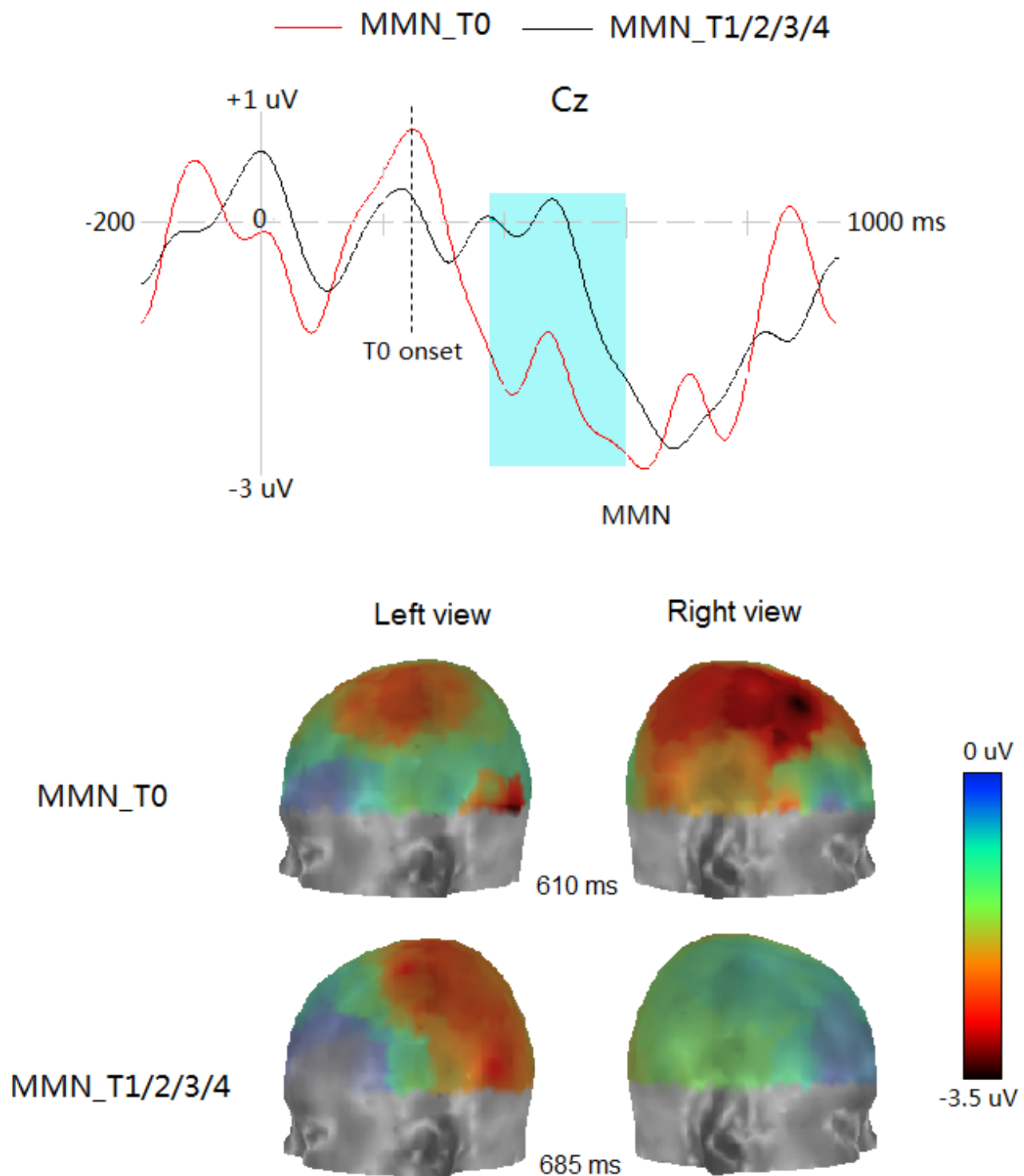
Fig. 1. The MMN components elicited by T0 and T1/2/3/4 deviant stimuli and the scalp distribution of the peak amplitudes of the MMN components, respectively. The blue color highlighting means the significant difference between MMN of T0 (MMN_T0) and MMN of T1/2/3/4 (MMN_T1/2/3/4).

To our knowledge, this is the first report to show that T0 can elicit enhanced MMN than canonical tone (T1/2/3/4) did. It is accepted that, as an automatic change-detection neural response, the MMN component of ERPs can be used to explore the accuracy of auditory discrimination of speech, not only reflecting the accuracy of behavior discrimination but also the sensory memory traces of previous stimuli, which lays the foundation for change detection [20]. It is imperative to use MMN as an index and explore the long-term memory representation of higher-order language phenomena, e.g.,
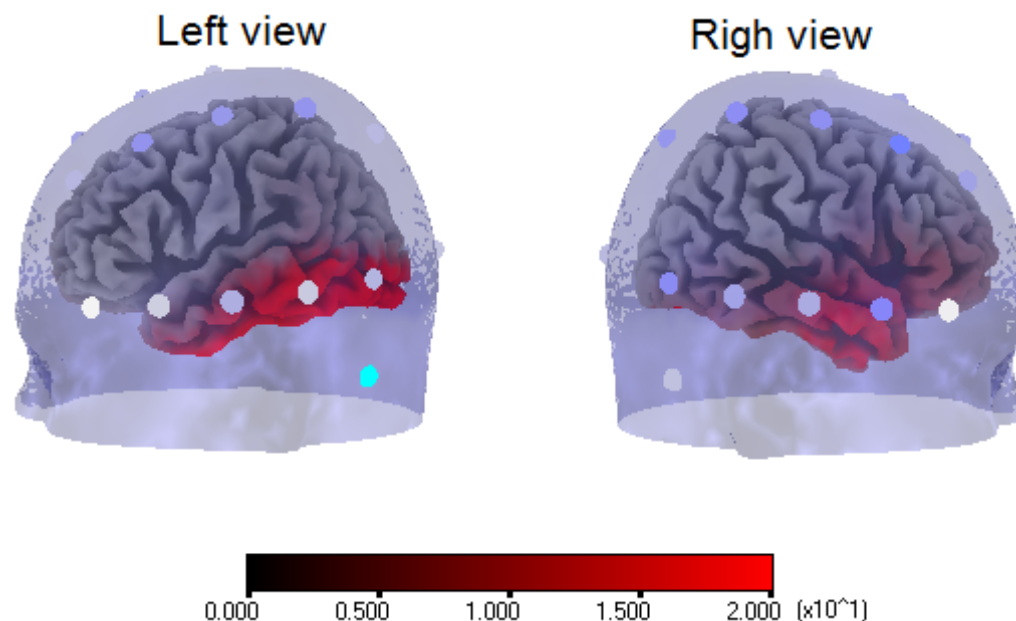
Fig. 2. The cortex current density reconstruction of the neutral-tone MMN (MMN_T0). The cortical generators of MMN_T0 were analyzed using the sLORETA method (standardized low-resolution brain electromagnetic tomography) with a standardized boundary element method volume conductor model and the Montreal Neurological Institute stereotactic coordinate, which could produce standardized current density images with zero localization error.

the memory traces of native syllables [23, 24]. Although the present study presented familiar congruent tones for adult native speakers of Mandarin Chinese, i.e., canonical tone of the first word and neutral tone of the second word (T1/2/3/4 + T0), the incongruent condition (T1/2/3/4 + T1/2/3/4) was infrequent indeed. Therefore, the present enhanced MMN for neutral tone further indicated the familiar effects of MMN based on long-term memory.

Interestingly, one recent study investigated the perception of T0 in infants learning Mandarin (tone language) and Dutch (stress language) and found that, after familiarizing to neutral tone sequences, Dutch infants distinguished T1T0 from T1T4 [12]. In contrast, Mandarin infants failed to distinguish the tone contrast. Dutch infants' persistent discrimination indicates that they may view neutral typical tonal contrast as lexical stress rather than tonal information. However, Mandarin-speaking infants' failure means that the representation of T0 is incomplete in their first year of life, so it may take longer than we expected to acquire tone categories [11].

The cortical source analysis of neutral-tone MMN exhibited the maximum current density at the left temporal lobe (Brodmann area 20, Inferior Temporal Gyrus, Temporal Lobe), in line with previous findings that the left superior temporal lobe was the major cortex generator of the MMN related to word processing [20, 28]. Generally, the amplitude of MMN recorded in the scalp was the largest in the fronto-central scalp region, and the generator sources revealed by equivalent current dipoles showed that the dominant MMN distribution in the front central lobe is mainly accounted for by the sum of the activities generated on bilateral

supra-temporal lobes [21]. Converging evidence revealed that at least two intracranial processes are involved in MMN: the bilateral supra-temporal process which produces the supra-temporal MMN subcomponent , and the right-hemisphere frontal process, which produces the frontal MMN subcomponent [20–22]. Presumably, the supra-temporal component is related to pre-perceptual change detection, while the frontal component is associated with an involuntary attentional switch concerning auditory changes. However, the generators of MMN related to language stimuli usually is left-lateralized [29].

Interestingly, there was evidence that the predominant MMN in the left hemisphere was only observed for the stimuli familiar to the central nervous system. The central nervous system has previously developed a memory network [30]. Based on the source analysis, we found evident left-temporal predominant neural-tone MMN without the frontal subcomponent, indicating the neutral tone's pre-perceptual memory-based change detection.

In sum, to explore the pre-attentive processing of Mandarin neutral tone, MMN was recorded using the deviant-standard-reverse paradigm that elicits different responses between deviant and standard stimuli. The neural-tone MMN was elicited with the cortex generator located at the left temporal lobe. These results provide new evidence for native speakers of Mandarin Chinese processing T0 automatically under non-attentional conditions and further indicate that T0 does exist as an automatically recognizable one in native Mandarin speakers' tone system, at least for adults.

Abbreviations

ERPs, event-related potentials; MMN, mismatch negativity.

Author contributions

Weijing Zhou and Lun Zhao designed the experiments and wrote the paper; Zhiyan Wang and Suwan Wang performed the experiments and the data analysis.

Ethics approval and consent to participate

The ethics committee of Yang Zhou University approved the study, and the written informed consent was obtained from all the participants before the testing.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Duanmu S. The phonology of standard Chinese. 2nd edn. New York: Oxford University Press. 2007.

[2] Feng SL. Interactions between morphology, syntax and prosody in Chinese. Beijing: Peking University Press. 1997.

[3] Lu J, Wang J. On defining "Qingsheng (neutral tone) ". Contemporary Linguistics. 2005; 7: 107-112.

[4] Zhu HY. Neutral tone' s character and its regulative principles. Applied Linguistics. 2009; 2: 34-41.

[5] Chao YR. A grammar of spoken Chinese. Berkeley: University of California Press. 1968.

[6] Li A. Phonetic correlates of neutral tone in different information structures. Contemporary Linguistics. 2017; 19: 348-378.

[7] Bao H, Lin M. Introduction of experimental phonetics. Beijing: Peking University Press. 2014.

[8] Cao J. A study on Mandarin lexical stress. Report of Phonetics Research. 2008; 20-29. (In Chinese)

[9] Chen A, Kager R. Discrimination of lexical tones in the first year of life. Infant and Child Development. 2016; 25: 426-439.

[10] Chen Y, Xu Y. Production of weak elements in speech-evidence from F0 patterns of neutral tone in standard Chinese. Phonetica. 2006; 63: 47-75.

[11] Deng D. Experimental study of Chinese prosodic word. Beijing: Peking University Press. 2010.

[12] Fan S, Li A, Chen A. Perception of lexical neutral tone among adults and infants. Frontiers in Psychology. 2019; 9: 322.

[13] Gao J, Li A. Production of neutral tone on disyllabic words by two-year-old Mandarin-speaking children. Studies on Speech Production. 2018; 8: 89-98.

[14] Lin T, Wang L. A course in phonetics. Beijing: Peking University Press. 2013.

[15] Liu L, Kager R. Perception of tones by infants learning a non-tone language. Cognition. 2015; 133: 385-394.

[16] Luo C, Wang J. An outline of common phonetics. Beijing: The Commercial Press. 2002.

[17] Wang, Y. The effects of pitch and duration on the perception of the neutral tone in standard Chinese. Acta Acustica. 2004; 29: 453-461.

[18] Wei G. Neutral tone and weak stress in Beijing Mandarin and the transcription of the Chinese phonetic alphabet of putonghua. Chinese Research. 2005; 6: 525-536. (In Chinese)

[19] Yang S. The synthesis rules on Mandarin neutral tone. Applied Acoustics. 1989; 10: 12-18.

[20] Zhong X, Wang B, Yang Y, Lü S. The perception of stress in prosodic words of standard Chinese. Acta Psychologica Sinica. 2001; 6: 481-488.

[21] Yip M. Tone. Cambridge: Cambridge University Press. 2002.

[22] Maddieson I. Tone. The world atlas of language structures online. In M. S. Dryer & M. Haspelmath (eds.) Leipzig: Max Planck Institute for Evolutionary Anthropology. 2013.

[23] Näätänen R, Paavilainen P, Rinne T, Alho K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical Neurophysiology. 2008; 118: 2544-2590.

[24] Rinne T, Alho K, Ilmoniemi RJ, Virtanen J, Näätänen R. Separate time behaviors of the temporal and frontal mismatch negativity sources. NeuroImage. 2000; 12: 14-19.

[25] Rinne T, Antila S, Winkler I. Mismatch negativity is unaffected by top-down predictive information. Neuroreport. 2001; 12: 2209-2213.

[26] Pulvermüller F, Kujala T, Shtyrov Y, Simola J, Tiitinen H, Alku P, et al. Memory traces for words as revealed by the mismatch negativity. NeuroImage. 2001; 14: 607-616.

[27] Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, et al. Electrophysiological markers of voice familiarity. The European Journal of Neuroscience. 2006; 23: 3081-3086.

[28] Atkinson RC, Shiffrin RM. Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (eds.) The psychology of learning and motivation: advances in research and theory (pp. 89-195). New York: Academic Press. 1968.

[29] Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS. A standardized boundary element method volume conductor model. Clinical Neurophysiology. 2002; 113: 702-712.

[30] Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods and Findings in Experimental and Clinical Pharmacology. 2003; 24: 5-12.