



## OPEN ACCESS

## EDITED BY

Caicai Zhang,  
The Hong Kong Polytechnic University,  
Hong Kong SAR, China

## REVIEWED BY

Jiaqiang Zhu,  
Hong Kong Polytechnic University,  
Hong Kong SAR, China  
Fei Chen,  
Hunan University, China

## \*CORRESPONDENCE

Chun-Hsien Hsu  
✉ kevinhsu@ncu.edu.tw

RECEIVED 01 March 2024

ACCEPTED 16 August 2024

PUBLISHED 11 September 2024

## CITATION

Hsu C-H, Cheong T-H and Huang W-J (2024)  
Exploring the impact of tonal inventory on  
speech perception across languages: a study  
of MMN responses in tonal language  
speakers.  
*Front. Psychol.* 15:1394309.  
doi: 10.3389/fpsyg.2024.1394309

## COPYRIGHT

© 2024 Hsu, Cheong and Huang. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Exploring the impact of tonal inventory on speech perception across languages: a study of MMN responses in tonal language speakers

Chun-Hsien Hsu<sup>1\*</sup>, Tong-Hou Cheong<sup>1</sup> and Wen-Jun Huang<sup>2</sup>

<sup>1</sup>Institute of Cognitive Neuroscience, National Central University, Taoyuan, Taiwan, <sup>2</sup>Department of Hakka Language and Social Sciences, National Central University, Taoyuan, Taiwan

Previous research on the perception of segmental features of languages has established a correlation between the phoneme inventory of a language and its speakers' perceptual abilities, as indexed by discrimination tasks and Mismatch Negativity (MMN). Building on this background, the current study elucidated the relationship between perceptual ability and tonal inventory by utilizing two tonal languages. Two groups of participants were included in the present experiment: Mandarin speakers and Hakka-Mandarin speakers. Onset latency analysis revealed a significant difference in the Mandarin syllable condition, with Hakka-Mandarin speakers demonstrating earlier MMN latency than Mandarin speakers. This suggests a more efficient auditory processing mechanism in Hakka-Mandarin speakers. Both groups, however, showed similar MMN latency in the Hakka syllable condition. The interaction between language background and syllable type indicates that other factors, such as syllable sonority, also influence MMN responses. These findings highlight the importance of considering multiple phonemic inventories and syllable characteristics in studies of tonal perception.

## KEYWORDS

MMN, lexical tones, tonal language, phonological inventory, speech perception

## 1 Introduction

In language comprehension, speech perception is a critical skill involving the discrimination of both segmental and suprasegmental features. Research utilizing electroencephalography (EEG) has demonstrated that the size of a phoneme inventory can significantly influence the amplitude of the mismatch negativity response (MMN; [Brunellière et al., 2011](#); [Hacquard et al., 2007](#); [Zhang et al., 2005](#)). MMN is a unique ERP used to detect distinguishable changes in acoustic features within a stream of sound and is not influenced by attention ([Näätänen et al., 1993, 1997](#)). It is assumed that MMN activity indexes an automatic, pre-attentive auditory processing mechanism, and it varies depending on several factors, including the participants' language experiences and acoustic features of the stimuli. [Hacquard et al. \(2007\)](#)'s study demonstrated that speakers of languages with larger vowel inventories (like French) showed larger MMN amplitudes compared to speakers of languages with smaller vowel inventories (like Spanish). This indicates that participants' language backgrounds, and particularly the size of their vowel inventories, influence how they perceive vowel changes. Their finding was consistent with [Teles and Huey \(2020\)](#)' demonstration that speakers of

languages with larger vowel inventories (e.g., English or French) would expand the space of vowel dispersion relative to speakers with smaller vowel inventories (e.g., Spanish), such that they would produce the target phonemes acoustically far away from one another. Taken together, the effects of phoneme inventory size on speech production and perception imply speakers of languages with complex vowel systems would exhibit enhanced perceptual sensitivity to vowels compared to those with simpler vowel systems.

While existing studies have explored the role of segmental features in both tonal and non-tonal languages, there is the question of whether the organization of tonal representation can have an effect on speech perception. The accurate perception of suprasegmental information is also crucial in language comprehension, as it can provide listeners with important linguistic cues, such as affective-prosodic cues and prosodic phrasing. This is particularly relevant in tonal languages, such as Chinese and Thai, where pitch differences are used to differentiate words (Chandrasekaran et al., 2007; Huang and Johnson, 2011). Therefore, this study aimed to explore whether the perception of lexical tones is processed in the same way regardless of the size of the tonal inventory of participants' language. This prediction was based on the finding that the size of the phonological inventory correlates with the ability to produce and perceive spoken sounds in monolinguals and bilinguals.

Instead of using behavioral measurements, the preset study applied the event-related potentials (ERP) method and measured the MMN responses to syllables. Numerous studies of speech perception have focused on MMN. The MMN paradigm typically involves a rapidly presented stream of repeated standard sounds occasionally interrupted by rare deviant sounds. MMN activity can be measured by comparing ERP responses to the deviant sound with those to the standard sound or by comparing ERP responses to the deviant sound in an MMN experiment with those to the same sound in an equal-probability control block (Jacobsen and Schröger, 2001). In addition to the amplitude of ERP activity, ERP studies also have demonstrated that the delay in the latency of MMN activity was associated with the insufficiency of phonological perception during the processing of linguistic stimuli (Alonso-Búa et al., 2006; Cheng et al., 2021; Zhang et al., 2005). For example, native Japanese speakers often struggle to differentiate between the English phonemes /r/ and /l/, both of which are mapped to Japanese /l/. Zhang et al. (2005) used magnetoencephalography (MEG) to record MMN in response to /r/ and /l/ sounds in native Japanese and native American English listeners. The study found that native Japanese listeners were less sensitive to the phonemic /r-l/ difference than native American English listeners, and their MMN amplitudes and latency were significantly smaller and longer, respectively.

In the case of tonal inventory, the five-scale tone representation scheme is a method used to describe the pitch contour of lexical tones in tonal languages (Chao, 1968). This scheme employs a numerical scale from 1 to 5, where each number represents a specific pitch level, with 1 being the lowest and 5 the highest. Using this scheme, the tonal contour of a syllable can be described by a sequence of these numbers (Figure 1A). For example, Mandarin Chinese syllables have four lexical tones (Duanmu, 2007; Li et al., 2023): the high level tone (Tone 1, or 55-tone, according to the five-scale tone representation scheme), high rising contour tone (Tone 2; 35-tone), low falling-raising contour tone (Tone 3; 214-tone), and high falling contour tone (Tone 4/ 51-tone). In addition to the full Tone 3

(214-tone), the half-Tone 3 (21-tone), a reduced form of the 214-tone frequently used in natural speech, is an allophonic variant of the traditional 214-tone (Fu and Lee, 2022; Lu and Lee-Kim, 2021; Zhang and Lai, 2010). Chandrasekaran et al. (2007) used two experimental blocks with different tonal contrasts. In one block, the participants frequently heard the syllable /yi/ with 55-tone and occasionally heard /yi/ with 214-tone. In the other block, the standard stimulus was /yi/ with 35-tone. The results showed that native Mandarin speakers' MMN responses to the 55/214 contrast were larger than their MMN responses to the 55/35 contrast, indicating that MMN amplitudes are correlated with the acoustic similarity between pairs of standard and deviant sounds. Native English speakers' MMN did not demonstrate the effect of tonal contrast on MMN. Furthermore, while native Mandarin speakers' MMN to the 55/214 contrast was larger than native English speakers' MMN to the 55/214 contrast, there was no significant group difference in the 55/35 contrast. These findings suggest that native tonal speakers are more sensitive to the height dimension than to the contour dimension, and that nontonal speakers' pitch perception does not appear to be significantly dependent on the height versus contour distinction.

As previously mentioned, it is not clear whether the size of a tonal language's inventory would affect the ability of speakers to perceive and differentiate pitch. To address this gap, the current study aimed to compare MMN responses to tonal contrasts in two groups of speakers: Mandarin Chinese and Hakka (Hailu)-Mandarin Chinese bilingual speakers to investigate whether there is a correlation between the complexity of the tonal system and the acuity of pitch perception. Hailu Hakka and Mandarin Chinese have similar basic vowels and word-initial consonants, and Hailu Hakka has more liaison consonants. They both have three syllable structures, including CV (Consonant-Vowel), CVV (Consonant-Vowel-Vowel) and CVC (Consonant-Vowel-Consonant) forms. Regarding the lexical tones of Hailu Hakka, there are seven distinct tones (Huang and Yu, 2022). Among these, four tones are similar to the lexical tones of Mandarin (Figure 1B): the high-level tone (55-tone), which corresponds to Mandarin's Tone 1; the high rising contour tone (35-tone), which resembles Mandarin's Tone 2; the low falling tone (21-tone), akin to the half-Tone 3 in Mandarin; and the high falling tone (52-tone), which parallels Mandarin's Tone 4. According to the review by Huang and Yu (2022), the low falling tone in Hakka has been coded as both a 31-tone and a 21-tone in different studies. This variation arises due to the different methods of normalization and analysis employed across these studies. To ensure consistency and avoid confusion, the low falling tone in Hakka was referred to as the 21-tone in the present study. Additionally, Hakka has unique tones that differentiate it from Mandarin: the middle-level tone (33-tone); the short high-level tone (55-tone), which is similar to Mandarin's Tone 1 but shorter in duration; and the short middle falling tone (42-tone), characterized by a mid-level pitch that falls to a lower pitch with a shorter duration. Building on the findings of Hacquard et al. (2007) regarding the impact of vowel inventory size on perceptual ability, one could posit that a similar correlation exists between the complexity of a tone system in tonal languages and the perceptual abilities of its speakers. That is, the additional tones in Hailu Hakka suggests a more complex tonal system, potentially indicating that speakers of Hailu Hakka have a more nuanced perceptual process sensitive to dynamic changes in tone. Therefore, one might expect that Mandarin participants may be less sensitive to the tonal changes as compared with Hakka-Mandarin speakers and exhibit reduced or delayed MMN response.

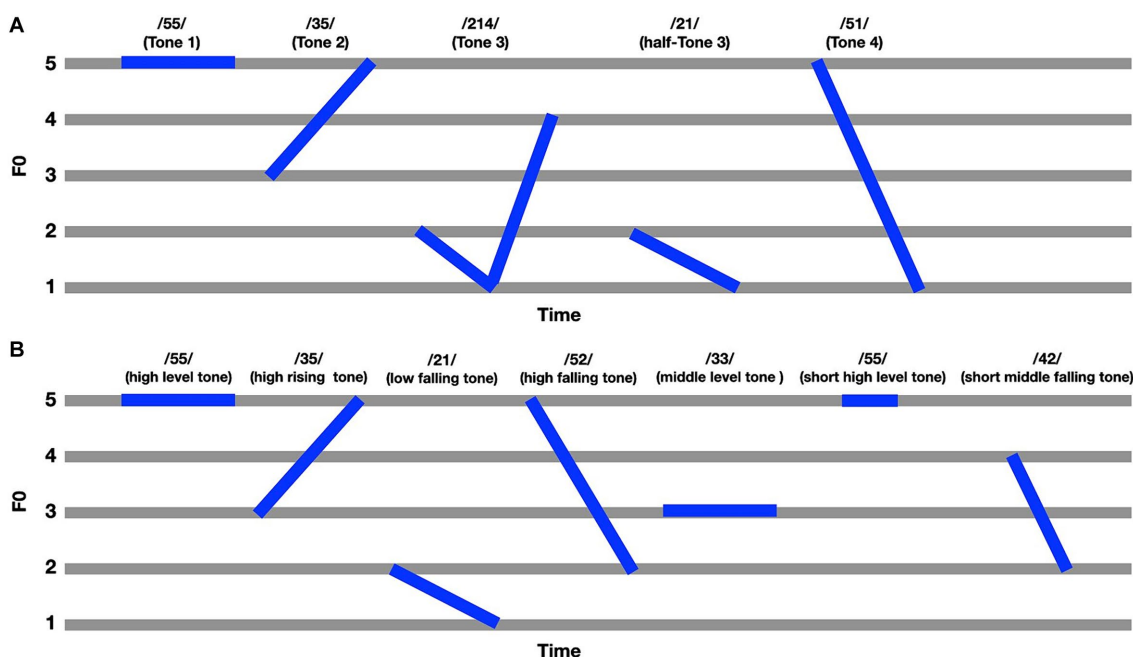


FIGURE 1

Representation of lexical tones in Mandarin Chinese (A) and Hailu Hakka (B) utilizing a five-level scale for tone marks (Chao, 1968). The digits on the left indicate the pitch level, where 1 corresponds to the lowest and 5 to the highest pitch. (A) The lexical tones of Mandarin using the five-level scale as described by Duanmu (2007); (B) The lexical tones of Hailu Hakka according to the study by Huang and Yu (2022).

## 2 Materials and methods

### 2.1 Participants

A total of 17 native Mandarin speakers and 16 Hakka-Mandarin bilinguals (aged 18–30 years) were recruited to participate in the MMN experiment. All participants spoke Taiwanese Mandarin. Mandarin speakers were undergraduate and graduate students at National Central University. They majored in engineering, earth science, life science, and English or French literature. All Hakka-Mandarin bilinguals were raised in communities where Hailu Hakka and Mandarin Chinese were the primary languages of daily conversation and were pursuing undergraduate or graduate studies at the College of Hakka Studies at National Central University, where Hailu Hakka served as the main language of instruction and classroom discourse. All of the participants had normal hearing and normal or corrected-to-normal vision. The current study was approved by the Human Subject Research Ethics Committee of National Taiwan University.

### 2.2 Stimuli

The experiment was conducted using two distinct sets of speech stimuli. The first set consisted of two Mandarin syllables, specifically /zu/ with 55-tone and 21-tone. The second set of stimuli consisted of the Hakka syllable /so/ with 55-tone and 21-tone. The selected Mandarin syllables are real words in Mandarin and are not words or morphemes in Hailu Hakka. The selected Hakka syllables are real words in Hakka and are not real words or morphemes in Mandarin. When recording these syllables, the native speakers were instructed to read a carrier

sentence in their respective native languages with the target syllable. These Mandarin syllables were sourced from a speech dataset (Sinica COSPRO 08\_M054) referenced in Tseng et al. (2005). The speaker was an adult male native speaker of Taiwanese Mandarin. The carrier sentence for the Mandarin syllables was “講話時動不動就會提到\_” (whenever talking frequently mentions\_). The speaker of Hakka syllables was a male native Hailu Hakka speaker who did not know Mandarin. These Hakka syllables were recorded in a phonological lab at Department of Hakka Language and Social Sciences and have been used in the study of Huang and Yu (2022). The carrier sentence for the Hakka syllables was “佢唸\_盡正” (He pronounces\_very accurately). The syllables were then normalized to a duration of 350 ms and intensity of 70 dB using Praat. Speech waveforms and acoustic parameters of the stimuli are shown in Figure 2 and Table 1, respectively. Hakka syllable /so/ with 55-tone and Mandarin syllable /zo/ with 55-tone are very similar in F0 contour and direction. Likewise, Hakka syllable /so/ with 21-tone and Mandarin syllable /zo/ with 21-tone are very similar in F0 contour and direction.

### 2.3 Procedure

Jacobsen and Schröger (2001) suggested that a control procedure would allow control of the state of refractoriness during the oddball paradigm. Accordingly, each participant was required to undergo four experimental blocks in the present study, consisting of a Mandarin-control block, Mandarin-MMN block, Hakka-control block, and Hakka-MMN block. During the experiment, participants were presented with spoken syllables of one language in each experimental block. In Mandarin blocks, participants heard the Mandarin Chinese syllable /zu/ with 55- and 21-tone. In Hakka blocks, the participants

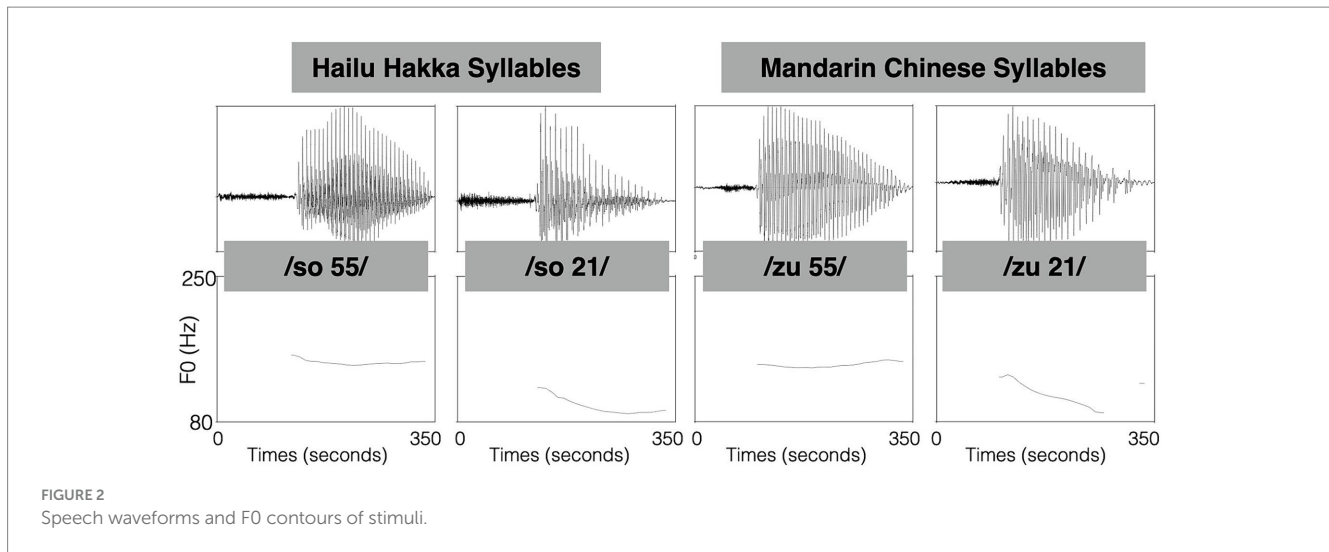


TABLE 1 Voice onset time (VOT), the first three formant frequencies and F0 range for each stimuli.

	Hakka	Hakka	Mandarin	Mandarin
	/so 55/	/so 21/	/zu 55/	/zu 21/
VOT (ms)	122	122	94	99
F1 (Hz)	624	645	402	393
F2 (Hz)	1,030	1,059	851	764
F3 (Hz)	2,815	2,813	2,962	2,691
F0 range (Hz)	146–155	92–131	142–152	90–134

heard the syllable /so/ with 55- and 21-tone. Each experimental block comprised 500 trials. In the control blocks, syllables were randomly presented in 55-tone or 21-tone with equal probabilities ( $p=0.5$ ). In MMN blocks, the 55-tone and 21-tone syllables were randomly presented for 100 ( $p=0.2$ ) and 400 ( $p=0.8$ ) trials, respectively. The order of the experimental blocks was counterbalanced across participants. Each trial began with the presentation of a syllable lasting 350 ms (70 dB), followed by a 400-ms inter-trial interval. The syllables were presented using two loudspeakers. While participating in the experiment, participants watched a movie without sound or subtitles.

## 2.4 Data recording and preprocessing steps

EEG data were recorded using 32 Ag/AgCl electrodes (QuickCap, Neuromedical Supplies, Sterling, United States). The electrodes were online-referenced to the average of the left and right mastoids for offline analysis. The EEG was continuously recorded and digitized at a rate of 1,024 Hz, and the signal was amplified using a Grael 4 K EEG amplifier with a band-pass filter at DC–409 Hz. Electrode impedances were kept below 5 k $\Omega$ . Eye movements and blinks were monitored using supraorbital and infraorbital electrodes, and electrodes in the external canthi.

Instead of using conventional ERP analyses, we utilized the Hilbert-Huang transformation (Hsu et al., 2016; Huang et al., 1998) for the offline analysis, because HHT can provide better resolution. In brief, HHT is a two-step method for analysis of nonlinear and nonstationary signals. The first step is empirical mode decomposition (EMD). EMD

is a data-driven, adaptive method that decomposes a signal into a finite number of intrinsic mode functions (IMFs) determined using an iterative sifting process that separates the signal into high and low frequency components. This decomposition method is adaptive and can automatically adjust to the signal's frequency content. Once the signal is decomposed into IMFs, the instantaneous phase of each IMF can be calculated using the direct quadrature transform. The instantaneous frequency can then be obtained by calculating the time derivative of the instantaneous phase. This approach allows for better frequency resolution as it estimates the frequency content of the signal at each time point, rather than using methods such as the convolution integral method used in Fourier transform and wavelet transform. Therefore, HHT is useful in the analysis of non-linear and non-stationary signals.

In the present study, EEG data were analyzed in the following manner based on the procedure described by Hsu et al. (2016). Continuous EEG data were epoched with 100 ms pre-stimulus intervals and 600 ms post-stimulus intervals. The pre-stimulus interval (–100 to 0 ms) was used for baseline correction. Trials were rejected if they were contaminated by voltage variations larger than 100  $\mu$ V in amplitude. We then decomposed each EEG segment into seven IMFs using the masked empirical mode decomposition (Quinn et al., 2021) and obtained event-related modes (ERMs) by averaging IMFs across trials (Al-Subari et al., 2015a,b). Similar to noise removal using filters, ERMs were obtained by summing the IMFs based on their instantaneous frequencies and then averaging across trials. The correlation between MMN components and frequency bands is well known, and Kalyakin et al. (2007) has suggested that EEG activity between 2 and 8 Hz reflects MMN. Previous studies of MMN



responses using the HHT method confirmed that EMD can be used to analyze EEG signals and employ IMFs with frequencies ranging between 2 and 8 Hz to estimate MMN activity (Cong et al., 2009; Hsu et al., 2016). Therefore, this study focused on IMFs with frequencies between 2 and 8 Hz to extract MMN-related activities.

## 2.5 Statistical analysis

### 2.5.1 Mass univariate cluster-based permutation tests

In statistical analyses, the MMN effect was measured by comparing the ERPs to the deviant stimuli of the MMN block and ERPs to the same stimuli of the control (equal-probability) block. The grand-averaged waveforms for the 55-tone syllables in the MMN and control blocks are presented in Figures 3A,B. Through visual inspection of the data, a negative-going component peaking at around 300 ms was observed, and this component was more prominent in the

waveforms of the MMN blocks than in those of the control blocks. We used mass univariate cluster-based permutation tests as recommended by Maris and Oostenveld (2007) to evaluate the significance of MMN activity (ERPs of the 55-tone stimuli in the MMN block versus ERPs of the same stimuli in the control block) of each participant group (Mandarin participants and Hakka-Mandarin participants) and each syllable (Mandarin syllable and Hakka syllable). This cluster-based nonparametric approach is recommended to control Type I error rates in electrophysiology experiments where precise latencies and scalp distribution are unknown *a priori*. The mass univariate cluster-based permutation tests were run using the Eelbrain package (version 0.39.8). The general procedure for the cluster-based test was as follows: a paired t-statistic (deviant minus control, one tail) was calculated at each time point and channel. Spatial-temporal clusters were then formed from test statistics that were contiguously significant (uncorrected  $p=0.01$ ) through time and channels. For each cluster, the cluster mass statistic was computed, which was the sum of all  $t$  values in the cluster. To determine the

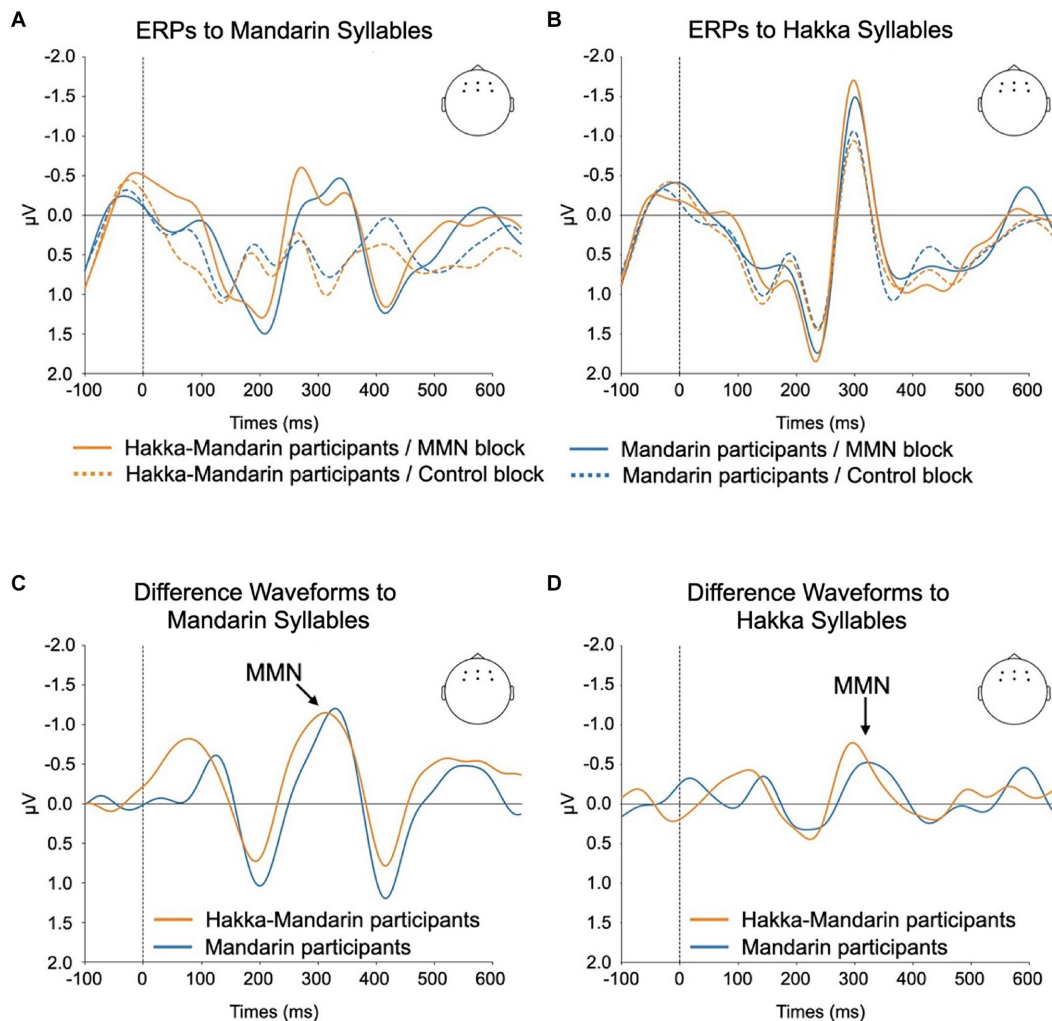


FIGURE 3

ERP waveforms elicited by speech stimuli averaged across six electrodes positioned on the anterior scalp. (A) displays the raw ERP waveforms in response to Mandarin syllables, while (B) illustrates the raw ERP waveforms elicited by Hakka syllables. (C,D) show difference waveforms for Mandarin and Hakka syllables, respectively. These difference waveforms are calculated by subtracting the ERP elicited by the 55-tone stimulus in the control block from the ERP of the identical stimulus in the MMN block.

reliability of these clusters, the actually observed cluster-level test statistics were compared against the null distribution based on 10,000 random permutations of the condition labels.

In addition, we also analyzed the MMN activity obtained by calculating ERP difference waves through a subtraction method (Figures 3C,D). Specifically, the ERP elicited by the 55-tone stimulus in the control block was subtracted from the ERP of the identical stimulus in the MMN block. Then, the same mass univariate cluster-based permutation tests were applied to evaluate the effect of participant groups (MMN activity of Hakka-Mandarin participants minus MMN activity of Mandarin participants) and the effect of syllable types (MMN activity of Mandarin syllable minus MMN activity of Hakka syllable).

### 2.5.2 Onset latency of MMN activity

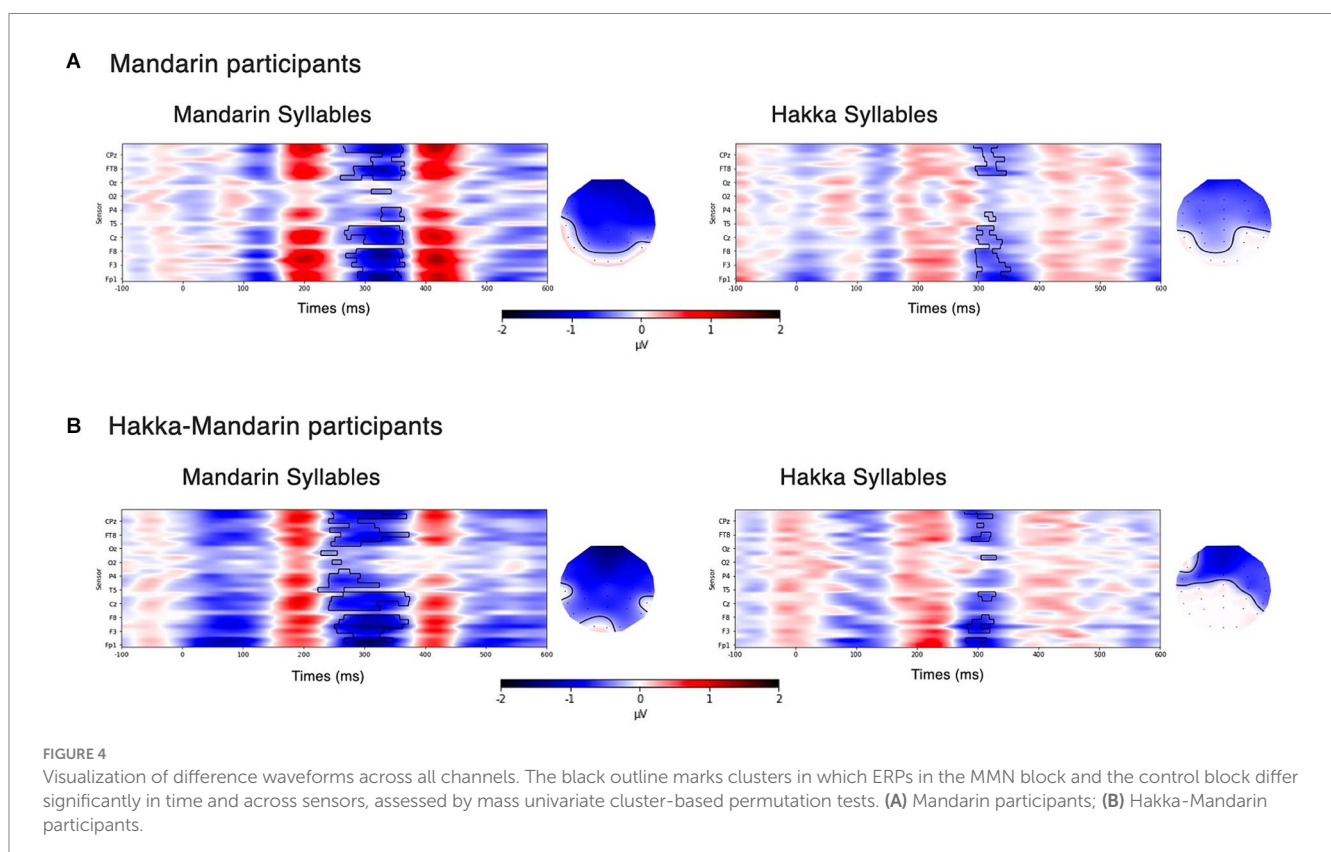
Since cluster-based permutation tests on ERPs do not establish precise effect onsets or offsets (Sassenhagen and Draschkow, 2019), we tested the onset latency of MMN activity using the method described below. Difference waves in six electrodes (F3, FC3, Fz, FCz, F4 and FC4) distributed on the frontal scalp were used for the analysis of the onset latency. The latency of MMN activity was defined by finding the most negative peak from 200 ms to 400 ms in the difference wave (deviant condition minus control condition) and then working backward in the waveform until 50% of that peak voltage was reached. This approach has been highly recommended for quantifying the onset of an activity while applying to difference waves (Kiesel et al., 2008; Luck et al., 2009). For statistical analyses, the onset latency of MMN were analyzed with the linear mixed-effects model (Bates, 2005) with two random factors (random intercepts for participants and electrodes). The use of mixed-effects

models with electrodes as a crossed random effect could omit strong effects mainly influenced by one electrode instead of all electrodes of interest. Language types (Hakka syllables versus Mandarin Syllable), participant groups (Hakka-Mandarin bilinguals versus Mandarin speakers), and their interaction were fixed factors. Latency was analyzed in R (Version 3.5.2) and RStudio (Version 1.1.463). The linear mixed-effects model was run using the “lmer” function as implemented in the lme4 package for R (Version 1.1–21). Reported *p*-values were calculated based on Satterthwaite’s method as implemented in the lmerTest package in R (Version 3.1–3). *Post hoc* comparisons were carried out using the “glht” function (the multcomp package, Version 1.4–15) with Bonferroni correction.

## 3 Results

### 3.1 MMN effect on ERPs in each language groups

The results reported below were considered significant at a level of  $p < 0.05$ . Statistical results of the mass univariate analysis are shown in Figure 4. Overall, the results demonstrated that ERP amplitudes to 55-tones differed significantly between deviant and control conditions. Please note that these analyses do not establish precise effect onsets or offsets (Sassenhagen and Draschkow, 2019). Latencies were reported here as descriptive statistics. For Mandarin participants (Figure 4A), the MMN activity was observed between 260 and 366 ms (cluster-level  $p < 0.001$ ) in the Mandarin syllable condition. On the other hand, their ERPs to Hakka syllables showed a significant spatial-temporal cluster



between 292 to 353 ms (cluster-level  $p < 0.05$ ). For Hakka-Mandarin participants (Figure 4B), the MMN activity was observed between 224 and 375 ms (cluster-level  $p < 0.001$ ) in the Mandarin syllable condition, and their ERPs to Hakka syllables showed a significant spatial-temporal cluster between 280 to 331 ms (cluster-level  $p < 0.05$ ).

### 3.2 Effects on MMN difference waveforms

While comparing the difference waves between Mandarin participants and Hakka-Mandarin participants, the mass univariate analysis did not reveal any significant spatial-temporal cluster. Finally, for the effect of syllable types, the mass univariate analysis also did not reveal any significant spatial-temporal cluster.

### 3.3 Effects on the latency of MMN activity

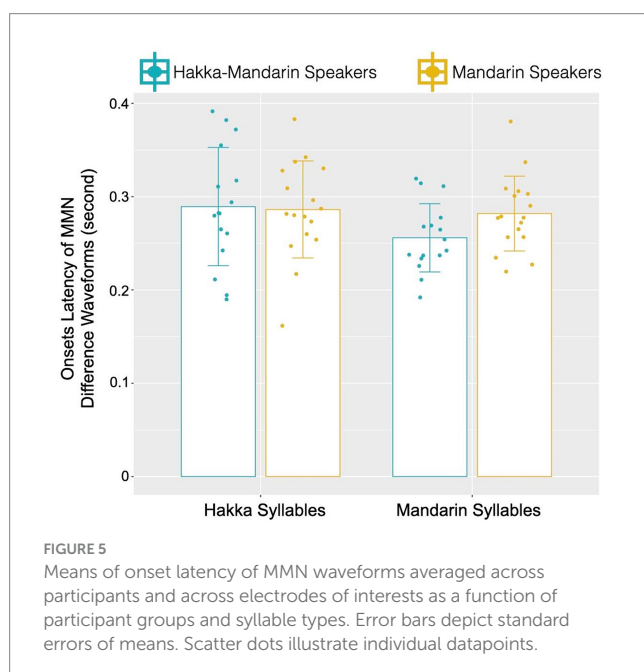
The linear mixed-effects model analysis showed a significant main effect of syllable types (Hakka syllables versus Mandarin Syllable) on the latency of MMN ( $\beta = -0.033$ ,  $SE = 0.006$ ,  $p < 0.0001$ ), and the main effect of participant groups was not significant ( $\beta = -0.003$ ,  $SE = 0.012$ ,  $p = 0.328$ ). That is, Mandarin sounds elicited earlier latency of MMN (269 ms) than Hakka sounds (287 ms). In addition, there was a significant two-way interaction between syllable types and participant groups ( $\beta = 0.029$ ,  $SE = 0.009$ ,  $p < 0.01$ ), which showed that the simple main effect of participant groups was significant in Mandarin syllable condition ( $\beta = -0.026$ ,  $SE = 0.013$ ,  $p < 0.05$ ) but not in Hakka syllable condition ( $\beta = 0.003$ ,  $SE = 0.013$ ,  $p > 0.05$ ). Specifically (Figure 5), Hakka-Mandarin participants' MMN latency (255 ms) was earlier than Mandarin participants' MMN (281 ms) in the Mandarin syllable condition, and both participant groups showed similar MMN latency in the Hakka syllable condition (Hakka-Mandarin participants: 289 ms; Mandarin participants: 286 ms). Likewise, the simple main effect of syllable types was significant in

Hakka-Mandarin participants ( $\beta = 0.034$ ,  $SE = 0.007$ ,  $p < 0.01$ ) but not in Mandarin participants ( $\beta = 0.004$ ,  $SE = 0.006$ ,  $p > 0.05$ ).

## 4 Conclusion

The present study found that ERPs to syllables in a deviant condition overall showed MMN activity. This finding was consistent with previous studies of MMN activity to tonal contrast which suggested that MMN activity is a robust index for the perception of pitch height in tonal language. As regarding the effect of tonal inventory, the working hypothesis for the present study was based on the assumption drawn in prior studies on vowel inventory. Specifically, Hacquard et al.'s (2007) study on MMN activity to vowel contrast demonstrated that the inventory organization of vowels would have an effect on perceived similarity which can be measured by the strength of MMN response. Building on the findings of the impact of phoneme inventory size on perceptual ability, one could posit that a similar correlation would exist between the complexity of a tone system in tonal languages and the perceptual abilities of its speakers. Note that the tonal inventories of Hailu Hakka are greater than those of Mandarin Chinese. Therefore, it was hypothesized that the MMN of Hakka-Mandarin speakers would be larger or earlier than that of Mandarin participants.

Although the present results for amplitude data did not show a significant difference between Hakka-Mandarin participants and Mandarin participants, the analysis of onset latency did show a significant difference between participant groups in the Mandarin syllable condition, and the pattern of MMN latencies aligned with the hypothesis. Our results showed that Hakka-Mandarin speakers demonstrated earlier MMN latency than Mandarin speakers in processing Mandarin syllables. In the literature, MMN is viewed as a reflection of an automatic, pre-attentive auditory processing mechanism, so the present observation of delayed MMN in Mandarin participants provides insights into the speed and efficiency of their auditory change detection. On the other hand, both participant groups showed similar MMN latency in the Hakka syllable condition. It may appear somewhat unexpected that the effect of tonal inventory was found only for the Mandarin syllable condition. The interaction between participants' language background (Hakka-Mandarin speakers versus Mandarin speakers) and types of syllables (Mandarin syllable /zu/ and Hakka syllable /so/) on the latency of MMN might indicate that other factors also play a crucial role on MMN responses to tonal contrast. One possibility is that the sonority of syllable also had an effect here. Since the stimuli used in the presented experiment were CV syllables without consonant clusters, it is assumed that the quality of the vowel would be a factor that influence the perceived sonority of the syllable. In the sonority hierarchy, mid vowels like /o/ are often considered to be more sonorous than high vowels like /u/. There is evidence suggesting that the perception of suprasegmental features, such as tonal contrasts, is more accurate in regions of higher sonority (Niebuhr, 2007; Zhang, 2002). Given this, the Hakka syllable /so/ could result in similar MMN latencies for both Hakka-Mandarin speakers and Mandarin speakers because the high sonority enhances the robustness of its auditory processing. On the other hand, individuals with multiple phonemic inventories may have more refined auditory processing capabilities, which could affect their MMN responses. Therefore, the experiences with multiple phonemic



inventories could enhance Hakka-Mandarin speakers' perceptual ability for the less sonorous syllable /zu/, leading to earlier MMN latencies compared to Mandarin speakers.

One might suggest that the observed tonal inventory effect could be attributed to a training effect, given that Hakka-Mandarin speakers are enrolled in undergraduate or graduate programs where Hailu Hakka is the primary language of instruction. However, if training were a concern, it is important to consider the relationship between lexical familiarity and MMN activity to words. Since bilingual individuals divide their language use between two languages, their frequency of use in each language is typically lower than that of monolinguals who use a single language. This reduced frequency of use would result in lower familiarity with vocabulary in each language. Studies using vocabulary fluency tasks have indicated that bilingual adults and children often experience difficulty in word production across languages (for a review, see Bialystok et al., 2009). Given that lexical familiarity also influences MMN to spoken words (Pulvermüller and Shtyrov, 2006), it could be expected that Hakka-Mandarin speakers might exhibit reduced or delayed MMN activation compared to Mandarin speakers. That is, if the training effect were significant, Hakka-Mandarin speakers would likely show reduced or delayed MMN responses due to lower familiarity. The fact that they do not exhibit such patterns suggests that the observed effects may be due to tonal inventory and language background.

It is worth noting that the Hakka syllables /so/ used in the present experiment are not real syllables in Mandarin Chinese, and the Mandarin syllables /zu/ are not real syllables in Hakka. By using syllables that are unique to each language, it becomes easier to attribute differences in perceptual processing to the participants' language background. For example, Kuo et al. (2016) investigated the effect of phoneme inventory on speech perception, and their results showed that bilingual children have a heightened phonological awareness while processing consonants that were shared between languages. Their results suggested that bilingual children would benefit from experiencing phonological segments in more variable contexts. This enhanced experience allows bilingual children to better dissociate phonological segments from their contexts, giving them an advantage in processing shared onsets. In the present study, the results demonstrated the effect of tonal inventory in processing unique syllables. Our approach avoids explanations based on shared phonological representations in richer and more variable contexts, which could also enhance speech perception abilities. Nevertheless, the present results were also in line with Kuo et al. (2016)'s finding that bilingual individuals develop a heightened awareness of language structure due to their exposure to a wider variety of phonemic contrasts.

Another observation to mention is that Mandarin speakers exhibited MMN activity to Hakka syllable /so/ which is not a real word or real morpheme in Mandarin Chinese. We speculated that this finding could reflect the organization of the phonological representation in tonal language speakers, particularly demonstrating that tonal representations are stored and processed parallel to segmental features (Malins and Joannisse, 2010). That is, MMN is a brain response to violations of a rule established by a sequence of auditory stimuli. Therefore, in the context of language, MMN to tonal changes is associated with how elements of syllables are processed and

stored. The present study's finding that Mandarin speakers exhibit MMN to tonal changes in a non-native syllable suggests that their phonological system is attuned not only to the segmental features of their native language but also to its tonal aspects. The tonal patterns used in the present experiment are native tones, which Mandarin speakers are familiar with. This familiarity might allow the mapping of these tonal patterns onto existing tonal exemplars in their lexicon, even though the segmental part of the stimuli is non-native. This finding is consistent with other studies on perceptual integration of tones and syllables in non-native speech perception (Choi and Tsui, 2023; Lin and Francis, 2014).

Although the present study provides implications of the tonal inventory effect on speech perception, some issues still need to be investigated in the future. The present study did not utilize tools such as the Language History Questionnaire (LHQ) to gather detailed information about the participants' linguistic backgrounds (Li et al., 2020). Using an LHQ in future studies would allow for a more comprehensive profile of each participant's language experience, which can help to provide a richer context for interpreting the results. Secondly, this study did not account for the participants' musical backgrounds, which could have an influence on auditory processing and speech perception. Musical training is known to enhance auditory discrimination skills. Future research should include an assessment of the participants' musical backgrounds to determine whether musical experience contributes to differences in MMN responses and to isolate the effects of linguistic experience from those of musical training. Finally, while MMN activity is associated with the diversity of tonal representations, a caveat emerges when considering the influence of the size of the tonal space. This aspect remains ambiguous, particularly in light of speech production studies that have not found notable differences in the fundamental frequency (F0) space across tonal languages (Kuang, 2013). Therefore, while it is clear that MMN activity is responsive to variations in tones, the extent to which the size of the tonal space itself affects this neural response has not been well-established.

In conclusion, this study provides valuable insights into how speakers of languages with more intricate tonal systems, such as Hailu Hakka with its seven distinct tones and variations in pitch height, contour, and duration, exhibit enhanced perceptual sensitivity to tonal contrasts compared to speakers of languages with simpler tonal systems. These findings should be interpreted with caution as the language history data might be insufficient. Despite of this limitation, the results support the idea that a linguistically rich environment shapes perceptual processes and fosters greater perceptual acuity.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by Human Subject Research Ethics Committee of National Taiwan University. The studies



were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

C-HH: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. T-HC: Writing – review & editing, Software, Methodology, Formal analysis, Data curation. W-JH: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Funding acquisition, Conceptualization.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from Minister of Education to C-HH and W-JH, and from Taiwan National Science and Technology Council (112-2628-H-008-002 and 108-2636-H-008-001) to C-HH.

## References

- Alonso-Búa, B., Díaz, F., and Ferraces, M. (2006). The contribution of AERPs (MMN and LDN) to studying temporal vs. linguistic processing deficits in children with reading difficulties. *Int. J. Psychophysiol.* 59, 159–167. doi: 10.1016/j.ijpsycho.2005.03.020
- Al-Subari, K., Al-Baddai, S., Tomé, A., Goldhacker, M., Faltermeier, R., and Lang, E. W. (2015b). EMDLAB: a toolbox for analysis of single-trial EEG dynamics using empirical mode decomposition. *J. Neurosci. Methods* 253, 193–205. doi: 10.1016/j.jneumeth.2015.06.020
- Al-Subari, K., Al-Baddai, S., Tomé, A. M., Volberg, G., Hammwöhner, R., and Lang, E. W. (2015a). Ensemble empirical mode decomposition analysis of EEG data collected during a contour integration task. *PLoS One* 10:e0119489. doi: 10.1371/journal.pone.0119489
- Bates, D. (2005). Fitting linear mixed models in R. *R News* 5, 27–30.
- Bialystok, E., Craik, F. I., Green, D. W., and Gollan, T. H. (2009). Bilingual minds. *Psychol. Sci. Public Interest* 10, 89–129. doi: 10.1177/1529100610387084
- Brunellière, A., Dufour, S., and Nguyen, N. (2011). Regional differences in the listener's phonemic inventory affect semantic processing: a mismatch negativity (MMN) study. *Brain Lang.* 117, 45–51. doi: 10.1016/j.bandl.2010.12.004
- Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Res.* 1128, 148–156. doi: 10.1016/j.brainres.2006.10.064
- Chao, Y. R. (1968). *Language and symbolic systems*, vol. 260. Cambridge: Cambridge University Press.
- Cheng, Y.-Y., Wu, H.-C., Shih, H.-Y., Yeh, P.-W., Yen, H.-L., and Lee, C.-Y. (2021). Deficits in processing of lexical tones in mandarin-speaking children with developmental language disorder: electrophysiological evidence. *J. Speech Lang. Hear. Res.* 64, 1176–1188. doi: 10.1044/2021\_JSLHR-19-00392
- Choi, W., and Tsui, R. K.-Y. (2023). Perceptual integrality of foreign segmental and tonal information: dimensional transfer hypothesis. *Stud. Second. Lang. Acquis.* 45, 1056–1073. doi: 10.1017/S0272263122000511
- Cong, F., Sipola, T., Huttunen-Scott, T., Xu, X., Ristaniemi, T., and Lyytinen, H. (2009). Hilbert-huang versus morlet wavelet transformation on mismatch negativity of children in uninterrupted sound paradigm. *Nonlinear Biomed. Phys.* 3, 1–8. doi: 10.1186/1753-4631-3-1
- Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford: Oxford University Press.
- Fu, Y., and Lee, Y. C. (2022). The production of tone 3 by advanced Korean learners of mandarin. *Linguistic Res.* 39, 213–233. doi: 10.17250/khisli.39.1.202203.008
- Hacquard, V., Walter, M. A., and Marantz, A. (2007). The effects of inventory on vowel perception in french and spanish: an MEG study. *Brain Lang.* 100, 295–300. doi: 10.1016/j.bandl.2006.04.009

## Acknowledgments

The authors would like to express their sincere gratitude to Neil Muggleton for his invaluable assistance in proofreading the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Hsu, C.-H., Lee, C.-Y., and Liang, W.-K. (2016). An improved method for measuring mismatch negativity using ensemble empirical mode decomposition. *J. Neurosci. Methods* 264, 78–85. doi: 10.1016/j.jneumeth.2016.02.015

Huang, T., and Johnson, K. (2011). Language specificity in speech perception: perception of mandarin tones by native and nonnative listeners. *Phonetica* 67, 243–267. doi: 10.1159/000327392

Huang, E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Royal Soc. London* 454, 903–995. doi: 10.1098/rspa.1998.0193

Huang, W.-J., and Yu, H.-M. (2022). Acoustic study of Taiwanese Hailu Hakka Tone values in isolation. *Global Hakka Stud.* 19, 1–38.

Jacobsen, T., and Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology* 38, 723–727. doi: 10.1111/1469-8986.3840723

Kalyakin, I., González, N., Joutsensalo, J., Huttunen, T., Kaartinen, J., and Lyytinen, H. (2007). Optimal digital filtering versus difference waves on the mismatch negativity in an uninterrupted sound paradigm. *Dev. Neuropsychol.* 31, 429–452. doi: 10.1080/87565640701229607

Kiesel, A., Miller, J., Jolicœur, P., and Brisson, B. (2008). Measurement of ERP latency differences: a comparison of single-participant and jackknife-based scoring methods. *Psychophysiology* 45, 250–274. doi: 10.1111/j.1469-8986.2007.00618.x

Kuang, J. (2013). The tonal space of contrastive five level tones. *Phonetica* 70, 1–23. doi: 10.1159/000353853

Kuo, L. J., Uchikoshi, Y., Kim, T. J., and Yang, X. (2016). Bilingualism and phonological awareness: re-examining theories of cross-language transfer and structural sensitivity. *Contemp. Educ. Psychol.* 46, 1–9. doi: 10.1016/j.cedpsych.2016.03.002

Li, C., Al-Tamimi, J., and Wu, Y. (2023). Tone as a factor influencing the dynamics of diphthong realizations in standard mandarin. In Radek Skarnitzl & Jan Volin (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 1876–1880).

Li, P., Zhang, F., Yu, A., and Zhao, X. (2020). Language history questionnaire (LHQ3): an enhanced tool for assessing multilingual experience. *Biling. Lang. Cogn.* 23, 938–944. doi: 10.1017/S1366728918001153

Lin, M., and Francis, A. L. (2014). Effects of language experience and expectations on attention to consonants and tones in English and mandarin Chinese. *J. Acoust. Soc. Am.* 136, 2827–2838. doi: 10.1121/1.4898047

Lu, Y.-A., and Lee-Kim, S.-I. (2021). The effect of linguistic experience on perceived vowel duration: evidence from Taiwan mandarin speakers. *J. Phon.* 86:101049. doi: 10.1016/j.wocn.2021.101049

- Luck, S. J., Kappenman, E. S., Fuller, R. L., Robinson, B., Summerfelt, A., and Gold, J. M. (2009). Impaired response selection in schizophrenia: evidence from the P3 wave and the lateralized readiness potential. *Psychophysiology* 46, 776–786. doi: 10.1111/j.1469-8986.2009.00817.x
- Malins, J. G., and Joannis, M. F. (2010). The roles of tonal and segmental information in mandarin spoken word recognition: an eyetracking study. *J. Mem. Lang.* 62, 407–420. doi: 10.1016/j.jml.2010.02.004
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0
- Näätänen, R., Paavilainen, P., Titinen, H., Jiang, D., and Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology* 30, 436–450. doi: 10.1111/j.1469-8986.1993.tb02067.x
- Niebuhr, O. (2007). The signalling of German rising-falling intonation categories—the interplay of synchronization, shape, and height. *Phonetica* 64, 174–193. doi: 10.1159/000107915
- Pulvermüller, F., and Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79, 49–71. doi: 10.1016/j.pneurobio.2006.04.004
- Quinn, A. J., Lopes-dos-Santos, V., Dupret, D., Nobre, A. C., and Woolrich, M. W. (2021). EMD: empirical mode decomposition and hilbert-huang spectral analyses in python. *J. Open Source Softw.* 6:2977. doi: 10.21105/joss.02977
- Sassenhagen, J., and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* 56:e13335. doi: 10.1111/psyp.13335
- Teles, L., and Huey, O. (2020). “The effects of the French vowel inventory on vowel 1102 production in spanish speakers,” in *University of Rochester Working Papers in the Language Sciences*. ed. P. Guekguezian (Cambridge, MA: MIT Working Papers in Linguistics), 8, 15–32.
- Tseng, C.-Y., Cheng, Y.-C., and Chang, C.-H. (2005). Sinica COSPRO and toolkit—corpora and platform of mandarin Chinese fluent speech. In *Proceedings of the 8th Oriental COCODA*, pp. 23–28.
- Zhang, J. (2002). The effects of duration and sonority on contour tone distribution – A typological survey and formal analysis. New York, NY: Routledge.
- Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., and Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage* 26, 703–720. doi: 10.1016/j.neuroimage.2005.02.040
- Zhang, J., and Lai, Y. (2010). Testing the role of phonetic knowledge in mandarin tone sandhi. *Phonology* 27, 153–201. doi: 10.1017/S0952675710000060