Processing of acoustic and phonological information of lexical tones at pre-attentive and attentive stages

Yicheng Rong, Yi Weng, and Gang Peng*

Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China;

***Corresponding author**: Gang Peng

E-mail: gpeng@polyu.edu.hk

Address: Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

Processing of acoustic and phonological information of lexical tones at pre-attentive and attentive stages

While Mismatch Negativity (MMN) and P300 have been found to correlate with the processing of acoustic and phonological information involved in speech perception, there is controversy surrounding how these two components index acoustic and/or phonological processing at pre-attentive and attentive stages. The current study employed both passive and active oddball paradigms to examine neural responses to lexical tones at the two stages in Cantonese speakers, using the paradigm of categorical perception (CP) where the between- and within-category deviants share the same acoustic distance from the standard but differ in the involvement of phonological information. We failed to observe a CP effect in P300, which might indicate that this component doesn't necessarily index phonological processing, while MMN does, as reflected by the finding of a greater MMN amplitude elicited from the between-category than within-category deviant. Nevertheless, phonological processing might be overridden by acoustic processing among participants who are sensitive to pitch.

Keywords: acoustic information; phonological information; Mismatch Negativity; P300; categorical perception; lexical tones

Introduction

Categorical perception (CP) lays the solid foundation for efficient speech perception by allowing continuous speech variations to be segmented into discrete categories. Earliest studies revealing the CP mechanism of speech perception focused on segments, where the perception of consonants has been found to exhibit a typical CP pattern (Liberman et al., 1957) while that of isolated vowels is far more gradient (Fry et al., 1962). Subsequent studies have found that suprasegmental components can also be perceived in a CP manner. For instance, lexical tones which contrast semantic meanings in tone languages can also be perceived categorically among native speakers when differences in pitch slope direction is involved (e.g., one level and one contour tones or two contour tones with opposite directions; Gandour, 1983; Hallé et al., 2004; Howie, 1976; Peng et al., 2010; W. S.-Y. Wang, 1976; Xu et al., 2006). Recently, there has been a growing tendency to investigate the neurocognitive correlates underlying the CP of lexical tones in typically developing individuals (Xi et al., 2010; Yu et al., 2014, 2017; Zheng et al., 2012) and clinical populations (Chen & Peng, 2020; X. Wang et al., 2017). Among these studies, the time course of acoustic and phonological processing was of particular interest. However, the processing of acoustic and phonological information during pre-attentive and attentive stages is still understudied.

Categorical perception of lexical tones

The CP paradigm is the most frequently used apparatus to assess the ability to perceive lexical tones in a categorical manner. Typically, the CP paradigm includes an identification task and a discrimination task (McMurray, 2022), where a synthetic speech continuum from one phonological category to another is used to simulate continuously changing sound variations. In the identification task, perceivers are required to allocate the sound stimuli into one of two categories. If two phonological categories are perceived categorically, perceivers will show consistency in the selection of a category when identifying sounds close to the two ends of the continuum, giving rise to step-like boundaries (Repp, 1984). The discrimination task can be implemented with different approaches. For instance, the AX approach requires perceivers to judge whether the sequentially presented two sounds are the same or different. Usually, there are peaks in discrimination function around category boundaries, that is perceivers show high sensitivity in perceiving the differences between sound pairs across the category boundaries which are labelled as the between-category contrasts. In addition, perceivers display low sensitivity in discriminating sound pairs at the two continuum ends, as the two sounds of each pair here are usually assigned to the same category. This cluster of sound pairs is called within-category contrasts. Apart from being treated with distinctive levels of sensitivity, within- and between-category contrasts engage different levels of processing: low-level acoustic processing and high-level phonological processing. Given that the within-category contrasts are considered variations of the same category, they only differ acoustically rather than phonologically. In contrast, the discrimination of between-category contrasts, where two sounds of each pair belong to different categories, involves differences at both acoustic and phonological levels (Yu et al., 2020).

The presence of CP seems to highly rely on the existence of phonemic contrast in the perceivers' native languages (Hallé et al., 2004, Peng et al., 2010). One example is that speakers of non-tonal languages were found to show reduced degree of CP of lexical tones (e.g., the boundary slope was smooth, and the discrimination peak was less sharp and might be far from the boundary) due to a lack of knowledge of phonological information brought by tone language experience (Hallé et al., 2004; Peng et al., 2010; W. S.-Y. Wang, 1976; Xu et al., 2006). In addition, poor language experience, which can refer to limited exposure to language or difficulties in processing or understanding language, may underlie a delay in the development of CP of lexical tones. This relationship has been observed in clinical populations, such as individuals with autism spectrum disorder (ASD) who showed wider boundary widths and lower sensitivity in discriminating between-category contrasts compared to typically developing age-matched peers, and their performance was similar to that of younger individuals (Chen & Peng, 2021; Rong et al., 2023).

Neural mechanisms underlying categorical perception of lexical tones

While the CP of lexical tones has been extensively investigated, the corresponding neural mechanisms of this process still needs to be clarified. As the processing of both acoustic and phonological information is involved, the paradigm of CP offers a window to examine the differences between low-level acoustic processing and high-level phonological processing. Several studies have compared the two levels of processing by comparing within- and between-category contrasts using the electrophysiological (EEG) measurements (Xi et al., 2010; Yu et al., 2014; L. Zhang et al., 2012; Zheng et al., 2012). The results of these EEG studies suggest that two event-related potentials (ERP) components obtained from the oddball paradigm, Mismatch Negativity (MMN) and P300, appear to correlate with CP. However, there is controversy about how MMN and P300 index acoustic and/or phonological processing at the pre-attentive and attentive stages, respectively.

Some scholars held the view that only the low-level acoustic information could be processed at the pre-attentive stage, and that the functional processing of high-level phonological information occurs at a later attentive stage (Luo et al., 2006). However, more recent EEG studies have provided evidence that phonological processing takes place at the earlier pre-attentive stage, suggesting that it can be automatic without attention (Xi et al., 2010; Yu et al., 2014, 2017). Specifically, Xi et al. (2010) and Yu et al. (2014) found that the MMN amplitude of the response to the between-category deviant was significantly greater relative to the within-category deviant in Mandarin speakers. Similar results were observed in speakers of another tone language, Cantonese (Yu et al., 2017). However, it is important to note that in Yu et al., (2017), the between-category deviant was acoustically farther from the standard than the within-category deviant. Specifically, the maximum difference in F0 between within-category deviant and standard was around 20 Hz while the maximum acoustic distance of between-category deviant to standard was nearly 120 Hz. It seemed more difficult to detect the within-category deviant due to the small acoustic difference. Previous research suggests that with increasing task demands, the MMN amplitude tends to be smaller (Polich, 2007). Therefore, it is not surprising that Yu et al. (2017) observed a significantly smaller MMN amplitude elicited from their within-category deviant whose detection is higher demanding relative to the between-category deviant. The CP effect in MMN amplitude in Cantonese speakers observed by Yu et al. (2017) might be the product of larger acoustic difference of between-category deviant to standard rather than the involvement of phonological processing. Therefore, one aim of the current study is to examine the acoustic and phonological processing by Cantonese speakers at the pre-attentive stage using between- and within-category deviants that share the same acoustic distance from the standard.

The acoustic and phonological cues have also been revealed to modulate the attention-related ERP component — P300 in the active oddball paradigm where focal attention is involved (Neisser, 1967). Unlike the passive oddball paradigm, the active one requires perceivers to attend to the auditory stimuli and make a response (e.g., by key-press) once they detect a deviant sound. Currently, there is some inconsistency in the relationship between phonological processing and P300. Zheng et al. (2012) proposed that the P300 amplitude could index the use of phonological cues when discriminating speech sounds, which was supported by their finding among Cantonese speakers which showed between-category deviant elicited significantly greater P300 amplitude relative to the equally spaced within-category deviant. However, the CP effect in terms of P300 amplitude (between > within) was not observed in another group of participants in their study whose native language was Mandarin. One possible explanation for the absence of CP effect in Mandarin speakers is that they are less sensitive to pitch changes compared to their Cantonese-speaking peers given the less dense tone system of Mandarin than Cantonese (Peng, 2006). By contrast, Chen and Peng (2020) comparing CP of lexical tones between Mandarin-speaking amusics and controls found significantly greater P300 amplitude of response to the between-category deviant relative to the within-category one among the participants, be they neurologically healthy controls or amusics. The inconsistency might be caused by a larger step size used in Chen and Peng (2020) relative to Zheng et al. (2012), which allowed Mandarin speakers in Chen and Peng (2020) to detect the deviants relatively easier. In addition, it should not be ignored that the different P300 amplitudes elicited from the between- and within-category deviants observed in previous studies might be driven by the different rates of physical responses (i.e., pressing keys) to the two types of deviants. Specifically, in the active oddball paradigm, participants were required to respond to the deviant stimuli by pressing a button or making some other physical responses. Physical response has been believed to be one of several factors that can influence the P300 component (Hillyard & Kutas, 1983; Verleger et al., 2005, 2016). One issue that can arise in the active oddball paradigm is that the detection rates of the different types of deviants (i.e., between- and within-category deviants) may not be equal. For example, participants may be better at detecting between- than within-category deviants. If there are fewer physical hits to the within-category deviants, this could pose possible artifacts on the P300 component. This is because the physical response can potentially affect the P300 component. Specifically, if participants are unable to detect within-category deviants reliably, it may lead to fewer physical responses to these deviants. Consequently, this could have an impact on the P300 component. To address this issue, the current study tried to rule out the possible artifacts on the P300 component by using a refined experimental material with the detectability of two types of deviants controlled. Specifically, we used a seven-step stimuli continuum that was based on two naturally produced sounds and the sounds were selected to ensure that the acoustic distance between them was sufficient to yield comparable detection rates between Stimulus #1 and Stimulus #7 when Stimulus #4 was the standard, one of which must be a within-category deviant and the other between-category deviant depending on the boundary position.

To summarize, the current study would compare both peak latency and amplitude of MMN and P300 evoked by between- and within-category deviants at both pre-attentive and later attentive stages with the following question was particularly pursued: whether MMN and P300 could faithfully index the case of speech perception involving phonological information.

Materials and methods

Power analysis

To determine the required sample size for our study, we performed a power analysis based on a pilot study involving 7 participants who were not included in the final experiment. The average boundary position and slope were 4.38 and -1.76 respectively. Our analysis focused on the interaction between Session (speech vs. non-speech) and Deviant Type (Stimulus #1 vs. Stimulus #7), using ANOVA as the statistical test (see the experimental procedure section for details of the experiment design). We utilized the "superpower" package in R to perform the power analysis (Lakens & Caldwell, 2021), beginning with the ANOVA_design function to construct a design with two within-subject factors. We derived the necessary parameters for the function from the pilot data, including the mean and standard deviation of each condition and the correlation for each possible comparison. Next, we employed the ANOVA_exact function to conduct an exact power analysis with an alpha level set to 0.05. The analysis yielded a power of 0.67 for the interaction effect. Finally, we generated a power curve using the plot_power function, which indicated that a sample size of 11 was recommended to achieve a power of 0.9 for detecting the interaction effect. Therefore, we aimed to include at least 11 subjects. Subsequent to the recruitment of additional participants, we observed that the same deviant stimulus served distinct roles, acting as the between-category deviant for one subset of participants, and the within-category deviant for another subset. In light of this finding, we chose to categorize participants into two distinct groups.

Participants

We recruited 45 native Hong Kong Cantonese speakers. They did not come from pre-defined groups, but rather were recruited based on their willingness and availability to participate in the study. One person failed to finish all tasks, and one was excluded due to her inability to complete the identification task. In addition, we excluded one participant who gave identical responses across all trials in the discrimination task because we suspected that he might not have taken the task seriously. Among the remaining 42 participants, 22 were categorized as Group A (9 female, mean age = 22.24 ± 2.74) and 20 were categorized as Group B (10 female, mean age = 22.05 ± 2.27) based on their boundary position measured by the identification task. Therefore, our standard stimulus was allocated into different lexical tone categories by these two groups of Cantonese participants, albeit they are native to the same language. Correspondingly, the same deviant stimulus acted as the between-category deviant for one group and the within-category deviant for the other group. This means that the processing of the same deviant stimulus involves phonological information for one group but not for the other one. Comparing these two groups of participants offers a precious window to investigate the relationship between phonological processing and ERP components (MMN and P300).

The two groups of participants were matched in their age (Welch's $t(39.66) = .246$, $p = .807$) and sex ($\chi^2(1) = .349$, $p = .554$). One participant in Group A was identified as ambidextrous using a modified Cantonese version of the Edinburgh handedness inventory (Oldfield, 1971) and the rest participants were all right-handed. They had minimal musical experience and no self-reported hearing impairment, speech and language-related disabilities, or brain injuries. None of them majored in linguistics or psychology. The study was approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University. The participants all read and signed a written informed consent form prior to the commencement of the experiment and received monetary compensation after the experiment.

Materials

The pitch contours ranging from Cantonese Tone 1 (high-level tone) to Tone 2 (high-rising tone) were embedded in two types of sound materials: real word (speech) and iterated rippled noise (IRN; non-speech). The Cantonese monosyllabic words /ji55/ (醫, "doctor") and /ji25/ (倚, "to lean on") were recorded by a native female speaker (16000 Hz/16 bit). The duration and intensity of the two recorded sounds was normalized to 300 ms and 70 dB using Praat, respectively. Based on the pitch templates with Tone 1 and Tone 2, a lexical tone continuum containing seven stimuli was constructed using Praat (see Figure 1a). For the non-speech sound materials, seven additional F0 trajectory-matched IRN stimuli were synthesized using 64 iteration steps (Swaminathan et al., 2008) which had a comparable spectrotemporal complexity as the speech sounds (see Figure 1b). However, speech stimuli were real words characterized by their semantic and phonetic content, while IRN stimuli did not contain semantic or phonetic content.

(Figure 1 about here)

The experimental procedure

For both speech and non-speech sound materials, stimuli #1, #4, and #7 were chosen. The two sets of sounds were presented in an oddball paradigm, with stimuli #4 being the standard and stimuli #1 and #7, whose pitch differences from the standard were physically equated, being the two types of deviants. There were two conditions: a passive listening condition and an active listening condition. To prevent the transfer of active attention to tonal changes in the passive condition, the active condition was always presented after the passive condition (Zhang & Shao, 2018). For each condition, there were two sessions: speech and non-speech. The order of the two sessions was counterbalanced among the participants. In the passive condition (speech and non-speech sessions), participants were required to ignore the played sounds while watching a sound-muted movie. Each session would begin with six warming-up trials playing the standards (Stimulus #4) with a 500-ms inter-stimulus-interval (ISI), followed by 1,000 test trials, including 800 standard trials and 200 deviant trials (100 for each type). Keeping the ISI constant in the passive condition helped participants' brains to establish a regular pattern of stimulus presentation, which allowed them to allocate resources efficiently and process the incoming information more effectively at pre-attentive stage. In the active condition, participants were required to attend to the sounds and press both buttons of a mouse simultaneously using their two thumbs once they caught a deviant from the standard. Similarly, each session of the active condition would begin with six warming-up trials (Stimulus #4), followed by 600 test trials with 480 standard trials and 120 deviant trials (60 for each type) which were presented with ISI jittering in the range from 800 to 1,200 ms. Jittering the ISI would make the timing of stimuli less predictable, which could help to maintain participants' engagement and attention and to reduce unconscious rhythmic responses in the participants' neural activity. In common practice, it can be observed that fixed ISIs are frequently adopted in passive oddball designs (e.g., Yu et al., 2014; 2017), while jittered ISIs are more likely to be employed in active oddball designs (e.g., Chen & Peng, 2020; D'Agostini et al., 2023). To avoid causing fatigue in the active condition, each session was evenly divided into five sub-sessions. There was a short break after each sub-session. In both passive and active conditions, test trials were presented pseudo-randomly on the premise that every two deviant trials were separated by at least three standard trials. The three standard trials directly following the deviant trials were excluded from analyses. In the two conditions, EEG data were recorded. The participants were reminded to keep their physical movements and eye blinks as low as possible before each session to reduce artifacts.

After the recordings, behavioural identification and discrimination post-tests were administered to measure the boundary position so as to determine the within- and between-category deviants for each participant. All the participants completed the identification task followed by the discrimination task. In the identification task, participants were required to select "醫" if they judged the sound to be Tone 1 and to select "倚" if they judge the sound to be Tone 2 after hearing a sound (Stimulus #1 to #7). In total, there were 63 trials, with each stimulus repeated nine times and played randomly. In the discrimination task, after hearing a pair of sounds, participants were required to judge the two sounds were same or different. There were 95 trials (19 pairs × 5 repetitions). The 19 pairs included seven same pairs (Stimuli #1-#1, #2-#2, … #7-

#7), six different pairs in forward order (Stimuli #1-#2, #2-#3, … #6-#7) and six different pairs in backward order (Stimuli #2-#1, #3-#2, … #7-#6).

All the tasks were administered to the participants in a soundproof booth, and the sound stimuli were presented to them through a pair of foam ear insets.

Data analysis

EEG data acquisition and analysis

The EEG signals were recorded via a SynAmps 2 amplifier (NeuroScan, Charlotte, NC, U.S.A.) with a 64-channel Quik-Cap Neo Net cap carrying Ag/AgCl electrodes placed according to the extended international 10–20 system. Vertical eye blinks and horizontal eye movements were monitored by electrodes around the eyes. The impedance between the online reference electrode (located between Cz and CPz) and any recording electrodes was kept below 5 kΩ. Alternating current signals with a band frequency from 0.05 Hz to 400 Hz were continuously recorded and digitized with a 24-bit resolution at a sampling rate of 1000 Hz.

Pre-processing of EEG signals was implemented using EEGLAB (Delorme & Makeig, 2004). The EEG signals were first filtered with a bandpass filter ranging from 0.1 to 30 Hz (Tanner et al., 2015). Epochs of 700 ms and 900 ms time-locked to the onset of target sounds were extracted for the passive and active conditions, respectively. For each epoch, there was a 100 ms pre-stimulus interval serving as baseline correction. Epochs with amplitudes exceeding ± 100 µV or ocular artifacts were excluded from analyses. The rest of the epochs were re-referenced to the mastoids and the average reference for the passive condition (Yu et al., 2014; 2017) and the active condition (Maurits et al., 2006; D'Agostini et al., 2023), respectively. The mean acceptance rate was 79.98% (SD = 11.07%) in the speech session and 79.59% (SD = 12.27%) in the non-speech session for the passive condition. As for the two sessions of active condition (speech and non-speech), the mean acceptance rates were 88.68% (SD = 11.39%) and 88.42% (SD = 11.50%), respectively.

The current study focused on the peak latency and mean amplitude of the two components: MMN in the passive condition and P300 in the active condition. Considering the global field power obtained in the two conditions of the current study (passive condition: Figure 2a; active condition: Figure 2b) and time windows of the two components in previous studies (Näätänen et al., 2007; Zheng et al., 2012), we chose 100-250 ms as the time window for MMN and 320-700 ms for P300. As for the MMN component, the electrode locations were constrained to the region of interest according to previous MMN studies, which was further confirmed by the topographic distribution map in the current study (Figure 2c). In total, 15 fronto-central electrodes where the MMN amplitude was expected to peak were selected: F1, F3, FC1, FC3, C1 and C3 on the left hemisphere, Fz, FCz and Cz located along the midline, and F2, F4, FC2, FC4, C2 and C4 on the right hemisphere. For each session (speech and non-speech), MMN for each type of deviant (Stimuli #1 and #7) was obtained by subtracting the standard ERP from the corresponding deviant ERP. MMN peak latency was defined as the timing point corresponding to the minimal point of the 2^{nd} order polynomial fitted curve to the deviant-standard difference wave within the specified time window (100-250 ms). For each participant, the mean MMN amplitude for each type of deviant in each session was obtained from the 150 ms time interval centred on the peak. As for the P300 component, based on the topographic distribution map of P300 amplitudes (Figure 2d), we selected 10 electrodes (left: P1, P3, P5 and PO3; middle: Pz and POz; right: P2, P4, P6 and PO4) for analysing this component, which is also in line with the literature that P300 was prominent at the parieto-occipital site (Chen & Peng, 2020). Similarly, P300 was obtained by subtracting the ERP to the standard from that to the deviant. P300 peak latency, calculated by the 2nd order polynomial curve-fitting algorism within the time window (320-700 ms), was the time point corresponding to the highest point of the fitted curve. Mean P300 amplitude of each type of deviant in each session was obtained by calculating the average amplitude within the 380 ms time interval for every participant.

(Figure 2 about here)

As for the behavioural data obtained in the active condition, hit rate of each type of deviant in each session was calculated for each participant. For example, the hit rate of deviant Stimulus #7 in the speech session was calculated as the ratio of pressing the mouse button when this type of deviant was played to its total number (i.e., 60). The active condition would yield four hit rates for each participant, with two for speech session and two for non-speech session. In each session, one is for Stimulus #1 and the other is for Stimulus #7.

Post-test behavioural data analysis

For the identification task, boundary slop and position were calculated using a logistic regression equation as below (1), where P_1 represents the percentage of Tone 1 responses, b_0 is the intercept and b_1 is the slope of the identification function (Xu et al., 2006). A steeper slope is an indicator of a high degree of CP. In addition, x is stimulus number in the continuum, that is 1 to 7 in the current study. Boundary position (x_{bc}), the identification midpoint dividing the two tonal categories, is estimated when P_1 is equal to 50%.

$$log_e\left(\frac{P_1}{1-P_1}\right) = b_0 + b_1x \quad (1)$$

Boundary position was measured for each participant to determine his or her within- and between-category deviants. Specifically, if a participant showed a boundary position higher than four, Stimulus #1 was his or her within-category deviant while Stimulus #7 was the between-category deviant, and the participant would be labelled as Group A. The boundary position higher than four indicates that Stimulus #4 (the standard) is categorically perceived as /ji55/ "醫" corresponding to Stimulus #1, which means that the deviant Stimulus #1 is the within-category deviant while the deviant Stimulus #7 (/ji25/ "倚") is the between-category deviant. In contrast, for a participant who showed a boundary position lower than four, that is Stimulus #4 (the standard) was categorized as /ji25/ "倚" (Stimulus #7), Stimulus #7 belonged to the within-category deviant and Stimulus #1 belonged to the between-category deviant. He or she would be labelled as Group B.

For the discrimination task, the 19 tone pairs were divided into six comparison units (units 1-2, 2-3, 3-4, 4-5, 5-6, 6-7), and each comparison unit contained four types of tone pairs. For example, in the comparison unit 1-2, the four types are stimuli pairs #1-#1, #2-#2, #1-#2, and #2-#1. For each comparison unit, a mean accuracy of the four tone pairs was calculated. In total, there were six discrimination accuracies for the six comparison units. Next, the comparison units were divided into between- and within-category types for each participant in reference to the obtained boundary position (Xu et al., 2006). Take a participant who showed a boundary position of 3.78, the comparison unit 3-4 was the between-category type, and the accuracy of this unit would be the between-category discrimination accuracy for this participant. His or her within-category discrimination accuracy would be obtained by averaging the discrimination accuracy of the rest five comparison units.

Results

Post-test identification and discrimination results

For the results of identification task, the boundary position and slope were calculated for each individual participant in each group. The average boundary positions and slopes for Groups A and B were then obtained by averaging those individual identification data, as shown in Table 1. Figure 3 displays the overall identification curves for the participants in the two groups. As Figure 3 represents the percentage of Tone 1 responses to each stimulus averaged across participants in each group, the boundary positions in Figure 3 may not be exactly the same as the average boundary positions presented in Table 1.

(Table 1 about here)

(Figure 3 about here)

Analysis of variance (ANOVA) in R, employing the afex package, was used to compare the two groups in boundary slope and position. ANOVA on boundary slope revealed no significant effect of group, suggesting that the boundary slope did not differ between Group A and Group B, $F(1, 40) = 0.01$, MSE = .05, $p = .916$. ANOVA on boundary position revealed a significant effect of group, with the boundary position in Group A was closer to Tone 2 while that in Group B was closer to Tone 1, $F(1, 40) = 51.61$, MSE = .05, $p < .001$. The results showed that the participants in Group A would perceive the standard (Stimulus #4) as Tone 1 (deviant Stimulus #1), which meant Stimulus #1 was within-category deviant while Stimulus #7 was between-category deviant for this group of participants. For those in Group B, in contrast, the standard (Stimulus #4) would be perceived as Tone 2 – deviant Stimulus #7, which therefore was their within-category deviant, while Stimulus #1 was the between-category deviant.

For the discrimination task, the between- and within-category accuracy among the two groups are shown in Figure 4. Two-way repeated measures ANOVAs were performed on discrimination accuracy with Group (A vs. B) as a between-subjects factor and Category Type (within- vs. between-category) as a within-subject factor. Statistical analysis revealed significant main effects of Category Type ($F(1, 40) = 19.52$, MSE = .01, $p < .001$) and Group ($F(1, 40) = 4.77$, MSE = .01, $p = .035$). There was no significant two-way interaction of Group × Category Type, $F(1, 40) = .01$, MSE = .01, $p = .939$. For the significant main effect of Category Type, the between-category tone pairs generated a much higher accuracy than the within-category ones in both Group A and B. Relative to Group A, Group B had a significantly higher discrimination accuracy.

(Figure 4 about here)

Electrophysiological results

MMN amplitude and latency

As for the data obtained in the passive oddball paradigm, Figure 5 shows the difference waves of the two deviants at three representative electrodes elicited from participants in Group A and Group B in the speech and non-speech sessions. Four-way repeated measures ANOVAs were performed for the mean amplitude and peak latency of MMN component, respectively. There were three within-subject factors – Hemisphere (left, middle vs. right), Session (speech vs. non-speech) and Deviant Type (Stimulus #1 vs. Stimulus #7), and one between-subjects factor – Group (A vs. B). Post hoc comparisons were conducted using the lsmeans package with Bonferroni adjustment (Lenth, 2016) to verify whether there were CP effects in the MMN component at the three brain areas (left, middle, and right) of participants in Group A and Group B in the two sessions (speech and non-speech), that is the simple effect of Stimulus #1 – #7 differences within any level of the other three factors.

(Figure 5 about here)

Figure 6 plots the mean MMN amplitudes at the locations of the selected electrodes on the midline and the left and right hemispheres for speech and non-speech Stimuli #1 and #7 among two groups of participants. Statistical analysis revealed a main effect of Deviant Type, $F(1, 40) = 5.22$, MSE = 3.07, $p = .028$, with Stimulus #7 eliciting a greater MMN amplitude than Stimulus #1. There was a two-way interaction of Group × Session, $F(1, 40) = 4.44$, MSE = 9.81, $p = .041$, and a four-way interaction of Deviant Type × Group × Session × Hemisphere, $F(1.47, 58.84) = 3.68$, MSE = .15, $p = .044$. Other effects did not reach significance. The two-way interaction effect was analysed under different groups to examine whether there was a difference between speech and non-speech sessions in the mean MMN amplitude within each group. Post hoc pairwise comparisons showed that the deviants in the speech session did not elicit greater MMN amplitude than those in the non-speech session, and vice versa in Group A ($t = -1.60$, SE = .39, $p = .118$) and Group B ($t = 1.39$, SE = .40, $p = .172$). In addition, we compared Group A and Group B separately within each session and found a greater MMN amplitude elicited from Group B than from Group A in the speech session ($t = 2.80$, SE = .33, $p = 0.008$), while there was no group difference in the non-speech session ($t = -.55$, SE = .45, $p = .589$). Regarding the four-way interaction, post hoc comparisons were conducted to examine whether there was a CP effect in terms of MMN amplitude (i.e., significant Stimuli #1 – #7 difference) at different brain areas of participants in different groups in different sessions. The results of participants in Group A showed that, in the speech session, the MMN amplitude of the response to the between-category deviant (Stimulus #7) were significantly greater relative to the within-category one (Stimulus #1) at all the three brain areas – left ($t = 3.68$, SE = .25, $p =$ .001), middle ($t = 3.03$, SE = .29, $p = .004$), and right ($t = 3.05$, SE = .31, $p = .004$), but not in the non-speech session (*p*s > .05). For the participants in Group B, in neither session was there significant difference in MMN amplitude for the processing of between-category (Stimulus #1) and within-category (Stimulus #7) deviants (*p*s > .05).

(Figure 6 about here)

For the peak latency of the MMN difference waves, statistical analysis revealed significant main effects of Session, $F(1, 40) = 11.60$, MSE < .01, $p = .002$, and Deviant Type, $F(1, 40) = 6.55$, MSE = .01, $p = .014$. No other main or interaction effects reached significance (all *p*s > .05). The significant main effect of Session indicates that for both Group A and Group B the peak latency observed in the speech session (mean = 184 ms) was earlier compared to that in the non-speech session (mean = 204 ms) regardless of deviant type and brain area. Regarding the main effect of Deviant Type, in both speech and non-speech sessions, the peak latency of Stimulus #7 (mean = 186 ms) was found to be shorter than that of Stimulus #1 (mean = 202 ms) in both groups of participants regardless of brain area. Neither group exhibited a significant CP effect in terms of the peak latency, and since the peak latency of difference waves is not of interest, the results regarding this parameter will not be discussed in detail.

P300 amplitude and latency

Figure 7 shows the hit rates of participants in Group A and Group B for detecting the two deviants (Stimulus #1 and Stimulus #7) in speech and non-speech sessions of the active oddball paradigm. A three-way Group (Group A vs. Group B) × Session (speech vs. non-speech) × Deviant type (Stimulus #1 vs. Stimulus #7) repeated measures ANOVA was conducted for the dependent variable, hit rate. There were a two-way interaction of Group × Session, $F(1, 40) = 4.68$, MSE < .01, $p = .036$, and a two-way interaction of Session × Deviant type, $F(1, 40) = 4.99$, MSE < .01, $p = .031$. No other main or interaction effects were significant. For the significant interaction of Group × Session, post hoc comparisons with Bonferroni adjustment showed that the hit rate in the speech session did not significantly differ from that in the non-speech session in either group (*p*s > .05) and the hit rate of Group A did not significantly differ from that of Group B in either session (*p*s > .05). For the significant interaction of Session × Deviant type, post hoc comparisons revealed the differences in hit rate between Stimulus #1 and Stimulus #7 was not significant in either session (*p*s > .05). In addition, the hit rate of Stimulus #1 in the speech session did not significantly differ from that in the non-speech session ($p > .05$), while the hit rate of Stimulus #7 in the speech session was significantly higher than that in the non-speech session, $t(40) = 2.11$, SE = .01, $p = .042$.

(Figure 7 about here)

As for the EEG data obtained in the active oddball paradigm, Figure 8 shows the difference waves of the two deviants at three representative electrodes in the speech and non-speech sessions for both groups of participants. Four-way repeated measures ANOVAs were performed for the mean amplitude and peak latency of P300 component. Group (A vs. B) served as a between-subjects factor, while Hemisphere (left, middle vs. right), Session (speech vs. non-speech) and Deviant Type (Stimuli #1 vs. #7) were within-subject factors. Post hoc comparisons were conducted, when necessary, to verify whether there were Stimulus #1 – #7 differences at the three brain areas (left, middle, and right) of participants in Group A and Group B in speech and non-speech sessions.

(Figure 8 about here)

Figure 9 shows the mean P300 amplitudes at the locations of the selected electrodes on the midline and the left and right hemispheres for speech and non-speech Stimuli #1 and #7 among two groups of participants. Results showed significant main effects of Session, $F(1, 40) = 5.95$, MSE = 4.00, $p = .019$, and Hemisphere $F(1.91, 76.23) = 38.16$, MSE = 1.56, $p < .001$. In addition, there was a three-way interaction of Deviant Type × Group × Session, $F(1, 41) = 4.22$, MSE = 1.80, $p = .047$. No other effects reach the significance level. The significant main effect of Session suggests that at all the selected electrodes, the processing of non-speech deviants (mean = 3.27 μV) elicited a larger P300 amplitude than the processing of speech deviants (mean = 2.83 μV). For the significant main effect of Hemisphere, post hoc pairwise comparisons showed that the P300 amplitude at electrodes located along the midline (mean = 3.71 μV) was significantly greater than those at the left (mean = 2.85 μV; $t = 6.57$, SE = .13, $p < .001$) and right (mean = 2.60 μV; $t = 9.16$, SE = .12, $p < .001$) hemispheric recording sites. The two hemispheres did not significantly differ from each other in P300 amplitude, $t = 1.71$, SE = .15, $p = .287$.

(Figure 9 about here)

Next, the three-way interaction was further analyzed in terms of different sessions by conducting two two-way ANOVAs on Deviant Type × Group separately for the speech and non-speech sessions. First, in the speech session, there was a significant interaction of Deviant Type × Group, $F(1, 40) = 4.55$, MSE = .42, $p = .039$, while no other effects reach the significance level. Post hoc comparisons showed that the processing of Stimulus #1 did not elicit a larger P300 amplitude than the processing of Stimulus #7, and vice versa in Group A ($t = 1.68$, SE = .20, $p = .100$) and Group B ($t = -1.35$, SE = .21, $p = .172$). Second, in the non-speech session, the results revealed no significant main or interaction effects (ps > .05).

For the peak latency of the P300 difference waves, statistical analysis revealed a significant two-way interaction of Session × Deviant Type, $F(1, 40) = 12.77$, MSE < .01, $p < .001$, and there was no other significant effect (all *p*s > .05). The difference between the two types of deviants was examined under speech and non-speech sessions. In the speech session, the peak latency of Stimulus #7 (mean = 498 ms) was found to be shorter than that of Stimulus #1 (mean = 529 ms) regardless of group and brain area, $t = -2.66$, SE = .01, $p = .011$. However, the peak latency of the two types of deviants did not significantly differ from one another in the non-speech session, $t = 1.05$, SE = .01, $p = .299$. As no CP effect in terms of P300 latency was observed, the results will not be discussed further.

Discussion

The current study investigated the fine-grained lexical tone perception at pre-attentive and attentive stages using passive and active oddball paradigms to illuminate the processing of acoustic and phonological information of lexical tones at the two stages. Tone category (within/between-category) was manipulated along a lexical tone continuum (word /ji/ with Tone 1 to Tone 2), as the response to within-category deviant has been believed to primarily rely on acoustic processing while the between-category deviant provides a phonological contrast (Chen & Peng, 2020; Xi et al., 2010; Yu et al., 2017; L. Zhang et al., 2012; Zheng et al., 2012). A post-test identification task revealed that a group of participants (Group A) perceived Stimulus #4 (standard) as the same category as Stimulus #1 (deviant) and categorized the equally spaced Stimulus #7 (deviant) into a distinct category (see the solid line in Figure 3), suggesting the first type of deviant (Stimulus #1) was within-category deviant and the second type (Stimulus #7) was between-category one for these participants. For another group of participants

(Group B), the result of identification task indicated that Stimulus #7 was within-category deviant and Stimulus #1 was between-category deviant (see the dash line in Figure 3). At the pre-attentive stage, we found the response of Group A to the between-category deviant elicited significantly greater MMN amplitudes than their response to the within-category one in the speech session (see Figure 6). This is consistent with the result of discrimination task, where this group of participants had a significantly higher between-category discrimination accuracy than within-category one (see Figure 4). The results of Group A indicate that MMN can index the case of speech perception involving phonological information. However, the difference between within- and between-category deviants in MMN amplitude was not observed in Group B in either session, although there was behavioural evidence showing they had an enhanced sensitivity to between-category discrimination than to within-category discrimination. At the attentive stage, the results showed that the P300 amplitude elicited from either group in response to between-category deviant was not significantly higher relative to within-category deviant even in the speech session, suggesting phonological information did not modulate the P300 amplitude in the current study.

Fine-grained lexical tone perception at the pre-attentive stage

In the passive oddball paradigm, a CP effect in terms of MMN amplitude (i.e., a greater MMN amplitude elicited from between-category deviant) was observed in one group of Hong Kong Cantonese speakers. The finding of the CP effect at the pre-attentive stage suggests that MMN amplitude can act as a neural marker of phonological processing, which converges with previous EEG studies examining CP of lexical tones in Mandarin speakers (Xi et al., 2010; Yu et al., 2014) and Cantonese speakers (Yu et al., 2017). However, the CP effect in MMN amplitude (between > within) observed among

Cantonese speakers by Yu et al. (2017) should be interpreted with caution as it might be driven by the different acoustic distances from the standard to the two types of deviants. Specifically, the between-category deviant employed in Yu et al. (2017) is acoustically farther from the standard than the within-category deviant. Thus, the CP effect in MMN amplitude observed among Cantonese speakers by the authors might just be the product of larger acoustic difference from between-category deviant to standard rather than the involvement of phonological processing. Employing two deviants that share the same acoustic distance from the standard, the current study offers evidence from Cantonese speakers that MMN can serve as an indicator of the processing of phonological information during CP of lexical tones.

However, the CP effect in MMN amplitude was absent in another group of participants in the current study, which was mainly driven by their relatively large MMN response to the within-category deviant (see Figure 6). This might be attributed to that the participants in Group B were highly sensitive to acoustic pitch difference. Specifically, a higher discrimination accuracy was found in this group of participants than those in Group A (see Figure 4). The acoustic processing might override the phonological processing in Group B due to their high sensitivity, giving rise to the absence of CP effect (between > within) in terms of MMN amplitude. The influence of sensitivity to acoustic information on CP of lexical tones has also been detected in clinical populations (Chen & Peng, 2020; X. Wang et al., 2017). For instance, Wang et al. (2017) examined categorical lexical tone perception in individuals with ASD using a passive oddball paradigm. While participants without ASD in their study exhibited enhanced MMN responses to the between-category deviant relative to the within-category one, showing a CP pattern, weakened CP of lexical tones was observed in participants with ASD, as MMN amplitude elicited from within-category deviant was atypically larger. One reason for the abnormality is that individuals with ASD are characterized with an over-developed neural network for low-level perceptual processing, such as acoustic processing (Mottron et al., 2006, 2013), which has been supported by a great amount of EEG evidence that non-linguistic pitch changes elicited more negative MMN amplitude in individuals with ASD relative to those without ASD (Ferri et al., 2003; Gomot et al., 2002, 2011; Lepistö et al., 2005). The high sensitivity in discriminating within-category deviant was believed to underlie the absence of CP pattern in ASD. In Chen and Peng (2020) focusing on individuals with congenital amusia, compared with their controls, amusics showed reduced responses to both within- and between-category deviants at behavioural and neural levels, resulting in poor CP performance. The results of Chen and Peng (2020) suggest that low sensitivity to acoustic pitch changes in this population underlie their impoverished phonological processing skill. These findings are the evidence of "feed-forward mechanisms", which proposes that low-level acoustic processing has an impact on high-level phonological processing (Binder, 2000; Scott & Wise, 2004), but do not seem to corroborate the idea that phonological processing would become dominant at the cost of acoustic processing through "feedback mechanisms" when phonological information is present (Hickok & Poeppel, 2007). In addition, the role of acoustic processing in phonological processing has been found to be modulated by attention (Dunn et al., 2008; Strange, 2011; Whitehouse & Bishop, 2008; C. Zhang & Shao, 2018). For acoustic and phonological processing of lexical tones at a stage with focal attention see the next section.

Fine-grained lexical tone perception at the attentive stage

In both sessions of the active oddball paradigm, the two types of deviants elicited comparable P300 amplitudes, which was observed in both groups of Hong Kong

Cantonese speakers. In other words, in neither group was there an enhanced P300 response to the between-category deviant that is believed to involve phonological information. A similar result has also been observed among Mandarin speakers in mainland China by Zhang et al. (2012) and Zheng et al. (2012). The absence of CP effect in terms of P300 amplitude suggests that this ERP component does not necessarily index phonological processing.

One thing to note is that although the P300 amplitudes of responses to the between-and within-category deviants were found to be comparable, Zhang et al. (2012) claimed that P300 could reflect phonological processing. This was based on the finding of greater P300 amplitude of the response to the between-category deviant at the electrodes in the left hemisphere compared to the right ones given they believed that the left recording sites were the key brain areas for the processing of phonological information. However, it is called into question since there is inconsistency over the hemisphere lateralization for phonological and acoustic processing of lexical tones (Luo et al., 2006; Yu et al., 2020). Our findings do not seem to show any evidence of a left lateralization for the processing of phonological cue in lexical tones, or right lateralization for the acoustic processing. In addition, EEG technology with high temporal resolution does not seem to be able to provide sufficient spatial information to identify the relationship between observed ERP components and brain lateralization.

The results of current study seem to contradict some previous studies showing enhanced P300 response to the between-category deviant relative to the within-category one among speakers of tone languages (Chen & Peng, 2020; Zheng et al., 2012). We speculate that one possible explanation for this discrepancy is that the rates of detecting the two types of deviants used in the current study are almost the same, whereas the number of detected between-category deviant employed in previous research is significantly larger than that of within-category deviant (Chen & Peng, 2020; Zheng et al., 2012). In other words, the response rate to the between-category deviant was significantly higher than that to the within-category deviant. Given that P300 component has been found to be related to decision making, an active process of deciding on how to make a response to a present stimulus (Hillyard & Kutas, 1983; Verleger et al., 2005, 2016). Different response (by key-press) rates to between- and within-category deviants might contribute to the different P300 amplitudes elicited from the two types of deviants observed in previous studies.

However, we acknowledge that our study might have lacked sufficient statistical power to detect more subtle differences in P300 amplitude. Statistical power is an important consideration in experimental research, as it determines the ability of a study to avoid false negatives. In our study, the sample size might have limited our statistical power, particularly in detecting higher-level interactions. Therefore, we recommend that future studies with larger sample sizes should be conducted to confirm our findings. Another limitation of the current study was the relatively large acoustic distance between standards and deviants. If the acoustic distance between standards and deviants becomes less distinguishable, it is possible that the P300 component may show a CP effect indexing phonological information. Further research is needed to explore this question using multiple sets of standard and deviant stimuli with varying acoustic distances. Specifically, we plan to use stimuli with smaller acoustic distances while ensuring that the detection rates of between-category and within-category deviants remain comparable. We believe that this approach will provide a more nuanced understanding of the relationship between phonological processing and the P300 component. In addition, the sensitivity to acoustic pitch difference was reflected by the discrimination task only. Since the sensitivity to pitch change appears to be an important factor when examining the neural mechanisms underlying CP of lexical tones, we recommend that future studies incorporate comprehensive tests of pitch sensitivity to better understand its potential influence on the neural responses to within- and between-category tonal stimuli and to further elucidate the relationship between pitch sensitivity and the CP of lexical tones.

Conclusion

The present study investigated whether MMN and P300 could faithfully index the case of speech perception involving phonological information. Our finding that the response of one group of participants to the between-category deviant evoked a significantly greater MMN amplitude relative to the within-category deviant only in the speech session might suggest that MMN could act as an indicator of phonological processing at the pre-attentive stage. However, the CP effect in MMN amplitude was not observed in another group of participants, indicating that phonological processing might be overridden by acoustic processing among participants who are highly sensitive to pitch. In addition, the absence of CP effect in terms of P300 might suggest that this component, obtained at the attentive stage, does not necessarily index the processing of phonological cue. The findings regarding MMN and P300, taken together, reveal that attention might modulate the role of acoustic and phonological information in the fined-grained lexical tone perception.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Acknowledgments

This study was supported by a grant from the Research Grants Council of Hong Kong (GRF: 15610321).

References:

Binder, J. (2000). The new neuroanatomy of speech perception. *Brain*, *123*(12), 2371–2372. https://doi.org/10.1093/brain/123.12.2371

Chen, F., & Peng, G. (2020). Reduced sensitivity to between-category information but preserved categorical perception of lexical tones in tone language speakers with congenital amusia. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.581410

Chen, F., & Peng, G. (2021). Categorical perception of pitch contours and voice onset time in Mandarin-speaking adolescents with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4468–4484. https://doi.org/10.1044/2021_JSLHR-20-00725

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dunn, M. A., Gomes, H., & Gravel, J. (2008). Mismatch negativity in children with autism and typical development. *Journal of Autism and Developmental Disorders*, *38*(1), 52–71. https://doi.org/10.1007/s10803-007-0359-3

D'Agostini, M., Burger, A. M., Jelinčić, V., von Leupoldt, A., & Van Diest, I. (2023). Effects of transcutaneous auricular vagus nerve stimulation on P300 magnitudes and salivary alpha-amylase during an auditory oddball task. *Biological Psychology*, 108646. https://doi.org/10.1016/j.biopsycho.2023.108646

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680. https://doi.org/10.1016/S1388-2457(03)00153-6

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171–189.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175. https://doi.org/10.1016/S0095-4470(19)30813-7

Gomot, M., Blanc, R., Clery, H., Roux, S., Barthelemy, C., & Bruneau, N. (2011). Candidate electrophysiological endophenotypes of hyper-reactivity to change in autism. *Journal of Autism and Developmental Disorders*, *41*(6), 705–714. https://doi.org/10.1007/s10803-010-1091-y

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: Electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(5), 577–584. https://doi.org/10.1111/1469-8986.3950577

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. https://doi.org/10.1016/S0095-4470(03)00016-0

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology*, *34*(1), 33–61. https://doi.org/10.1146/annurev.ps.34.020183.000341

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

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science, 4*(1), 2515245920951503.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1). https://doi.org/10.18637/jss.v069.i01

Lepistö, T., Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1–2), 147–157. https://doi.org/10.1016/j.brainres.2005.10.052

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, *103*(51), 19558–19563. https://doi.org/10.1073/pnas.0607065104

Maurits, N. M., Scheeringa, R., Van Der Hoeven, J. H., & De Jong, R. (2006). EEG coherence obtained from an auditory oddball task increases with age. *Journal of Clinical Neurophysiology, 23*(5), 395–403. https://doi.org/10.1097/01.wnp.0000219410.97922.4e

McMurray, B. (2022). *The myth of categorical perception*. https://doi.org/https://doi.org/10.31234/osf.io/dq7ej

Mottron, L., Bouvet, L., Bonnel, A., Samson, F., Burack, J. A., Dawson, M., & Heaton, P. (2013). Veridical mapping in the development of exceptional autistic abilities. *Neuroscience & Biobehavioral Reviews*, *37*(2), 209–228. https://doi.org/10.1016/j.neubiorev.2012.11.016

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A syllabus-based comparative study of 27 Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*(1), 135–154.

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 243–335). Academic Press.

Rong, Y., Weng, Y., Chen, F., & Peng, G. (2023). Categorical perception of Mandarin lexical tones in language-delayed autistic children. *Autism, 27*(5), 1426–1437. https://doi.org/10.1177/13623613221138687

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, *92*(1–2), 13–45. https://doi.org/10.1016/j.cognition.2002.12.002

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, *52*(8), 997–1009. https://doi.org/10.1111/psyp.12437

Verleger, R., Grauhan, N., & Śmigasiewicz, K. (2016). Is P3 a strategic or a tactical component? Relationships of P3 sub-components to response times in oddball tasks with go, no-go and choice responses. *NeuroImage*, *143*, 223–234. https://doi.org/10.1016/j.neuroimage.2016.08.049

Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19*(3), 165–181. https://doi.org/10.1027/0269-8803.19.3.165

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, *280*(1), 61–72. https://doi.org/10.1111/j.1749-6632.1976.tb25472.x

Wang, X., Wang, S., Fan, Y., Huang, D., & Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, *7*(1), 43254. https://doi.org/10.1038/srep43254

Whitehouse, A. J. O., & Bishop, D. V. M. (2008). Do children with autism 'switch off' to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524. https://doi.org/10.1111/j.1467-7687.2008.00697.x

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223–231. https://doi.org/10.1016/j.neuroscience.2010.06.077

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

Yu, K., Wang, R., Li, L., & Li, P. (2014). Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00729

Yu, K., Wang, R., & Li, P. (2020). Native and nonnative processing of acoustic and phonological information of lexical tones in Chinese: Behavioral and neural correlates. In H.-M. Liu, F.-M. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition. Chinese language learning sciences* (pp. 79–99). Springer.

Yu, K., Zhou, Y., Li, L., Su, J., Wang, R., & Li, P. (2017). The interaction between phonological information and pitch type at pre-attentive stage: An ERP study of lexical tones. *Language, Cognition and Neuroscience*, *32*(9), 1164–1175. https://doi.org/10.1080/23273798.2017.1310909

Zhang, C., & Shao, J. (2018). Normal pre-attentive and impaired attentive processing of lexical tones in Cantonese-speaking congenital amusics. *Scientific Reports*, *8*(1), 8420. https://doi.org/10.1038/s41598-018-26368-7

Zhang, L., Xi, J., Wu, H., Shu, H., & Li, P. (2012). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *NeuroReport*, *23*(1), 35–39. https://doi.org/10.1097/WNR.0b013e32834e4842

Zheng, H.-Y., Minett, J. W., Peng, G., & Wang, W. S.-Y. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes*, *27*(2), 184–209. https://doi.org/10.1080/01690965.2010.520493

Table 1. Boundary position and slope of identification in Group A and Group B.

Identification function	Group A	Group B
Boundary position (SD)	4.41 (0.36)	3.77 (0.18)
Identification slope (SD)	-1.80 (0.18)	-1.81 (0.26)

Figures:

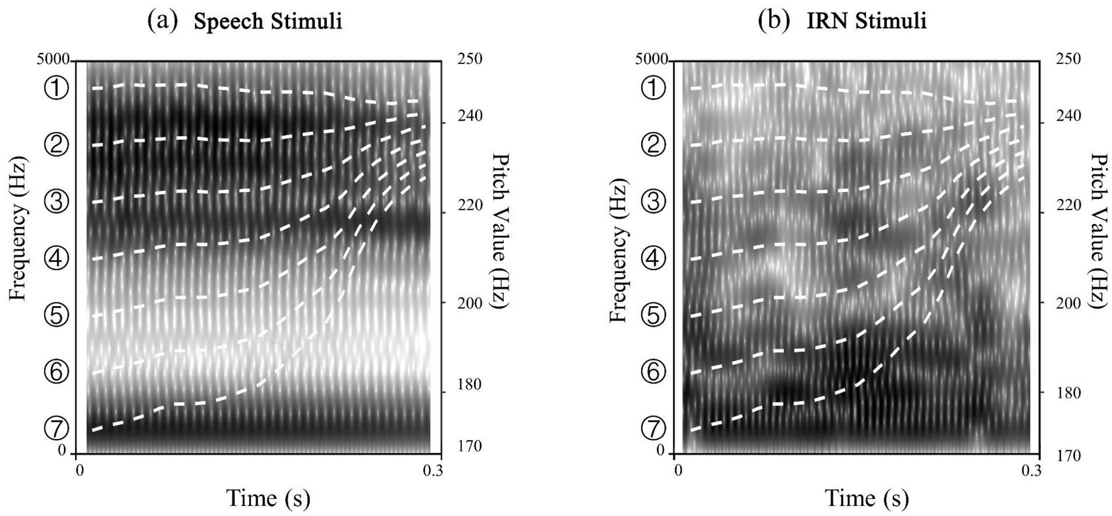

Figure 1. Schematic diagram of pitch contours embedded in (a) real word (speech) and (b) IRN (non-speech). The right-side y-axis indicates the fundamental frequency in Hz. The white curves indicate the seven pitch contours along each continuum.

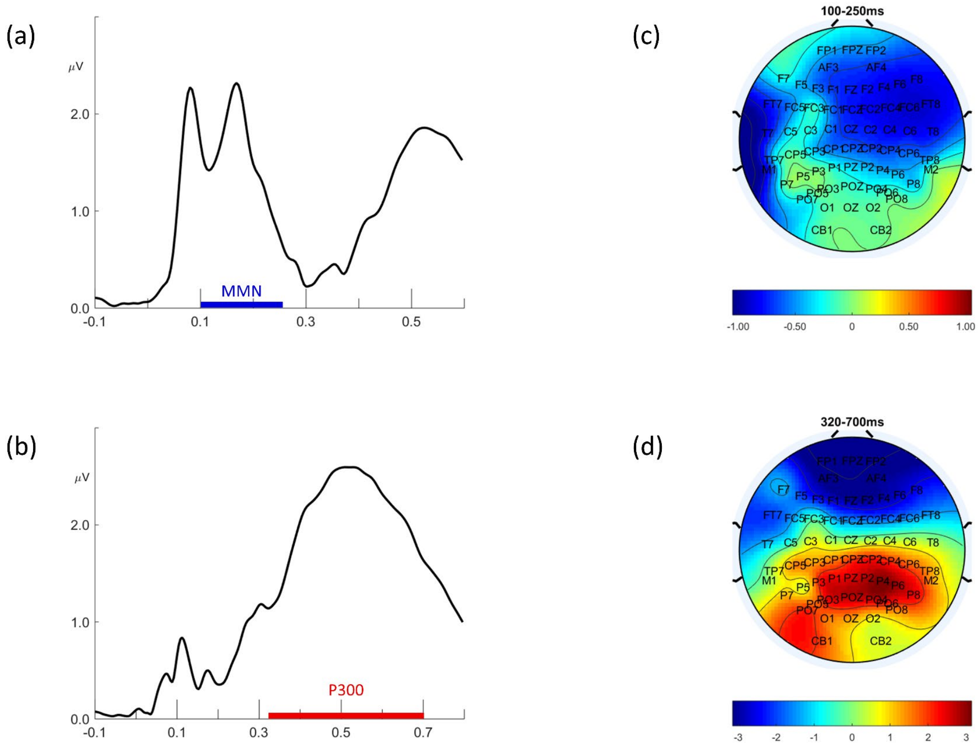

Figure 2. (a) Global field power for the passive condition averaged across all participants and stimuli and across both speech and non-speech sessions. (b) Global field power for the active condition averaged across participants, stimuli, and sessions. (c) Topographic distribution map of MMN amplitudes (100 – 250 ms). (d) Topographic distribution map of P300 amplitudes (320 – 700 ms).

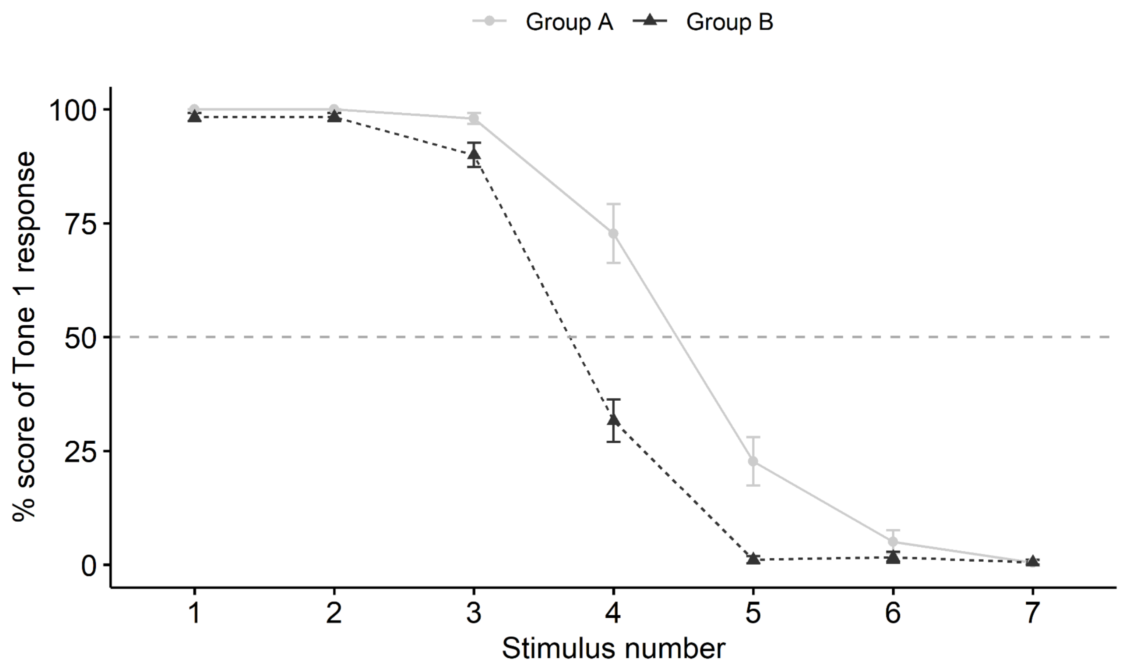

Figure 3. The identification curves of Tone 1 responses in Group A and Group B. Error bars: ± 1 standard error.

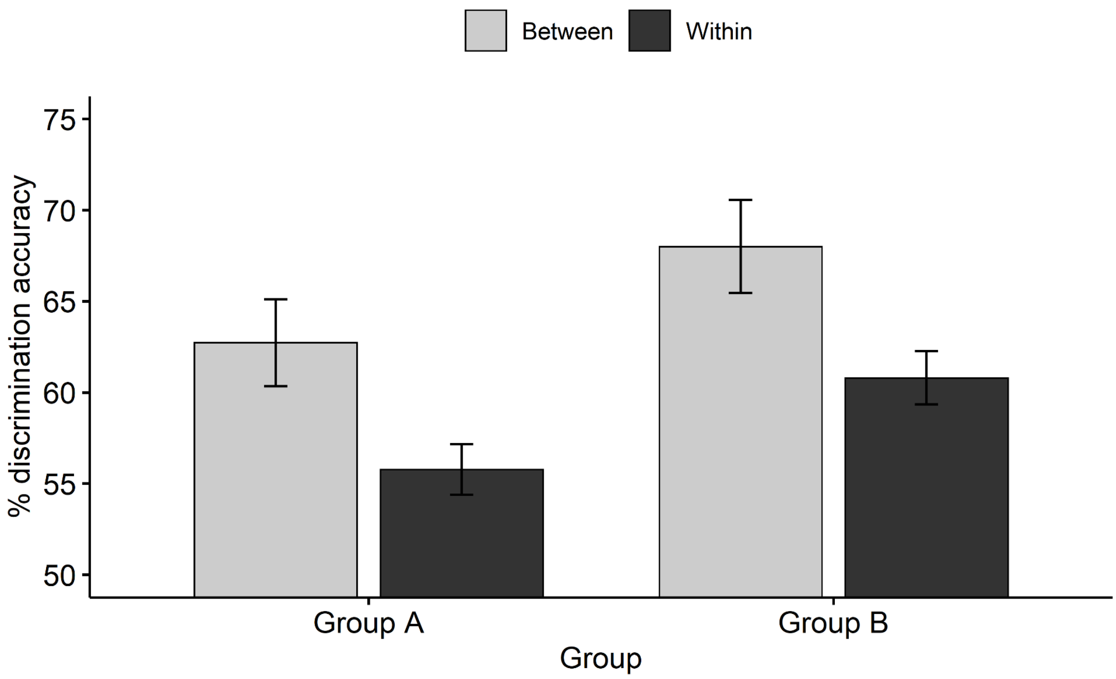

Figure 4. Discrimination accuracy of the between- and within-category tone pairs for Group A and Group B. Error bars: ± 1 standard error.

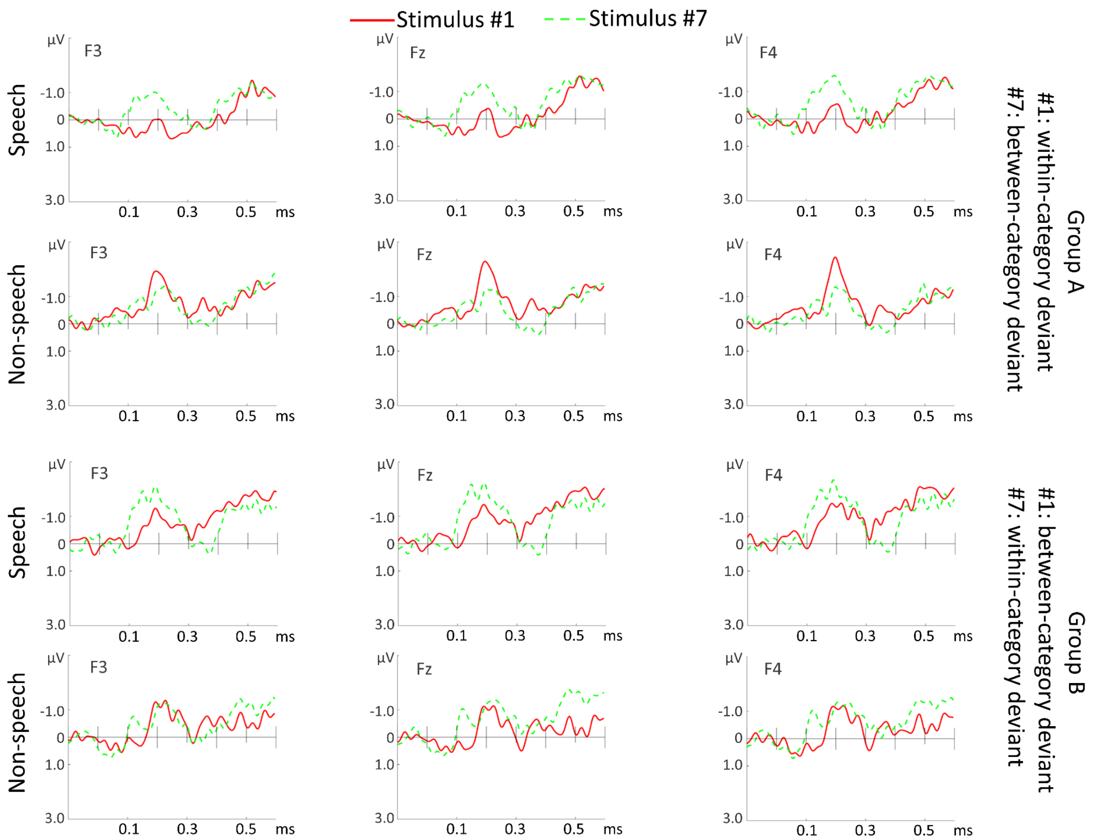

Figure 5. Difference waves obtained in the passive condition at three representative electrodes: F3 (left), Fz (middle), and F4 (right) in speech and non-speech sessions for Stimulus #1 (red lines) and Stimulus #7 (green lines) in Group A and Group B.

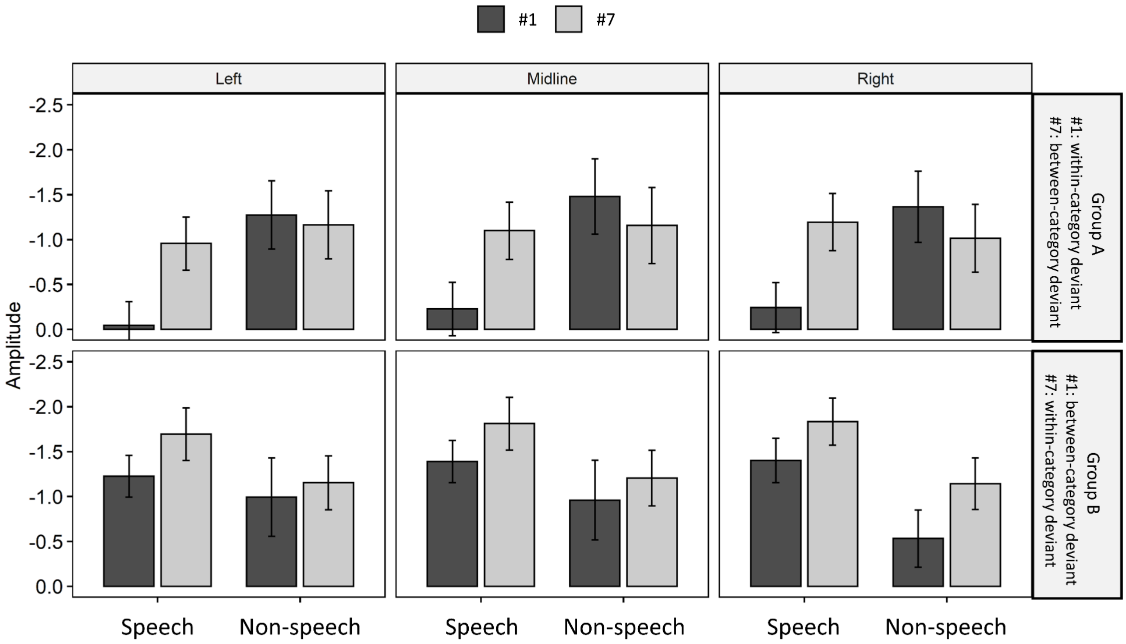

Figure 6. Mean MMN amplitude of difference waves elicited from Stimuli #1 and #7 at selected electrodes in speech and non-speech sessions for Group A and Group B.

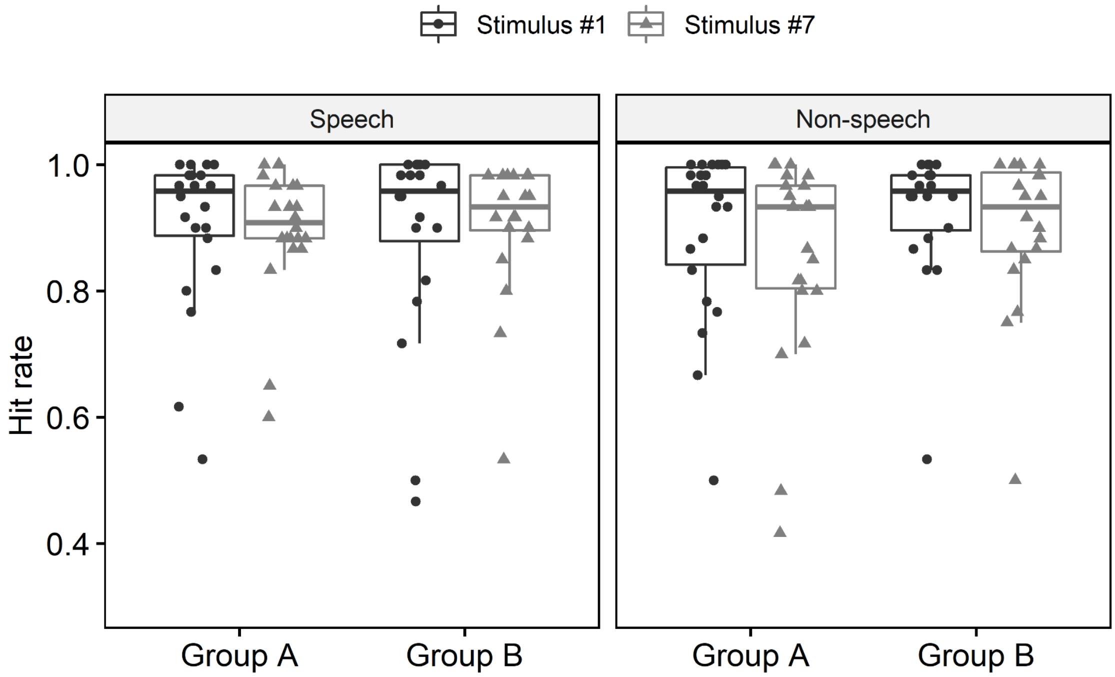

Figure 7. Hit rates of Stimulus #1 and Stimulus #7 deviants in non-speech and speech sessions for Group A and Group B. Error bars: ± 1 standard error.

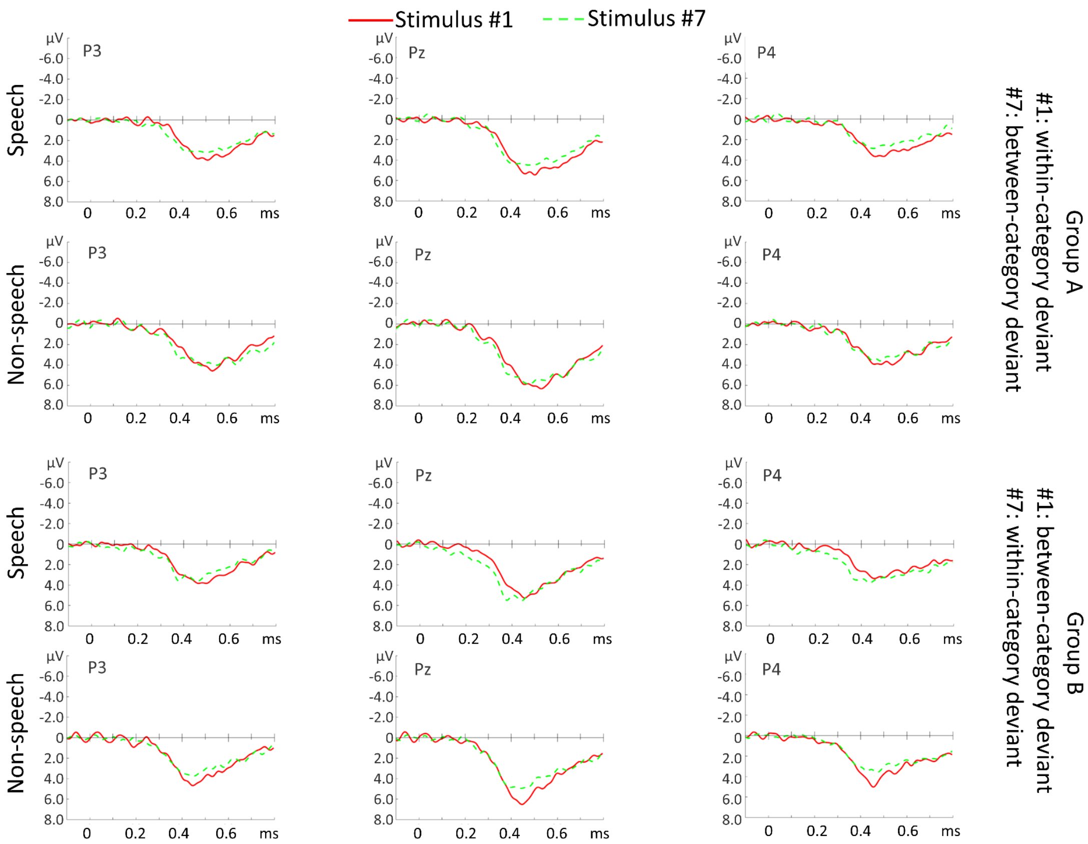

Figure 8. Difference waves obtained in the active condition at three representative electrodes: P3 (left), Pz (middle), and P4 (right) in speech and non-speech sessions for Stimulus #1 (red lines) and Stimulus #7 (green lines) in Group A and Group B.

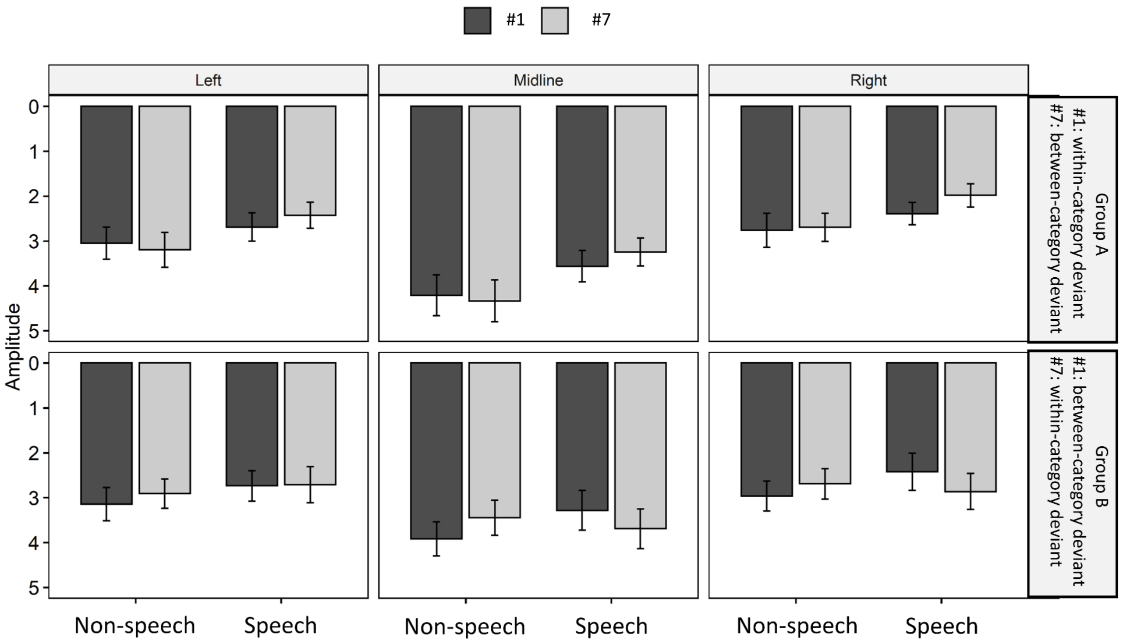

Figure 9. Mean P300 amplitude of difference waves elicited from Stimuli #1 and #7 at selected electrodes in speech and non-speech sessions for Group A and Group B.