https://doi.org/10.1016/j.bandl.2021.105029

Neural responses in novice learners' perceptual learning and generalization of lexical tones: The effect of training variability

Zhen Qin[a,*], Minzhi Gong[b], Caicai Zhang[c,*]

[a] Division of Humanities, The Hong Kong University of Science and Technology

[b] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong

[c] Research Centre for Language, Cognition and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

E-mail addresses: hmzqin@ust.hk (Z. Qin), minzhigong@cuhk.edu.hk (M. Gong), caicai.zhang@polyu.edu.hk (C. Zhang).

∗ Corresponding authors at: Division of Humanities, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong (Z. Qin). Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Yuk Choi Road, Hung Hom, Hong Kong (C. Zhang).

Abstract

The acoustics of lexical tones are highly variable across talkers, and require second-language (L2) learners' flexibility in accommodating talker-specific tonal variations for successful learning. This study investigated how tone training with high vs. low talker-variability modulated novice learners' neural responses to non-native tones. A passive oddball paradigm tested Mandarin-speaking participants' neural responses to Cantonese low-high and low-mid tonal contrasts in the pretest and posttest. Participants were trained using a tone identification task with feedback, either with high or low talker-variability. The results of mismatch negativity (MMN) showed no group difference in the pretest whereas the high-variability group demonstrated greater neural sensitivity to the low-high tonal contrast produced by a novel talker and a trained talker in the posttest. The finding provides (tentative) novel evidence that training variability may benefit perceptual learning of the relatively easy tone pair and facilitate the formation of talker-independent representations of non-native tones by novice learners.

(150 words)

Keywords

Training variability; Cantonese level tones; perceptual learning; talker generalization; mismatch negativity; late discrimination negativity

1. Introduction

The speech signal contains great variability, which poses considerable difficulty for second language (L2) leaners' perceptual learning of speech sound categories. For instance, multiple sources of variations have been noted in the acoustic signal of lexical tones, including talker, gender and tonal context (e.g., C. Zhang & Chen, 2016). Such variations place a demand on second-language (L2) learners to extract the abstract representations from tonal exemplars across different talkers to accommodate talker-specific tonal variations (K. Zhang et al., 2018), in order to distinguish different tone categories successfully. Recent training studies have debated the role of training exposure to talker variability (i.e., training variability) in perceptual learning of speech sounds at a behavioral level (Fuhrmeister & Myers, 2020; Perrachione et al., 2011). An open question is whether, and if so how, training variability influences perceptual learning of non-native tones and generalization to new tokens produced by novel talkers (i.e., talker generalization) at a neural level. In this paper, we will investigate the effect of training variability on Mandarin-speaking participants' neural responses, as indexed by the mismatch negativity (MMN) and late discrimination negativity (LDN), in their perceptual learning of Cantonese level tones produced by trained talkers and a novel talker.

1.1. Training variability in tone learning

The variability of training materials is assumed to benefit learners and has been tested in perceptual learning of segments (Bradlow et al., 1997, 1999; Lively et al., 1994) and prosodic categories such as lexical tones (Wayland & Guion, 2004; Wiener et al., 2020). However, the findings are mixed in terms of its beneficial impact. Studies have found the beneficial effects of exposing learners to variability, for example, of the /ɹ/-/l/ contrast, during training (Bradlow et al., 1997; Lively et al., 1994). Training variability is often introduced by presenting training tokens

of /ɹ/ and /l/ produced by multiple talkers or occurring in different phonological contexts (e.g., vowels following /ɹ/ or /l/) (Bradlow et al., 1997, 1999). In the case of perceptual learning of lexical tones, Wang and her colleagues showed that perceptual training using Mandarin tone stimuli produced by multiple talkers facilitated English-speaking participants' perception of Mandarin tones produced by the trained talkers, generalization to tones produced by novel talkers, as well as long-term retention of non-native tones (e.g., six months after training) (Wang et al., 1999). The finding suggests that training variability might have facilitated learners' focus on perceptual cues of lexical tones (i.e., pitch height: higher or lower tones; pitch contour: level, falling, or rising tones), which are generalizable across talkers and facilitated tone learning over a long time period. Given the high variability of lexical tones across (and within) talkers (Peng, 2006a), the tone variability induced by different talkers (i.e., talker variability) during training seems to be critical for learners' abstraction of tone representations.

On the other hand, many tone training studies used tone stimuli produced by multiple talkers but challenged whether the positive effect of high-variability training was universal relative to low-variability training (Dong et al., 2019; Sadakata & McQueen, 2014). First, the (beneficial) effect of training variability is influenced by how variability is implemented during training (Perrachione et al., 2011). Perrachione et al. (2011) trained English-speaking participants to use Mandarin tones in identifying pseudowords and manipulated the degree of trial-to-trial variability in different types of high-variability training. The results showed that talker-blocked training (in which stimuli produced by a single talker were presented in one block and stimuli from multiple talkers were introduced across blocks) elicited a faster learning rate as well as a better learning outcome than the talker-mixed training (in which stimuli produced by multiple talkers were presented within a block).

Another factor that might account for the mixed findings in the literature is the target learners who were trained or tested (C. B. Chang & Bowles, 2015). Many of the earlier studies that found the beneficial effect of training variability had tested learners who had prior experience with the L2 and might be more capable of dealing with training variability of L2 sounds (Bradlow et al., 1997; Lively et al., 1994; Wang et al., 1999). However, studies that trained novice learners (i.e., naïve listeners) on non-native tones often did not reveal the beneficial effect of training variability (Dong et al., 2019; Perrachione et al., 2011). A different but related issue in the current tone learning studies is that most of the aforementioned studies investigated perceptual learning of Mandarin tones with pitch contour contrasts by novice learners who speak non-tonal languages (e.g., English and Dutch) without prior exposure to lexical tones (Dong et al., 2019; Perrachione et al., 2011; Sadakata & McQueen, 2014). Different from contour tones in Mandarin, Cantonese has multiple level tones, including T1 (high-level tone), T3 (mid-level tone) and T6 (low-level tone), which are primarily distinguished by fine-grained pitch height differences. Level tones with less dynamic contour changes are more susceptible to the influence of talker variability than contour tones (Peng, 2006a), and thus constitute an important investigation case to further understand the effect of training variability on level-tone learning through talker generalization.

The third factor which may account for the mixed findings of high-variability training is training time (Fenn et al., 2013; Fuhrmeister & Myers, 2020). While most previous training studies did not control at what time the participants were trained, recent studies suggested that evening training facilitated better retention of newly-learned sound contrasts by promoting generalization across talkers than morning training did (Earle & Myers, 2015; Xie et al., 2018). For instance, Earle and Myers (2015) showed that while English listeners trained in the evening

improved significantly in identifying the novel Hindi sound stimuli produced by an *untrained* talker (but not those produced by a trained talker), those trained in the morning did not show such a pattern. In the case of tone learning, Qin and Zhang (2019) showed that Mandarin listeners trained in the evening showed an improved trend in identifying the level tones produced by both the trained and untrained talkers. Again, those trained in the morning did not show such a pattern. In short, previous studies have found that high-variability speech training, when conducted in the evening, has the potential to benefit perceptual learning of lexical tones through talker generalization. We will then examine perceptual learning of (Cantonese) level tones, produced by trained and untrained talkers.

1.2. Training variability in neural processing of tones

While training variability has been tested in many behavioral studies on lexical tones, it remains unclear how training variability will affect learners' neural responses to non-native tonal contrasts after training. It is important to note that participants would need to discriminate or identify tone stimuli consciously in behavioral studies, so their behavioral responses may have been affected by factors of attention, working memory and others. In contrast, event-related potentials (ERPs) are a good method to study the pre-attentive (or unconscious) processing of lexical tones when the auditory stimuli are presented to participants without their focal attention. Testing pre-attentive processing of lexical tones using ERPs is more informative than only recording behavioral responses (e.g., discrimination accuracy) because perceptual changes may only occur at the unconscious level, for instance, in a training study (Lu et al., 2015). For instance, the MMN, a frontal negative ERP component occurring about 100–300 ms after stimulus onset, has been used as a tool to assess the pre-attentive ability to distinguish lexical tones by native and non-native listeners (see Näätänen, 2001 for an overview). The MMN is

elicited by infrequent stimuli that deviate from frequently presented (standard) stimuli in pitch or other phonetic cues (e.g., duration, voice onset time), and a larger MMN amplitude and/or an earlier MMN peak indicate a greater sensitivity to these cues (Näätänen, 2001; Tuninetti et al., 2017). The changes of MMN amplitude and/or peak latency can be observed even before changes in behavioral discrimination performance (Tremblay et al., 1998). Thus, the pre-attentive response, MMN, provides a sensitive tool to test the neural mechanisms underlying native and non-native tone discrimination (Chandrasekaran et al., 2007a; Kaan et al., 2007).

A few studies have employed the MMN to examine the processing of Mandarin tones by native Mandarin-learning children (Lee et al., 2012) and adult learners who speak non-tonal languages, for example, English (Liu et al., 2018; Yu et al., 2019). However, fewer studies have used the MMN to investigate the effect of laboratory training on neural responses to non-native tones by adult learners who speak tonal languages. Kaan and her colleagues used a passive oddball paradigm to investigate the effects of L1 backgrounds (i.e., Mandarin versus English) and perceptual identification training (i.e., before and after training) on the pre-attentive processing of Thai tones as indexed by the MMN (Kaan et al., 2007, 2008). The ERP results showed that the Mandarin and English-speaking participants achieved different training outcomes, which was attributed to the effect of L1. After training, the English listeners showed an increased MMN (150-300 ms). Interestingly, the MMN increase was not observed after training for the Mandarin listeners, who only showed a decreased late negativity (500-700 ms). The group difference was further modulated by tonal contrasts, in that no group difference was found with respect to the Thai low-mid tonal contrast (i.e., a tone pair that was perceptually trained vs. a high-low tonal tone pair that was not), which was attributed to a large MMN amplitude in both groups before training. To our knowledge, Lu et al., is the only ERP study

which used MMN (and late negativity in a time window of 500-800 ms) to examine the effect of different training methods (i.e., perception-only training versus perception-plus-production training) on English listeners' pre-attentive processing of non-native (Thai) tones (Lu et al., 2015). The behavioral results showed that English-speaking participants in both training groups were able to generalize from the trained stimuli to the untrained stimuli in novel phonetic contexts (i.e., syllables) regarding their discrimination performance. However, the MMN results did not yield a difference after training, suggesting a similar effect of the perception-only and the perception-plus-production training on the pre-attentive processing of non-native tones.

In addition to the MMN, the late negativity (mentioned above in Kaan and her colleagues' research), likely to be the late discriminative negativity (LDN), is a negative wave which could follow an MMN and often occurs around 500 ms after the onset of auditory stimuli (Cheour et al., 2001). Although the cognitive function of the late negativity remains debated (e.g., whether the MMN and late negativity have the same underlying mechanism), the late negativity was often reported to reflect additional processing of the stimuli, for instance, when the salient features of the stimuli are hard to detect (Bishop et al., 2011) or when the stimuli are newly encountered (Zachau et al., 2005). In the studies on lexical tones, the late negativity has been suggested to be associated with the transfer of the newly-encountered tone regularity into long-term memory, that is, a higher level of tone abstraction (Cheour et al., 2001). It was also suggested to reflect the reorientation of attention after involuntary attention to deviant tone stimuli (Lu et al., 2015). Importantly, the late negativity was reported to become smaller in amplitude after training in several tone training studies, potentially suggesting an effect of perceptual learning on more efficient neural transfer or attentional reorientation to lexical tone changes (Kaan et al., 2007, 2008). Since the decreased negativity was associated with improved discrimination, we followed

the tentative interpretation in (Chen et al., 2018) that the late negativity is a discriminative neural response, that is, the LDN. While the previous ERP research has suggested the effect of L1 on the MMN and LDN as well as the efficacy of perception-only training on the processing of both trained and untrained tone stimuli, little ERP research (to our knowledge) to date has investigated the effect of training variability on the neural processing of non-native tones, especially in novice learners with tonal L1 backgrounds. A MMN study, which are sensitive in revealing unconscious changes after training, may be well suited for informing the debate on the effect of training variability and deepening our understanding of the changes of neural sensitivity to non-native tones.

1.3. The current study

While many studies have employed the MMN (and LDN) to test neural processing of contour tones by novice learners with non-tonal L1 backgrounds (Chandrasekaran et al., 2007b), it is less clear how training variability modulates the neural responses to level tones by novice learners with tonal L1 backgrounds, and their neural responses to tones produced by trained and novel talkers (i.e., talker generalization) after training. Therefore, the present study investigated the effect of training variability on neural responses by focusing on Mandarin-speaking novice learners' perceptual learning of Cantonese level-level tonal contrasts. On the one hand, with a tonal L1 background, Mandarin speakers are familiar with the use of pitch patterns and their variability in the lexical domain, meaning that they may have some competence in handling training variability (Wayland & Guion, 2004; K. Zhang et al., 2018). On the other hand, Mandarin speakers rely more on pitch contour cues than pitch height cues in pitch perception as a result of the influence of their contour tone system (Gandour, 1983). The tone system places a demand on Mandarin speakers to learn to differentiate fine-grained variations in a less-familiar

dimension (i.e., pitch height) and generalize the learned contrasts to new talkers (Qin & Jongman, 2016). These aspects of Mandarin-speaking novice learners make them a valuable case in studying the effect of training variability on tone learning.

The aim of the present study is to examine whether, and if so how, training variability influences Mandarin-speaking novice learners' neural processing of Cantonese level tones by testing pre-attentive (MMN) and late, potentially attentive neural responses (LDN). Level-tone stimuli produced by trained and novel (untrained) talkers are used to assess talker generalization. If there is a beneficial effect of high-variability training on talker generalization in Mandarin-speaking novice learners, we expect learners receiving high-variability training to show a more pronounced MMN (e.g., a larger amplitude) and/or a more decreased LDN (e.g., a smaller amplitude) than learners receiving low-variability training for tone stimuli, especially for stimuli produced by the untrained talker.

Another aim of the current study is to investigate which tone pairs (i.e., tonal contrasts) are more likely to yield the effect of training variability. Some tone pairs are more easily confused than others because of their acoustic salience. For instance, Chandrasekaran et al. (2007b) tested the effect of L1 on the pre-attentive processing of different Mandarin tone pairs depending on the acoustic salience, that is, an easy tone pair (T1, a high level tone vs. T2, a high rising tone) with larger acoustic differences and a difficult tone pair (T2, a high rising tone vs. T3, a falling-rising tone) with smaller acoustic differences. The results showed that the Mandarin listeners demonstrated a larger MMN amplitude than the English listeners to the easy tone pair which was acoustically salient (Chandrasekaran et al., 2007a). Since training may also modulate the neural processing of tone pairs differently depending on their acoustic salience (Kaan et al., 2007; Wang et al., 1999), two level-tone pairs with large acoustic differences (i.e., an easy tone pair)

and small acoustic differences (i.e., a difficult tone pair) are included in the current study. Given the reported perceptual difficulty by non-native Mandarin listeners in differentiating Cantonese level tones (Qin & Jongman, 2016), the effect of training variability in terms of MMN is predicted to show different results for the tone pairs with an effect more likely to be found for the easy tone pair which is acoustically salient. In a nutshell, we predict an interaction of training groups with other factors (e.g., tonal contrasts) on the MMN, instead of a simple effect of training groups, given the mixed findings in the literature (Chandrasekaran et al., 2007b; Kaan et al., 2007). A decreased LDN, which might also interact with tonal contrasts, is predicted after training based on the previous findings (Kaan et al., 2007, 2008).

To test these predictions, we adopted a pretest-training-posttest design to compare the neural responses in novice learners with tonal L1 backgrounds. Two training groups of Mandarin-speaking participants received Cantonese-tone identification training in either a high-variability or a low-variability condition. The participants were all trained in the evening using a talker-blocked fashion to achieve the optimal learning outcome. Stimuli of easy and difficult tone pairs, produced by trained and novel talkers, were used to assess the effect of training variability on early (pre-attentive, MMN) and late (possibly attentive, LDN) neural responses.

2. Methods and materials

2.1. Participants

Forty Mandarin-speaking participants were recruited for the experiment in Hong Kong. They were all native Mandarin speakers. And they were novice learners with minimal exposure to Cantonese (length of residence in Hong Kong shorter than thirteen months; no classroom learning of Cantonese). All the participants identified Beijing Mandarin (i.e., Putonghua) to be their L1, alone or together with another Mandarin Chinese variety. None of them knew any

Southern Chinese dialect/language (e.g., Shanghainese). None of them had received more than three years of music lessons in any musical instrument including vocal training or reported a history of hearing impairment and neurological disorders. All participants were college students. The participants gave written informed consent and were paid for their participation. The testing took place at the Speech and Language Science Lab at the Hong Kong Polytechnic University. Although a power analysis[1] was conducted to determinate the sample size in the current study, it is acknowledged that larger sample sizes were recently recommended to attain sufficient statistical power in neuroscience research (Button et al., 2013; Gelman & Carlin, 2014), which should be considered in future studies (see discussion below).

The participants were randomly and equally assigned into either a high-variability training group (mean age: 25.5 [standard deviation (SD): 2.4], 12 females) or a low-variability training group (mean age: 25.0 [standard deviation (SD): 2.6], 13 females). The two groups were crucially matched on age, gender and musical, pitch and cognitive aptitude prior to training (see below). Since learners' pretraining aptitude might interact with the influence of training variability (Perrachione et al., 2011; Qin et al., 2021), a set of behavioral pretests was conducted to ensure that the two groups were matched at their musical and pitch aptitude as well as their cognitive abilities at the group level before the training session. Specifically, the two groups had similar performance in the following pretests:

(1) the pitch-related subtests of the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003) which were used to measure the participants' musical aptitude in terms of the mean accuracy (high-variability group: 0.83 [SD: 0.07]; low-variability group: 0.84 [SD: 0.08]) (Chen et al., 2016; Cui & Kuang, 2019);

[1] A sample size calculation (N = 20 in each group) for repeated-measures analysis of variances (ANOVA) was conducted in advance with α value set at 0.05, power of 0.92, and a large effect size estimated from the data of our previous tone training study (Qin & Zhang, 2019).

(2) a pitch threshold test adopted from Qin, Zhang, and Wang (2021) which was used to assess the participants' pitch height processing abilities of speech tones (high-variability group: 1.5 [SD: 2.5]; low-variability group: 1.6 [SD:2.3]) and non-speech tones (high-variability group: 1.9 [SD: 3.0]; low-variability group: 1.7 [SD: 3.1]) in semitones;

(3) a pitch memory span test adopted from Williamson and Stewart (2010) which was used to quantify the participants' short-term pitch memory span (high-variability group: 6.4 [SD: 1.3]; low-variability group: 5.8 [SD:1.4]) in terms of the number of tones which can be recalled; and

(4) the subsets of the Test of Everyday Attention (TEA) (Ou & Law, 2017), which were used to measure the participants' attentional capacity (i.e., selective and sustained attention and attentional switching) in the auditory mode (high-variability group: 11.3 [SD: 1.6]; low-variability group: 11.7 [SD: 1.6]) and the visual mode (high-variability group: 10.9 [SD: 1.7]; low-variability group: 10.3 [SD: 1.7]) in terms of the count of correct trials.

2.2. Stimuli

The training (and behavioral pretest) stimuli were three Cantonese level tones, /55/ T1 (a high-level tone), /33/ T3 (a mid-level tone), and /22/ T6 (a low-level tone) carried by ten base syllables[2] (/jan/, /ji/, /jau/, /jiu/, /fan/, /fu/, /ngaa/, /si/, /se/ and /wai/) in isolation, illustrated in Fig. 1. Each tone is labeled using Chao's (1968) tone letters, which are in the range of 1–5, with 5 referring to the highest pitch and 1 referring to the lowest pitch. All thirty words are meaningful in Cantonese.

[2] Both syllable and talker variations are sources of variability in lexical tone perception. Given a well-documented talker normalization process in lexical tone perception (Peng, 2006b; C. Zhang & Chen, 2016), this study opted for manipulating talker variability, but used the same set of (different) syllables across training conditions.

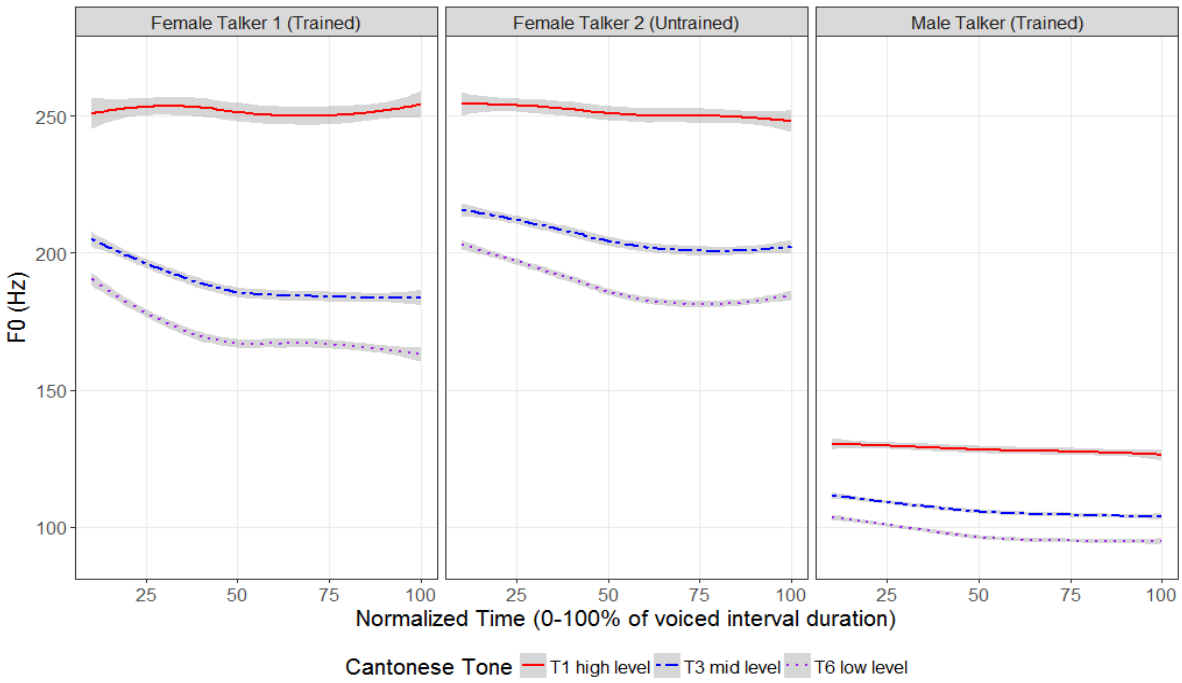

Fig. 1. Tonal contours of the three Cantonese level tones produced by three native speakers of Hong Kong Cantonese. Tonal contours were measured using ten measurement points and produced by a trained female talker (left) whose stimuli were used in the high- and low-variability training conditions, a trained male talker (right) whose stimuli were used in the high-variability training condition alone, and an untrained female talker (middle) whose stimuli were used in the (neural) posttest alone.

Two female talkers and a male talker of Hong Kong Cantonese recorded three repetitions of each target word in a sound attenuated booth on a PC workstation connected with an Azden ECZ990 microphone (Azden, Mt Arlington, NJ). The recordings were made at a sampling rate of 44100 Hz with 16 bits per sample. The stimuli were normalized in duration to 500 ms (a value similar to the duration of naturally produced stimuli), and their mean acoustic intensity was scaled to 70 dB using Praat (Boersma & Weenink, 2018).

To increase the (token) variability of tone stimuli in the high-variability training condition, two tokens for each target word were chosen from the three repetitions by the investigators based on their intelligibility and pronunciation accuracy. While two tokens were used in the high-

variability training condition, only one token of each target word was used in the low-variability training condition. As for the key manipulation of talker variability, as illustrated in Fig. 1, female talker 1 who had a wide pitch range was used in both the high-variability and low-variability training conditions, and the male talker was used in the high-variability training condition alone to introduce talker variability in this condition. Two talkers of different genders were used in the high-variability training to ensure that the training had inter-talker differences but did not hinder the initial learning due to too much (talker) variability (Y. S. Chang et al., 2017). Lastly, female talker 2 was used in the (neural) posttest alone to assess talker generalization for both groups (see further details below). T3 produced by the female talker 1 has roughly the same pitch as T6 produced by the female talker 2. Their stimuli will lead to wrong categorization without talker generalization/normalization.

To examine neural changes induced by training, a passive oddball paradigm was used to test the participants' mismatch negativity (MMN) in a (neural) pretest and posttest. The pretest included a subset of the training stimuli, namely the syllable /ji/ carrying the three Cantonese level tones (T1 – /ji55/ 'doctor', T3 – /ji33/ 'meaning', T6 – /ji22/ 'two') produced by female talker 1. The posttest additionally included the same three words with syllable /ji/ produced by untrained female talker 2, to assess talker generalization. The three tones from each talker were grouped into two pairs: T6-T1 (/ji22/-/ji55/, the low-high tonal contrast; T61 hereafter) and T6-T3 (/ji22/-/ji33/, the low-mid tonal contrast; T63 hereafter). The T61 pair had a large pitch height difference (low-level vs. high-level tone), whereas the T63 pair had a small pitch difference (low-level vs. mid-level tone). One token from each talker that was judged by a Cantonese-speaking phonetician as representative of each intended tone was used in the neural pretest and

posttest. The chosen token of each tone, produced by the trained female talker, in the neural tests was identical to that used in the behavioral pretest and training.

2.3. Procedure

Participants completed a pretest, training, and a posttest at a similar time each evening over three days within the same week. The design allows for an (immediate) overnight consolidation of tone learning and minimizes the potential effect of circadian rhythm on the performance between the pretest and posttest (Qin & Zhang, 2019). As shown in Table 1, a neural pretest and a behavioral pretest were conducted to test the participants' pretraining tonal sensitivity in the evening of the first day (Day 1). The training and the neural posttest were then scheduled on two consecutive days in the week: the two groups of participants went through a high- and a low-variability training session, respectively, in the evening (Day 2), and they were sent home for overnight sleep right after the training; a neural posttest was then conducted to retest the participants' neural responses in the evening of the next day (Day 3).

Table 1 Overview of timing in the experiment protocol.

Days	Day1	Day 2	Day 3
Time	**7-10PM**	**7-9 PM**	**7-10PM**
Sessions	**Pretest session**	**Training session**	**Posttest session**
High-variability Training Group	1. MMN pretest (talker 1) 2. AX Discrimination 3. MBEA 4. Pitch threshold task 5. Pitch memory task 6. TEA	ID training (high-variability; talkers 1&3)	MMN posttest (talkers 1&2)
Low-variability Training Group	1. MMN pretest (talker 1) 2. AX Discrimination 3. MBEA 4. Pitch threshold task 5. Pitch memory task 6. TEA	ID training (low-variability; talker 1 only)	MMN posttest (talkers 1&2)

The neural pretest was conducted using a passive oddball paradigm similar to previous studies (Lee et al., 2012; C. Zhang & Shao, 2018). The low tone – T6 was assigned as the common standard, and either T1 or T3 was used as the deviant in the T61 and T63 block, respectively. In each block, the standard T6 was presented frequently at a probability of 0.85, and the deviant (T1 or T3) was presented infrequently at a probability of 0.15. A total of 510 standards and 90 deviants were binaurally presented through earphones to the participants in each block. The standards and deviants were presented pseudo-randomly, such that the first eight stimuli of a block were always standards and any two adjacent deviants were separated by at least two standards. The inter-stimulus interval was 800ms. The participants watched a self-selected muted movie with subtitles and were instructed to ignore the auditory stimuli. Each

block lasted about eight minutes. There was a short break after each block. The order of the two blocks (2 tone pairs) was counterbalanced (two lists counter-balancing the order of tone pairs) among the participants as much as possible within and across the two training groups.

The behavioral pretest was conducted using an AX (same-different) discrimination task to measure the participants' initial discrimination sensitivity of T61 and T63 tone pairs. The participants were instructed to distinguish whether the two tones they heard belonged to the same or different tone categories by pressing one of two buttons (left arrow and right arrow) indicating "the same" or "different", respectively, on the keyboard. The two tones in each pair were carried by the same syllable. The inter-stimulus interval was 1000 ms. No feedback was given. An equal number of AA pairs (the same tone within each pair) and AB pairs (different tones within each pair) were used to counterbalance the two types of tone pairs. To ensure that the participants discriminated tones based on a change in the identity of tone categories, two acoustically different tokens of the same tone were used in each AA pair. The presentation order of two tones in each AB pair was counterbalanced in different trials. A total of 80 tokens (2 pairs * 2 orders * 2 AA/AB types * 10 syllables) were presented in a random order to the participants in one block.

In the training, a forced-choice identification (ID) task of the three Cantonese level-tone categories was administered to the participants. During the training, the participants were instructed to identify each tone (T1-High, T3-Mid and T6-Low) after hearing the auditory stimuli by pressing three buttons (1, 3, and 6) in a self-paced fashion. Written feedback ("Correct" in green or "Incorrect. The correct answer is…" in red) was given immediately after every trial. The participants were instructed to learn to categorize the three tones based on feedback and achieve the best performance they could in this session. Talker variability was manipulated in different training groups: in the low-variability training group, stimuli produced by the trained

female talker alone, comprising a total of 600 tokens (1 talker * 3 tones * 10 syllables * 1 token * 20 repetitions), were used in the training; in the high-variability training group, stimuli produced by the trained female and male talkers together, also comprising a total of 600 tokens (2 talkers * 3 tones * 10 syllables * 2 tokens * 5 repetitions), were used in the training. The tokens produced by the trained female talker were presented auditorily to the participants with 60 tokens (3 tones * 10 syllables * 2 repetitions) repeated across ten blocks in the low-variability condition. To achieve a good training outcome (Perrachione et al., 2011), talker variability was introduced across blocks, with blocks alternating between the two talkers and five repeated blocks of 60 tokens produced by each talker (3 tones * 10 syllables * 2 tokens) in the high-variability condition. The AX discrimination task took approximately 5-10 minutes, and the training took approximately 30-40 minutes. The AX discrimination and the training identification tasks were both conducted using the Paradigm software (Perception Research Systems, Inc. http://www.paradigmexperiments.com/).

The neural posttest[3] was also conducted using the passive oddball paradigm after training. The procedure was similar to the neural pretest with the exception that the tone stimuli produced by the untrained female talker were also used to assess talker generalization. There was a total of four blocks (2 tone pairs × 2 talkers) in the posttest. The order of the blocks was also counterbalanced in the neural posttest (six lists counter-balancing the order of tone pairs and talkers) among the participants as much as possible within and across the two training groups.

[3] The behavioral posttest was not included in the design because Mandarin listeners' discrimination performance did not change after tone identification training according to our previous experiments (Qin & Zhang, 2019).

2.4. EEG recording and preprocessing

EEG signals were recorded via a SynAmps 2 amplifier (NeuroScan, Charlotte, NC) with a cap carrying 64 Ag/AgCI electrodes placed at specific locations according to the extended international 10-20 system. The horizontal eye movements (HEOG) were recorded by electrodes placed on the outer canthi of each eye; vertical eye movements (VEOG) were recorded by electrodes vertically placed above and below the left eye. Impedance between the reference electrode (located between Cz and CPz) and any recording electrode was kept below 10 kΩ. The EEG signals were continuously recorded and digitized with a 24-bit resolution at a sampling rate of 1000 Hz.

All processing was conducted with EEGLAB toolbox (version 2019.1) (Delorme & Makeig, 2004) and ERPLAB toolbox (version 8.10) (Lopez-Calderon & Luck, 2014) in Matlab (MathWorks, Natick, MA). The recorded raw signals of EEG data were first re-referenced to the left and right mastoid and filtered with a 0.01-30 Hz band-pass filter (C. Zhang & Shao, 2018). Then the continuous recordings were segmented into 1100 ms epochs, from a 100 ms pre-stimulus baseline to 1000 ms after stimulus onset. No bad channels were visually identified. Independent component analysis (ICA) was then conducted for removing ocular artifacts such as blinks and eye movements, which were identified visually based on the activity power spectrum, scalp typography and activity over trials (Chen et al., 2018). Next, two artifact detection procedures were performed (Meng et al., 2020). The first one was employed for blink detection: epochs containing peak-to-peak amplitude exceeding 70 μV were rejected; the second one was for eye movement detection: epochs with covariance between the step function and the data greater than 27.5 μV were excluded from averaging. All channels were included in the artifact detection. The mean acceptance rate was matched between the high-variability training group

and the low-variability training group in the pretest (high-variability group: 50.4% [SD: 23.1%]; low-variability group: 50.8% [SD: 25.2%]) and in the posttest (high-variability group: 61.5% [SD: 26.1%]; low-variability group: 62.8% [SD: 23.8%]). The acceptance rate of each group was comparable to that of previous EEG studies using similar processing methods (Liu et al., 2018; Meng et al., 2020). No participant was excluded due to the low acceptance rate (Meng et al., 2020). Difference waves were obtained by subtracting the waveforms of the standards from that of each deviant.

2.5. Data analysis

Three electrodes – F3, Fz and F4 where the MMN was expected to peak were selected for the analysis of MMN activities following the standard practice used in the MMN studies (Chen et al., 2018; Yu et al., 2019). Consistent with the existing literature, the peak latency of MMN, was detected within the time window of 100-300 ms in each condition for each individual participant (Yu et al., 2019). The amplitude of MMN was calculated with a time window of ± 20 ms centered on the detected MMN peak at the Fz electrode in each condition for each participant (Chen et al., 2018; Yu et al., 2019). To reduce variability at a single electrode, the mean amplitude and peak latency of MMN averaged across the three chosen electrodes (F3, Fz, F4)[4] for the two training groups in each condition were submitted for further statistical analyses. As for the LDN analysis, following previous studies measuring the LDN responses to lexical tone changes (Kaan et al., 2007, 2008; Lu et al., 2015), multiple electrodes were selected as follows: F3, F5, F7, FC3, and FC5 (left), Fz and FCz (midline) and F4, F6, F8, FC4, and FC6 (right). The LDN did not have a clear peak and it had a broader scalp distribution (e.g., distributed in both

[4] Since the MMN lateralization is not the focus of the present study, the electrode site (F3, Fz and F4) was not included in the statistical analyses. Exploratory analyses of MMN with electrode as one of the within-subject factors also confirmed that it did not interact with other factors such as training groups.

left and right hemispheres) than the MMN based on visual inspection of the difference waves (Fig. 2 and 3) and scalp distribution (Fig. 5). Thus, we did not identify a LDN peak latency, and did not divide them into different regions[5]. In line with previous studies (Chen et al., 2018; Lu et al., 2015), the mean amplitude of LDN was analyzed by averaging the amplitude within a time-window of 500-700 ms across the selected electrodes in different regions.

The MMN and LDN activities were analyzed using repeated-measures ANOVAs (and t-tests) using IBM SPSS Statistics for Windows, version 25 (IBM Corp., Armonk, N. Y., USA). For all ANOVAs, Greenhouse-Geisser corrections were applied where sphericity criteria were not met. While a significance level of $p < .05$ was used for all analyses, Bonforroni correction was used to counteract the problem of multiple comparisons in the post-hoc analyses whenever applicable[6]. Effect size was calculated following Lakens (2013) and reported for reference of power analyses in future studies.

3. Results

Two sets of analyses conducted on the tone discrimination performance of the behavioral pretest and the tone identification performance of the training, respectively, confirmed that the two groups did not differ in their ability to discriminate tones before training and that they both improved during training. These analyses are reported in the Supplementary Materials.

3.1. Results of the MMN

Fig. 2 plots the difference waves of T61 for both training groups in each condition, and Fig. 3 plots the difference waves of T63 for both training groups in each condition (see Fig. 2 and 3 in

[5] Exploratory analyses of the LDN with electrode region as one of the within-subject factors confirmed that it did not interact with other factors, so the factor was taken out to simplify the analysis.
[6] The cases which did not reach statistical significance after adjusting *p*-values were reported.

the Supplementary Materials for the grand-averaged ERP response to the standard and deviant stimuli). Fig. 4 illustrates the scalp topography of the MMN peak amplitude at corresponding peak latencies in each condition for each training group and tone pair. See Table 1 in the Supplementary Materials for the MMN mean amplitude and peak latency for each training group and tone pair in each condition.

3.1.1. MMN mean amplitude

Two sets of analyses were conducted on the MMN mean amplitude. The first analysis aimed to examine whether there are neural changes in the posttest vs. pretest, and whether the neural changes are modulated by the type of training (high- vs. low-variability training). This analysis therefore focused on the stimuli of the trained female talker, which were presented to the participants in both pretest and posttest. The second analysis aimed to compare the two training groups on their effects on generalization to a novel, untrained talker after training as indexed by the MMN activities. For this reason, the second analysis focused on the posttest to compare the MMN responses to the trained vs. untrained talker.

For the first analysis, a three-way repeated-measures ANOVA with Group (high-variability, low-variability) as the between-subject factor, and Test (pretest, posttest), and Tone pair (T61, T63) as two within-subject factors was conducted to test the effect of training on MMN responses to the stimuli produced by the trained talker. The interaction between Group, Test, and Tone pair was significant (F [1, 36] = 4.76, p = .036, μ^2_p = 0.12). Post-hoc analyses for T61 tokens revealed that the high-variability training group exhibited a larger MMN amplitude (more negative) than the low-variability training group (for the stimuli produced by the trained talker) in the neural posttest (t [38] = − 2.32, p = .026, Cohen's d_s = 0.73; marginally significant after Bonferroni correction .05/2=.025), but not in the neural pretest. While the low-variability

training group showed a smaller MMN amplitude at the posttest than at the pretest (t [19] = −2.35, p = .030, Cohen's d_z = 0.53; marginally significant after Bonferroni correction .05/2=.025), such a difference was not found for the high-variability training group. Post-hoc analyses for T63 tokens did not show any significant effect.

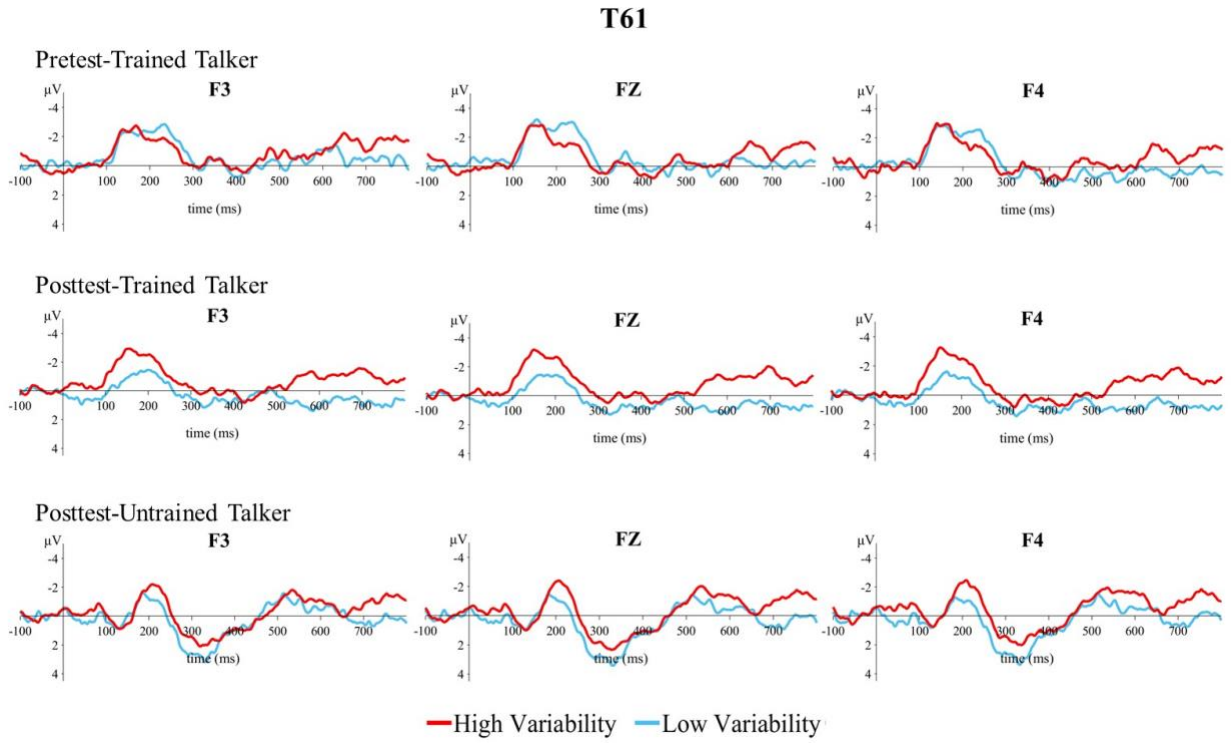

Fig. 2. Difference waves (deviant-standard) of the T61 pair for the high-variability training group (red) and the low-variability training group (blue) in the neural pretest and posttest. The stimuli produced by female talker 1 (trained) were used in the neural pretest whereas the stimuli produced by female talker 1 (trained) and female talker 2 (untrained) were used in the neural posttest.

For the second analysis, a three-way Group (high-variability, low-variability) × Talker (trained, untrained) × Tone pair (T61, T63) repeated-measures ANOVA was conducted. The results revealed a marginally significant effect of Group (F [1, 36] = 3.40, p = .07), with the high-variability training group having a numerical trend of larger MMN amplitude than the low-variability training group. Importantly, the interaction between Group, Talker, and Tone pair was significant (F [1, 36] = 4.98, p = .032, μ^2_p = 0.12). Post-hoc analyses on T61 tokens revealed a

main effect of Group (F [1, 38] = 6.03, p = .019, μ^2_p = 0.14), where the high-variability training group exhibited a larger MMN amplitude than the low-variability training group irrespective of stimuli produced by the trained and untrained talkers in the neural posttest. Post-hoc analyses for T63 tokens revealed a marginally significant interaction between Group and Talker (F [1, 36] = 3.40, p = .07; not significant after Bonferroni correction), with the high-variability training group having a numerical trend of larger MMN amplitude than the low-variability training group for stimuli produced by the untrained talker.

3.1.2. MMN peak latency

Likewise, two sets of analyses similar to those on the MMN mean amplitude were conducted on the MMN peak latency. First, a three-way Group × Test × Tone pair repeated-measures ANOVA was conducted to test the effect of training on MMN responses to the stimuli produced by the trained talker. The results did not show either a main effect or a significant interaction effect.

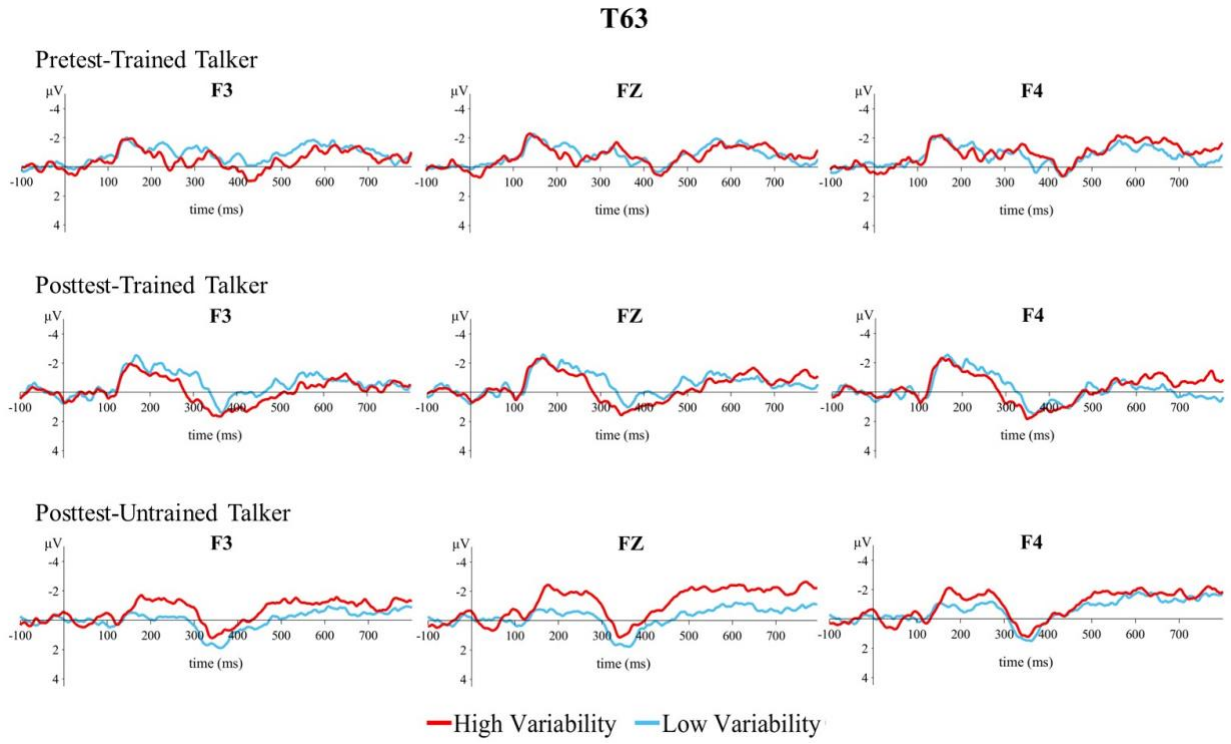

Fig. 3. Difference waves (deviant-standard) of the T63 pair for the high-variability training group (red) and the low-variability training group (blue) in the neural pretest and posttest. The stimuli produced by female talker 1 (trained) were used in the neural pretest whereas the stimuli produced by female talker 1 (trained) and female talker 2 (untrained) were used in the neural posttest.

Another three-way Group × Talker × Tone pair repeated-measures ANOVA was conducted to test talker generalization after training in terms of the MMN peak latency. A main effect of Talker was found (F [1, 38] = 7.63, p = .009, μ^2_p = 0.17), where the stimuli produced by the trained talker yielded an earlier MMN peak than the stimuli produced by the untrained talker. Moreover, the interaction between Group and Talker was also significant (F [1, 38] = 5.73, p = .022, μ^2_p = 0.13). Post-hoc analyses showed that the interaction was driven by an earlier MMN peak for the stimuli produced by the trained talker than that produced by the untrained talker found for the high-variability training group (t [1, 19] = −4.57, p < .001, Cohen's d_z = 1.02), but not for the low-variability training group.

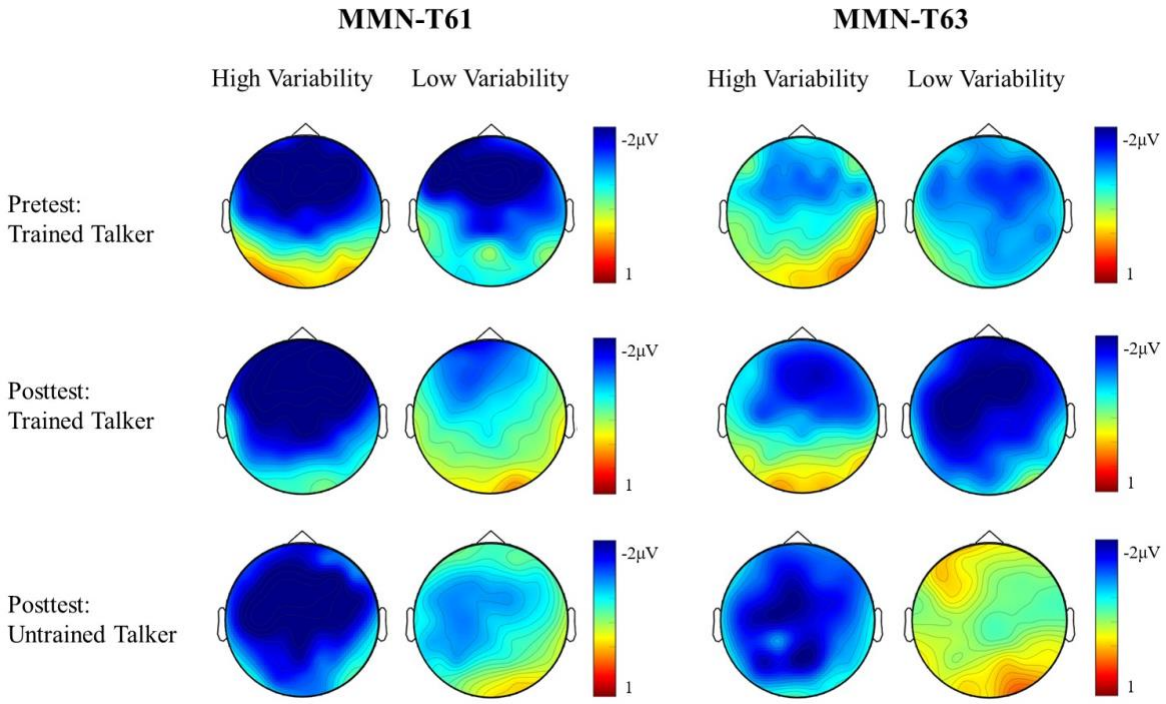

Fig. 4. Scalp distribution of MMN at the corresponding *peak* in each condition by the high-variability training group and the low-variability training group for the T61 (left) and T63 (right) pair in the neural pretest and posttest.

To summarize, these results of MMN amplitude and peak latency showed that while the high- and low-variability training groups had comparable MMN responses before training, the high-variability training group showed a larger MMN amplitude than the low-variability training group, specifically for the T61 contrast, irrespective of stimuli produced by the trained and untrained talkers after training. In addition, the high-variability training group showed an earlier MMN peak to stimuli produced by the trained talker than those produced by the untrained talker, whereas the low-variability training group did not. The results of MMN responses potentially suggest a beneficial effect of training variability on learners' early (pre-attentive) neural responses through the process of talker generalization.

3.2. Results of the LDN

In addition to the MMN, a late negativity component, likely to be the LDN (Chen et al., 2018; Cheour et al., 2001), was observed around 500 ms after stimulus onset (the duration of the tone stimuli is 500 ms). Fig. 5 shows the scalp topography of the LDN amplitude between 500 ms and 700 ms for each training group and tone pair in each condition. See Table 2 in the Supplementary Materials for the LDN difference amplitude for each training group and tone pair in each condition.

Similar to the analyses on the MMN above, two analyses were conducted on the LDN amplitude. First, a three-way Group × Test × Tone pair was conducted to test the effect of training on LDN responses to the stimuli produced by the trained talker. A main effect of Test was found (F [1, 36] = 5.64, p = .023, μ^2_p = 0.14), where the posttest elicited a smaller LDN amplitude (less negative) than the pretest for stimuli produced by the trained talker. Another three-way Group × Talker × Tone pair repeated-measures ANOVA was conducted to test talker generalization in LDN responses to the stimuli produced by the trained and untrained talkers in the neural posttest. A main effect of Talker was found (F [1, 34] = 5.71, p = .023, μ^2_p = 0.14), where the stimuli produced by the untrained talker elicited a larger LDN amplitude than the stimuli produced by the trained talker in the posttest.

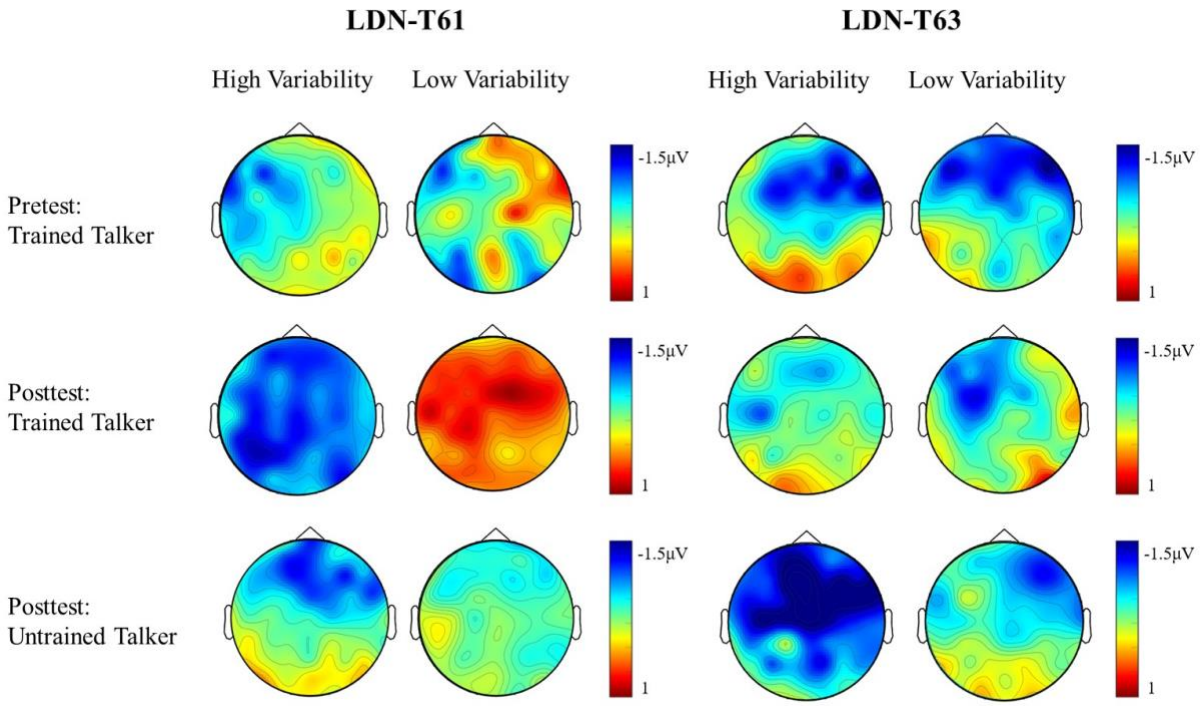

Fig. 5. Scalp distribution of LDN (500-700 ms) by the high-variability training group and the low-variability training group for the T61 (left) and T63 (right) pair in the neural pretest and posttest.

These results of LDN amplitude revealed a decreased LDN after training, potentially suggesting learners' reduced difficulty in discriminating Cantonese level tones after training (Kaan et al., 2007, 2008). The results also showed a more negative LDN for the stimuli produced by the untrained talker than for those produced by the trained talker in the posttest, probably suggesting learners' greater difficulty in discriminating untrained (i.e., unfamiliar) materials than trained materials after training. However, the results of LDN did not show an effect of training variability on learners' late (possibly attentive) neural responses.

4. Discussion

The present study investigated whether, and if so how, training variability influences Mandarin-speaking novice learners' neural processing of Cantonese level tones. Two tone pairs,

T61 and T63, were included to further investigate which tone pair would be more likely to yield the effect of training variability. The Mandarin-speaking participants were trained using either a high-variability or a low-variability training method. Their MMN and LDN were tested in a passive oddball paradigm for stimuli produced by the trained and untrained talkers in the neural pretest and posttest. We predicted that learners receiving high-variability training would show a larger MMN amplitude and/or a more decreased LDN than learners receiving low-variability training, but the effect of training variability would be modulated by the easy (T61) and difficult (T63) tone pairs depending on their acoustic salience. The results of MMN provided support for the formulated hypotheses, with the high-variability training group showing a more negative MMN to the stimuli produced by the trained and untrained talkers for the T61 contrast, but not for the T63 contrast. On the other hand, the results of LDN did not show an effect of training variability but revealed a decreased LDN after training and to the trained talker.

First and foremost, the results of MMN showed that training variability affected learners' pre-attentive neural responses to non-native tonal contrasts and tentatively supported the beneficial effect of high-variability training. Specifically, while the high- and low-variability training groups had similar behavioral and MMN responses before training, the high-variability training group showed a larger MMN amplitude, suggesting a greater sensitivity, than the low-variability training group (for the T61 contrast[7]) after training. Importantly, the effect was found for stimuli produced by the trained and untrained talkers. The high-variability group further demonstrated an earlier-peaking MMN to the trained talker as compared to the untrained talker, a pattern not found in the low-variability group. These findings suggest that the high-variability training, which involved talker (between-talker) variability and also to some extent token

[7] The statistical significance of the effect, for stimuli produced by the trained talker, differed when correction for multiple comparisons was applied.

(within-talker) variability, might have helped Mandarin-speaking participants to focus on the correct cue of pitch height which is used to differentiate Cantonese level tones and generalize their perceptual learning to tokens produced by a novel talker during their pre-attentive processing. The present study complemented the existing behavioral studies of tone training (Wang et al., 1999), and provided neural evidence that high-variability training enhanced learners' pre-attentive sensitivity to Cantonese level tones, and facilitated generalization of these tones across talkers.

Different from some behavioral studies which did not find the beneficial effect of high-variability training at the group level (Dong et al., 2019; Sadakata & McQueen, 2014), the MMN results of the current study showed the benefit of high-variability training not only for (T61) stimuli produced by the trained talker and but also for (T61 and, probably, T63) stimuli produced by the untrained talker. Besides testing neural versus behavioral responses, one important difference between the present study and these behavioral studies is that this study allowed for overnight consolidation, which often results in better retention of the learned information as well as increased generalization to new tokens at the behavioral and neural levels (Earle et al., 2017; Earle & Myers, 2015), by training (and testing) the participants in the evening hours followed by one night's sleep. In other words, the high-variability training, together with memory consolidation processes (Qin & Zhang, 2019), might have enhanced the retention of learned phonetic information and facilitated the transfer of episodic tone information from an acoustic-sensory-based trace to a more talker-independent representation of lexical tones as revealed by talker generalization (Fenn et al., 2013). In addition, the training paradigm, which introduced talker variability differently in previous studies, might also account for the mixed findings regarding high-variability training. While the current study decreased trial-to-trial variability by

training participants in a talker-blocked fashion and using two talkers to achieve a faster learning rate as well as better learning outcome of the high-variability training (Perrachione et al., 2011), other studies introduced talker-variability within a block and/or used more talkers for the high-variability training (Dong et al., 2019; Sadakata & McQueen, 2014).

Importantly, as predicted, the effect of training variability showed different results of MMN for the T61 and T63 tone pairs. The high-variability training group showed a larger MMN amplitude than the low-variability training only for T61 tokens which are acoustically salient (Chandrasekaran et al., 2007a, 2007b). In contrast, the beneficial effect of high-variability training was not significant for T63 tokens which are less acoustically salient (Mok et al., 2013), and this requires further interpretation. For T63 tokens, while the high-variability training group exhibited a numerical trend of larger MMN amplitude than the low-variability training group for stimuli produced by the untrained talker, the two groups had similar MMN responses for stimuli produced by the trained talker in the neural posttest. One possible explanation is that the high-variability training facilitated perceptual learning of the difficult tone pair to a smaller degree, that is, in terms of generalization to stimuli produced by a novel talker alone (the numerical trend needs to be confirmed using a larger sample of participants). The perceptual difficulty of the T63 pair can be attributed to its small acoustic salience. The Cantonese mid-level and low-level tones (T3 and T6), which only differ in the fine-grained pitch height differences, were found to be difficult even for native listeners to distinguish and were reported to undergo a merging process (Fung & Lee, 2019; Mok et al., 2013). On the other hand, the high-variability training benefited perceptual learning of the easy tone pair across-the-board given its acoustic salience. The finding of acoustic salience in terms of training variability is consistent with previous findings that the

acoustic salience of tonal contrasts modulated the effect of L1 background (Chandrasekaran et al., 2007b, 2007a), and possibly provided a new direction for future MMN research.

Note that no training improvement was found in the MMN (posttest versus pretest). The results of MMN are aligned with the results of Kaan et al. (2007, 2008) which did not find an increased MMN for Mandarin-speaking participants (but did for English-speaking participants) after training. One straightforward explanation is that a training-induced change of MMN might require a longer period of training, together with overnight consolidation, through which the learners can develop more stable representations of non-native tone categories (Earle & Myers, 2015). While previous studies often included multiple training sessions over days (Wang et al., 1999), this study trained participants using a one-day training session which lasted for 30-40 mins. Mandarin listeners, who initially did not use pitch height differences under the influence of their native contour-tone language (Qin & Jongman, 2016), may need more training to learn to use the new cue, especially when distinguishing the tone pair (i.e., T63) with small acoustic differences (Mok et al., 2013). It is also worth noting that whereas no change in MMN amplitude was found for the high-variability training group, a reduced MMN[8] was found for the low-variability training group. The results may in part support the account of longer training spanning multiple days, because the low-variability training group may need more exposure to tone stimuli before stabilizing their pre-attentive processing of non-native tones (K. Zhang et al., 2018). This pattern also ties well with the role of overnight consolidation in the stabilization of learned phonetic information, as discussed above. High-variability training, immediately followed by overnight consolidation, appears to induce better retention of tone stimuli produced by the trained talker (for the T61 pair) than low-variability training in their neural processing, despite

[8] The statistical significance of the effect differed when correction for multiple comparisons was applied.

the same amount of intense but brief training (Bradlow et al., 1999; Lively et al., 1994). Along the same line, the high-variability group exhibited better pre-attentive processing of stimuli produced by the untrained talker (for the T61 pair and probably the T63 pair) than the low-variability group at this early learning stage. This result may also be attributed to the facilitative effect of high-variability training on the retention of learned phonetic information and the generalization to a novel talker (Bradlow et al., 1997; Wang et al., 1999).

In contrast to the results of MMN, the results of LDN did not show an effect of training variability. While the high- and low-variability training groups had similar LDN responses before training, they both showed a decreased LDN for stimuli produced by the trained talker after training and a larger amplitude of LDN for stimuli produced by the untrained talker, irrespective of tone pairs. Different from the MMN which taps into the pre-attentive processing of lexical tones, the LDN may reflect the transfer of the newly-encountered tone regularity into long-term memory (Chen et al., 2018) and/or the reorientation of attention after involuntary attention to deviant tone stimuli (Lu et al., 2015). In either case, the decreased LDN confirmed that Cantonese level-tone contrasts became easier to detect by Mandarin-speaking learners after training in line with previous studies on tone training (Kaan et al., 2007, 2008).

In addition, the finding of Zachau et al. (2005) showed that detecting regularities of tone patterns in unfamiliar stimuli might increase the difficulty of transfer to long-term memory, and hence increase the amplitude of the LDN. Thus, the larger amplitude of LDN found for stimuli produced by the untrained talker suggests a greater difficulty processing novel stimuli than trained stimuli. It should be acknowledged that the late negativity was interpreted as LDN, but it remains unclear whether the LDN found in this study accounted for discrimination performance after training at the behavioral level. To fully examine the nature of LDN, it will be worth

exploring the association of a possible training-induced improved discrimination behaviorally and decreased LDN after training (Kaan et al., 2007, 2008). Future studies will be necessary to assess the presence of discrimination change and its association with LDN change, for example, by including an AX discrimination posttest.

Finally, there are some limitations with the present study that should be addressed by future studies. First, while the sample size in the current study is comparable to similar learning studies (Lu et al., 2015), it is admittedly small. Following the discussion of the limitation of small sample sizes in neural (and behavioral) language learning research (Brysbaert, 2020; Button et al., 2013), future studies are recommended to include a larger sample to replicate the current findings, and to test whether the results can be generalized to other learner groups (e.g., experienced learners; learners of other L1 backgrounds). Second, only a single session of training was conducted, and for the high-variability training group only two talkers were included to ensure that the training did not hinder initial learning due to (too much) talker-induced variability (Y. S. Chang et al., 2017). Future studies should include more talker variability in multiple training sessions, in order to examine whether a learning effect for the perceptually challenging tone pair (e.g., T63) would arise with more training (and more variability). Last, while the participants' pretraining aptitude was matched between the two training groups in the present study, it was not systematically manipulated in terms of the training condition. Recent studies have shown that high-aptitude learners benefited more from the high-variability training whereas low-aptitude learners might benefit more from the low-variability training (Perrachione et al., 2011; Sadakata & McQueen, 2014). Therefore, it would be useful to compare two subgroups of participants with higher and lower aptitude in the high-variability training condition using a larger sample of participants. The present study therefore

calls for future tone training studies to further examine the interaction between pretraining aptitude and training variability at the (pre-attentive) neural level (Fuhrmeister & Myers, 2020, 2021).

To conclude, the present study adds to our knowledge on the effect of training variability on tone learning, demonstrating that the high-(talker) variability training may enhance Mandarin-speaking novice learners' early pre-attentive neural responses to the easy non-native level-tone contrast, but not late neural responses to the level-tone stimuli. The finding of MMN, together with the decreased LDN, provided tentative neural evidence that high-variability training might have benefited Mandarin-speaking participants' neural sensitivity to certain Cantonese level tones (T61) and facilitated their perceptual learning of these tones across talkers. The finding further suggested that the acoustic salience of tonal contrasts and the nature of neural responses, together with other factors (e.g., consolidation and trial-to-trial variability), might modulate the beneficial effect of high-variability training. Further research should provide a refined investigation of the effect of training variability in relation to learners' individual aptitude at the neural level with a larger sample.

CRediT Author Statement

Zhen Qin: Conceptualization, Methodology, Data curation, Formal analysis, Writing-Original draft, Funding acquisition. **Minzhi Gong**: Investigation, Data curation, Formal analysis. **Caicai Zhang**: Conceptualization, Methodology, Writing- Reviewing and Editing.

Acknowledgements

This research was supported by a *Language Learning* Early Career Research Grant and Start-up Fund at the Division of Humanities, the Hong Kong University of Science and Technology awarded to Zhen Qin. The authors would like to thank Weijie Tan for her help in the early stages of the project.

References

Bishop, D. V. M., Hardiman, M. J., & Barry, J. G. (2011). Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. *Developmental Science*, *14*(2), 402–416. https://doi.org/10.1111/j.1467-7687.2010.00990.x

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer [Computer program]*. *Version 6.0.43*. Retrieved 8 September 2018.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify english /r/and /1/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English / r / and / l /: IV. Some effects of perceptual learning on speech production . *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Brysbaert, M. (2020). Power considerations in bilingualism research: Time to step up our game. In *Bilingualism*. https://doi.org/10.1017/S1366728920000437

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007a). Experience-dependent neural

plasticity is sensitive to shape of pitch contours. *NeuroReport*, *18*(18), 1963–1967. https://doi.org/10.1097/WNR.0b013e3282f213c5

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, *1128*(1), 148–156. https://doi.org/10.1016/j.brainres.2006.10.064

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *The Journal of the Acoustical Society of America*, *138*(6), 3703–3716. https://doi.org/10.1121/1.4937612

Chang, Y. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126. https://doi.org/10.1121/1.4976037

Chao, Y. R. (1968). A grammar of spoken Chinese = Zhongguo hua de wen fa. In *Zhongguo hua de wen fa*. Berkeley : University of California Press.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, *31*(6), 751–760. https://doi.org/10.1080/23273798.2016.1156715

Chen, A., Peter, V., Wijnen, F., Schnack, H., & Burnham, D. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language*, *180–182*, 31–41. https://doi.org/10.1016/j.bandl.2018.04.006

Cheour, M., Korpilahti, P., Martynova, O., & Lang, A. H. (2001). Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and

infants. In *Audiology and Neuro-Otology* (Vol. 6, Issue 1, pp. 2–11). https://doi.org/10.1159/000046804

Cui, A., & Kuang, J. (2019). The effects of musicality and language background on cue integration in pitch perception. *The Journal of the Acoustical Society of America*, *146*(6), 4086–4096. https://doi.org/10.1121/1.5134442

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

Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, *2019*(8), e7191. https://doi.org/10.7717/peerj.7191

Earle, F. S., Landi, N., & Myers, E. B. (2017). Sleep duration predicts behavioral and neural differences in adult speech sound learning. *Neuroscience Letters*, *636*, 77–82. https://doi.org/10.1016/j.neulet.2016.10.044

Earle, F. S., & Myers, E. B. (2015). Overnight consolidation promotes generalization across talkers in the identification of nonnative speech sounds. *The Journal of the Acoustical Society of America*, *137*(1), EL91–EL97. https://doi.org/10.1121/1.4903918

Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2013). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition*, *128*(3), 280–286. https://doi.org/10.1016/j.cognition.2013.04.007

Fuhrmeister, P., & Myers, E. B. (2020). Desirable and undesirable difficulties: Influences of

variability, training schedule, and aptitude on nonnative phonetic learning. *Attention, Perception, and Psychophysics*, *82*, 2049–2065. https://doi.org/10.3758/s13414-019-01925-y

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language*, *215*, 104919. https://doi.org/https://doi.org/10.1016/j.bandl.2021.104919

Fung, R. S. Y., & Lee, C. K. C. (2019). Tone mergers in Hong Kong Cantonese: An asymmetry of production and perception. *The Journal of the Acoustical Society of America*, *146*(5), EL424–EL430. https://doi.org/10.1121/1.5133661

Gandour, J. T. (1983). Tone perception in far eastern-languages. *Journal of Phonetics*, *11*(2), 149–175.

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Kaan, E., Barkley, C. M., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. *BMC Neuroscience*, *9*, 53. https://doi.org/10.1186/1471-2202-9-53

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research*, *1148*(1), 113–122. https://doi.org/10.1016/j.brainres.2007.02.019

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A

practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(NOV). https://doi.org/10.3389/fpsyg.2013.00863

Lee, C. Y., Yen, H. ling, Yeh, P. wen, Lin, W. H., Cheng, Y. Y., Tzeng, Y. L., & Wu, H. C. (2012). Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia*, *50*(14), 3228–3239. https://doi.org/10.1016/j.neuropsychologia.2012.08.025

Liu, L., Ong, J. H., Tuninetti, A., & Escudero, P. (2018). One way or another: Evidence for perceptual asymmetry in pre-attentive learning of non-native contrasts. *Frontiers in Psychology*, *9*(MAR), 162. https://doi.org/10.3389/fpsyg.2018.00162

Lively, S. E., Pisoni, D. B., Yamada, R. A., Yoh'ichi, T., & Yamada, T. (1994). Training japanese listeners to identify english /r/ and /l/. iii. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*(4), 2076–2087. https://doi.org/10.1121/1.410149

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*(1 APR), 213. https://doi.org/10.3389/fnhum.2014.00213

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception - A behavioral and ERP study. *Brain Research*, *1624*, 28–44. https://doi.org/10.1016/j.brainres.2015.07.014

Meng, Y., Zhang, J., Liu, S., & Wu, C. (2020). Influence of different acoustic cues in L1 lexical tone on the perception of L2 lexical stress using principal component analysis: an ERP study. *Experimental Brain Research*, *238*(6), 1489–1498. https://doi.org/10.1007/s00221-

020-05823-w

Mok, P., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370. https://doi.org/10.1017/s0954394513000161

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(1), 1–21. https://doi.org/10.1111/1469-8986.3810001

Ou, J., & Law, S. P. (2017). Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. *Attention, Perception, and Psychophysics*, *79*(3), 945–963. https://doi.org/10.3758/s13414-017-1283-z

Peng, G. (2006a). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of mandarin and cantonese. In *Journal of Chinese Linguistics* (Vol. 34, Issue 1, pp. 134–154).

Peng, G. (2006b). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of mandarin and cantonese. *Journal of Chinese Linguistics*, *34*(1), 134–154.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of Musical Disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*(1), 58–75. https://doi.org/10.1196/annals.1284.006

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.

https://doi.org/10.1121/1.3593366

Qin, Z., & Jongman, A. (2016). Does Second Language Experience Modulate Perception of Tones in a Third Language? *Language and Speech*, *59*(3), 318–338. https://doi.org/10.1177/0023830915590191

Qin, Z., & Zhang, C. (2019). The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PLOS ONE*, *14*(12), e0221498. https://doi.org/10.1371/journal.pone.0221498

Qin, Z., Zhang, C., & Wang, W. S. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America*, *149*(1), 435–446. https://doi.org/10.1121/10.0003330

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*(NOV), 1318. https://doi.org/10.3389/fpsyg.2014.01318

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training. *NeuroReport*, *9*(16), 3557–3560. https://doi.org/10.1097/00001756-199811160-00003

Tuninetti, A., Chládková, K., Peter, V., Schiller, N. O., & Escudero, P. (2017). When speaker identity is unavoidable: Neural processing of speaker identity cues in natural speech. *Brain and Language*, *174*, 42–49. https://doi.org/10.1016/j.bandl.2017.07.001

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–

3658. https://doi.org/10.1121/1.428217

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681–712. https://doi.org/10.1111/j.1467-9922.2004.00283.x

Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do Explicit Instruction and High Variability Phonetic Training Improve Nonnative Speakers' Mandarin Tone Productions? *Modern Language Journal*, *104*(1), 152–168. https://doi.org/10.1111/modl.12619

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, *18*(6), 657–669. https://doi.org/10.1080/09658211.2010.501339

Xie, X., Earle, F. S., & Myers, E. B. (2018). Sleep facilitates generalisation of accent adaptation to a new talker. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2017.1369551

Yu, K., Li, L., Chen, Y., Zhou, Y., Wang, R., Zhang, Y., & Li, P. (2019). Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology*, *56*(11), e13448. https://doi.org/10.1111/psyp.13448

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: Early and late mismatch negativity to tone patterns. *NeuroReport*, *16*(18), 2015–2019. https://doi.org/10.1097/00001756-200512190-00009

Zhang, C., & Chen, S. (2016). Toward an integrative model of talker normalization. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(8), 1252–1268.

https://doi.org/10.1037/xhp0000216

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, K., Peng, G., Li, Y., Minett, J. W., & Wang, W. S. Y. (2018). The effect of speech variability on tonal language speakers' second language lexical tone learning. *Frontiers in Psychology*, *9*(OCT), 1–13. https://doi.org/10.3389/fpsyg.2018.01982