frontiers in Psychology

ORIGINAL RESEARCH
published: 24 May 2019
doi: 10.3389/fpsyg.2019.01178

The Role of Temporal Acoustic Exaggeration in High Variability Phonetic Training: A Behavioral and ERP Study

Bing Cheng[1], Xiaojuan Zhang[1], Siying Fan[1] and Yang Zhang[2]*

[1] English Department & Language and Cognitive Neuroscience Lab, School of Foreign Studies, Xi'an Jiaotong University, Xi'an, China, [2] Department of Speech-Language-Hearing Sciences, Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, United States

High variability phonetic training (HVPT) has been found to be effective in helping adult learners acquire non-native phonetic contrasts. The present study investigated the role of temporal acoustic exaggeration by comparing the canonical HVPT paradigm without involving acoustic exaggeration with a modified adaptive HVPT paradigm that integrated key temporal exaggerations in infant-directed speech (IDS). Sixty native Chinese adults participated in the training of the English /i/ and /ɪ/ vowel contrast and were randomly assigned to three subject groups. Twenty were trained with the typical HVPT paradigm (the HVPT group), twenty were trained under the modified adaptive approach with acoustic exaggeration (the HVPT-E group), and twenty were in the control group. Behavioral tasks for the pre- and post- tests used natural word identification, synthetic stimuli identification, and synthetic stimuli discrimination. Mismatch negativity (MMN) responses from the HVPT-E group were also obtained to assess the training effects in within- and across- category discrimination without requiring focused attention. Like previous studies, significant generalization effects to new talkers were found in both the HVPT group and the HVPT-E group. The HVPT-E group, by contrast, showed greater improvement as reflected in larger progress in natural word identification performance. Furthermore, the HVPT-E group exhibited more native-like categorical perception based on spectral cues after training, together with corresponding training-induced changes in the MMN responses to within- and across- category differences. These data provide the initial evidence supporting the important role of temporal acoustic exaggeration with adaptive training in facilitating phonetic learning and promoting brain plasticity at the perceptual and pre-attentive neural levels.

Keywords: acoustic exaggeration, HVPT, categorical perception, mismatch negativity, second language learning

OPEN ACCESS

Edited by:
Fei Chen,
Southern University of Science and Technology, China

Reviewed by:
Sari Ylinen,
University of Helsinki, Finland
Yan H. Yu,
St. John's University, United States

****Correspondence:***
Yang Zhang
zhanglab@umn.edu;
zhang470@umn.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *24 January 2019*
Accepted: *06 May 2019*
Published: *24 May 2019*

Citation:
Cheng B, Zhang X, Fan S and Zhang Y (2019) The Role of Temporal Acoustic Exaggeration in High Variability Phonetic Training: A Behavioral and ERP Study. Front. Psychol. 10:1178. doi: 10.3389/fpsyg.2019.01178

INTRODUCTION

Learning non-native speech sounds can be a particularly challenging task. For example, the distinction of the English vowel contrast /i/-/ɪ/ is exceptionally difficult for many native Chinese speakers who learn English as a second language (L2) (Flege et al., 1997; Wang, 1997; Wang and Heuven, 2006; Cheng and Zhang, 2013; Liu et al., 2014; Huang et al., 2018). The literature has well documented that high variability phonetic training (HVPT) is effective on improving non-native speech perception. This improvement can generalize to new contexts and talkers, and transfer to production (Logan et al., 1991; Lively et al., 1993; Strange, 1995; Bradlow and Pisoni, 1997; Wang et al., 1999; Iverson et al., 2005; Sadakata and Mcqueen, 2013). However, not all L2 learners appear to benefit from HVPT; the effectiveness is limited by learners' perceptual abilities, first language (L1) background, and the nature of speech categories to be learned (Iverson and Evans, 2009; Perrachione et al., 2011; Sadakata and Mcqueen, 2014; Grenon et al., 2019). Although earlier studies showed the efficacy of HVPT only with the identification training protocol, a recent study demonstrated the feasibility of both identification and discrimination training with similar improvements (Shinohara and Iverson, 2018). Some studies further indicated the benefits of combining systematic temporal/spectral exaggeration with adaptive training in the HVPT paradigm (Zhang et al., 2000, 2009; Grenon et al., 2019) while achieving limited success in overcoming the native language interference. But the effects of this modified and integrated HVPT approach have not been directly compared with those without introducing the temporal acoustic exaggeration.

The present investigation aimed to compare the two HVPT protocols in training adult Chinese speakers to learn the English vowel contrast /i/-/ɪ/. We included two training groups and one control group to test the efficacy and generalizability of the integrated training approach combining the features of high variability and temporal acoustic exaggeration with adaptive learning motivated by infant-directed speech (IDS), which has been previously shown to promote adult L2 perceptual learning (Iverson et al., 2003; Zhang et al., 2009). We also supplemented the behavioral results with electrophysiological data to examine the training-related neural plasticity in acquiring L2 speech categories, which can provide insights into the nature of underlying mechanisms of non-native phonetic learning in adulthood at both attentive and pre-attentive levels.

Non-native Phonetic Training Methods

Developmental speech perception studies have shown a strong native language neural commitment process early in life that facilitates L1 learning and constrains the infant's ability to perceive non-native speech contrasts (Kuhl, 2000). Despite the limitations from L1 interference, adult learners can improve their L2 perception and production at least for some non-native speech sounds, indicating that the adult brain has more neural plasticity than was ever believed (Pisoni et al., 1982; Strange and Dittmann, 1984; Jamieson and Morosan, 1986; Logan et al., 1991; Lively et al., 1993, 1994; Bradlow and Pisoni, 1997; Wang et al., 1999). Based on the assumption that adult perceptual mechanisms can be retuned, auditory training studies have utilized a variety of short-term intensive laboratory training methods to improve non-native speech perception, which can result in robust learning in terms of generalization to new phonetic contexts and new talkers (Francis et al., 2000; McCandliss et al., 2002; Pruitt et al., 2006). However, it seems that effective generalization may depend on whether learners encounter sufficiently variable stimuli during training. For example, Strange and Dittmann (1984) trained native Japanese speakers to learn the English contrast /r/-/l/ using the stimuli along a synthetic continuum of *rock-lock*. The results showed that despite improvements in discrimination and identification after training on the trained synthetic stimuli and on a novel synthetic continuum, the training effects did not appear to transfer to naturally-produced minimal pairs. Later on, Logan et al. (1991) also trained native Japanese speakers to perceive the English /r/-/l/ contrast, and they introduced highly variable stimuli including multiple natural minimal pairs produced by multiple talkers (i.e., HVPT). The results revealed improvements in both trained and untrained stimuli. Lively et al. (1993) replicated the generalization effect of HVPT. When using the same HVPT paradigm with the training stimuli produced by a single talker in a follow-up experiment, they found improvements for the trained talker but not for a new talker, which indicates an important role for high talker variability in phonetic training.

Since those early attempts, the effectiveness of HVPT has been reported on a variety of phonetic contrasts with some studies showing benefits of long-term retention and transfer of learning from perception to production (Lively et al., 1994; Bradlow and Pisoni, 1997; Bradlow et al., 1999; Sakai and Moorman, 2018). However, more recent phonetic training research has revealed some mixed results of HVPT training effects. On the one hand, a great number of studies showed that training with highly variable stimuli was more effective for L2 phonetic learning than low-variability training (Wang et al., 1999; Hardison, 2003; Brosseau-Lapré et al., 2013; Sadakata and Mcqueen, 2013; Zhang et al., 2018). On the other hand, other studies demonstrated that variability may hinder learning difficult L2 contrasts in some circumstances (Wayland and Guion, 2004; Wade et al., 2007; Chang and Bowles, 2015; Antoniou and Wong, 2016; Giannakopoulou et al., 2017). For example, Wade et al. (2007) showed that variability diminishes learning effects on highly confusable English vowels, relative to less confusable ones. Additionally, variability appears to diminish learning effects for novice learners (Wayland and Guion, 2004; Chang and Bowles, 2015), but it still is possible for novice learners to benefit from variability if they have strong perceptual abilities (Perrachione et al., 2011; Sadakata and Mcqueen, 2014; Antoniou and Wong, 2015). These results suggest that the effectiveness of HVPT partly depends on the learner characteristics and the nature of the to-be-learned categories.

One notable aspect in the recent HVPT studies was the efforts to optimize improvements by combining different training methods that are considered to have a positive influence on the underlying processes for phonetic perception. Given that highly variable input may initially present a challenge for the acquisition of difficult L2 sounds, the present study followed up our pilot work (Zhang and Cheng, 2011; Cheng and Zhang, 2013) to incorporate some key exaggerated characteristics of infant-directed speech into the HVPT paradigm. For the purpose of the current study, we chose to test training effects of the two training paradigms (HVPT without acoustic exaggeration versus the modified HVPT-E with exaggeration) on the perception of English /i/ and /ɪ/ by Chinese adults. IDS, also referred to as "motherese," is the exaggerated speech style used to address infants, with slower speaking rate, simplified language, exaggerated formant structure, frequent repetitions, and exaggerated prosody, which is presumed to help infants form speech categories (Fernald and Kuhl, 1987; Maye et al., 2002; Werker et al., 2007). However, not all aspects of IDS can facilitate speech learning or are intended to promote speech learning (McMurray et al., 2013). For instance, exaggerations in pitch range and contour do not necessarily facilitate phonetic categorization (Trainor and Desjardins, 2002; Kitamura and Burnham, 2003). Nevertheless, some studies have found that when the training input stimuli were systematically manipulated with temporal and spectral exaggeration, adult L2 perceptual learning could be greatly enhanced (Zhang et al., 2000, 2009; Iverson et al., 2005). The critical notion here is that by exaggerating acoustic differences between stimuli, researchers can make the contrast more discriminable, which can be incorporated in a scaffolding structure for step-by-step adaptive training based on the individual learners' perceptual ability to facilitate phonetic training (Zhang and Cheng, 2011). For example, Zhang et al. (2009) used a computer-assisted training program featuring acoustic exaggeration and multi-talker variability to train Japanese adults who had limited English exposure to learn the English /r/ and /l/ categories. The results showed that the trainees obtained significant improvement in identification with generalization to untrained synthetic and natural stimuli. As Iverson et al. (2005) and Zhang et al. (2009) did not include a training group without the use of acoustic exaggeration for comparison, the tentative claims about the important role of acoustic exaggeration in the HVPT paradigm have not been directly verified. The present study represents our first attempt to fulfill this gap to assess the contribution of acoustic exaggeration to Chinese adult learners' perception of the difficult English /i/-/ɪ/ contrast by comparing training effects of the canonical HVPT paradigm with naturally-produced input (i.e., input produced in multiple contexts by multiple speakers) and the modified HVPT paradigm with acoustically-exaggerated input delivered in an adaptive fashion.

Behavioral Measures of Training Effects

Researchers argue that exposure to high variability input helps learners ignore phonetically irrelevant information and retain long-term memory representations of relevant phonetic features, thus facilitating generalization from a trained set to a novel set (Deng et al., 2018). Although considerable training research has demonstrated training-induced improvement of accuracy in identification and discrimination performance, the training-induced improvements do not necessarily indicate that the leaners have formed more robust representations of L2 speech categories. For a stringent test of speech identification and discrimination performance and the effects of transfer of learning on retuning category representations, it is important to assess the learners' categorical perception (CP) in pre- and post-training tests with a synthetic speech continuum that was not used in the training protocol (Zhang et al., 2009; Zhang and Cheng, 2011).

Categorical perception is the phenomenon in which the categories possessed by an organism influences the organism's perception (Liberman et al., 1957; Eimas et al., 1971; Kuhl and Miller, 1978; Steinschneider et al., 1995; Zhang et al., 2005; Kuhl et al., 2006; Goldstone and Hendrickson, 2010). As such, perception is "warped" in the way that differences between objects belonging to different categories are amplified, while differences between objects falling into the same category are deemphasized. Thus it is by CP that the perceptual system transforms relatively linear sensory signals into non-linear internal representations. Although there appears to be an innate basis for CP, the boundaries along acoustic continuua between the speech categories can be modified or even lost as a consequence of learning (Roberson et al., 2000). There is also strong evidence demonstrating that CP can be induced by learning alone (Goldstone, 1994; Livingston et al., 1998). Under a spatial metaphor with axes defined by values on relevant dimensions (e.g., Nosofsky, 1986), the perceptual space may undergo a "warping" process in categorization training by moving members of the same category closer together and thus rendering them to be less discriminable. Therefore, examining CP of the target L2 speech sounds with a synthetic continuum in pre- and post- tests can provide feasible fine-grained measures to assess the training effects, which would be a more direct index of learning-induced changes in category representations than simple performance improvement in identifying the naturally produced speech stimuli.

The prototypical pattern of CP is characterized by a phonetic boundary effect, which shows a sudden membership shift between two categories in the identification function and a distinct peak at the category boundry in the discrimination function. Thus CP is operationalized as poor within-category discrimination and significantly better across-category discrimination, despite the fact that the physical differences are equal (Liberman et al., 1957). Early studies reported the prototypical pattern of CP for consonants with enhanced discrimination for across-category pairs of stimuli and diminished discrimination for within-category paris of stimuli in the speech continuum, and some failed to demonstrate categorical effects for vowels (e.g., Liberman et al., 1957; Fry et al., 1962). More recent research, however, was able to show that categorical-like effects are not completely absent for vowels, but the pattern tends to be weaker for vowels than that of consonants (Minagawa Kawai et al., 2005; Altmann et al., 2014). Since learning experience fundamentally shapes the way the brain represents and processes the category structures of speech sounds (Livingston et al., 1998), it is predicted that successful L2 phonetic learning would ideally lead to more robust categorical representations and native-like CP behavior. As a consequence, one would expect to see the sudden shift of category memberships in the identification function as well as enhanced sensitivity at the category boundary without raising within-category sensitivity in the discrimination function.

Neurophysiological Measures of Training Effects

In addition to behavioral measures, the current study aimed to assess training-induced brain plasticity by measuring the mismatch negativity (MMN) to investigate neural mechanisms underlying non-native speech learning. Previous research showed that training can induce significant changes in neurophysiological responses, suggesting that the adult brain is not strictly bound by the native language neural commitment early in life and that memory traces can be developed in order to encode new phonetic representations (Tremblay et al., 1997; Zhang and Wang, 2007; Zhang et al., 2009; Deng et al., 2018). In this regard, the MMN response, identified as a component of event-related potentials (ERPs), is a valuable tool for investigating speech perception and the formation of memory traces for newly learned speech categories at the pre-attention level. The MMN is typically elicited within the oddball paradigm when a change occurs in a stimulus stream with the repetitive standard stimulus replaced by infrequent deviant stimulus differing slightly in various parameters from standard stimulus. No focused attention on detecting the stimulus change is required. The MMN quantification requires subtracting the ERP of the frequent standard stimulus from that of the rare deviant stimulus with a negative peak occurring in the time window of 100 to 250 ms following the onset of detectable acoustic change in the deviant stimulus (for a review, see Näätänen, 2001). Previous research has shown varying (i.e., from non-significant to significant) degrees of association between MMN responses and behavioral discrimination performance under different experimental conditions (e.g., Näätänen et al., 1993; Tiitinen et al., 1994; Kujala et al., 2001; Novitski et al., 2004; Chen and Sussman, 2013; Yu et al., 2017). There were also different findings regarding whether the MMN can reflect the category boundary effect in categorical perception (e.g., Ylinen et al., 2006; Xi et al., 2010). For instance, Ylinen et al. (2006) suggested that the status of the deviant relative to the phoneme boundary did not affect the MMN amplitude; that is, the MMN responses to within- and across- category differences did not differ regardless of whether the listeners demonstrated CP for the target sounds or not. On the contrary, in studying the neurophysiological correlates of categorical perception of Chinese lexical tones, Xi et al. (2010) found that the across-category deviants elicited larger MMN than that of the within-category deviants in native speakers of Chinese, and this phenomenon was not observed in the non-native speakers. Despite the controversies, speech research has indeed documented that the MMN can reflect learning-induced changes with enhanced MMN amplitude and reduced MMN latency (Menning et al., 2002; Kujala and Näätänen, 2010).

There are at least two advantages of using the MMN in a speech training study. First, the MMN is measured without requiring the participant's attention, which allows a more objective test of the perceptual changes at the pre-attentive level. Second, the MMN can index not only the final behavioral outcome, but also capture the significant progress on the way toward a successful outcome, since the MMN may emerge prior to behavioral manifestation of a significant change (Näätänen et al., 1982; Näätänen and Gaillard, 1983; Alho and Sinervo, 1997). If the MMN reflects training effects and corresponds to behavioral measures, it can be predicted that changes in the MMN amplitudes elicited by across-category differences from the pre-test to the post-test should be significantly larger, while the MMN elicited by within-category differences should not change significantly. This learning process may also be accompanied with shorter MMN latency for the across-category discrimination but not for the within-category discrimination.

Current Study

The current study adds to the speech training literature by examining the role of temporal acoustic exaggeration in the HVPT paradigm to train Chinese adults to learn the English /i/-/ɪ/ vowel contrast. The English tense and lax front unrounded vowels (/i/-/ɪ/ contrast) have been shown to be exceptionally difficult for Chinese speakers. Acoustically, English /i/ has a lower first formant (F1) and higher second formant (F2) than /ɪ/, and /i/ is also typically longer than /ɪ/ in duration. Unlike English, Mandarin Chinese has only one single category of /i/ which is likewise characterized by its low F1 and high F2 (Hillenbrand et al., 1995). According to the Perceptual Assimilation Model (Best, 1995), the fact that the two categories of English vowels /i/-/ɪ/ assimilate into the single Mandarin Chinese /i/ category predicts the difficulty for Chinese adults to learn the contrast. Empirical research evidenced this difficulty. Native English speakers can use both duration and formant frequency cues for distinguishing the two sounds (Mermelstein, 1978; Whalen, 1989; Grenon et al., 2019), and they predominantly rely on the spectral cues (Mermelstein, 1978; Hillenbrand et al., 2000). On the contrary, at least at the initial learning stage, English-as-a-second-language (ESL) learners rely dominantly on duration cues rather than spectral cues in both perception and production (Escudero and Boersma, 2004; Morrison, 2005; Wang and Heuven, 2006; Yang, 2011; Liu et al., 2014).

The HVPT approach with acoustic modification has been found to be effective in changing L2 cue weighting of non-native vowel perception. For example, Ylinen et al. (2010) used the HVPT technique with modified acoustic stimuli to equate durations between vowels to train adult Finnish native speakers who relied more on duration to identify the English /i/-/ɪ/ contrast before training, thus forcing the listeners to use spectral cues. Results showed significant improvement in identifying both natural and duration-modified stimuli. Giannakopoulou et al. (2013) replicated the effects with training native Greek ESL learners to distinguish the same contrast. However, previous research has highlighted the limiting role of the L1 background, indicating the necessity to test the efficacy of HVPT for L2 learners with different L1s (Iverson and Evans, 2009; Grenon et al., 2019). In addition, previous studies have established the positive effects of talker variability in facilitating second language phonetic learning (e.g., Logan et al., 1991; Hardison, 2003; Zhang et al., 2009). The present study continued this line of work with an aim to investigate the efficacy of the modified HVPT paradigm combined with temporal acoustic exaggeration to enhance Chinese L2 learners' reliance on spectral cues of the /i/-/ɪ/ contrast as measured by behavioral CP effects and MMN responses on the duration-equated stimuli. In particular, we were interested in examining differences in three measurable outcomes in pre- and post- tests, i.e., (1) whether the two training paradigms would show generalization to novel words and new talkers, (2) whether the learner would develop more native-like categorical perception of the /i/-/ɪ/ contrast based on spectral cues, and (3) whether the training paradigms would produce the predicted MMN changes for non-native phonetic learning.

To this end, we embedded the phonetic learning task in word stimuli, in which training requires identifying minimal pair contrasts that are synchronized with visual cues of the speakers' articulatory motion in video clips on screen (Zhang and Cheng, 2011). We compared training effects of HVPT-E with acoustically-exaggerated input versus HVPT without acoustically-exaggerated input, with overall exposure trials matched across conditions. Our measures of training effects in the pre- and post- tests were threefold: (1) identification of the target vowels in naturally spoken words to measure training and generalization effects; (2) identification and discrimination of synthetic stimuli for categorical perception tests based on spectral cues alone to measure changes in perceptual sensitivity for across- and within- category discrimination; (3) the MMN responses that correspond to the training-induced changes for across- and within- category discrimination at the pre-attentive level.

MATERIALS AND METHODS

Participants

Sixty native Chinese speakers at Xi'an Jiaotong University participated in this study. Participants were volunteers aged between 18 and 36 years old. They were recruited for participation following the approval and guideline of Institutional Review Board for Biomedical Research at Xi'an Jiaotong University, China. Written informed consent was obtained from each participant with hourly monetary compensation for their participation. The participants were all right-handed and had no history of speech, language or hearing problems or disorders. All of them had studied English for at least 6 years prior to attending college and were taking English courses at the university. No one had the experience of living in an English-speaking country or community for over a month. Participants were randomly assigned to three subject groups, 20 in the HVPT group, 20 in the HVPT-E group, and 20 in a control group that did not receive training.

Stimuli

In both HVPT and HVPT-E paradigms, 40 different tokens (20 minimal pairs) were trained. A total of 840 trials were included, with 120 trials in each session (7 sessions). In the canonical HVPT paradigm, training stimuli were naturally spoken English words containing the target phonemes of /i/ and /ɪ/. Six native American English speakers (3 males and 3 females) participated as talkers for the training stimuli. Digital video (with audio track) recordings of four native speakers (2 females, 2 males) were used as visual cues in the training program, and the other two (audio track only) were adopted in the progress-monitoring quizzes following each training session. In the modified paradigm (HVPT-E), the natural productions were further synthesized in Praat to be acoustically modified with four levels of temporal exaggeration: 300, 208, 144 and 100% (with no exaggeration). The video frames of each word token in the training program were edited in Final Cut Studio to match the duration at the four levels.

The stimuli adopted in the pre- and post- test included natural word stimuli and synthetic target phoneme stimuli. The natural word stimuli were used in the natural word identification test, recorded by four native American English speakers, two males and two females, all of whom were new to the trainees. Altogether, 160 natural word stimuli were used (20 words × 4 talkers × 2 times). Among the twenty words, 10 words (5 minimal pairs) were selected from the training words and other 10 words (5 minimal pairs) were untrained words.

Synthesized phoneme stimuli were an eleven-step continuum between /i/ and /ɪ/, previously used in the identification and discrimination tests as well as the electrophysiological test to examine categorical perception in native speakers and non-native speakers of English (Cheng and Zhang, 2013). The /i/ and /ɪ/ sounds were first recorded in the "h_d" context in a word list read by a male native English speaker at 44.1 kHz sampling rate and then were digitally processed (overlap-add method) to have an equal duration of 170 ms and fade-in and fade-out time of 10 ms using Sound Forge (SoundForge9, Sony Corporation, Japan). For the /i/ sound, the F1 and F2 frequencies are 355 and 2346 Hz, respectively. For the /ɪ/ sound, F1 and F2 were 435 and 2006 Hz. These two stimuli were then employed as the two endpoints to create the eleven-step continuum using a morphing technique in the STRAIGHT package (Kawahara et al., 1999) on the MATLAB platform (Mathworks Corporation, United States). The stimuli on the continuum were normalized to have equal average RMS (root mean square) intensity.

Procedures

Pre/Post-Tests

The sixty participants were randomly assigned to three groups, two training groups with 20 in the HVPT group and 20 in the HVPT-E group and one control group with 20 participants. Identical tests were conducted one week before and one week after training. To verify that participants' perceptual performance of the /i/-/ɪ/ contrast in three groups were not significantly different before training, we conducted one-way ANOVA for the behavioral data in the pre-test. No significant differences between three groups were found in any of the natural word identification [$F(2,57) = 0.882$, $p = 0.419$, $\eta^2_p = 0.030$ for the overall natural word identification], synthetic phoneme identification [$F(2,57) = 0.494$, $p = 0.613$, $\eta^2_p = 0.017$ for the boundary slope; $F(2,57) = 0.626$, $p = 0.538$, $\eta^2_p = 0.021$ for the boundary location; $F(2,57) = 0.286$, $p = 0.752$, $\eta^2_p = 0.010$ for the boundary width] and discrimination [$F(2,57) = 0.027$, $p = 0.973$, $\eta^2_p = 0.001$ for the across-category pair discrimination; $F(2,57) = 2.192$, $p = 0.121$, $\eta^2_p = 0.071$ for the within-category pair discrimination] tests. The MMN data were additionally collected from the HVPT-E group before and after training.

The EEG recording for the HVPT-E group was administrated in an electronically and acoustically shielded room. Continuous EEG data were recorded (sampling rate = 500 Hz; bandwidth = 0.3–30 Hz) using the Net Station System with Net 400 amplifier and a 64-channel Geodesic Sensor Net cap (EGI Inc., United States). All the 64 electrodes had impedances below 5 kΩ during recording. The default setting in the EGI system was used with the vertex electrode Cz as the reference. The participants were instructed to ignore the presented sounds of /i/-/ɪ/ continuum while watching a self-selected muted movie on the 20-inch TV monitor placed three meters away from the participant. The stimuli were 3, 7, and 11 from the 11-step synthesized speech continuum. Those stimuli were chosen based on our previous cross-language pilot study (Cheng and Zhang, 2013). In Cheng and Zhang (2013), Step 3 was 100% identified by 10 native speakers of American English as the phoneme /i/ while Step 7 and Step 11 were identified as the phoneme /ɪ/ with accuracy rates of 96.5 and 100%, respectively. Stimulus presentation followed the Double Oddball Paradigm in Xi et al. (2010), which was implemented in Eprime 2.0 (Psychology Software Tools Inc., United States). In this paradigm, an MMN response can be elicited when the repeated presentation of the standard stimulus (Step 7) is interrupted by either deviant stimulus (Step 3 or Step 11). An across-category stimulus pair (Steps 3 and 7) and a within-category stimulus pair (Step 7 and 11) were chosen as two contrasts which occurred pseudo-randomly; that is, the deviant stimuli were designed not to be presented consecutively. The standard stimulus occurred at the frequency of 80% (960 trials) of all trials and the two deviant stimuli occurred each at 10% (120 trials for each deviant). The inter-stimulus interval ranged from 700 to 800 ms. Sound stimuli were presented binaurally at 75 dB SPL via Etymotic Research ER-1 Insert Earphones. The EEG recording session lasted approximately 30 min (excluding the preparation time).

In order to test training and generalization effects to new talkers and new phonetic contexts (new minimal pair words) in which the target phonemes occur, the natural word identification task used 10 trained words and 10 untrained words, all produced by four new native American English speakers. Each word was presented eight times. In the test, participants were asked to judge whether the word they heard included /i/ or /ɪ/ by clicking the icons of the two phonemes on the screen which were represented in International Phonetic Alphabet (IPA). All participants had learned these IPA symbols as part of their English curriculum since middle school. The experimenter verified that each participant recognized the /i/ and /ɪ/ symbols correctly in association with a target minimal pair "beat" vs. "bit" prior to training.

Both identification and discrimination tasks were used for CP tests with the synthetic stimuli. The identification test required participants to label vowel stimuli. Each stimulus (the eleven steps) from the continuum was randomly presented 20 times. The discrimination test required participants to judge whether pairs of presented stimuli were the same or different. Two different sound pairs (across-category pair: Step 3 and Step 7; within-category pair: Step 7 and Step 11) were presented in random order at 10 times, respectively. For each of the two sound pairs, there were four types in the form of AB, BA, AA, and BB, with AB and BA representing different phonemes in the reverse order and AA and BB representing foil trials with the same phoneme.

Training Procedures

As described earlier, each trainee was randomly assigned to one of the two training groups: the HVPT group and the HVPT-E group. The only difference in the two training paradigms was the training input with or without acoustic exaggeration, with all other variables matched across conditions. Both training protocols included 7 sessions (**Table 1**). It took the HVPT-E group approximately 60 to 90 min to complete training (including the quizzes in between training sessions), and the HVPT group 50 to 80 min, depending on each participant's pace and response speed.

TABLE 1 | Talker numbers and stimuli in each session of the training program.

Session	Talker Numbers	Stimuli	
		HVPT-E	HVPT
Session 1	1	exaggerated 300%	natural
Session 2	2	exaggerated 300%	natural
Session 3	3	exaggerated 300%	natural
Session 4	4	exaggerated 300%	natural
Session 5	4	exaggerated 208%	natural
Session 6	4	exaggerated 144%	natural
Session 7	4	No exaggeration 100% (natural)	natural

The participants in the two training groups were asked to complete the training at their own pace in a sound-treated booth. By clicking either one of the target phoneme icons on the screen, the trainee saw a talker utter a word containing the clicked phoneme in the visual video in the center of the screen (Zhang and Cheng, 2011). For each icon, there were 60 words containing the phoneme ready to be clicked. Trainees may click either icon of phoneme on the screen randomly. After a total of 120 clicks, a progress quiz of 10 words was held. If the accuracy rate was above 90%, the trainee would automatically move on to the next session. If not, they were required to do the training session one more time followed by the quiz. After the repeat session, the trainees could elect to move on to the next level even though the accuracy rate failed to reach 90%.

Data Collection and Analysis

Behavioral Data

In the natural word identification, percent correct accuracy of all tested words (including the trained words and the untrained words) and the untrained words pre-and post-training were compared using repeated measures ANOVA to examine whether there was significant training-induced improvement and generalization. Further simple effect tests and *post hoc* two-tailed t-tests (if needed) were conducted to verify the effects with each group. In analyzing the identification performance for the synthetic continuum, the boundary location, slope and width were calculated with a probit model to fit the individual identification curve (Finney, 1971). The location of the boundary was defined as the point where half the trials of the stimuli were identified as /i/, and the other half was identified as /ɪ/. The boundary width was calculated as the distance between the 25th and 75th percentiles in the fitted identification curve with the probit analysis (Hallé et al., 2004). In order to account for the asymptotic property of the probit model, 0.1 replaced 0% and 99.9 replaced 100% in the individual identification data points. The boundary slope, location and width in the pre- and post- tests were compared between pre- and post- tests using repeated measures ANOVA (group × training) to determine whether training led to significant changes in the identification function. The discrimination data of synthetic continuum was analyzed in terms of percent correct accuracy (Xu et al., 2006). There were four types of comparisons for any given pair of sounds in the form of AB, BA, AA and BB, where stimuli A and B were separated by four steps on the continuum. The step pairs AB and BA included different phonemes while the pairs AA and BB were foil trials with the same phoneme. Discrimination accuracy (P) for each sound pair was determined by the formula of P = P("S"/S) × P(S)+P("D"/D) × P(D), with P("S"/S) representing the percentage of correct "same" responses to the "same" pairs (AA or BB), and P("D"/D) representing the percentage of correct "different" responses to the "different" pairs (AB or BA).

EEG Data

The MMN response was the component of interest in the ERP experiment to measure within- and across- category discriminatory sensitivity. The offline ERP analysis was conducted with BESA (Version 6.1, MEGIS Software GmbH, Germany). The artifacts of the raw data were first corrected to minimize influences of horizontal and vertical eye movements. The auto-correction parameters for horizontal electrooculogram (HEOG) and vertical electrooculogram (VEOG) were 100.0 and 150.0 μV, respectively. Trials with amplitudes beyond ±50 μV were rejected. The bandpass filter was at 0.5–30 Hz, and the data were re-referenced with common average reference. The ERP epoch length was set at 700 ms, including a pre-stimulus baseline of 100 and 600 ms after the onset of the stimulus.

After pre-processing, the MMN responses were derived by subtracting the ERPs for the standard stimulus from the ERPs for the deviant stimuli. Three electrode sites were defined with grouped electrodes in order to improve signal to noise ratio of the data. Left-site electrodes included F3, FC3 and C3, the mid-site electrodes included Fz, FCz, and Cz, and the right-site electrodes included F4, FC4, and C4. The electrode grouping was based on visual inspection of the topographical potential distribution of the MMN responses. A similar approach was used in previous studies (Zhang et al., 2011). The MMN amplitude for each electrode site was quantified by averaging data points within the time window of 40 ms around the MMN peak for each deviant on the individual subject basis. This adaptive window quantification was adopted in consideration of large inter-subject variability in the MMN peak latencies. Individuals' peaks were looked for 100–300 ms after the stimulus onset. At least 100 accepted deviant trials after artifact correction (120 trials in total) were included in each condition for each subject. For the MMN data collected from the HVPT-E group, repeated measures ANOVA test was conducted to examine effects of training (pre-test vs. post-test), stimulus type (across-category vs. within-category), and electrode site (left, mid, and right). In case of multiple comparisons (e.g., electrode site) in the ANOVA test, the reported *p*-values were Greenhouse-Geisser corrected.

RESULTS

Behavioral Results

Natural Word Identification

In natural word identification, the percent correct scores of the three participant groups before and after training were compared using repeated measures of ANOVA with training as a within-subject factor and group as a between-subject factor. As expected, there was a significant group effect in the learning outcomes [$F(2,57) = 5.732$, $p = 0.005$, $\eta^2_p = 0.167$]. There was also a significant group × training interaction effect [$F(2,57) = 5.065$, $p = 0.009$, $\eta^2_p = 0.151$]. Simple effect tests revealed that the identification accuracy of the HVPT-E group significantly improved after training with an increase of 9.7% ($p < 0.001$) (**Figure 1**). The HVPT group showed a significant improvement in overall natural word identification accuracy as well ($p = 0.003$), with an increase of 5.2%. In contrast, the control group before and after training did not change significantly ($p = 0.315$). Further comparison between the two training groups showed that the HVPT-E group showed greater improvement as reflected in larger progress in natural word identification performance produced by new talkers ($p = 0.021$).

Transfer of learning was examined with untrained word identification (**Figure 1**). Notably, equivalent amount of gain in the pre- and post- tests was found in the HVPT-E group with an increase of 9.2%. Repeated measures ANOVA showed significant main effects for training [$F(1,57) = 17.056$, $p < 0.001$, $\eta^2_p = 0.230$] and group [$F(2,57) = 3.372$, $p = 0.041$, $\eta^2_p = 0.106$]. Further comparisons showed significant differences between the HVPT-E group and the HVPT group ($p = 0.038$) as well as between the HVPT-E group and the control group ($p = 0.022$).

Identification of the Synthetic /i/-/ɪ/ Continuum

Figure 2 depicts the identification performance of the three groups for the /i/-/ɪ/ continuum in the pre- and post- tests. The boundary slope, location and width in the pre- and post-tests of three groups were compared with repeated measures ANOVA test. Significant group × training interaction effects were found in the boundary slope [$F(2,57) = 6.213$, $p = 0.004$, $\eta^2_p = 0.179$]. Simple effect test results showed that significantly steeper boundary slopes were observed after training in the HVPT-E group ($p < 0.001$) and the HVPT group ($p = 0.004$), in contrast with no significant changes in the control group ($p = 0.313$). Regarding the boundary width, there was a significant main effect for training [$F(1,57) = 7.317$, $p = 0.009$, $\eta^2_p = 0.114$]. But there were no significant changes in the boundary location before and after training. As the synthetic stimuli were controlled for duration, both the HVPT-E and the HVPT group demonstrated training-induced effects on synthetic phoneme identification solely based on spectral cues, as reflected by significantly steeper boundary slopes after training. Notably, there was a significant group effect in the slope change [$F(2,57) = 7.034$, $p = 0.002$, $\eta^2_p = 0.198$]. Further comparison between the two training groups showed that the HVPT-E group had larger training-induced improvement than the HVPT group

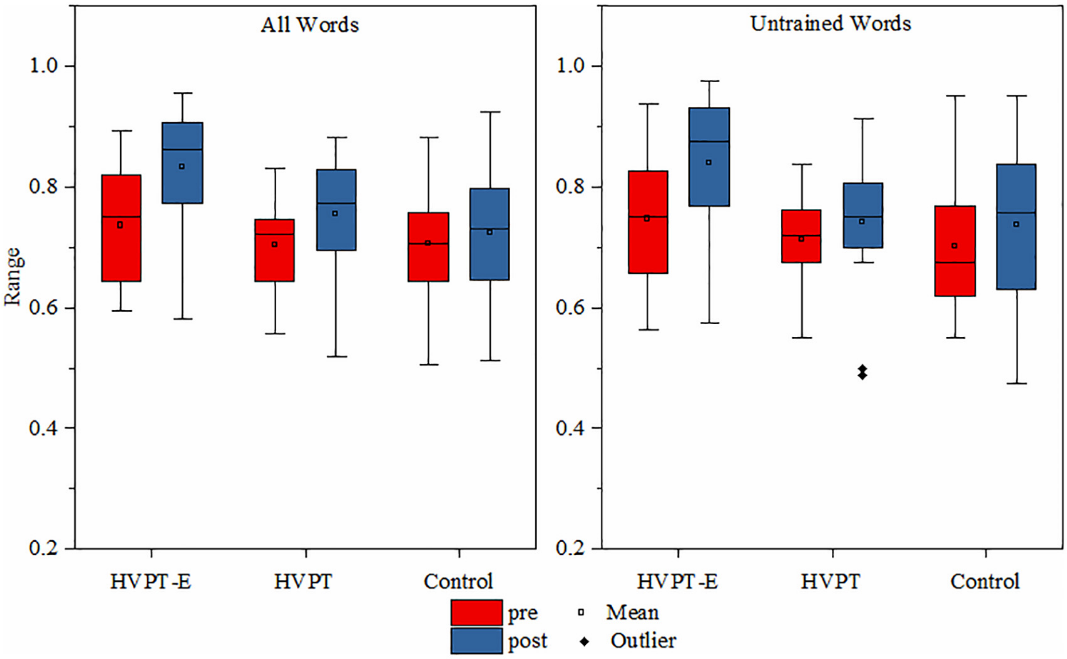

FIGURE 1 | Boxplots for natural word identification of the three subject groups in the pre- and post- tests.

with more robust boundary effect, as indicated by the steeper slope in the identification function ($p = 0.007$).

Discrimination of Across- and Within- Category /i/-/ɪ/ Stimuli

The percent correct accuracy data of the across-category pair (Step 3 vs. Step 7) and the within-category pair (Step 7 vs. Step 11) discrimination in the pre- and post- tests were compared using repeated measures ANOVA (group × training) (**Figure 3**). A significant group × training interaction effect was found in the across-category pair discrimination [$F(2,57) = 3.973$, $p = 0.024$, $\eta^2_p = 0.122$]. Simple effect test revealed that only the HVPT-E group had significant improvement after training in discriminating across-category stimuli ($p = 0.001$). In stark contrast, no significant pre-post change was observed in the HVPT group ($p = 0.861$) or the control group ($p = 0.888$). Contrary to our expectation, the within-category pair discrimination data also showed a significant main effect for training [$F(2,57) = 4.366$, $p = 0.041$, $\eta^2_p = 0.071$].

ERP Results From the HVPT-E Group

Figure 4 depicts the grand-average ERP waveforms of 20 participants in the HVPT-E group for the across-category deviant, the within-category deviant, and the standard in the pre- and post- tests. In the MMN amplitude data (**Figure 5**), repeated measures ANOVA revealed significant effects for the main factors of training (pre vs. post) [$F(1,19) = 5.278$, $p = 0.033$, $\eta^2_p = 0.217$], deviant type (across vs. within) [$F(1,19) = 13.447$, $p = 0.002$, $\eta^2_p = 0.414$], and electrode site (left, mid, right) [$F(2,38) = 7.826$, $p = 0.002$, $\eta^2_p = 0.356$].

In the MMN latency data (**Figure 6**), repeated measures ANOVA showed a significant main effect of training [$F(1,19) = 4.678$, $p = 0.044$, $\eta^2_p = 0.196$] and a significant interaction effect of deviant type and training [$F(1,19) = 5.407$, $p = 0.031$, $\eta^2_p = 0.222$]. Simple effect tests further showed that the latency of the across-category deviant did not significantly change in the pre-post comparison across the left, mid and right electrode sites ($p = 0.680$). However, the MMN peak latency for the within-category deviant was significantly delayed after training when compared with that of the pre-test [$F(1,19) = 8.288$, $p = 0.010$, $\eta^2_p = 0.198$].

DISCUSSION

The current study tested the role of temporal acoustic exaggeration by comparing the training effects of the modified HVPT paradigm integrating the characteristic of temporal exaggeration in IDS with that of the canonical HVPT design. Forty native Chinese adult learners of English were assigned to two training groups in the self-paced training program that included seven sessions lasting approximately 50 to 90 min, and another 20 subjects served as the control group with no training. Twenty out of the 40 trainees heard naturally-produced words containing the target phonemes (e.g., *sheep* vs. *ship*) by four native American English talkers during the training (the HVPT group) while the other 20 participants heard acoustically-exaggerated words by the same four talkers during the training (the HVPT-E group) with an adaptive training design. The word types, presentation order and frequencies were matched across conditions. We administered pre- and post- tests to

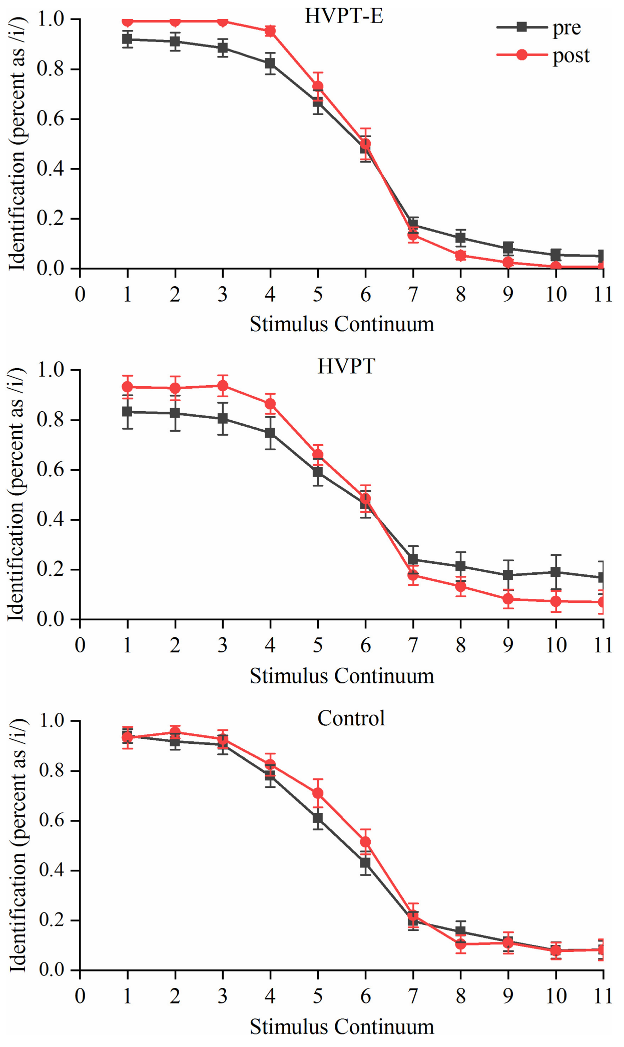

FIGURE 2 | Identification functions of the /i/-/i/ synthetic continuum for the three groups in pre- and post- tests. The data points in the plots represent means + standard error (SE).

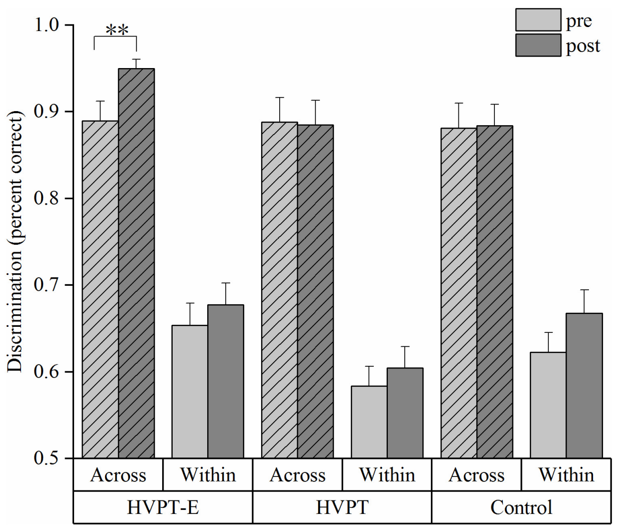

FIGURE 3 | Discrimination accuracy of across- and within- category stimulus pairs by the three groups. The bar plots represent means + standard error (SE) (***p* < 0.01).

assess training effects, including a natural word identification test, synthetic phoneme identification and discrimination tests, and an ERP test for the HVPT-E group. We predicted improvements both in the HVPT-E and HVPT group with greater improvement in the HVPT-E group, given the literature on benefits of acoustic exaggeration and HVPT in enhancing non-native phonetic learning (Logan et al., 1991; Lively et al., 1993; Iverson et al., 2003; Zhang et al., 2009). Consistent with our prediction, the behavioral results showed the advantage of the HVPT design. Both the HVPT group and the HVPT-E group showed significant improvement in naturally spoken words identification produced by new talkers as well as significant changes in synthetic phoneme identification based on spectral cues. Critically, data of the HVPT-E group revealed greater amount of improvement in words identification produced by new talkers and effective generalization to new words (i.e., new phonetic contexts). Further, the HVPT-E group exhibited more native-like CP reflected in both synthetic phoneme identification and discrimination behavior in the utilization of spectral cues. Thus consistent with our previous pilot study (Cheng and Zhang, 2013), the results highlighted the benefit of temporal acoustic exaggeration and adaptive training in training Chinese adults to perceive the English /i/-/ɪ/ vowel contrast.

Consistent with previous research (e.g., Lively et al., 1993), our behavioral data showed that the both training approach (i.e., HVPT and HVPT-E) improved the performance of Chinese adults in perceiving the English /i/-/ɪ/ vowel contrast, which generalized to new talkers. The small amount of gain in the HVPT group suggests that other factors such as learners' perceptual abilities, L1 backgrounds, and the nature of speech categories to be learned that may limit the efficacy of the HVPT paradigm (Iverson and Evans, 2009; Perrachione et al., 2011; Sadakata and Mcqueen, 2014). For example, Wade et al. (2007) suggested that benefits of HVPT varied across vowel categories, and could disappear for highly confusable vowels. The English /i/-/ɪ/ distinction is reported to be generally hard for a great number of L2 learners regardless of amount of L2 experience (Bohn and Flege, 1997; Flege et al., 1997; Cebrian, 2006; Kondaurova and Francis, 2008; Escudero et al., 2012). Additionally, according to the PAM, the English /i/-/ɪ/ distinction could be particularly hard for Chinese adults because they are both assimilated into a single Mandarin Chinese /i/ category (Best, 1995). Thus, the limited effectiveness of HVPT on Chinese learners' improvement and generalization is not utterly surprising.

The greater learning effects in the HVPT-E group provided evidence for the need to incorporate the temporal exaggeration in the HVPT paradigm to aid the participants in perceiving the non-native /i/-/ɪ/ contrast. It has been argued that learners' ability to transfer what they have learned about the contrast distinction

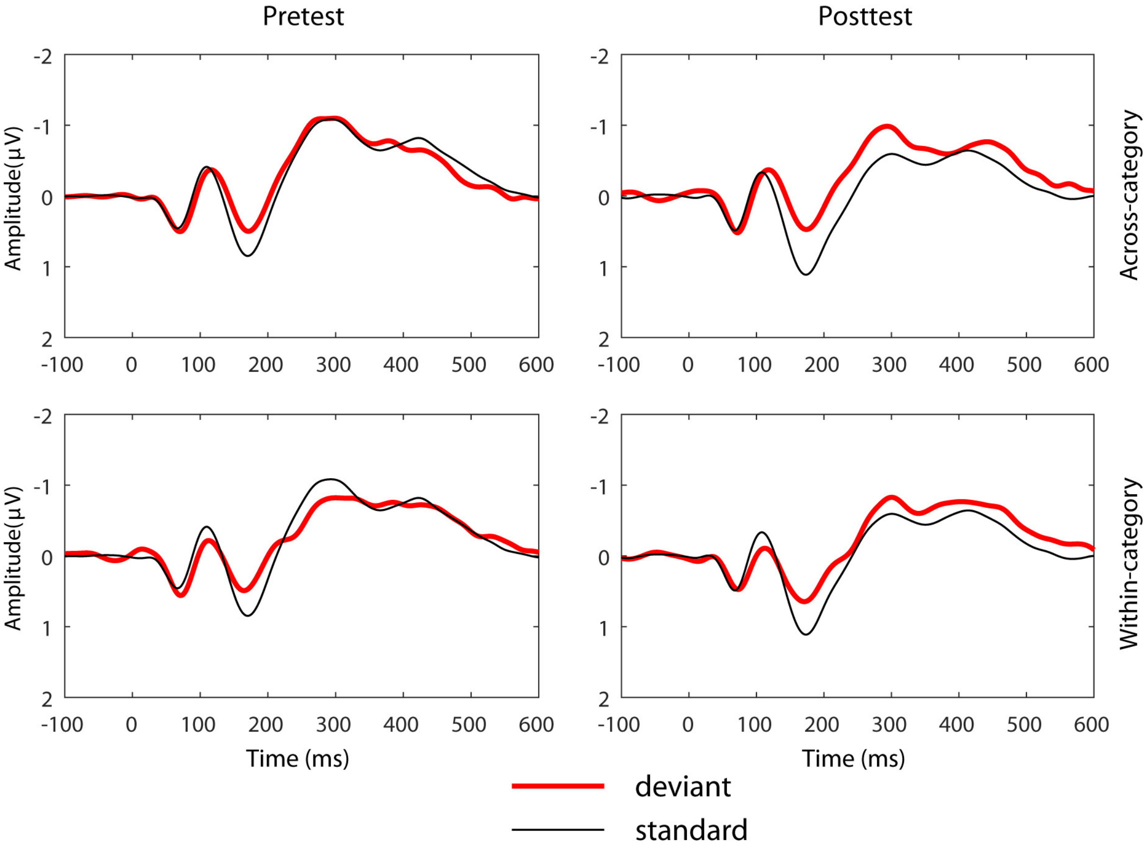

FIGURE 4 | Pre- and post- training grand mean ERP waveform data at the mid-central site for the across- and within- category contrasts for the HVPT-E group.

to other new stimuli is a result of extracting category-relevant information from training input and developing phonologically constant categorical representations, which can accommodate and be facilitated by a great range of exemplars (e.g., Logan et al., 1991). Developing categorical representations for speech sounds need to involve changing the perceptual weighting of different acoustic cues that contrast the phonetic categories, especially when certain cues critical for native speakers may be weighted as secondary by L2 learners of that language (Bion et al., 2006; Zhang et al., 2009; Ylinen et al., 2010; Giannakopoulou et al., 2013). However, the specific learning mechanism underlying category acquisition may not be well captured by tests with naturally spoken stimuli which have rich redundancy of acoustic cues to indicate phonemic contrasts. For instance, Bion et al. (2006) examined both perception and production of English front vowels by seventeen Brazilian L2 learners. On perception tests with natural speech stimuli, many subjects got high scores comparable to those of native English speakers, whereas these L2 learners' performance was much lower than the native English speakers in tests with synthetic stimuli that controlled vowel duration length and varied only in spectral cues. It is interesting to note that the pretest results for natural and synthetic stimuli in our study showed a reversed pattern with the relatively higher difficulty with the naturally spoken words. In particular, the literature has well documented that to distinguish the /i/-/ɪ/ contrast, native English speakers rely predominantly on the spectral cues (Mermelstein, 1978; Hillenbrand et al., 2000). In contrast, L2 learners of English, Chinese included, tend to make their judgment based on duration cues (Grenon et al., 2019). Therefore, although the two trained groups showed significant improvement in natural word identification, it is hard to determine whether these trainees learned to use primary spectral cues to identify the target vowels without looking at the evidence from the identification and discrimination data from the duration-controlled synthetic stimuli.

Further evidence on whether the trainees had formed more robust abstract tense and lax categories comes from assessment of the learners' categorical perception that taps into a fine-grained examination of transfer of learning and training-induced changes in the utilization of the spectral cues. In the CP tests in which only spectral cues were available, both the HVPT group and the HVPT-E group showed significantly steeper slope after training, revealing a more abrupt membership shift between the two categories of /i/-/ɪ/. However, only the HVPT-E group further demonstrated a significant training-induced improvement in the across-category discrimination accompanied by no significant change in the within-category discrimination, indicating enhanced sensitivity at the category boundary in the HVPT-E group. Like the control group, the HVPT group showed no significant pre-post changes in the across- and

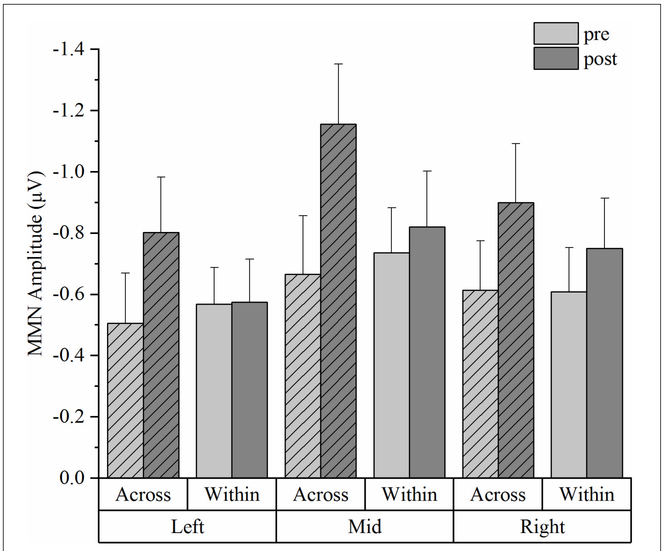

FIGURE 5 | Mismatch negativity (MMN) mean amplitudes from the HVPT-E group for the across- and within- category deviants in pre- and post- tests across the three electrode sites (the left site vs. the mid site vs. the right site). The bars represent means + standard error (SE).

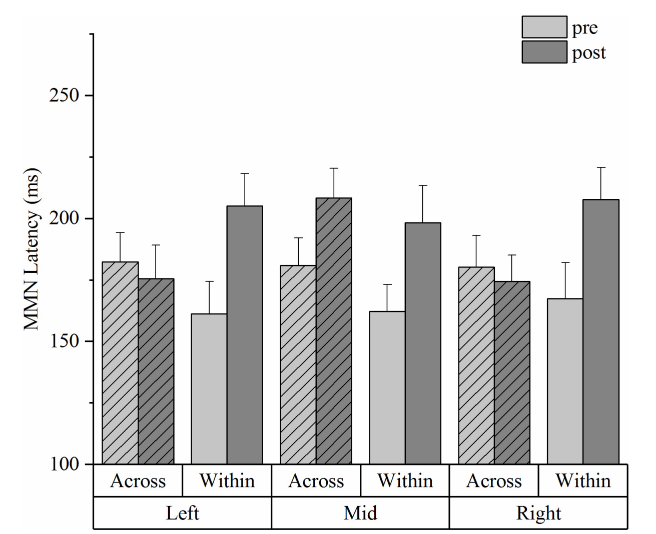

FIGURE 6 | Mismatch negativity mean latencies from the HVPT-E group for the across- and within- category deviants in pre- and post- tests across the three electrode sites (the left site vs. the mid site vs. the right site). The bars represent means + SE.

within- category discrimination data. The group differences in the pre- and post- test results with the synthetic stimuli showed that the HVPT method integrated with temporal acoustic exaggeration succeeded in improving the learners' ability to attend to and utilize the primary spectral features with more native-like categorical perception of the /i/-/ɪ/ contrast. As duration was strictly controlled in the synthetic stimuli, the significant training effects in the HVPT-E group reflected a true training-induced change in perceptual weighting of the spectral cues that is not attributable to enhanced sensitivity to the secondary duration feature.

Previous studies also reported on changing L2 cue weighting by modified HVPT techniques (Ylinen et al., 2010; Giannakopoulou et al., 2013; Grenon et al., 2019). For example, Ylinen et al. (2010) trained adult Finnish native speakers who relied on duration to identify the /i/-/ɪ/ contrast before training by using the HVPT with modified acoustic stimuli which had equal durations between the vowels. Their training data showed significant improvement in identification scores for both natural and duration-equated stimuli. More recent research by Hu et al. (2016), however, showed that phonetic training does not have to resort to the HVPT method. By controlling secondary cues in the input stimuli, vowel perception training in one single phonetic context produced by one single talker could also significantly change listeners' perceptual weighting strategy. Specifically, equalizing vowel duration in the training stimuli without HVPT successfully reduced native Chinese listeners' reliance on the duration cue and improved their use of spectral cues in identifying the English /i/-/ɪ/ vowels. This line of speech training studies supports the experience-driven "attention to dimension" (or "A2D") models of speech perception (McClelland, 2001; Francis and Nusbaum, 2002; Zhang et al., 2009), which consider perceptual learning as a process of specific changes in attentional distribution by reallocating the learners' attention on the relevant acoustic dimension which is critical for the L2 phonetic contrasts. For new phonetic categories to be learnt, the perceptual dimensions that are relevant to the category formation should be perceptually "stretched" while irrelevant dimensions should be "shrunk." Thus the underlying assumption is that the training experience with the absence of duration cues may force the listeners to pay attention to other perceptual cues (e.g., spectral cues) for non-native vowel perception (Ylinen et al., 2010; Giannakopoulou et al., 2013; Hu et al., 2016). In our modified HVPT-E approach, the varying levels of temporal exaggeration for both target vowels provided more variable range of duration than in the natural stimuli, which lead to native Chinese listeners' increased reliance on spectral cues for English /i-ɪ/ contrast. Although increased variability could be detrimental to efficient discrimination in that as clusters of exemplars increase in size, effective borders of the clusters will shrink or overlap (Eaves et al., 2016), we speculate that the introduction of irrelevant variability along duration dimension could encourage the learners to resort to other more stable cues (i.e., spectral cues) and achieve effects similar to those of the inhibitory training methods (Francis et al., 2000; Holt and Lotto, 2006; Kondaurova and Francis, 2010). For example, Kondaurova and Francis (2010) compared the effectiveness of three training methods (i.e., adaptive training with controlled duration, inhibition training with variable duration, and prototype training) on training native Spanish listeners to perceive English vowel contrast /i/-/ɪ/. They showed that inhibition training was more effective than the other two methods in terms of withdrawing attention from vowel duration. Assuming that reduced reliance on duration cues would make the weight of spectral cues relatively heavier than that of duration cues for perceiving the /i-ɪ/ contrast, it remains to be tested whether the modified HVPT by adding variance along irrelevant dimension could actually reduce learners' reliance on duration cues, thus helping Chinese learners reach the native-like preference of cue weighting in perceiving English vowel contrast /i/-/ɪ/.

Corroborating evidence from the ERP data in the HVPT-E group demonstrated significant training effects in the stage of pre-attentive cortical processing of the speech sounds, which are reflected by increased MMN amplitude for the across-category deviant after training and increased MMN latency for the within-category deviant after training. The training-induced MMN enhancement for the across-category deviant not only confirmed the behavioral training results of sharpened categorical perception but also demonstrated the important role of acoustic exaggeration in promoting neural plasticity for L2 phonetic category acquisition in adulthood. Our results are consistent with previous research showing enhanced MMN responses for training-induced improvement in speech perception (Kraus et al., 1995; Menning et al., 2002; Zhang et al., 2009). More importantly, the MMN enhancement effect in the pre-post comparison was observed only for detecting across-category differences (Winkler et al., 1999; Sharma and Dorman, 2000; Nenonen et al., 2003), which was accompanied by delayed MMN responses for detecting within-category acoustic differences.

One puzzling phenomenon is that while the post-test MMN results for across-category and within-category contrasts reflect native-like behavioral categorical perception in the HVPT-E group, the pre-test MMN data did not appear to show the same pattern consistent with the behavioral results because the pre-test MMN responses for the across-category contrast did not differ from those for the within-category contrast. According to Ylinen et al. (2006), the MMN component may not necessarily reflect the phoneme boundary effect but index prototypicality of the stimuli. We find this interpretation applicable to our pre-test MMN data. For the Chinese adult subjects, both steps 7 and 11 in the synthetic /i-ɪ/ continuum could be treated as non-prototypical /i/ sounds as the phoneme /ɪ/ does not exist in Mandarin Chinese. By contrast, Step 3 was heard as the /i/ sound, which exists in the Chinese vowel inventory. Thus, the so-called across-category contrast (Step 3 vs. Step 7) prior to training might be treated by the Chinese listeners at the pre-attentive level as acoustic difference between a prototypical vowel and a non-prototypical vowel whereas the within-category contrast (Step 7 and Step 11) might be treated as acoustic difference between two non-prototypical sounds. In this regard, the post-test MMN changes in HVPT-E group could also be viewed as fundamental training-induced changes in evaluating stimulus prototypicality relevant to the L2 phonemic contrast at the pre-attentive level, which would give rise to more native-like categorical perception results (Kronrod et al., 2016).

The MMN results in our study are also in accordance with previous evidence that cue weighting is language-specific in establishing long-term representations of phonetic categories. The enhanced MMN responses on detecting across-category differences in the duration-controlled synthetic vowel stimuli indicate that the training resulted in fundamental changes in cortical representations and automatic processing of the spectral cues that are important for the speech contrast. These categorical representations may experience repeated reinforcement from the training input in the form of attentional reallocation to the critical L2 features to become permanent. This is consistent with the Native Language Neural Commitment (NLNC) theory, which considers phonetic learning as an implicit self-reinforcing computational process. During L1 acquisition, the self-reinforcing process leads to neural commitment with increased sensitivity and efficiency to process the phonological patterns of the native language (Kuhl et al., 2008). This theory also claims that neural commitment can be reversible in adulthood with enriched exposure that can induce substantial plasticity for L2 learning. Improved post-training performance in both HVPT and HVPT-E groups and the MMN responses of the HVPT-E group in our study lend support to the theory by demonstrating substantial neuroplasticity in adulthood, which can be harnessed by proper treatment of the input acoustic properties and delivery mechanism. It is also important to note that there was a large scale of inter-participant variability in the MMN data. According to Kuhl et al. (2008), the degree of plasticity in L2 learning hinges on the stability of the underlying perceptual representations. Short periods of perceptual training in a laboratory setting might not be adequate to affect some participants' neural structures due to the instability of the phonetic representations for the L2 sounds and interference from their L1 phonetic representations. This calls for more fine-grained research to probe the differences in individual learners in a longitudinal design.

The behavioral data together with the MMN results provided strong evidence for the method of HVPT integrating temporal exaggeration to aid the learners in forming more native-like categories of the English /i/-/ɪ/ contrast, which has been shown in previous research (e.g., Zhang et al., 2009). However, the previous studies did not specifically separate the relative contributions of acoustic exaggeration and HVPT. To our knowledge, this is the first study to highlight the specific role of acoustic exaggeration in improving participants' perception of the /i/-/ɪ/ contrast, particularly in terms of categorical perception based on the critical spectral cues with or without attentional focus. Research has shown that acoustic exaggeration may enhance the cues distinguishing the linguistic features of the native language from very early in infancy (Ratner and Luberoff, 1984; Kuhl et al., 1997; Liu et al., 2003; Zhang et al., 2011). In adults, Uther et al. (2012) showed hyperarticulation of vowels elicited larger MMN response in both native and non-native speakers of English, suggesting that acoustic exaggeration could increase neural sensitivity to speech contrasts in second language learners, which may facilitate phonetic learning. Our study confirmed that acoustic exaggeration can be incorporated in training materials to help non-native adult learners establish more robust abstract sound categories along relevant dimensions. Additionally, it is worth noting that in contrast to the previous training studies mostly lasting long hours or even weeks (Bradlow and Pisoni, 1997; Callan et al., 2003; Zhang et al., 2009), the modified HVPT-E protocol required much less time to achieve equivalent amounts of gain. Presumably, the greater acoustic variety and characteristic details can facilitate the formation of prototypical representations of the phonetic category in both L1 and L2 acquisition (Kuhl et al., 2008). But it remains to be tested whether such L2 training effects are sustainable and transferable from perception to production in the long term. It is also unknown how generalizable the HVPT approach with acoustic exaggeration is for different types of L2 vowel and consonant contrasts and whether it is applicable to clinical populations such as those with severe hearing loss (Miller et al., 2016).

One inherent confound to the interpretation of the observed advantage in the HVPT-E group in our study is that temporal acoustic exaggeration provided more exposure time to the vowels as the HVPT-E stimuli were 200% longer than the HVPT stimuli in the beginning of the training program. However, it is noteworthy that according to the training records, the total training time for an average participant in the HVPT-E group was approximately 10 min more than the average person in the HVPT group. Given the L2 phonetic training literature, it seems unlikely that an extra 10 min of training itself would result in significant differences between the HVPT-E group and HVPT group. Further study could be designed to verify this speculative statement here by controlling the stimulus exposure time in the training sessions. A second limitation is that as the ERP data in the current study were only from the HVPT-E group, a full examination of the group differences in the ERP data could not be conducted to strengthen the findings at the neurophysiological level. A third important limitation of the current study is that unlike our previous work (Cheng and Zhang, 2013) with participants who did not major in English, the participants in the current report were all studying English in a highly selective university in China. They appeared to show relatively high (near-native) level of categorical perception before training, as indicated by much higher across-category discrimination accuracy compared to within-category discrimination in all three groups. Future studies can be conducted to include a wider range of individual L2 proficiency to determine the generalizability and effectiveness of the HVPT-E approach.

SUMMARY

This study provided direct evidence that high variability phonetic training with temporal acoustic exaggeration was more effective than the canonical HVPT. Both behavioral and electrophysiological data indicated significant training and generalization effects of the temporally-exaggerated HVPT on the Chinese ESL learners' perception of English /i/-/ɪ/ contrast. The results demonstrate great plasticity of non-native phonetic learning in adulthood induced by enriched input in a software training program, which has important implications for second language pedagogy as well as theories on language learning.

ETHICS STATEMENT

This study was approved from the Institutional Review Board for Biomedical Research at Xi'an Jiaotong University (IRB Code Number: 2018-553). Written informed consent was obtained from each participant.

AUTHOR CONTRIBUTIONS

BC and YZ conceived the study. BC, XZ, and YZ wrote the manuscript. All authors designed the study, collected and analyzed the data, read and approved the manuscript and agreed to be accountable for all aspects of the work.

FUNDING

The research was supported in part by grants from the National Social Science Fund of China (15BYY005). YZ additionally received support from University of Minnesota's Brain Imaging Grant to work on the project.

REFERENCES

Alho, K., and Sinervo, N. (1997). Preattentive processing of complex sounds in the human brain. *Neurosci. Lett.* 233, 33–36. doi: 10.1016/s0304-3940(97)00620-4

Altmann, C. F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., and Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia* 64, 13–23. doi: 10.1016/j.neuropsychologia.2014.09.006

Antoniou, M., and Wong, P. C. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *J. Acoust. Soc. Am.* 139, 271–278. doi: 10.1121/1.4939736

Antoniou, M., and Wong, P. C. M. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *J. Acoust. Soc. Am.* 138, 571–574. doi: 10.1121/1.4923362

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. W. Strange (Baltimore: MD: York Press).

Bion, R. A., Escudero, P., Rauber, A. S., and Baptista, B. O. (2006). "Category formation and the role of spectral quality in the perception and production of English front vowels," in *Proceeding of the Ninth International Conference on Spoken Language Processing*, Pittsburgh, PA, 1363–1366.

Bohn, O., and Flege, J. E. (1997). "Perception and production of a new vowel category by adult second language learners," in *Second-Language Speech: Structure and Process*, eds. A. James and J. Leather (Berlin: de Gruyter), 53–74.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Percept. Psychophys.* 61, 977–985. doi: 10.3758/bf03206911

Bradlow, A. R., and Pisoni, D. B. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101, 2299–2310. doi: 10.1121/1.418276

Brosseau-Lapré, F., Rvachew, S., Clayards, M., and Dickson, D. (2013). Stimulus variability and perceptual learning of nonnative vowel categories. *Appl. Psycholinguist.* 34, 419–441. doi: 10.1017/s0142716411000750

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., and Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage* 19, 113–124. doi: 10.1016/s1053-8119(03)00020-x

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *J. Phon.* 34, 372–387. doi: 10.1016/j.wocn.2005.08.003

Chang, C. B., and Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *J. Acoust. Soc. Am.* 138, 3703–3716. doi: 10.1121/1.4937612

Chen, S., and Sussman, E. S. (2013). Context effects on auditory distraction. *Biol. Psychol.* 94, 297–309. doi: 10.1016/j.biopsycho.2013.07.005

Cheng, B., and Zhang, Y. (2013). Neural plasticity in phonetic training of the /i-I/ contrast for adult Chinese speakers. *J. Acoust. Soc. Am.* 134:4245. doi: 10.1016/j.neuroimage.2017.01.042

Deng, Z., Chandrasekaran, B., Wang, S., and Wong, P. C. M. (2018). Training-induced brain activation and functional connectivity differentiate multi-talker and single-talker speech training. *Neurobiol. Learn. Mem.* 151, 1–9. doi: 10.1016/j.nlm.2018.03.009

Eaves, B. S. Jr., Feldman, N. H., Griffiths, T. L., and Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychol. Rev.* 123, 758–771. doi: 10.1037/rev0000031

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (1971). Speech perception in infants. *Science* 171, 303–306. doi: 10.1126/science.171.3968.303

Escudero, P., and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Stud. Second Lang. Acquis.* 26, 551–585. doi: 10.1017/S0272263104040021

Escudero, P., Simon, E., and Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: acoustic similarity predicts and explains cross-linguistic and L2 perception. *J. Phon.* 40, 280–288. doi: 10.1016/j.wocn.2011.11.004

Fernald, A., and Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behav. Dev.* 10, 279–293. doi: 10.1016/0163-6383(87)90017-8

Finney, D. J. (1971). *Probit Analysis*, 3rd Edn. Cambridge: Cambridge University Press.

Flege, J. E., Bohn, O., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowel. *J. Phon.* 25, 437–470. doi: 10.1006/jpho.1997.0052

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Percept. Psychophys.* 62, 1668–1680. doi: 10.3758/bf03212164

Francis, A. L., and Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 349–366. doi: 10.1037//0096-1523.28.2.349

Fry, D. B., Abramson, A. S., Eimas, P. D., and Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Lang. Speech* 5, 171–189. doi: 10.1177/002383096200500401

Giannakopoulou, A., Brown, H., Clayards, M., and Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ* 5:e3209. doi: 10.7717/peerj.3209

Giannakopoulou, A., Uther, M., and Ylinen, S. (2013). "Phonetic and orthographic cues are weighted in speech sound perception by second language speakers: evidence from Greek speakers of English," in *Proceedings of Meetings of the Acoustical Society of America*, Vol. 20, Winchester.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *J. Exp. Psychol. Gen.* 123, 178–200. doi: 10.1037//0096-3445.123.2.178

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical Perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 69–78. doi: 10.1002/wcs.26

Grenon, I., Kubota, M., and Sheppard, C. (2019). The creation of a new vowel category by adult learners after adaptive phonetic training. *J. Phon.* 72, 17–34. doi: 10.1016/j.wocn.2018.10.005

Hallé, P. A., Chang, Y. C., and Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *J. Phon.* 32, 395–421. doi: 10.1016/s0095-4470(03)00016-0

Hardison, D. M. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Appl. Psycholinguist.* 24, 495–522. doi: 10.1017/s0142716403000250

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099–3111. doi: 10.1121/1.411872

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some effects of duration on vowel recognition. *J. Acoust. Soc. Am.* 108, 3013–3022. doi: 10.1121/1.1323463

Holt, L. L., and Lotto, A. J. (2006). Cue weighting in auditory categorization: implications for first and second language acquisition. *J. Acoust. Soc. Am.* 119, 3059–3071. doi: 10.1121/1.2188377

Hu, W., Mi, L., Yang, Z., Tao, S., Li, M., Wang, W., et al. (2016). Shifting perceptual weights in L2 vowel identification after training. *PLoS One* 11:e0162876. doi: 10.1371/journal.pone.0162876

Huang, D., Yu, L., Wang, X., Fan, Y., Wang, S., and Zhang, Y. (2018). Distinct patterns of discrimination and orienting for temporal processing of speech and nonspeech in Chinese children with autism: an event-related potential study. *Eur. J. Neurosci.* 47, 662–668. doi: 10.1111/ejn.13657

Iverson, P., and Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am.* 126, 866–877. doi: 10.1121/1.3148196

Iverson, P., Hazan, V., and Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. *J. Acoust. Soc. Am.* 118, 3267–3278. doi: 10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57. doi: 10.1016/S0010-0277(02)00198-1

Jamieson, D. G., and Morosan, D. E. (1986). Training non-native speech contrasts in adults: acquisition of the English /delta/-/theta/ contrast by francophones. *Percept. Psychophys.* 40, 205–215. doi: 10.3758/bf03211500

Kawahara, H., Masuda-Katsuse, I., and De Cheveigne, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds1. *Speech Commun.* 27, 187–207. doi: 10.1016/s0167-6393(98)00085-5

Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/s15327078in0401_5

Kondaurova, M. V., and Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *J. Acoust. Soc. Am.* 124, 3959–3971. doi: 10.1121/1.2999341

Kondaurova, M. V., and Francis, A. L. (2010). The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods. *J. Phon.* 38, 569–587. doi: 10.1016/j.wocn.2010.08.003

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., and Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *J. Cogn. Neurosci.* 7, 25–32. doi: 10.1162/jocn.1995.7.1.25

Kronrod, Y., Coppess, E., and Feldman, N. H. (2016). A unified account of categorical effects in phonetic perception. *Psychon Bull. Rev.* 23, 1681–1712. doi: 10.3758/s13423-016-1049-y

Kuhl, P. K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11850–11857. doi: 10.1073/pnas.97.22.11850

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686. doi: 10.1126/science.277.5326.684

Kuhl, P. K., Conboy, B. T., Sharon, C. C., Denise, P., Maritza, R. G., and Tobey, N. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Kuhl, P. K., and Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *J. Acoust. Soc. Am.* 63, 905–917. doi: 10.1121/1.381770

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., and Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9, F13–F21. doi: 10.1111/j.1467-7687.2006.00468.x

Kujala, T., Kallio, J., Tervaniemi, M., and Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clin. Neurophysiol.* 112, 1712–1719. doi: 10.1016/s1388-2457(01)00625-3

Kujala, T., and Näätänen, R. (2010). The adaptive brain: a neurophysiological perspective. *Prog. Neurobiol.* 91, 55–67. doi: 10.1016/j.pneurobio.2010.01.006

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368. doi: 10.1037/h0044417

Liu, C., Jin, S., and Chen, C. (2014). Durations of American English vowels by native and non-native speakers: acoustic analyses and perceptual effects. *Lang. Speech* 57, 238–253. doi: 10.1177/0023830913507692

Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* 6, F1–F10. doi: 10.1111/1467-7687.00275

Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.* 94, 1242–1255. doi: 10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am.* 96, 2076–2087. doi: 10.1121/1.410149

Livingston, K. R., Andrews, J. K., and Harnad, S. (1998). Categorical perception effects induced by category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 732–753. doi: 10.1037/0278-7393.24.3.732

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report. *J. Acoust. Soc. Am.* 89, 874–886. doi: 10.1121/1.1894649

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn. Affect. Behav. Neurosci.* 2, 89–108. doi: 10.3758/cabn.2.2.89

McClelland, J. L. (2001). "Failures to learn and their remediation: a Hebbian account," in *Mechanisms of Cognitive Development: Behavioral and Neural Perspectives*, eds J. L. McClelland and R. Siegler (Mahwah, NJ: Erlbaum), 97–121.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., and McEchron, W. (2013). Infant directed speech and the development of speech perception: enhancing development or an unintended consequence? *Cognition* 129, 362–378. doi: 10.1016/j.cognition.2013.07.015

Menning, H., Imaizumi, S., Zwitserlood, P., and Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learn. Mem.* 9, 253–267. doi: 10.1101/lm.49402

Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *J. Acoust. Soc. Am.* 63, 572–580. doi: 10.1121/1.381756

Miller, S. E., Zhang, Y., and Nelson, P. B. (2016). Efficacy of multiple-talker phonetic identification training in postlingually deafened cochlear implant listeners. *J. Speech Lang. Hear. Res.* 59, 90–98. doi: 10.1044/2015_JSLHR-H-15-0154

Minagawa Kawai, Y., Mori, K., and Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *J. Cogn. Neurosci.* 17, 1376–1385. doi: 10.1162/0898929054985482

Morrison, G. S. (2005). An appropriate metric for cue weighting in L2 speech perception: response to Escudero and Boersma (2004). *Stud. Second Lang. Acquis.* 27, 597–606. doi: 10.1017/S0272263105050266

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, 1–21. doi: 10.1111/1469-8986.3810001

Näätänen, R., and Gaillard, A. W. K. (1983). The orienting reflex and the N2 deflection of the event-related potential (ERP). *Adv. Psychol.* 10, 119–141. doi: 10.1016/s0166-4115(08)62036-1

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., and Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport* 4, 503–506. doi: 10.1097/00001756-199305000-00010

Näätänen, R., Simpson, M., and Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biol. Psychol.* 14, 53–98. doi: 10.1016/0301-0511(82)90017-5

Nenonen, S., Shestakova, A., Huotilainen, M., and Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cogn. Brain Res.* 16, 492–495. doi: 10.1016/s0926-6410(03)00055-7

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57. doi: 10.1037//0096-3445.115.1.39

Novitski, N., Tervaniemi, M., Huotilainen, M., and Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cogn. Brain Res.* 20, 26–36. doi: 10.1016/s0926-6410(04)00032-1

Perrachione, T. K., Jiyeon, L., Ha, L. Y. Y., and Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *J. Acoust. Soc. Am.* 130, 461–472. doi: 10.1121/1.3593366

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 297–314. doi: 10.1037//0096-1523.8.2.297

Pruitt, J. S., Jenkins, J. J., and Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *J. Acoust. Soc. Am.* 119, 1684–1696. doi: 10.1121/1.2161427

Ratner, N. B., and Luberoff, A. (1984). Cues to post-vocalic voicing in mother–child speech. *J. Phon.* 12, 285–289.

Roberson, D., Davies, I., and Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *J. Exp. Psychol. Gen.* 129, 369–398. doi: 10.1037//0096-3445.129.3.369

Sadakata, M., and Mcqueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: evidence from Japanese geminates. *J. Acoust. Soc. Am.* 134, 1324–1335. doi: 10.1121/1.4812767

Sadakata, M., and Mcqueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Front. Psychol.* 5:1318. doi: 10.3389/fpsyg.2014.01318

Sakai, M., and Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Appl. Psycholinguist.* 39, 187–224. doi: 10.1017/s0142716417000418

Sharma, A., and Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *J. Acoust. Soc. Am.* 107, 2697–2703. doi: 10.1121/1.428655

Shinohara, Y., and Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *J. Phon.* 66, 242–251. doi: 10.1016/j.wocn.2017.11.002

Steinschneider, M., Schroeder, C. E., Arezzo, J. C., and Vaughan, H. G. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang.* 48, 326–340. doi: 10.1006/brln.1995.1015

Strange, W. (1995). "Phonetics of second-language acquisition: past, present, future," in *Proceedings of the 18th International Congress of Phonetic Sciences*, eds K. Elenius, and P. Branderud (Stockholm: KTH-Stockholm University), 76–83.

Strange, W., and Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Percept. Psychophys.* 36, 131–145. doi: 10.3758/bf03202673

Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372, 90–92. doi: 10.1038/372090a0

Trainor, L. J., and Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychon. Bull. Rev.* 9, 335–340. doi: 10.3758/bf03196290

Tremblay, K., Kraus, N., Carrell, T. D., and Mcgee, T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J. Acoust. Soc. Am.* 102, 3762–3773. doi: 10.1121/1.420139

Uther, M., Giannakopoulou, A., and Iverson, P. (2012). Hyperarticulation of vowels enhances phonetic change responses in both native and non-native speakers of English: evidence from an auditory event-related potential study. *Brain Res.* 1470, 52–58. doi: 10.1016/j.brainres.2012.06.041

Wade, T., Jongman, A., and Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica* 64, 122–144. doi: 10.1159/000107913

Wang, H., and Heuven, V. J. V. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. *Linguist. Neth.* 23, 237–248. doi: 10.1075/avt.23.23wan

Wang, X. (1997). *The Acquisition of English Vowels by Mandarin ESL Learners: A Study of Production and Perception*. Master thesis, Simon Fraser University, Burnaby.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am.* 106, 3649–3658. doi: 10.1121/1.428217

Wayland, R. P., and Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: a preliminary report. *Lang. Learn.* 54, 681–712. doi: 10.1111/j.1467-9922.2004.00283.x

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103, 147–162. doi: 10.1016/j.cognition.2006.03.006

Whalen, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Percept. Psychophys.* 46, 284–292. doi: 10.3758/bf03208093

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., et al. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–642. doi: 10.1017/s0048577299981908

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

Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120, 1063–1074. doi: 10.1121/1.2213572

Yang, X. (2011). An investigation of Chinese college student's perceptual patterns with the English /ɪ/-/i/ contrast. *Contemp. Foreign Lang. Stud.* 2, 39–44.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., and Naatanen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Res.* 1072, 175–185. doi: 10.1016/j.brainres.2005.12.004

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., et al. (2010). Training the brain to weight speech cues differently: a study of Finnish second-language users of English. *J. Cogn. Neurosci.* 22, 1319–1332. doi: 10.1162/jocn.2009.21272

Yu, Y. H., Shafer, V. L., and Sussman, E. S. (2017). Neurophysiological and behavioral responses of mandarin lexical tone processing. *Front. Neurosci.* 11:95. doi: 10.3389/fnins.2017.00095

Zhang, L., Wang, J., Hong, T., Li, Y., Zhang, Y., and Shu, H. (2018). Mandarin-speaking, kindergarten-aged children with cochlear implants benefit from natural f0 patterns in the use of semantic context during speech recognition. *J. Speech Lang. Hear. Res.* 61, 2146–2152. doi: 10.1044/2018_JSLHR-H-17-0327

Zhang, Y., and Cheng, B. (2011). "Brain plasticity and phonetic training for English-as-a-second-language learners," in *English as a Second Language*, ed. D. J. Alonso (Hauppauge, NY: Nova Science Publishers), 1–50.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., et al. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Dev. Sci.* 14, 566–581. doi: 10.1111/j.1467-7687.2010.01004.x

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Kotani, M., et al. (2000). "Neural plasticity revealed in perceptual training of a Japanese adult listener to learn American/l-r/contrast: a whole-head magnetoencephalography study," in *Proceedings of 6th International Conference on Spoken Language Processing*, Vol. 3, Beijing, 953–956.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., et al. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *Neuroimage* 46, 226–240. doi: 10.1016/j.neuroimage.2009.01.028

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., and Tohkura, Y. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage* 26, 703–720. doi: 10.1016/j.neuroimage.2005.02.040

Zhang, Y., and Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism* 10, 147–160. doi: 10.1017/s1366728907002908

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