High-Variability Phonetic Training Benefits Lexical Tone Perception: An Investigation on Mandarin-Speaking Pediatric Cochlear Implant Users

Hao Zhang,a Hongwei Ding,a and Yang Zhangb,c

Purpose: Lexical tone perception is known to be persistently difficult for individuals with cochlear implants (CIs). The purpose of this study was to evaluate the efficacy of high-variability phonetic training (HVPT) in improving Mandarin tone perception for native-speaking children with CIs.

Method: A total of 28 Mandarin-speaking pediatric CI recipients participated in the study. Half of the children with CIs received a five-session HVPT within a period of 3 weeks. Identification and discrimination of lexical tones produced by familiar talkers (used during training) and novel talkers (not used during training) were measured before, immediately after, and 10 weeks after training termination. The other half untrained children served as control for the identical pre- and posttests.

Results: Lexical tone perception significantly improved in both trained identification task and untrained discrimination task for the trainees. There was also a significant effect in transfer of learning to perceiving tones produced by novel talkers. Moreover, training-induced gains were retained for up to 10 weeks after training. By comparison, no significant pre–post changes were observed in the control group.

Conclusion: The results provide the first systematical assessment for the efficacy of the HVPT protocol for Mandarin-speaking pediatric CI users with congenital hearing loss, which supports the clinical utility of intensive short-term HVPT in these children’s rehabilitative regimens.

As an epitome of success in modern neural prostheses, the cochlear implant (CI) provides unprecedented acoustic accessibility with a myriad of potential for individuals with profound hearing loss (Kral et al., 2019, 2016; Moore & Shannon, 2009). While the CI has been remarkably successful in partially restoring auditory sensation, highly impoverished input is delivered to the recipient’s central auditory system due to the limitations of the contemporary CI multichannel technology (Moore & Shannon, 2009). In particular, the coarse spectral–temporal properties of the auditory input are much degraded with respect to pitch encoding, which poses a unique challenge for CI users.

Pitch information plays a pivotal role in spoken Chinese understanding. As a tonal language, Mandarin Chinese makes phonemic use of four types of pitch variations—a high-flat pitch for Tone 1 (T1), a midrising pitch for Tone 2 (T2), a falling–rising pitch for Tone 3 (T3), and a high-falling pitch for Tone 4 (T4; Chao, 1948). For instance, depending on the use of the four lexical tones, the Chinese syllable /ma/ can respectively mean “mother,” “hemp,” “horse,” or “scold” (W. S.-Y. Wang, 1973). Not surprisingly, a great number of reports have documented persistent difficulty with lexical tone perception in Mandarin-speaking recipients (e.g., Chen et al., 2014; Peng et al., 2017, 2004; H. Zhang, Zhang, Ding, & Zhang, 2020; H. Zhang, Zhang, Peng, et al., 2020; Zhou et al., 2013).

It is necessary to develop efficacious auditory training programs for this clinical population to make effective use of the degraded sound input to promote speech perceptual performance. This study on Mandarin-speaking pediatric CI recipients aimed to evaluate the benefits of a

Disclosure: The authors have declared that no competing interests existed at the time of publication.
popular laboratory-based speech training protocol in improving their lexical tone perception. Auditory training refers to a sound-based rehabilitative intervention to improve trainees’ speech and hearing abilities via a variety of listening exercises (Sweetow & Sabes, 2006). A body of previous studies have demonstrated that formal auditory training can improve overall outcomes of speech perception in children with CIs (X. Cheng et al., 2018; Good et al., 2017; Ingvalsen et al., 2014; Kronenberger et al., 2011; Lo et al., 2020; Mishra et al., 2015; Roman et al., 2016; Wu et al., 2007). In a recent study, X. Cheng et al. (2018) implemented a melodic contour identification (MCI) training in pediatric Mandarin-speaking CI users over a period of 8 weeks. Nine pitch patterns were involved in the MCI training (i.e., rising, rising–flat, rising–falling, flat–rising, flat, flat–falling, falling–rising, falling–flat, and falling), in which children with CIs were instructed to identify the melodic contours using a closed-set, nine-alternative forced-choice (9-AFC) procedure. The results suggest that intensive music training could contribute to significant improvements in both the trained MCI task and untrained task of lexical tone identification (X. Cheng et al., 2018). However, it warrants further investigations to examine whether such training-induced gains in lexical tone identification are sustainable over a long period after training completion. It is noteworthy that a previous study by Wu et al. (2007) revealed that the speech training–related improvements in lexical tone identification retained for up to 2 months in a cohort of 10 Mandarin-speaking children with hearing impairment (seven CI recipients and three hearing aid users). However, the evaluation of training efficacy in Wu et al. was limited to the trained task. Further research is needed to evaluate the generalization of the training benefits or transfer of learning to novel tasks/stimuli. In addition, protocols adopted in these studies were relatively lengthy, requiring either daily training or long-term training sessions sustaining for months. This situation inevitably placed heavy burdens on the trainees, which, in turn, posed a great challenge to full compliance with the training procedure. Therefore, efforts are needed to develop and assess efficient and effective training techniques for pediatric CI users under the stringent assessment criteria of robust generalization and reliable retention.

The high-variability phonetic training (HVPT) approach is a widely acclaimed efficient and effective speech training method in second-language learning research. This laboratory-based training approach is typically efficient with a time span of 5–10 days. There is compelling evidence that such limited learning experience is effective to induce tangible gains in speech perception (see Ingvalsen & Wong, 2016, for a review), which can retain over 3 months or longer (Bradlow et al., 1999; Iverson & Evans, 2009). The success of HVPT is not limited to the trained talkers and speech stimuli; rather, the training-related benefits transfer to novel talkers and untrained stimuli (Logan et al., 1991). Key features of HVPT are the use of target words produced by multiple talkers (i.e., high acoustic variability) with the to-be-learned speech contrast placed in different phonological contexts (i.e., high phonological variability). In addition, trained listeners are instructed to make a closed-set forced-choice identification (as opposed to discrimination) response, and trial-by-trial response feedback is instantly offered with trial/session repetitions when deemed necessary (B. Cheng et al., 2019; Y. Zhang et al., 2009). The identification task was advocated because it could provide the opportunity to suppress attention to irrelevant acoustic dimensions and increase auditory sensitivity to primary acoustic differences between categories without raising the sensitivity to within-category differences. This approach of sensitivity allocation is believed more effective than training with a discrimination task (Lively et al., 1993; Pisoni & Lively, 1995).

HVPT has been successfully applied to the learning of nonnative phonetic contrasts that are particularly difficult for the adult second-language learners (see Ingvalsen & Wong, 2016, for a review). At the outset, the HVPT paradigm was developed to help native Japanese speakers learn the English phonetic contrast of /t/-/l/ (e.g., Bradlow et al., 1999, 1997; Iverson et al., 2005; Lively et al., 1993, 1994; Logan et al., 1991). The paradigm has since been extended to the learning of Mandarin tones to native English speakers (Dong et al., 2019; Ingvalsen et al., 2013; Perrachione et al., 2011; Y. Wang et al., 2003; Y. Wang & Kuhl, 2003; Y. Wang et al., 1999; Wong & Perrachione, 2007). For example, Y. Wang et al. (1999) implemented the HVPT paradigm in eight American learners of Mandarin over the course of 2 weeks. During training, the participants were instructed to identify the four tone types in naturally produced monosyllabic Mandarin words from four native Mandarin-speaking talkers. The results indicated that tone identification improvements induced from HVPT could generalize to novel stimuli and talker. Moreover, the HVPT-induced gains retained a long period (up to 6 months) post-training (Y. Wang et al., 1999).

The success of HVPT protocols with second-language learners inspired efforts to adopt this approach for training adult CI users. Initial efforts evaluated the efficacy of HVPT in postlingually deafened adults with CIs (Miller et al., 2016a, 2016b). A pretest–intervention–posttest design was implemented with four 2-hr sessions of identification training over a period of 2 weeks. The results demonstrated that the multitalker identification training promoted identification of speech contrasts of both /ba/-/da/ and /wa/-/ja/ in a group of nine trained adults with CIs who were deafened postlingually. The findings are encouraging because they suggest that substantial plasticity in speech perception exists even in adulthood for potential enhancement for phonetic representation in CI recipients. However, the postlingually deafened adults had normal auditory learning experience in childhood, which is crucially different from the congenitally deaf children with CIs. The adult CI users need to learn how to remap the neural activation patterns provided by CI stimulations onto mental representations for their previously acquired speech and language patterns (Boothroyd, 2010). By contrast, the prelingually deafened CI users are not equipped with prior phonological knowledge
and have to acquire the phonemic and phonotactic patterns totally through electric hearing. In addition, lexical tone perception presents a persistent difficulty for CI users due to the well-known technical limitations of the device in encoding pitch information. Whether the HVPT method can be extended to improve lexical tone perception in pediatric Mandarin-speaking CI recipients awaits further empirical investigation.

This study was designed to assess the benefits of HVPT, which has not been systematically implemented and tested in tonal language speakers with CIs. There were three specific aims: (a) to evaluate the identification training effects on the trained materials, (b) to investigate the transfer of learning effects to novel stimuli, and (c) to examine the retention of the benefits over a relatively longer period posttraining. If the HVPT approach is successful, there would be three expected outcomes: (a) The children with CIs who received HVPT would show significant improvement in their lexical tone identification from pre- to posttest, whereas the control children who did not receive the training would not have significant pre–post changes. (b) The identification training–induced benefits would generalize to novel speech stimuli as well as the discrimination task. (c) The benefits would retain for months after the training protocol terminated. Findings of the study will contribute to the development of efficacious programs for aural intervention and speech rehabilitation for pediatric CI recipients who are tonal language speakers.

Method
Participants
A total of 28 native Mandarin-speaking children with CIs (\(M_{\text{age}} = 4.98\) years, 15 girls and 13 boys) were recruited from the Shanghai Rehabilitation Center of the Deaf Children. All pediatric CI users were born with bilateral severe-to-profound hearing impairment. They were unilaterally implanted before 3.5 years of age and had been using their CI devices for at least 1.5 years. All children used spoken standard Mandarin Chinese as the main mode of communication. Parental report indicated no history of psychiatric disorders or brain injuries. In addition, the Hiskey–Nebraska Test of Learning Aptitude (Hiskey, 1966; Yang et al., 2011) was adopted to confirm normal nonverbal intelligence for these children with CIs. The demographic information for each child participant is shown in Table 1. All children with CIs for this study used their own CI devices in a sound-treated therapy room. Informed consent was obtained in accordance with the Ethics Committee of School of Foreign Languages, Shanghai Jiao Tong University.

Experimental Design
The experimental design and schedule are schematically represented in Figure 1. The recruited children with CIs were quasirandomly assigned into training group (\(n = 14\)) and control group (\(n = 14\); cf. Mishra et al., 2015). The quasirandom assignment allowed a close match between the two groups in terms of baseline performance in lexical tone perception and relevant demographic characteristics, including chronological age, age at implantation, CI experience, and nonverbal intelligence. In addition, this subject allocation arrangement could take practical constraints into consideration. For instance, it could maximally avoid scheduling conflicts and time commitment issues to facilitate participation of the potential trainees in the multiple required training and test sessions. The training group (age range: 4.23–5.67 years, \(M_{\text{age}} = 4.91\) years) participated in five HVPT sessions and were encouraged to complete the pretest, posttest, and follow-up test on lexical tone perception. By contrast, the control group (age range: 4.08–5.83 years, \(M_{\text{age}} = 5.06\) years) did not receive training and were assessed with the same pre- and posttests with the same time frame as the training group. Table 2 displays the subject profiles of the two groups. All 28 children with CIs completed the pre- and posttests, whereas eight out of 14 trained children participated in the follow-up retention test approximately 10 weeks after training completion.

Test Stimuli
The test stimuli in lexical tone perception were four Mandarin tones naturally produced with the /i/ syllable. The stimuli were recorded at a sampling rate of 44.1 kHz (16 bit) in a sound-attenuated booth from two males (M1 and M2) and two females (F1 and F2) who are native Mandarin-speaking adults. Each speaker was required to produce each tone 5 times in citation form clearly, resulting in 80 test stimuli in total. All stimuli were normalized with the root-mean-square intensity level of 65 dB SPL while keeping the natural duration of each stimulus.

Test Procedure
The lexical tone perception test was measured 3 times (i.e., pretest, posttest, and follow-up test). The pretest served as baseline measurement of lexical tone perception performance for each participant, and the posttest was presented after 3 weeks of the pretest completion. For the retention test, an interval of 10 weeks was set following the posttest. For the pre- and posttests, both identification and discrimination tasks were administered in a random and counterbalanced order across the children with CIs. All auditory stimuli in the two tasks were delivered in the free field via a loudspeaker (JBL CM220), which was placed approximately 1.2 m from the listener.

In the identification task, a 4-AFC paradigm was adopted to identify each stimulus as T1, T2, T3, or T4. Children with CIs were instructed to choose a matching picture of the presented tonal stimulus from four pictures of driving cars on the laptop screen. Initially, the experimenter explained the relationship for the sound–picture matching to the participant, with the picture of a car driving on a level road indicating T1, a car driving on a rising road indicating T2, a car driving on a dipping road indicating T3, and a car driving on a falling road indicating T4.
Meanwhile, an AX paradigm was used in the discrimination task to instruct children with CIs to determine whether the contrastive tone pairs were the “same condition” or “different condition.” Four “same” pairs (i.e., T1–T1, T2–T2, T3–T3, and T4–T4) and six “different” pairs (i.e., T1–T2, T1–T3, T1–T4, T2–T3, T2–T4, and T3–T4) were constructed with an interstimulus interval of 500 ms for each pair. Likewise, the children with CIs were instructed to perform sound–picture matching. An explicit explanation was offered initially to help the participant to match the picture of two apples with the same condition and to match the picture of an apple and an orange with the different condition. Screenshots of the lexical tone identification and discrimination tasks are shown in Figure A1.

Before the actual test, a practice session with trial-by-trial feedback was provided to ensure that all children with CIs could follow the requirements. To avoid carryover, practice stimuli were produced by a new female talker.

Table 1. Demographic information of the children with cochlear implants (CIs).

<table>
<thead>
<tr>
<th>Subject (sex)</th>
<th>Group</th>
<th>CA (years)</th>
<th>Speech processor</th>
<th>Speech strategy</th>
<th>CI side</th>
<th>Age at CI (years)</th>
<th>CI duration (years)</th>
<th>H-NTLA score</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1 (F)</td>
<td>Training</td>
<td>5.42</td>
<td>Nucleus6</td>
<td>ACE</td>
<td>Right</td>
<td>3</td>
<td>2.42</td>
<td>106</td>
</tr>
<tr>
<td>t2 (F)</td>
<td>Training</td>
<td>5.33</td>
<td>Naida</td>
<td>HiRes-Optima</td>
<td>Right</td>
<td>1.04</td>
<td>4.29</td>
<td>114</td>
</tr>
<tr>
<td>t3 (F)</td>
<td>Training</td>
<td>4.67</td>
<td>Freedom</td>
<td>ACE</td>
<td>Right</td>
<td>1.17</td>
<td>3.5</td>
<td>96</td>
</tr>
<tr>
<td>t4 (F)</td>
<td>Training</td>
<td>4.63</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>2.13</td>
<td>2.5</td>
<td>103</td>
</tr>
<tr>
<td>t5 (M)</td>
<td>Training</td>
<td>5</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>0.92</td>
<td>4.08</td>
<td>104</td>
</tr>
<tr>
<td>t6 (M)</td>
<td>Training</td>
<td>4.33</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.67</td>
<td>2.66</td>
<td>110</td>
</tr>
<tr>
<td>t7 (M)</td>
<td>Training</td>
<td>5.67</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.67</td>
<td>4</td>
<td>109</td>
</tr>
<tr>
<td>t8 (M)</td>
<td>Training</td>
<td>5.5</td>
<td>Nucleus5</td>
<td>ACE</td>
<td>Left</td>
<td>0.95</td>
<td>4.55</td>
<td>106</td>
</tr>
<tr>
<td>t9 (F)</td>
<td>Training</td>
<td>4.97</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.58</td>
<td>3.38</td>
<td>105</td>
</tr>
<tr>
<td>t10 (F)</td>
<td>Training</td>
<td>5.03</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.14</td>
<td>3.89</td>
<td>100</td>
</tr>
<tr>
<td>t11 (F)</td>
<td>Training</td>
<td>4.35</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.04</td>
<td>3.3</td>
<td>126</td>
</tr>
<tr>
<td>t12 (M)</td>
<td>Training</td>
<td>4.23</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1</td>
<td>3.23</td>
<td>102</td>
</tr>
<tr>
<td>t13 (M)</td>
<td>Training</td>
<td>4.63</td>
<td>Nucleus5</td>
<td>ACE</td>
<td>Left</td>
<td>1.38</td>
<td>3.24</td>
<td>114</td>
</tr>
<tr>
<td>t14 (M)</td>
<td>Training</td>
<td>5</td>
<td>Nucleus6</td>
<td>ACE</td>
<td>Right</td>
<td>1.5</td>
<td>3.49</td>
<td>125</td>
</tr>
<tr>
<td>c1 (F)</td>
<td>Control</td>
<td>5.58</td>
<td>Nucleus6</td>
<td>ACE</td>
<td>Left</td>
<td>3.33</td>
<td>2.25</td>
<td>106</td>
</tr>
<tr>
<td>c2 (F)</td>
<td>Control</td>
<td>5.83</td>
<td>Neptune</td>
<td>HiRes-Optima</td>
<td>Right</td>
<td>1</td>
<td>4.83</td>
<td>106</td>
</tr>
<tr>
<td>c3 (M)</td>
<td>Control</td>
<td>4.08</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>0.92</td>
<td>3.16</td>
<td>117</td>
</tr>
<tr>
<td>c4 (M)</td>
<td>Control</td>
<td>5.58</td>
<td>Naida</td>
<td>HiRes-Optima</td>
<td>Right</td>
<td>2.67</td>
<td>2.91</td>
<td>124</td>
</tr>
<tr>
<td>c5 (M)</td>
<td>Control</td>
<td>5.33</td>
<td>Nucleus6</td>
<td>ACE</td>
<td>Left</td>
<td>2.67</td>
<td>2.66</td>
<td>108</td>
</tr>
<tr>
<td>c6 (M)</td>
<td>Control</td>
<td>4.67</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1</td>
<td>3.67</td>
<td>103</td>
</tr>
<tr>
<td>c7 (M)</td>
<td>Control</td>
<td>4.67</td>
<td>Nucleus6</td>
<td>ACE</td>
<td>Right</td>
<td>1.42</td>
<td>3.25</td>
<td>124</td>
</tr>
<tr>
<td>c8 (F)</td>
<td>Control</td>
<td>5</td>
<td>Nucleus5</td>
<td>ACE</td>
<td>Left</td>
<td>1.37</td>
<td>3.63</td>
<td>108</td>
</tr>
<tr>
<td>c9 (F)</td>
<td>Control</td>
<td>4.36</td>
<td>Nucleus6</td>
<td>ACE</td>
<td>Right</td>
<td>1.41</td>
<td>2.95</td>
<td>96</td>
</tr>
<tr>
<td>c10 (M)</td>
<td>Control</td>
<td>5.17</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.21</td>
<td>3.95</td>
<td>113</td>
</tr>
<tr>
<td>c11 (M)</td>
<td>Control</td>
<td>5.25</td>
<td>OPUS1</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.71</td>
<td>3.54</td>
<td>125</td>
</tr>
<tr>
<td>c12 (M)</td>
<td>Control</td>
<td>4.89</td>
<td>OPUS2</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.35</td>
<td>3.54</td>
<td>118</td>
</tr>
<tr>
<td>c13 (M)</td>
<td>Control</td>
<td>4.87</td>
<td>OPUS1</td>
<td>FS4-P</td>
<td>Right</td>
<td>1.13</td>
<td>3.74</td>
<td>108</td>
</tr>
<tr>
<td>c14 (F)</td>
<td>Control</td>
<td>5.64</td>
<td>OPUS1</td>
<td>FS4-P</td>
<td>Right</td>
<td>2.67</td>
<td>2.98</td>
<td>109</td>
</tr>
</tbody>
</table>

Note. CA = chronological age; H-NTLA = Hiskey–Nebraska Test of Learning Aptitude; F = female; ACE = advanced combination encoder; HiRes-Optima = high-resolution optima; FS4-P = fine structure processing strategy; M = male.

Figure 1. Schematic diagram of the experimental design. HVPT = high-variability phonetic training.
which were not included in the tests or training. In total, 80 stimuli were randomly presented to each child within two blocks for the test session of the identification task. On the other hand, we selected 12 “same” trials and 12 “different” trials from each speaker in the discrimination task, which were presented in a random order in two blocks with a total of 96 contrastive tone pairs. In order to avoid fatigue effect, the two tasks were measured on two separate days, and a 2-min break was offered between blocks of each task.

Training Stimuli
The training sessions used four Mandarin tones naturally produced with three vowels (i.e., /i/, /a/, and /u/), resulting in 12 specific syllables as training stimuli. The training stimuli were recorded from 10 native Mandarin-speaking adults (five males and five females), with each syllable reproduced 5 times. Two speakers (F1 and M1) who produced the test stimuli were also included for recording the training stimuli. A total of 600 lexical tone training items were obtained from multiple talkers and in different phonological contexts.

Training Protocol
Children with CIs from the training group were introduced to complete five sessions of phonetic identification training over the course of 3 weeks in the rehabilitation center. A computer-based training program was developed (cf. Miller et al., 2016a, 2016b). The training program included a 4-AFC paradigm for sound–picture matching, the procedure of which was similar as in the lexical tone identification task. However, as opposed to the test sessions of lexical tone identification, trial-by-trial feedback and repetition were provided accordingly for each item in the training sessions. One repetition was offered for items with correct identification responses, and two repetitions were offered for those with incorrect responses. All training stimuli were delivered via a loudspeaker at approximately 65 dB SPL.

Training began with two unique talkers (F1 and M1) in the first session, with 120 training items implemented in six blocks. Blocked high-variability stimulus delivery was employed to control the variability within each training block (Dong et al., 2019; Perrachione et al., 2011). In addition, adaptive scaffolding was adopted in the training program (Y. Zhang et al., 2009). After completing all six training blocks, the trainees were required to take an identification quiz of 24 items with 12 training syllables. If the listener’s quiz performance exceeded 85% correct, they could succeed to the next training session that added two more talkers (one female and one male) in the training blocks. Otherwise, the trainee repeated the current training session. It should be noted that the quiz was presented after completion of six blocks in each training session, and two additional talkers (one male and one female) were added in the subsequent training sessions, until all 10 talkers were included in the last training session. All trained children with CIs completed the five sessions of HVPT at their own pace over a period of 3 weeks. The training protocol was performed on five center visits, with each training session completed per visit. Each training session lasted approximately 40 min, resulting in an average of 3.33 total training hours per trainee.

Statistical Analysis
The percentage of accuracy rates for the identification and discrimination tasks in three test sessions (i.e., pretest, posttest, and follow-up test) was computed for the children with CIs. For statistical analysis, the percentage correct was transformed into rationalized arc sine unit (rau), which was widely performed in evaluations of speech intelligibility performance to reduce saturation effect and restore homoscedasticity (Oleson et al., 2019; Studebaker, 1985; H. Zhang, Zhang, Ding, & Zhang, 2020; L. Zhang et al., 2018).

Statistical analyses were performed with the open source R platform (Version 3.6.1). Linear mixed-effects (LME) models were constructed with the package of lme4 (Bates et al., 2015) to evaluate lexical tone perception in different test sessions for children with CIs. The experimental conditions (e.g., group, test session, and tone type/tone pair) and interactions were treated as fixed factors. By-participant intercepts and maximal slopes that would converge (i.e., by-participant slopes for all experimental conditions) were entered as random effects (Barr et al., 2013). The most complex model with all fixed factors and random effects was compared with a simplified model that excluded a fixed factor in question using the analysis of variance function in lmerTest package (Kuznetsova et al., 2017). The Akaike information criterion was computed to check whether the fixed factor in question could contribute to an improved fit for the LME model. Eventually, the model with the lowest Akaike information criterion was selected as the best-fit model to estimate the significance of the fixed factors, with an alpha setting at .05. Based on the best-fit model, the analysis of variance function in lmerTest package was implemented to obtain F and p values of the significant fixed factors with the Satterthwaite method. Post hoc pairwise comparisons were performed for the significant fixed factors using the emmeans package (Lenth et al., 2018) with false discovery rate correction (Benjamini & Yekutieli, 2001) for multiple comparisons to obtain t ratios and p values.
Results
Effects of HVPT

The overall mean identification and discrimination scores in rau for each group in pre- and posttests are shown in Figure 2. As described in the Method section, LME models were created to evaluate effects of HVPT in lexical tone identification and discrimination with test session (pretest and posttest), tone type/tone pair (tone type in identification analysis and tone pair in discrimination analysis), and group (training group and control group) as fixed factors. In addition, by-participant intercepts and by-participant slopes for group, test session, and tone type/tone pair were treated as random effects.

The training group showed significant pre–post improvements on lexical tone identification, while the control group did not. Statistically, the best-fit model formula for identification analysis was as follows: lmer(Score ~ Group × Session + Type × Session + (1 + Group + Session + Type | Subject), REML = FALSE). The LME results revealed a significant interaction effect of Group × Test Session, F(1, 59) = 33.52, p < .001, indicating pre–post changes in lexical tone identification differed significantly between the two groups. Further post hoc comparisons indicated that significantly improved identification performance from the pre- to posttest sessions was shown in the training group, t(55) = 8.15, p < .001, but not in the control group, t(55) = 0.07, p = .95. Meanwhile, the difference on lexical tone identification between the two groups was significant in posttest session, t(27) = 4.83, p < .001, but not in pretest session, t(27) = -1.01, p = .38. Taken together, the results suggested that the training group and control group showed comparable baseline performance on lexical tone identification. Moreover, the pre–post improvements were only significant in the training group rather than in the control group. In addition, the two-way interaction of Test Session × Tone Type was significant, F(3, 131) = 3.76, p = .01, but not the three-way interaction of Group × Test Session × Tone Type, F(3, 132) = 1.67, p = .18. The results suggested that the pre–post changes were not equivalent for each tone type. Overall, the identification improvement from pre- to posttests was significant for T2, t(136) = 5.66, p < .001, and T3, t(136) = 3.26, p = .002, but not for T1 and T4 (all ps > .05).

Lexical tone discrimination from pre- to posttests improved significantly in the training group, but not in the control group. For discrimination analysis, the best-fit model formula was as follows: lmer(Score ~ Group × Session + Pair × Session + (1 + Group + Pair + Session | Subject), REML = FALSE). Statistical analysis for discrimination data revealed a significant Group × Test Session interaction, F(1, 33) = 17.78, p < .001. Post hoc tests indicated that the HVPT increased lexical tone discrimination scores from the pre- to posttest sessions in the trained children with CIs, t(33) = 7.75, p < .001, but not in the control children with CIs, t(33) = 1.93, p = .09. Meanwhile, the difference on discrimination performance between the training and control groups was not significant in pretest, t(21) = -0.28, p = .78, whereas this difference was significant in posttest, t(21) = 3.2, p = .009. The interaction and follow-up comparison results indicated that only the trained children with CIs improved lexical tone discrimination from pre- to posttests significantly, regardless of the two groups’ shared comparable performances in the pretest session. It should be noted that the three-way interaction of Group × Test Session × Tone Pair was insignificant, F(5, 210) = 1.95, p = .095, but only the two-way interaction of Test Session × Tone Pair was significant, F(5, 210) = 2.98, p = .01. This suggested that the discrimination changes from pre- to posttest were not equivalent among the tone pairs. Figure 3 shows the mean rau scores of lexical tone identification and discrimination basing on different Tone Types/Tone Pairs for the training group (Panels A and B) and the control group (Panels C and D).

Figure 2. Overall mean scores in rationalized arcsine unit (rau) of lexical tone identification and discrimination in pretest and posttest for the two groups. Error bars represent the standard errors.

To illustrate intersubject variability, the individual listeners’ results are displayed in Figure 4. Paired-samples t tests were performed to compare perceptual performance of each individual between pre- and posttests. In the training group, a significant training effect was revealed for t1,
t3, t5, t6, and t12 in lexical tone identification (all $p < .05$) and for t1, t4, t6, t10, t12, and t13 in lexical tone discrimination (all $p < .04$). By striking contrast, no subject in the control group showed significant difference between the two test sessions either in identification performance (all $p > .1$) or in discrimination performance (all $p > .08$).

**Transfer to Unfamiliar Talkers**

Figure 5 displays the average rau scores of lexical tone identification and discrimination in the trained children with CIs for the familiar talkers (i.e., F1 and M1, two talkers used in both test and training sessions) and unfamiliar talkers (i.e., F2 and M2, two talkers only used in test sessions but not in training sessions). To determine whether the training-induced gains could transfer to unfamiliar talkers, LME models were constructed using test session (pretest and posttest), tone type/tone pair (tone type in identification analysis and tone pair in discrimination analysis), and familiarity (familiar and unfamiliar) as fixed factors and using participant as a random factor.

Training-induced gains in lexical tone identification and discrimination were not confined to the familiar talkers but transferred to unfamiliar talkers. The best-fit model formula for identification analysis was as follows: lmer(Score ~ Familiarity × Type + Type × Session + (1 + Familiarity + Type + Session | Subject), REML = FALSE). A significant main effect of test session was found, $F(1, 16) = 52.07, p < .001$, with a significant identification improvement from pre- to posttests in both familiar talkers, $t(34) = 6.55, p < .001$, and unfamiliar talkers, $t(34) = 5.28, p < .001$. The results indicated that significant gains induced from the identification training were not constrained to lexical tone identification for the familiar talkers but transferred to unfamiliar talkers. In addition, the best-fit model formula for discrimination analysis was as follows: lmer(Score ~ Familiarity × Session + Pair × Session + (1 + Familiarity + Pair + Session | Subject), REML = FALSE). Similarly, the main effect of test session was significant in discrimination analysis, $F(1, 15) = 53.84, p < .001$. The training-related gains were observed in both familiar talkers, $t(34) = 4.8, p < .001$, and unfamiliar talkers, $t(34) = 7.18, p < .001$. The results suggested that identification training-induced benefits could...
generalize to lexical tone discrimination for both familiar talkers and unfamiliar talkers.

In order to preclude the possibility of talker specificity, additional LME models were constructed on identification and discrimination scores in the control children with CIs. Test session (pretest and posttest), tone type/tone pair (tone type in identification analysis and tone pair in discrimination analysis), and talker (F1M1 and F2M2) were used as fixed factors, and participant was treated as a random factor. For identification analysis, neither the main effect of test session, $F(1, 23) = 0.09, p = .77$, nor the interaction of Test Session $\times$ Talker, $F(1, 168) = 0.95, p = .33$, approached significance. Likewise, there was no significant effect of test session, $F(1, 16) = 2.5, p = .13$, or Test Session $\times$ Talker, $F(1, 266) = 1.18, p = .28$, in discrimination analysis. The results indicated that, as opposed to the trained children with CIs, control children did not show significant pre-post changes for either talker group of F1M1 or F2M2 in both lexical tone identification and discrimination.

**Retention of Training Effects**

Eight out of 14 trained children with CIs completed the pretest, posttest, and follow-up test sessions. Due to restrictions related to the outbreak of COVID-19, timely data collection from the other six participants could not be completed for the retention tests. Thus, data from the eight children were analyzed to assess the stability of training-related improvements in lexical tone perception. These children’s mean rau scores of lexical tone identification and discrimination in different test sessions are illustrated in Figure 6. To evaluate the retention of training effects, LME models were created with test session (pretest, posttest, and follow-up test) and tone type/tone pair (tone type in identification analysis and tone pair in discrimination analysis) as fixed factors and with participant as a random factor.

The trained children with CIs demonstrated reliable retention for the training-related benefits of lexical tone identification and discrimination. In identification analysis, the best-fit model formula was as follows: lmer(Score $\sim$ Type $\times$ Session + (1 + Type + Session | Subject), REML = FALSE). The results indicated a significant interaction effect of Test Session $\times$ Tone Type, $F(6, 72) = 3.06, p = .01$. Post hoc pairwise comparisons showed significantly improved identification performance in posttest over pretest and in follow-up test over pretest for T1, T2, and T3, $t > 2.24, ps < .05$. By contrast, the difference on lexical tone identification between posttest and follow-up test was insignificant for either tone type, $t < 1.2, ps > .23$. In addition,
the best-fit model formula for discrimination analysis was as follows: lmer(Score ~ Pair × Session + (1 + Pair + Session | Subject), REML = FALSE). Similarly, there was a significant interaction effect of Test Session × Tone Pair, $F(10, 112) = 2.35, p = .01$. In particular, post hoc comparisons revealed a significant improvement for discrimination of T2–T3 pair from pre- to posttest, $t(49) = 5.26, p < .001$, and from posttest to follow-up test, $t(49) = 6.29, p < .001$.

Figure 5. (A) Mean identification scores in rationalized arcsine unit (rau) of each tone type from familiar talkers and unfamiliar talkers for the trained children with cochlear implants (CIs). (B) Mean discrimination scores in rau of each tone pair from familiar talkers and unfamiliar talkers for the trained children with CIs. Error bars represent the standard errors.

Figure 6. (A) Mean identification scores in rationalized arcsine unit (rau) of each tone type in pretest, posttest, and follow-up test across the trained children completed all three test sessions. (B) Mean discrimination scores in rau of each tone pair in pretest, posttest, and follow-up test across the children with cochlear implants completed all three test sessions. Error bars represent the standard errors.
but not from posttest to follow-up test, \( t(49) = 0.91, p = .37 \). The results suggested that gains from the HVPT protocol in lexical tone identification and discrimination maintained for at least 10 weeks in the follow-up test session.

**Discussion**

The purpose of this study was to systematically examine the benefits of HVPT for lexical tone perception in native Mandarin-speaking children with CIs. A pretest, training, posttest, follow-up test experimental design was implemented among the trained children. The results were largely in accordance with our predictions based on prior studies. Significant training-related improvements were revealed for both trained task of lexical tone identification and untrained task of lexical tone discrimination. Moreover, robust transfer to unfamiliar talkers and reliable retention after a 10-week interval posttraining were demonstrated in the trainees. Our findings confirmed the efficacy of the HVPT approach in pediatric CI recipients and lent support to the intervention practice of formal identification training in this clinical population to improve their lexical tone perception.

**Significant Benefits for Trained and Untrained Tasks**

Overall improvements were observed for lexical tone identification in the training group after the five-session identification training. By contrast, there was no significant change from pre- to posttests in the control group without receiving the HVPT (see Figure 2). Given the fact that the two groups showed comparable baseline performance of lexical tone perception in the pretest session, it is reasonable to attribute the significant improvements in the trained children with CIs to the HVPT approach. The tangible training-induced benefits in the trained task of this study were largely consistent with a prior report (Miller et al., 2016a) that recruited a group of postlingually deafened adults with CIs in recognition of consonant contrasts of /ba/–/da/ and /wa/–/ja/.

Moreover, significant training-induced gains from the trained task of lexical tone identification transferred to untrained task of lexical tone discrimination. From the theoretical perceptive, current computational models posit that the key mechanism responsible for generalization is the dissociation of idiosyncratic cues (e.g., talker/context-related cues) in competing with phonetically relevant cues to develop abstract phonetic categories (Apfelbaum & McMurray, 2011; Ramscar & Baayen, 2013; Ramscar et al., 2010). The identification training paradigm that incorporates high acoustic and phonological variability is thought to improve abstract mental representations of the to-be-learned phonetic contrasts by forcing the trainees to make a category-level response on the identity of each training stimulus (Lively et al., 1993; Pisoni & Lively, 1995). The category-level responses require the listener to go through a normalization process that is supposed to enhance the perceptual sensitivity to between-categories phonetic differences, but not to within-category acoustic differences (Pisoni & Lively, 1995). If that were the case, the improved abstract tonal representations from the HVPT would be expected to benefit perceptual performance of both identification and discrimination of lexical tones in the trained children with CIs.

An intriguing question has been raised with respect to whether the training-induced benefits in lexical tone representations could derive mainly from the improved sensitivity of which acoustic cue, since Mandarin tones are represented by multiple acoustic cues (Tupper et al., 2020). Apart from the primary cue of fundamental frequency (F0; the acoustic correlate for pitch variation), the concurrent temporal signals as amplitude envelope and duration can contribute to Mandarin tone identification (Fu & Zeng, 2000; Fu et al., 1998; Whalen & Xu, 1992). In addition to F0 information, pediatric CI users could shift the cue-weighting patterns with more reliance on duration dimension over their normal-hearing (NH) peers in recognizing lexical tones (Deroche et al., 2019; Peng et al., 2017). In this study, the training benefits were significantly driven by improvements for T2 and T3 tone types in identification task and for T2–T3 tone pair in discrimination task from pre- to posttests (see Figure 3). T2 and T3 in Mandarin Chinese represent the most difficult tone contrast to discern because of their acoustic similarity (Shen & Lin, 1991; Shen et al., 1993). Prior studies demonstrated the recognition confusion between T2 and T3 in native pediatric CI users (e.g., Chen et al., 2014; Peng et al., 2017; H. Zhang, Zhang, Ding, & Zhang, 2020). In isolated speech, the most reliable contrastive feature to distinguish T3 from T2 is the timing of the inflection point that turns the pitch contour from falling to rising. Therefore, it is tempting to suppose that the intensive and repetitive identification listening exercises with explicit instructions on pitch patterns (concurrent pictures depicting pitch contours) could potentially sharpen perceptual acuity in pitch variations to discriminate the subtle pitch differences between the two tone types (Wiener et al., 2020).

In accordance with the well-documented remarkable heterogeneity among pediatric CI users, individual differences were ubiquitous in the amount of learning across the trainees in this study (see Figure 4). In this regard, six out of the 14 trained children (t2, t7, t8, t9, t11, and t14) obtained no significant training benefits in lexical tone perception. One plausible possibility for the insignificant training-related improvements in these trainees is the saturated or ceiling-level representations of specific tones (T1 and T4, in particular). The four Mandarin tones are not uniformly easy or difficult to recognize, resulting in large standard errors in overall perceptual outcomes. This situation could limit the room for potential amounts of learning in good performers. In addition, another possible explanation for the limited learning with some less successful children with CIs is related to listening strategies. A recent report by Peng et al. (2017) revealed that lexical tone identification performance for pediatric CI recipients with prelingual deafness...
was significantly correlated with their reliance on pitch contours, although they showed higher reliance on duration cues relative to their NH peers. It is suggested that some of the less successful CI listeners could have adapted to the less optimal listening strategies that might reduce learning (Miller et al., 2016a). Their strong reliance on duration patterns might be too strong to be overcome with our short-term identification training, leading to the lack of significant pre–post changes in poor performers with CIs.

**Generalization and Retention of Learning**

The observed training-induced gains in lexical tone perception were not confined to the trained talkers but transferred to unfamiliar talkers that were not involved in training materials (see Figure 5). Robust generalization to novel, unfamiliar talkers and speech plays an important role in evaluating the success or failure of a training protocol (Lively et al., 1993; Logan et al., 1991). The robust transfer to novel talkers in pediatric CI recipients was largely in line with the prior findings in postlingually deafened CI adults (Miller et al., 2016a), despite relatively greater gains for both familiar and unfamiliar talkers than the study by Miller et al., (2016a). Our results revealed perceptual improvements of 18.1% for familiar speech and 15.8% for unfamiliar speech as opposed to 11.5% for both familiarity types in Miller et al.’s et al., (2016a) research. Several candidate reasons were speculated to account for this discrepancy. This study evaluated perceptual performance of lexical tones that primarily depends on F0 variations, whereas Miller et al. (2016a) measured consonant identification (/b/, /d/, /v/, and /j/) with the primary acoustic cue of second formant (F2) transitions. F0 variations in our stimuli are embedded in the whole syllable, whereas F2 transitions in the previous study occurred only in the syllable onset. The acoustic salience factor, especially in terms of duration differences between F0 variations and F2 transitions, might contribute to different outcomes in lexical tone and consonant perception for the trainees. Moreover, in addition to the overall superiority and greater plasticity in speech learning capacities for children over adults (Gianakopoulou et al., 2013; Kuhl, 2004), the present training protocol incorporated much higher variability in training materials with both multiple talkers and phonological contexts. The high variability of training stimuli has been suggested to play a pivotal role in developing abstract representations of speech sounds for efficacious training paradigms (Lively et al., 1993; Shinohara & Iverson, 2018). There is a long-standing assumption that higher variable inputs contribute to more benefit in phonetic learning (Gianakopoulou et al., 2017). Additionally, the blocks in each training session of this study held materials with well-controlled variability that was constrained to the phonological contrast across different tone types. Blocked high-variability design has been demonstrated to be more efficient for the trainees, especially among the performers with relatively poor perceptual aptitude (Perrachione et al., 2011; Sadakata & McQueen, 2014). It follows that the higher variability with blocked design could potentially contribute to greater training-induced gains in lexical tone perception in children with CIs.

Furthermore, our data showed that the training-induced improvements sustained up to 10 weeks in both trained task of lexical tone identification and untrained task of lexical tone discrimination after the training had ceased by comparing the performance of the trained children with CIs at three different test sessions (see Figure 6). Reliable retention of the training benefits for a relatively long-term duration is an important indicator for the stability of learning post a training program (Henshaw & Ferguson, 2013). Nevertheless, a recent systematic review of auditory training revealed that prior studies have rarely assessed the retention of benefits in pediatric CI recipients (Rayes et al., 2019). One exception was the study by Wu et al. (2007), in which a cohort of 10 Mandarin-speaking children with hearing impairment (seven CI recipients and three hearing aid users), with an age range of 5.87–10.88 years, participated in a computer-assisted speech training for a period of 10 weeks. The previous study demonstrated an overall improvement of 15.1% in lexical tone identification, and the training-induced gains retained for up to 2 months in trained children with hearing impairment. It is encouraging to replicate an averaged degree of 17.1% improvement and a 10-week retention of tonal learning that are comparable to the findings of Wu et al. in our much younger and more homogeneous clinical participants of kindergarten-age children with CIs. In addition, the recruitment of a control group in this study could circumvent the potentially procedural-learning in the study by Wu et al. and ensure that the training-related improvements were not due to practice effects. Moreover, our findings indicated that the training-induced gains from identification task could generalize to the performance of lexical tone discrimination, whereas the training evaluation in Wu et al. was limited to the trained tasks. It is noteworthy that the training paradigms of the previous and present studies both incorporated multiple talkers and phonological contexts, with the former one obtaining four tones for six Mandarin vowels (/a/, /o/, /e/, /i/, /u/, and /ü/) from four native speakers while the latter one recording four Mandarin tones for three vowels (/a/, /i/, and /u/) from 10 native adults. It is reasonable to postulate that the HVPT could contribute to robust generalization and reliable retention for Mandarin tone perception in native children with CIs.

**Clinical Implications and Future Directions**

This study indicated that the HVPT could benefit lexical tone perception in Mandarin-speaking children with CIs. It is encouraging to validate the robust generalization and retention of training-induced gains in native pediatric CI recipients with prelingual deafness, in view of the well-documented limitations of current CI technology in pitch encoding. Although there was striking variability in the amount of learning across the trained children, the current findings have practical implications that lend support to the inclusion of tonal identification training with high-variable materials in pediatric
CI recipients’ rehabilitative regimens. The intensive and repetitive identification training has the potential to improve mental representations of lexical tones for this clinical population. In addition, the results that the significant training-induced benefits in lexical tone perception were driven by perceptual improvements in T2 and T3 would need to be taken into consideration. Rehabilitative practice could emphasize the two tones to improve the efficiency of training interventions (Peng et al., 2017).

Unlike several previous training studies on children with hearing impairment (X. Cheng et al., 2018; Fu et al., 2015; Hidalgo et al., 2017; Wu et al., 2007), this study enrolled a control group of pediatric CI users that received no explicit training to refrain from the possible practice effects from pre- to posttests. There was very little test–retest change in the control group, which helps eliminate the possibility that procedural learning or increased familiarity with the test procedure and stimuli could improve lexical tone perception. However, one major limitation of this study that should be acknowledged was the lack of NH peers. The NH baseline performance is necessary to evaluate whether the HVPT protocol could accrue optimal benefits in lexical tone learning for children with CIs. In addition to the NH controls, an active control intervention with passive exposure to the training stimuli should be considered in future work, which could help circumvent interference from potential placebo effects in evaluating training-induced gains (Henshaw & Ferguson, 2013). Another limitation was the relatively small subject sample, especially for the retention test that held only eight trainees due to restrictions related to the pandemic outbreak beyond our control. More subjects and their availability for longitudinal testing should be considered in future studies to obtain robust results. Moreover, this study adopted merely a high-variability identification training protocol. Prior cross-linguistic research assumed that high variability of the training materials was vital for Japanese speakers’ learning of English /r/–/l/, regardless if identification or discrimination training approach was adopted (Shinohara & Iverson, 2018). However, it should be noted that the study of Shinohara and Iverson (2018) did not actually test the assumption via directly comparing high versus low talker variability training paradigms. A handful of recent studies have shown a lack of benefits of phonetic training with high variability (speech from four talkers) over low variability (speech from a single talker) in nonnative speakers’ learning of phonetic contrasts (Dong et al., 2019; Giannakopoulou et al., 2017; Wiener et al., 2020). It is suggested that the increased encoding burden from multitalker variability compared to single-talker speech might offset the potential gains from high-variability exposure. Further research is needed to directly contrast the use of high- and low-variability training protocols to disentangle the power/specificity of the HVPT per se. This line of research could also help to determine the optimal dosage of variability and the appropriate prescription of training method for more effective and efficient intervention to facilitate lexical tone learning by children with CIs from tonal languages. It still remains indefinite to ascertain that the training-related improvements in representations of lexical tones are derived mainly from the trained children’s enhanced sensitivity to pitch information as Mandarin lexical tones in natural speech are indexed by multiple acoustic cues (Tupper et al., 2020). In this regard, further research is also needed using synthesized speech stimuli with well-controlled manipulations of acoustic cues (e.g., pitch, duration, and amplitude cues). Additionally, although the current study revealed across-the-board training-related benefits in lexical tone perception, it will be important to examine whether the HVPT approach can be extended to consonant and vowel training with entirely different primary acoustic cues in pediatric CI users.

Conclusions

This study provides confirmatory evidence that Mandarin-speaking pediatric CI recipients with congenital deafness can benefit from the HVPT in lexical tone perception. Training-induced gains were significant in both trained identification and untrained discrimination tasks, which could transfer to unfamiliar speech and retain for 10 weeks posttraining. These findings lend support to the inclusion of formal active training of lexical tones in intervention regimens for this clinical population.

Acknowledgments

This study was supported by grants from the Major Program of National Social Science Foundation of China (No. 18ZDA293) to H. Ding and Y. Zhang. Y. Zhang received additional support from the University of Minnesota’s Brain Imaging Grant and Grand Challenges Exploratory Research Grant for international collaboration. We thank the Shanghai Rehabilitation Center of the Deaf Children for the cooperation and assistance in implementing this study.

References


**Appendix**

**Figure A1.** Screenshots of the lexical tone identification task and the lexical tone discrimination task.