Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Article

Sustainable Benefits of High Variability Phonetic Training in Mandarin-speaking Kindergarteners with Cochlear Implants: Evidence from Categorical Perception of Lexical Tones

Hao Zhang [a], Wen Ma [a], Hongwei Ding [b, *] and Yang Zhang [c]

[a] Center for Clinical Neurolinguistics, School of Foreign Languages and Literature, Shandong University, Jinan, China
[b] Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
[c] Department of Speech-Language-Hearing Sciences and Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, USA
* Correspondence: hwding@sjtu.edu.cn (H. Ding), zhanglab@umn.edu (Y. Zhang).

Abstract

Objectives: Although pitch reception poses a great challenge for individuals with cochlear implants (CIs), formal auditory training (e.g., high variability phonetic training, HVPT) has been shown to provide direct benefits in pitch-related perceptual performances such as lexical tone recognition for CI users. As lexical tones in spoken language are expressed with a multitude of distinct spectral, temporal, and intensity cues, it is important to determine the sources of training benefits for CI users. The purpose of the present study was to conduct a rigorous fine-scale evaluation with the categorical perception (CP) paradigm to control the acoustic parameters and test the efficacy and sustainability of HVPT for Mandarin-speaking pediatric CI recipients. The main hypothesis was that HVPT-induced perceptual learning would greatly enhance CI users' ability to extract the primary pitch contours from spoken words for lexical tone identification and discrimination. Furthermore, individual differences in immediate and long-term gains from training would likely be attributable to baseline performance and duration of CI use.

Design: Twenty-eight prelingually deaf Mandarin-speaking kindergarteners with CIs were tested. Half of them received five sessions of HVPT within a period of three weeks. The other half served as control who did not receive the formal training. Two classical CP tasks on a tonal continuum from Mandarin Tone 1 (high-flat in pitch) to Tone 2 (mid-

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

rising in pitch) with fixed acoustic features of duration and intensity were administered before (pretest), immediately after (posttest), and 10 weeks post training termination (follow-up test). Participants were instructed to either label a speech stimulus along the continuum (i.e., identification task) or determine whether a pair of stimuli separated by zero or two steps from the continuum was the same or different (i.e., discrimination task). Identification function measures (i.e., boundary position and boundary width) and discrimination function scores (i.e., between-category score, within-category score, and peakedness score) were assessed for each child participant across the three test sessions.

Results: Linear mixed-effects (LME) models showed significant training-induced enhancement in lexical tone categorization with significantly narrower boundary width and better between-category discrimination in the immediate posttest over pretest for the trainees. Furthermore, training-induced gains were reliably retained in the follow-up test 10 weeks after training. By contrast, no significant changes were found in the control group across sessions. Regression analysis confirmed that baseline performance (i.e., boundary width in the pretest session) and duration of CI use were significant predictors for the magnitude of training-induced benefits.

Conclusions: The stringent CP tests with synthesized stimuli that excluded acoustic cues other than the pitch contour and were never used in training showed strong evidence for the efficacy of HVPT in yielding immediate and sustained improvement in lexical tone categorization for Mandarin-speaking children with CIs. The training results and individual differences have remarkable implications for developing personalized computer-based short-term HVPT protocols that may have sustainable long-term benefits for aural rehabilitation in this clinical population.

Keywords: high variability phonetic training (HVPT); categorical perception (CP); cochlear implant (CI); lexical tone; Mandarin-speaking kindergarteners; training-induced gains

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Abbreviations: CI = cochlear implant; CP = categorical perception; FDR = false discovery rate; F0 = fundamental frequency; H-NTLA = Hiskey–Nebraska test of learning aptitude; HVPT = high variability phonetic training; LME = linear mixed-effects; MCI = melodic contour identification; MMN = mismatch negativity; NH = normal hearing; PSOLA = Pitch-Synchronous Overlap Add; T1 = Tone 1; T2 = Tone 2; T3 = Tone 3; T4 = Tone 4; 2 AFC = two-alternative forced choice; 4 AFC = four-alternative forced choice; 9 AFC = nine-alternative forced-choice

1 Introduction

Cochlear implant (CI) is a prime example of successful modern neural prosthesis. It restores basic hearing in individuals with severe-to-profound hearing impairment, and facilitates the acquisition of spoken language in children with congenital deafness (Kral et al. 2016; 2019). Despite great strides in technological development, the signals delivered by CI are distorted and highly impoverished with spectral-temporal degradation (Moore & Shannon 2009). Due to the small number of electrodes and the large current spread of CIs, the fundamental frequency (F0) and harmonics cannot be sufficiently resolved (Oxenham 2008). CI users have to rely more on concurrent cues as duration and amplitude envelope to perceive weak pitch information (Luo et al. 2008; Peng et al. 2017; Kim et al. 2021). This poses a unique challenge for CI users in pitch encoding and decoding, which is critical for a tonal language such as Mandarin Chinese that makes use of four types of pitch variations for conveying different words, namely, Tone 1 (T1) with a high-flat pitch, Tone 2 (T2) with a mid-rising pitch, Tone 3 (T3) with a falling-rising pitch, and Tone 4 (T4) with a high-falling pitch (Chao 1948; Wang 1973). Presumably, the phonemic status of lexical tones would place great reliance on proper extraction of the pitch patterns for listeners to understand spoken Chinese.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

A body of studies evaluated lexical tone performance in Mandarin-speaking CI users, which consistently demonstrated remarkable deficits in their tonal perception and production (Tan et al. 2016; Chen & Wong 2017; Gao et al. 2021 for reviews). It is noticeable that recognition of naturally spoken Mandarin tones in quiet can be relatively robust in CI users, with an average accuracy of above 75% (Peng et al. 2004; Tao et al. 2015; Zhang et al. 2020a), possibly due to the available duration and intensity cues that co-vary with F0 contour. However, CI users' tonal recognition outcomes become significantly poorer when they have to rely solely on F0 information (Luo et al. 2009; Wang et al. 2011; Wang et al. 2012). For instance, Luo et al. (2009) showed that lexical tone recognition was only 63% correct when test syllables were normalized to the same duration and amplitude. Therefore, formal auditory rehabilitation is needed to help the CI recipients perceive the distinctions between the pitch patterns in different contexts spoken by different speakers, and various training approaches have been developed and tested with different outcomes (e.g., Wu et al. 2007; Cheng et al. 2018; Kim et al. 2021; Zhang et al. 2021a). The present study on Mandarin-speaking kindergarteners with CIs aimed to evaluate the immediate and long-term efficacy of a widely acclaimed laboratory-based phonetic training protocol in improving their lexical tone perception with stringent and fine-grained assessments.

1.1 Pitch Training with Melodic Contour Identification Aiming for Transfer of Learning

Several recent studies have documented that formal auditory training is potentially beneficial for CI users' pitch reception (Gfeller et al. 2015; Lo et al. 2015; Good et al. 2017; Cheng et al. 2018; Lo et al. 2020). For example, Cheng et al. (2018) introduced a melodic contour identification (MCI) training to pediatric Mandarin-speaking CI users over a period of eight weeks. Nine pitch patterns were involved in the MCI training protocol (i.e., rising, rising-flat, rising-falling, flat-rising, flat, flat-falling, falling-rising, falling-flat, and falling), in which the child participants were instructed to identify the melodic contours using a closed-set, nine-alternative forced-choice (9 AFC) procedure. The results

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

suggested that the intensive music training could contribute to significant improvements in both the trained MCI task and proximally related tasks including lexical tone identification and sentence recognition (Cheng et al. 2018). Given that pitch cues are shared essential attributes by music and speech (especially tonal languages) with documented cross-domain transfer effects (Wu et al. 2015; Nan et al. 2018; Deroche et al. 2019b; Torppa & Huotilainen 2019; Wiener & Bradley 2020; Zhang et al. 2020d; Chen et al. 2021; Zhu et al. 2021), it seems plausible to speculate that gains in music training could transfer to pitch-related performances in speech perception for CI users.

However, music-training regimens cannot guarantee improvement in pitch perception in general with automatic cross-domain transfer of learning. This has been demonstrated in a cohort of adult CI users who participated in the MCI training with the same 9 AFC procedure (Fuller et al. 2018). While the trainees accrued significant benefits in the trained MCI task, there was no significant transfer of learning to either word/sentence recognition or vocal emotion identification (Fuller et al. 2018). Several possible reasons may account for the inconsistent transfer effects of the MCI music training. Different age groups were tested in Cheng et al. (2018) and Fuller et al. (2018), with the former testing children while the latter measuring adults. Children are prone to show overall superiority over adults in speech learning capacities, possibly due to their greater brain plasticity (Kuhl 2004). From this perspective, it is not unexpected to observe different transfer effects of MCI training between the two age groups. In addition, the two prior studies used different tasks to assess the generalization of MCI learning. Lexical tone identification in Cheng et al. (2018) could be more proximally related to the training protocol than vocal emotion recognition in Fuller et al. (2018), considering the range of emotion categories and the possibility of more complex pitch variations of emotions over tones (Luo et al. 2007). The transfer of learning to vocal emotion recognition might require more training sessions or more effective training protocols.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

In addition to the lack of robust generalization across the studies, the training protocols that had been implemented were relatively lengthy, requiring either daily training or long-term training sessions lasting for months. The required time commitment could place heavy burdens on the individual trainees (pediatric participants in particular), which, in turn, poses a great challenge in their acceptance and consistent compliance to the training protocol in research or rehabilitation practice. Nevertheless, findings of the previous studies have considerable clinical implications as they all point to the existence of substantial plasticity in CI users to improve their pitch reception following auditory training. To overcome the limitations, further research is needed to develop and test more effective and efficient training techniques to enhance pitch decoding as well as phonological representations of lexical tones in pediatric CI users under the stringent assessment criteria of robust generalization and reliable retention.

1.2 High Variability Phonetic Training Targeting at Robust Perceptual Learning

High variability phonetic training (HVPT) is a well-established training approach in perceptual learning of second language speech, which utilizes naturally produced speech sounds by multiple talkers in varying phonological contexts. The HVPT paradigm is typically efficient, with a training period of approximately five to ten days. Such limited training experience is highly efficient to induce tangible gains in speech perception and production (Ingvalson & Wong 2016; Sakai & Moorman 2018; Zhang et al. 2021c for reviews). The training-induced perceptual learning has been shown to generalize to novel talkers and untrained stimuli, with the benefits being maintained over several months (Logan et al. 1991; Bradlow et al. 1999; Iverson & Evans 2009). Initially, the HVPT approach was developed to train native Japanese speakers on the English /r/-/l/ distinction, a phonetic contrast that is especially difficult for Japanese learners (e.g., Logan et al. 1991; Lively et al. 1993; Lively et al. 1994; Bradlow et al. 1997; Bradlow et al. 1999; Iverson et al. 2005; Zhang et al. 2009). Afterwards, HVPT has become standard in second language phonetic training, and has been

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

successfully extended to train native English speakers on Mandarin tones (e.g., Wang et al. 1999; Wang et al. 2003; Wong & Perrachione 2007; Perrachione et al. 2011; Ingvalson & Wong 2013; Dong et al. 2019; Wiener et al. 2020).

The success of HVPT in second language learners has encouraged attempts of extending this training approach to CI users. Initial efforts were administered to evaluate the efficacy of HVPT on recognition of speech contrasts of /ba/–/da/ and /wa/–/ja/ in a group of postlingually deafened adults with CIs (Miller et al. 2016a; 2016b). More importantly, recent research from our lab validated the training-induced benefits in lexical tone perception for Mandarin-speaking pediatric CI users following a five-session HVPT (Zhang et al. 2021a). However, the pediatric trainees' perceptual improvement in identifying or discerning naturally produced tonal tokens/contrasts do not necessarily indicate that they have developed more robust pitch perception of Mandarin tones. Unlike normal-hearing individuals whose primary cue for Mandarin tones is the pitch contour, pediatric CI recipients tend to rely more on the secondary cues of duration and intensity to make the lexical tone distinctions in both perception and production (Peng et al. 2017; Deroche et al. 2019a). The secondary temporal cues would also be helpful for Mandarin tone recognition (Whalen & Xu 1992; Fu et al. 1998; Fu & Zeng 2000; Luo & Fu 2004). For instance, the intensity (i.e., amplitude envelope) tends to change synchronously alone with the F0 variations, and the duration of different Mandarin tones in natural speech also shows comparatively distinguishable patterns, especially for T3 (the longest on average) and T4 (the shortest tone on average). A more recent study by Kim et al. (2021) showed slight (but not significant) pre-post improvement in the cue weighting of either F0 contour or amplitude envelope after a 5-day tone recognition training with enhanced-amplitude or natural-amplitude stimuli. In this regard, the HVPT-related benefits in lexical tone perception observed in our prior study (Zhang et al. 2021a) could be due to an improved adaptation in the temporal dimension as opposed to a sharpened acuity in detecting the pitch contour patterns for the target lexical tones in naturally produced word stimuli. It is necessary to have strict control on the test materials and training protocol in order to determine whether the HVPT

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

training induces more accurate phonological representations of the lexical tones based on the primary pitch contours among the pediatric trainees with CIs.

1.3 Rationales of the Present Study

A number of recent studies have demonstrated that pitch contours play a pivotal role in Mandarin-speaking CI children's lexical tone perception (Peng et al. 2017; Deroche et al. 2019a) and sentence recognition (Zhang et al. 2018; Huang et al. 2020). For example, with orthogonally manipulated F0 contour and duration, Peng et al. (2017) investigated the cue weighting of Mandarin tone identification in pediatric CI recipients and NH peers. Their findings showed that CI children's lexical tone recognition is primarily correlated with their reliance on F0 variations, despite their significantly higher reliance on duration patterns than that of NH listeners. Huang et al. (2020) further revealed that the contribution of F0 contours in correct sentence recognition was more salient in children with CIs than in age-matched counterparts with NH. Therefore, it would be desirable to develop training protocols for tonal language speakers with CIs that would encourage the trainees to enhance their auditory sensitivity to pitch variations.

In the training literature, the categorical perception (CP) paradigm has been incorporated and well-established for a fine-grained assessment on the transfer of perceptual learning to retuned representations of the to-be-learned phonetic categories (e.g., Zhang et al. 2009; Sadakata & McQueen 2014; Miller et al. 2016a; Cheng et al. 2019; Zhang et al. 2021b). The classical CP paradigm consists of identification and discrimination tasks of the speech stimuli from a well-controlled synthetic speech continuum (Liberman et al. 1957). Stimuli along the continuum vary with equal physical intervals of the primary acoustic cue from one phoneme to the other, while keeping other acoustic cues constant. A plethora of studies have synthesized tonal continua with systematic variations in the pitch contour while keeping duration and intensity of the stimuli constant, which demonstrated robust CP of Mandarin tones among native listeners

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

with normal hearing (NH) (e.g., Wang 1973; Xu et al. 2006; Peng et al. 2010; Xi et al. 2010; Zhang et al. 2012; Chen et al. 2017; Yu et al. 2019; Chen & Peng 2021; Ma et al. 2021; Zhu et al. 2021; Feng & Peng 2022). Moreover, the CP paradigm of Mandarin tones has been successfully extended to individuals with CIs (e.g., Luo et al. 2014; Peng et al. 2017; Zhang et al. 2019a; Zhang et al. 2020c), showing that CI users exhibited impaired but improvable lexical tone categorization/normalization. Statistically, CP can be operationalized as significantly better discrimination for stimulus contrasts across different phonetic categories than for contrasts within the same category with equivalent physical distances. It is a fundamental mechanism underlying the phonetic categorization, which is strongly influenced by listener's native language experience and auditory deprivation (Harnad 2003; Goldstone & Hendrickson 2010; Zhang 2016 for reviews). Encumbered by auditory deprivations in early life and CI device-related limitations in pitch encoding, kindergarteners with CIs encounter consistent difficultly in lexical tone development that waits for promotion via efficacious interventions (Tan et al. 2016; Gao et al. 2021 for reviews). In a nutshell, the classical CP paradigm provides an ideal stringent test for the pediatric CI users to assess the training-induced benefits on pitch perception of Mandarin tones in terms of generalizability and sustainability.

While the formal auditory training has been advocated as a promising way to glean benefits for CI users with solid statistical evidence at the group level (Rayes et al. 2019; Drouin & Theodore 2020; Stropahl et al. 2020 for reviews), few studies have delved into explaining the differences observed among the individual trainees. As a clinical population with ubiquitous heterogeneity, the CI recipients show enormous individual differences and variability with respect to their speech, language, and hearing performances (Kral & O'Donoghue 2010; Peterson et al. 2010; van Wieringen & Wouters 2015 for reviews). It is common to observe some "star" CI performers who can achieve stunning levels on par with their age-matched NH controls, at least in quiet listening condition, mixed with some others who receive very limited benefits with their CI devices and even struggle to acquire elementary spoken language skills. A myriad of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

factors putatively contributes to the well-documented individual variabilities. However, apart from a set of conventional variables including demographics and hearing history (e.g., the age at implantation, duration of CI use, and amount of residual hearing), it still remains understudied with regard to the tremendous individual variability in speech rehabilitation following implantation. This clinical problem has been a long-standing issue and challenge for future research on CI population (Pisoni et al. 2017). In the same vein, there is a pressing need to gain a better understanding of what and how individual patient profiles may account for the training outcomes, which calls for proper statistical modelling to tease apart potential predictors for the magnitudes of perceptual learning following structured auditory training.

1.4 Research Questions and Hypotheses of the Present Study

This study was designed with the motivation to provide a rigorous evaluation of the efficacy of HVPT for Mandarin-speaking kindergarteners with CIs. It included four main research questions: (a) whether HVPT could improve lexical tone categorization in pediatric CI users that generalizes to untrained speech stimuli; (b) whether training-induced benefits would retain over a relatively longer period; (c) whether the training protocol could reliably improve pitch perception of lexical tones by the pediatric CI users; and (d) whether the different amounts of training-induced improvement in lexical tone categorization could be predicted by any individual patient variables. The HVPT procedure with naturally spoken speech was introduced to the child participants, and identical CP tests of well-controlled synthesized lexical tone stimuli were conducted before, immediately after, and 10 weeks post training termination (i.e., pretest, posttest, and follow-up test). To verify the training effects, we also recruited a control group of CI children who did not receive the training and examined the identical test sessions with the same time frame as the training group. We hypothesized that (a) the HVPT would improve the trained children's lexical tone categorization

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

from pretest to posttest; (b) the transfer of perceptual learning would retain for months; (c) the training-related benefits were mainly attributable to enhanced pitch perception; and (d) specific profiles of the clinical pediatric trainees would be significant predictors for the magnitude of HVPT-induced benefits in lexical tone categorization. Findings of the study have considerable clinical implications for spoken language rehabilitation in pediatric CI recipients.

2 Method

2.1 Participants

A total of 28 kindergarten-aged, Mandarin-speaking children (15 females and 13 males) with CIs participated in this study. They had little knowledge of sign language and used spoken Mandarin Chinese as the dominant mode of communication. The inclusion criteria consisted of prelingual (diagnosed before one year old), bilateral severe-to-profound hearing loss, and unilateral cochlear implantation. The exclusion criteria included any psychiatric and developmental disorders. The mean chronological age at testing was 4.98 years (age range = 4.08 – 5.83 years), the mean age at implantation was 1.57 years (age range = 0.92 – 3.33 years), and the mean CI duration was 3.41 years (range = 2.25 – 4.83 years). In addition, all pediatric participants reported normal nonverbal intelligence, with significantly higher scores than the passing criteria of 84 in the Hiskey–Nebraska test of learning aptitude (H-NTLA) (Hiskey 1966; Yang et al. 2011). The demographic data along with CI device information are shown in Table 1. The investigation was implemented in a sound-treated therapy room at Shanghai Rehabilitation Center of the Deaf Children. Informed consent was received in accordance with the Ethics Committee of School of Foreign Languages, Shanghai Jiao Tong University.

[Insert Table 1 around here]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

The pediatric CI recipients were quasi-randomly assigned into training and control groups (c.f., Mishra et al. 2015), with each group consisting of 14 participants. The quasi-random assignment allowed a close match between the two groups on relevant demographic characteristics and baseline perceptual performance. Student's t tests revealed that the two groups did not differ in terms of chronological age (training mean = 4.91 years, control mean = 5.06 years, p = 0.41), age at CI (training mean = 1.44 years, control mean = 1.7 years, p = 0.32), CI experience (training mean = 3.47 years, control mean = 3.36 years, p = 0.67), or H-NTLA score (training mean = 109, control mean = 111, p = 0.34). It should be noted that 18 of the CI children were bimodal users (i.e., fitting an additional hearing aid in the contralateral ear), with each group consisting of 9 such participants. These bimodal users completed all the tests and training sessions without wearing their hearing aids. Subject profiles of the two groups are shown in Table 2.

[Insert Table 2 around here]

2.2 Experimental Design

The child participants were requested to complete three test sessions including the pretest, posttest, and follow-up test. The pretest served as baseline measurement on the CP of lexical tones for each pediatric CI user, and the posttest was implemented after three weeks of the pretest completion to evaluate the training-induced gains. For the follow-up retention assessment, an interval of 10 weeks was set after the posttest. The training group participated in five HVPT sessions over a period of three weeks between the pre- and post- tests. By contrast, the control group did not receive the formal training, and was assessed with the identical pre- and post- tests in the same time frames as the training group.

2.2.1 Test Stimuli and Procedures

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

The test stimuli in the CP of lexical tones were a synthesized continuum of seven stimuli going from T1 to T2. Mandarin monosyllables /i/ with T1 and T2 were recorded in plain and clear forms from a native female adult in a sound-attenuated booth at a sampling rate of 44.1 kHz (16 bit). The F0 parameters of the two syllable samples of T1 and T2 served as endpoints for the synthesis of the tonal continuum. At the first step, the two syllable tokens were digitally edited using Praat (Boersma & Weenink 2017), each having a duration of 400 milliseconds (ms) and an RMS intensity of 65 dB SPL. Second, F0 measurement was performed respectively across the whole duration of the syllable tokens of T1 and T2. The F0 distance between T1 and T2 was equally divided into six steps in Hz to derive a seven-stimulus tonal continuum. Specifically, the F0 contour of each tonal stimulus linearly transitioned from the onset (ranging from 160 to 250 Hz, with a step of 15 Hz) to the fixed offset at 250 Hz. Third, based on the syllable token of T1, the F0 contour was systematically replaced with the corresponding contour of each stimulus along the tonal continuum, using the Pitch-Synchronous Overlap Add (PSOLA) method in Praat (Boersma & Weenink 2017). The seven equidistant speech stimuli were generated with variations only in F0 contour, while keeping other acoustic cues constant. These stimuli share a relatively flat amplitude envelope, because they were re-synthesized on the basis of the same syllable token of T1 (see supplemental Appendix A). The schematic diagram of the F0 contours of the seven tonal stimuli along the speech continuum is illustrated in Figure 1. Stimulus No. 1 represents the prototypical T1 and No. 7 represents the prototypical T2. Feasibility of the tonal continuum and test protocols has been confirmed in our previous studies to examine the CP of Mandarin tones in pediatric CI users (Zhang et al. 2019a; 2020b; 2020c).

[Insert Figure 1 around here]

The kindergarten-aged participants were measured with the classical identification and discrimination tasks for the CP of speech sounds presented in a counter-balanced order. Both tasks were administrated in E-Prime 2.0 program (Psychology Software Tools Inc., USA) on a laptop. A loudspeaker (JBL CM220) was used to deliver the auditory stimuli

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

of the two tasks, which was placed in front of the listener with a distance of approximately 1.2 m. In the identification task, a two-alternative forced choice (2 AFC) paradigm was adopted. The participants were asked to recognize the tonal identity of the separately presented stimulus via choosing a matching picture of the tonal stimulus from the two pictures of driving cars. Initially, the experimenter explained the relationship for the sound-picture matching to each child participant: the picture of a car driving on a level road indicates T1 and the other picture of a car driving on a rising slope indicates T2. Meanwhile, an AX paradigm was employed in the discrimination task. The AX discrimination paradigm uses stimulus contrasts to direct the listener to judge whether stimulus X is the same as stimulus A without having to resort to a labelling strategy (Gerrits & Schouten 2004). Seven "same" contrasts (i.e. 1-1, 2-2, 3-3, 4-4, 5-5, 6-6, and 7-7) and ten "different" contrasts (i.e. 1-3, 3-1, 2-4, 4-2, 3-5, 5-3, 4-6, 6-4, 5-7, and 7-5) were constructed with an inter-stimulus interval of 500 ms for each contrast. In previous studies using the CP paradigm for lexical tones, the "different" contrasts were typically constructed with two stimuli separated by two steps along the tonal continuum (c.f., Peng et al. 2010; Chen et al. 2017; Chen & Peng 2021; Ma et al. 2021; Feng & Peng 2022). Given the ability of CI participants as reported in the literature and tested in our lab, they could barely hear the distinctions between two stimuli of the "different" contrasts if the separation was only one step apart (i.e., acoustic difference of 15 Hz). Likewise, the child listener was instructed to perform sound-picture matching for the current AX discrimination task. An explicit explanation was offered initially to help child participants match the picture that had two identical fruits (i.e., apple and apple) with the trial of "same" sounds, and the picture that had two different fruits (i.e., apple and orange) with the trial of "different" sounds.

Before each test session, a practice session with trial-by-trial feedback was provided for the two tasks, respectively, to ensure that the child participants could follow the instructions. In the identification practice, the two endpoints of the tonal continuum (i.e., stimulus No. 1 and 7) were selected as target stimuli, with each stimulus repeating four times

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

(eight practice trials in total). In the discrimination practice, four contrasts (i.e., 1-1, 7-7, 2-4, and 5-3) were used, with each contrast repeating twice (eight practice trials in total). The test sessions of the identification task were initiated after an accuracy of 90% was obtained for the practice session, whereas the initiation of discrimination tests required no such an accuracy criterion on the corresponding practice session. The accuracy criterion of 90% for identification task was set to guarantee that the child participants were familiar with the test procedure and were able to concentrate on the experimental task. This criterion was selected because our pilot study showed that the child participants could achieve a ceiling level in recognizing stimuli with T1 (with an accuracy of over 96%). In addition, several prior studies also adopted the criterion of 90% accuracy in practice to ensure that child participants could understand the tasks and follow instruction consistently (e.g., Chen et al. 2017; Ma et al. 2021; Feng & Peng 2022). The identification test used all seven stimuli of the tonal continuum, and there were two presentation blocks of 70 trials (10 trials for each sound). In the discrimination test, a presentation block consisted of 85 trials for the 17 constructed tonal contrasts (7 same contrasts and 10 different contrasts, 5 trials for each contrast), and two presentation blocks were used. The tonal stimuli/contrasts were presented in a random order. The participant's trial-by-trial sound-picture matching responses were logged by the experimenter via pressing the corresponding keys on the keyboard. In order to avoid fatigue, the two tasks were measured in two consecutive days, and a two-minute break was offered between blocks of each task.

2.2.2 Training Stimuli and Protocol

The training stimuli consisted of Mandarin monosyllables /i/, /a/, and /u/, which all were produced with four lexical tones. The 12 training syllables were obtained from 10 native speakers of Mandarin Chinese (5 males and 5 females), with each syllable recorded five samples per speaker. This resulted in a total of 600 training items with high variability features that incorporated multiple talkers and varied phonological contexts. All training stimuli were recorded in a sound-treated booth using an audio interface (Mbox Mini) coupled with a microphone (AKG C544L), at

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

a sampling rate of 44.1 kHz (16 bit). These stimuli were normalized with an RMS intensity level of 65 dB SPL, while keeping the natural duration of each stimulus. During training sessions, the auditory stimuli were presented through a loudspeaker (JBL CM220) that was placed approximately 1.2 m from the trained kindergartener with CIs.

The training group was introduced to complete five sessions of HVPT, and the training program was developed and implemented using E-Prime 2.0 on a laptop (c.f., Miller et al. 2016a; 2016b). Identification training was used in this study, because as opposed to discrimination training, identification training is more naturalistic and encourages listeners to attend more to the phonetically dependent differences across the stimuli, rather than the subtle independent differences within the abstract category (Lively et al. 1993; Pisoni & Lively 1995). A four-alternative forced-choice (4 AFC) paradigm was adopted in the identification training task to identify each stimulus as T1, T2, T3, or T4. A sound-picture matching task was administered, in which the trained children were instructed to choose a matching picture of the presented stimulus from four track pictures of driving cars (mimicking the pitch trajectory of the target lexical tones) on the laptop screen. The four pictures of driving cars are typical pedagogical materials used in formal school settings when introducing the four lexical tones in Mandarin Chinese, because they explicitly show the primary characteristics of pitch height and contour of different tone types. The relationship for the sound-picture matching was explicitly explained to the CI children before the implementation of each training session, 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. During the training sessions, trial-by-trial feedback was provided. In addition, trial repetition was given accordingly: one repetition was offered for trials with correct responses, and two repetitions were offered for those with incorrect responses.

In addition to the high-variability feature and the provision of feedback and repetition, adaptive scaffolding was incorporated into the training protocol by controlling the extent of talker variability within each training session/block

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

(Zhang et al. 2009). The first training session consisted of 120 training items from two talkers (one male and one female) implemented in six blocks. Afterwards, additional 120 training items from two new talkers (one male and one female) were added in the following training sessions, until all 10 talkers were incorporated in the last session. This protocol was employed because the blocked delivery with incremental variability has been demonstrated to be beneficial for phonetic learning, especially among the trainees with low pre-training aptitude (Perrachione et al. 2011; Sadakata & McQueen 2014; Fuhrmeister & Myers 2020). A quiz of 24 training items from the lately added two talkers (12 trials per talker) was presented for the trained children. If the child scored 85% or higher on the identification quiz, the trainee could proceed to the next training session. Otherwise, the trained child was required to repeat the current training session once and, thereafter, proceed to the next training session. In this procedure, the actual amount of training might differ across the trainees depending on how many repetition sessions they went through. Specifically, eight of them repeated Session 1, five of them repeated Session 2, three of them repeated Session 3, two of them repeated Session 4, but none repeated Session 5 (see supplemental Appendix B). It is noteworthy that Pearson's correlation analyses revealed that the actual amount of training sessions was not significantly correlated with training outcomes in terms of pre-post changes in either boundary width (r = -0.01, p = 0.98) or peakedness score (r = 0.2, p = 0.48), two parameters for the evaluation of the degree of CP (Chen & Peng 2021). The CI children from the training group completed the five HVPT sessions at their own pace over a period of three weeks. In general, the trained children completed one training session every two to four days.

2.3 Scoring and Statistical Analyses

The identification and discrimination data in three test sessions with identical test conditions (i.e., pretest, posttest, and follow-up test) were scored and analyzed for each participant. Two key parameters of the identification data were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

estimated from probit analysis (Finney 1971), including boundary position and boundary width, which are widely used to evaluate the identification performance in the CP paradigm (e.g., Peng et al. 2010; Shen & Froud 2016; Zhang et al. 2016; Chen et al. 2017; Zhang et al. 2017; Yu et al. 2019; Zhang et al. 2020c; Chen & Peng 2021; Ma et al. 2021; Feng & Peng 2022). The boundary position refers to the 50% crossover point of the identification curves, and the boundary width refers to the linear distance in terms of the stimulus step between 25% and 75% of the identification responses (Peng et al. 2010). The discrimination data were analysed with the d' measure, which is the sensitivity index and takes response bias into consideration (Macmillan & Creelman 2005; Shao et al. 2019). The d' is calculated as the z-score of the hit rate (correct responses to “different” condition) minus that of the false alarm rate (incorrect responses to “same” condition) for each stimulus contrast along the tonal continuum. Raw score adjustment was performed to avoid infinite values when the discrimination percentage accuracy was 0% or 100%. If that is the case, 0% is adjusted to 1% and 100% is adjusted to 99% to set effective floor value (d' = -4.65) and ceiling value (d' = 4.65) (Macmillan & Creelman 2005). More importantly, based on the boundary position of the identification responses for each participant, the discrimination data were divided into two types, including the between-category type and the within-category type (Jiang et al. 2012; Chen et al. 2017). The between-category discrimination score is the mean d' of stimulus contrasts spanning two tonal categories, and the within-category discrimination score is the mean d' of contrasts falling in the same category. For example, if the boundary position of a participant was 3.69, then for this child, the between-category score was the averaged d' of stimulus contrasts of 2-4 and 3-5, and the within-category score was the averaged d' of the contrasts of 1-3, 4-6, and 5-7. In addition, the peakedness score, which represents the magnitude of benefit of discrimination, was also calculated by subtracting the mean d' of within-category type from that of between-category type (Jiang et al. 2012; Chen et al. 2017). It reflects the degree of enhanced between-category discriminability relative to within-category discriminability for categorical perception. That is, a higher peakedness score means much greater sensitivity to the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

difference of between-category stimulus contrasts than that of within-category contrasts, which indicates that the perception of the target phonemes is more categorical.

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 assess the CP of lexical tones in different test sessions for the training and control groups. The ANOVA function in lmerTest package (Kuznetsova et al. 2017) was implemented to estimate *F* and *p* values of the significant fixed factors with α setting at 0.05. The emmeans package (Lenth et al. 2018) with false discovery rate (FDR) correction (Benjamini & Yekutieli 2001) was adopted to conduct post hoc multiple comparisons for the significant fixed factors to obtain *t* ratios, *p* values, and Cohen's *d* (for effect size evaluation). In addition, the package of Superpower (Lakens & Caldwell 2021) was used to estimate effect size (Cohen's *f* for fixed factors) and statistical power of the significant results.

To probe contributors to individual differences, linear regression models were constructed in R to examine the potential variables predicting the training-induced benefits of the CP of lexical tones in the trained children with CIs. Duration of CI use and baseline performance (i.e., CP results in the pretest session) acted as hypothesized predictors in the models. Evaluation of the tonal CP performance included benefit magnitude in terms of boundary width (i.e., difference of boundary width between pre- and post- tests) and in terms of peakedness score (i.e., difference of peakedness score between pre- and post- tests). The parameters of boundary width and peakedness score were selected because of their robust sensitivity in evaluating the degree of CP (Chen & Peng 2021). Separate regression models were created for each estimate of CP performance with all the hypothesized predictors added as fixed effects. Additional linear regression models were built to explore whether CI children's demographic characteristics could contribute to the individual differences in their CP performance of lexical tones in the pretest session, with implanted age and duration of CI used as hypothesized predictors. Similarly, separate models were created for the parameters of boundary

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

width and peakedness score. *F*-statistics and *p*-values for the fixed effects were assessed. Estimated coefficients (β), standard errors (*SE*), *t*-values, and *p*-values for the fixed effects were assessed with significance level $\alpha = 0.05$.

3 Results

3.1 Pre-post Test Effects of HVPT on Stimulus Identification and Discrimination

The grand averaged identification data (i.e., boundary position and boundary width) for the two groups in pre- and post- tests are collectively shown in Figure 2. To assess the training effects, LME models were constructed to analyze the perceptual pre-post changes on the CP of lexical tones for the two groups. The models were created on boundary position and boundary width separately with Test Session (pretest and posttest) and Group (training group and control group) as fixed factors, and with participant as a random factor. For boundary width analysis, there was a significant main effect of Test Session ($F(1, 28) = 14.06$, $p < 0.001$, Cohen's $f = 0.58$, power = 80.8%) and a significant interaction effect of Test Session by Group ($F(1, 28) = 13.06$, $p = 0.001$, Cohen's $f = 0.56$, power = 77.9%), but the main effect of Group was insignificant ($F(1, 28) = 1.69$, $p = 0.2$). Post hoc pairwise comparisons showed significantly narrower boundary width in the posttest than in the pretest for the training group ($t(30) = 5.02$, $p < 0.001$, Cohen's $d = 1.97$, power = 97.5%), but not for the control group ($t(30) = 0.09$, $p = 0.93$). Meanwhile, the training group showed significantly narrower boundary width than the control group in the posttest session ($t(48) = 2.81$, $p = 0.007$, $d = 1.57$, power = 29.1%), but not in the pretest session ($t(48) = -0.64$, $p = 0.53$). However, the boundary position analysis showed that neither main effect of Test Session ($F(1, 28) = 1.27$, $p = 0.27$), or Group ($F(1, 28) = 0.25$, $p = 0.62$), nor their interaction effect ($F(1, 28) = 0.05$, $p = 0.83$) was significant.

[Insert Figure 2 around here]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

The results of discrimination data (i.e., within-category score and between-category score) for each group from pre- to post- tests are illustrated in Figure 3. The LME models were constructed using Discrimination Type (within-category type and between-category type), Test Session (pretest and posttest) and Group (training group and control group) as fixed factors, and using participant as a random factor. The main effects of Test Session ($F(1, 84) = 14.34$, $p < 0.001$, $f = 0.68$, power = 31.7%) and Discrimination Type ($F(1, 84) = 70.37$, $p < 0.001$, $f = 1.51$, power = 99.9%) were significant, but not the main effect of Group ($F(1, 28) = 14.34$, $p = 0.33$). Moreover, the interaction effects of Test Session by Discrimination Type ($F(1, 84) = 3.51$, $p = 0.06$, $f = 0.34$, power = 27.4%) and Group by Test Session by Discrimination Type ($F(1, 84) = 3.42$, $p = 0.07$, $f = 0.28$, power = 20.3%) approached significance, but not the interaction effect of Group by Discrimination Type ($F(1, 84) = 0.37$, $p = 0.55$) or Group by Test Session ($F(1, 84) = 2.46$, $p = 0.12$). Further post hoc multiple comparisons indicated that between-category discrimination improved significantly from pre- to post- tests in the training group ($t(91) = 4.38$, $p < 0.001$, $d = 1.03$, power = 66.9%), but not in the control group ($t(91) = 1.08$, $p = 0.28$). In addition, discrimination score of between-category type was significantly higher than that of within-category type ($t(91) = 8.08$, $p < 0.001$, $d = 1.59$, power = 99.9%) across different groups and test sessions. However, the difference on within-category discrimination between the pretest and posttest sessions was insignificant either in the training group ($t(91) = 0.79$, $p = 0.43$) or in the control group ($t(91) = 1.06$, $p = 0.29$). Meanwhile, additional LME models were constructed on the discrimination scores of the training group, with Tonal Contrast (i.e., 1-3, 2-4, 3-5, 4-6, and 5-7), Test Session, and their interaction as fixed factors. Statistical results revealed that the pre-post discrimination improvement was significant in stimulus contrast 3-5 ($t(136) = 3.43$, $p = 0.001$, $d = 1.34$, power = 88.5%), and was approximately significant in contrast 2-4 ($t(136) = 1.98$, $p = 0.05$, $d = 0.78$, power = 44.8%). Intuitively, it may be confusing to observe a significant pre-post improvement in discrimination of stimulus contrast 3-5 ($p = 0.001$) but not in that of contrast 2-4 ($p = 0.05$), although both contrasts represent between-category type at the group level. However, we found that three of the trained children in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

the pretest and five of them in the posttest perceived contrast 2-4 as within-category type while inspecting individual data. The unstable status of stimulus contrast 2-4 across the trainees might contribute to its insignificant discrimination changes after training.

[Insert Figure 3 around here]

Overall, the results suggested that the training group and control group showed comparable baseline performance on the CP of lexical tones. However, the pre-post changes were only significant in the training group but not in the control group. Trained children with CIs enhanced lexical tone categorization following HVPT, showing a narrower boundary width (i.e., sharper identification slope) and a higher between-category score from pre- to post-tests.

To illustrate inter-subject variability, individual identification functions in conjunction with the corresponding boundary widths for each child participant in different test sessions (pretest, posttest, and follow-up test if applicable) are shown in Figure 4. In accordance with the well-established heterogeneity in pediatric CI users, a wide range of outcomes of the tonal CP was revealed for the kindergarten-aged children with CIs across all test sessions. However, all 14 trained children accrued training-induced gains with a relatively smaller boundary width in the posttest and follow-up test relative to in pretest.

[Insert Figure 4 around here]

3.2 Retention of Training Benefits

Each group had eight child participants who completed all three test sessions (i.e., subjects t1 to t8 form the training group and c1 to c8 form the control group in Table 1). Perceptual performances of these 16 pediatric CI users were analyzed to assess the stability of training-related improvements in the CP of lexical tones. These children’s

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

identification and discrimination performances are illustrated in Figure 5 based on different groups. To evaluate the retention of the HVPT benefits, LME models were constructed to analyze the perceptual differences between the two groups across different test sessions. For identification analysis, the models were created with Test Session (pretest, posttest, and follow-up test), and Group (training group and control group) as fixed factors, and with participant as a random factor. The results indicated a significant main effect of Test Session ($F(2, 32) = 7.51$, $p = 0.002$, $f = 0.54$, power = 67.6%) and a significant interaction effect of Test Session by Group ($F(2, 32) = 4.22$, $p = 0.02$, $f = 0.4$, power = 42.6%) on boundary width analysis, whereas the main effect of Group ($F(1, 16) = 1.73$, $p = 0.2$) was insignificant. Post hoc pairwise comparison showed that the boundary width became significantly narrower from the pretest to posttest sessions ($t(37) = 4.0$, $p < 0.001$, $d = 2.14$, power = 83.4%) and the changes remained significant from the pretest to follow-up test ($t(37) = 3.64$, $p = 0.001$, $d = 1.95$, power = 76%) in the training group. By contrast, the differences in boundary width across the three test sessions were insignificant in the control group (*ps* > 0.7). For discrimination analysis, the LME models were constructed using Discrimination Type (within-category type and between-category type), Test Session (pretest, posttest, and follow-up test) and Group (training group and control group) as fixed factors, and using participant as a random factor. The results showed significant main effects of Discrimination Type ($F(1, 80) = 39.42$, $p < 0.001$, $f = 1.49$, power = 92%) and Test Session ($F(2, 80) = 8.77$, $p < 0.001$, $f = 0.7$, power = 55.1%), but not Group ($F(1, 16) = 1.27$, $p = 0.28$). None of the interaction effects was significant (*Fs* < 0.65, *ps* > 0.5). Further analyses revealed that between-category discrimination improved significantly from pre- to post- tests ($t(91) = 2.44$, $p = 0.03$, $d = 1.09$, power = 39.4%) and from pre- to follow-up tests ($t(91) = 2.88$, $p = 0.01$, $d = 1.19$, power = 45.2%) in the training group, but not in the control group (*ps* > 0.2). Taken together, the training-induced improvements in the CP of lexical tones sustained to the follow-up test session in the training group, whereas the changes across the three test sessions were insignificant in the control group.

[Insert Figure 5 around here]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

3.3 Regression Analysis Results

The mixed effects linear regression models could take the mutual influence of multiple predictors into consideration, which was supposed to be more reliable than traditional correlation analysis like Pearson/Spearman's correlation (Koerner & Zhang 2017). The current linear regression results are collectively arranged in Table 3. The results indicated that baseline performance (i.e., boundary width in pretest session) (β = -0.89, SE = 0.11, t = -8.06, p < 0.001) duration of CI use (β = -0.27, SE = 0.12, t = -2.27, p = 0.04) were significantly associated with the training-related benefit magnitudes in identification parameter of boundary width. Adjusted R^2 was 0.83 for the formula including the variables of baseline performance and duration of CI use, which means that the two variables together could account for 83% of the variance for the trained children's pre-post changes of boundary width. The negative regression coefficients (β) indicated that the trained children with wider boundary width in pretest session or with longer use of their CI device tended to show sharper changes in identification slope. However, neither of the hypothesized predictors was significantly correlated with the benefit magnitudes of peakedness score (see Table 3). In addition, neither implanted age nor duration of CI use was significantly correlated with the CP performances in the pretest session (see supplemental Appendix C). Visual inspection of residual plots revealed that the residuals were normally distributed without any obvious deviations from homoscedasticity for each regression model.

[Insert Table 3 around here]

4 Discussion

One incentive for this study was to provide a stringent assessment on the effectiveness of HVPT in improving lexical tone perception in native Mandarin-speaking kindergarteners with CIs. For this purpose, identification and discrimination of tonal stimuli/contrasts along a synthesized speech continuum of T1 and T2 were measured in three

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

test sessions (i.e., pretest, posttest, and follow-up test). These duration- and intensity- normalized speech stimuli were never used in the training. The results largely supported our hypotheses and suggested that identification training approach robustly enhanced pitch perception of lexical tones independent of duration and intensity cues in pediatric CI recipients with prelingual deafness. Moreover, duration of CI use and baseline performance in pretest session could significantly predict the magnitude of pre-post changes in lexical tone categorization among the individual trainees. To the best of our knowledge, our report represents the first formal documentation of sustained benefits as shown in enhanced categorical perception (CP) of lexical tones following structured auditory training in pediatric CI users.

4.1 Robust Transfer and Reliable Retention of HVPT in Improving Lexical Tone Categorization in Children with CIs

In the present study, five sessions of HVPT significantly enhanced lexical tone categorization in Mandarin-speaking pediatric CI users. A critical marker of successful speech training is the robust transfer of perceptual learning to novel, untrained speech stimuli (Pisoni & Lively 1995). Moreover, the long-term objective of any formal training regimen targeting CI population is to improve performance beyond the immediate effect of practice alone (Pisoni et al. 2017). As lexical tone perception could rely heavily on duration and intensity cues other than the primary pitch contour patterns for CI users (Peng et al. 2017; Deroche et al. 2019a), the test stimuli of the three test sessions (i.e., pretest, posttest, and follow-up test) in the present study were designed to control for the duration and intensity cues and were never used in the training protocol. This experimental design allowed us to assess whether the HVPT protocol actually improved the trainees' sensitivity to and reliance on the pitch contour patterns for lexical tone perception and whether the training effects were sustainable.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

The current report focused on pre-post changes for identification and discrimination performances of a set of synthetic speech stimuli that were significantly different from the naturally produced stimuli of the original training protocol. The results revealed that the trained children showed significantly narrower identification boundary widths and higher between-category discrimination scores in the posttest relative to the pretest session. By contrast, there was no significant pre-post change for either identification or discrimination data in the control children who did not receive HVPT (see Figures 2 and 3). Narrower boundary widths and higher between-category scores are indicators of enhanced CP, which, respectively, correspond to steeper identification slopes and more prominent discrimination peaks. The hallmarks of CP consist of a sharp membership change of the two phonetic categories in the identification function and a distinct peak around the category boundary in the discrimination function (Liberman et al. 1957). It follows that the trained children with CIs perceived lexical tones more categorically following the intensive identification training. The results are also in line with the native language magnet theory with the training experience driving the perceptual space between phonemes to be "warped" more categorically, amplifying differences across different phonetic categories and deemphasizing differences within the same category (Kuhl 1991; Kuhl et al. 2008). As the speech continuum manipulates the key acoustic dimension while controlling other acoustic cues, CP represents a high-level phonological processing that underlies the perceptual abstraction of phonetic categories based on the primary acoustic cue of interest for the phonemic contrast. As the synthetic stimuli were never used in the training, the evident improvement in lexical tone categorization was indicative of robust transfer-of-training, rather than simple "practice effects" of the identification training protocol or memorization of the trained stimuli.

Apart from the significant pre-post improvement in between-category discrimination (*Mean* = 1.33, *SD* = 0.75 in pretest; *Mean* = 2.38, *SD* = 0.99 in posttest), there is an intriguing numerical (but not significant) increase in within-category discrimination (*Mean* = 0.72, *SD* = 0.52 in pretest; *Mean* = 0.91, *SD* = 0.35 in posttest) for the trained group (see

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Figure 3). This finding was largely consistent with several prior studies that showed slight and insignificant improvement in the amplitude of mismatch negativity (MMN) for within-category stimulus pairs after training (e.g., Miller et al. 2016a; Cheng et al. 2019). Exemplar-based models posit that matured listeners store the fine-grained phonetic details in memory, instead of discarding these details, and in turn incorporate the detailed within-category information into speech categorization (Goldinger et al. 1991; Johnson 1997; Goldinger 1998; Zhang & Chen 2016). For instance, recent eye-tracking research with visual-world paradigm demonstrated that within-category tonal information could influence sound categorization process and modulate lexical activation in the recognition of Mandarin-Chinese words by native adult listeners (Qin et al. 2019). Therefore, improved sensitivity to within-category differences could be beneficial for maintaining flexibility while coping with uncertainty (e.g., phonetic variances result from multi-talker variability), which might allow listeners to develop more confident (i.e., more categorical) perception of the phonemes they hear (McMurray et al. 2018). However, it should be noted that the perceptual enhancement for within-category difference in this study may not necessarily attribute to the HVPT protocol, but rather could be due to the test-retest practice effect, since the control group also exhibited a numerical increase in within-category discrimination from pretest (*Mean* = 0.67, *SD* = 0.55) to posttest (*Mean* = 0.92, *SD* = 0.45). Future visual-world eye-tracking research focusing on the investigation of within-category tonal information (c.f., Qin et al. 2019) can be conducted to test whether auditory training may improve deployment of the fine-grained information to facilitate lexical tone categorization.

Furthermore, the training-induced improvement in lexical tone categorization persisted 10 weeks after the training had ceased (see Figure 5). In a state-of-the-art review about the benefit of structured auditory training on hearing impaired individuals, only six out of 16 studies assessed long-term effects of auditory training, and half of the six studies observed sustained benefit over time after the termination of the training regimen (Stropahl et al. 2020). The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

retention effect in our study was consistent with a previous report, which also showed that improved lexical tone recognition from a 10-week computer-assisted speech training sustained for two months in a group of Mandarin-speaking hearing-impaired children (7 CI recipients and 3 hearing aid users) (Wu et al. 2007). The training protocol in Wu et al. (2007) also incorporated high variability training materials with four tones for six Mandarin vowels (/a/, /o/, /e/, /i/, /u/, and /ü/) recorded from four native speakers. However, the evaluation of training benefits was limited to the trained task of naturally produced tone recognition in the study by Wu et al. (2007). Our findings confirmed the efficacy of HVPT with a stringent assessment on the tonal CP performance, a testing paradigm with speech stimuli that were duration- and intensity- normalized unlike those naturally recorded speech sounds in the training sessions. In addition, the recruitment of a control group without receiving formal training in this study could help rule out the possibility of perceptual gains from the procedural learning in Wu et al. (2007) or from some form of test-retest effect.

Multiple features may have contributed to the robust learning effects observed in our study. The progressive and adaptive block-design structure of HVPT with the use of multiple talkers and phonological contexts in this study enhanced pitch perception of lexical tones in pediatric trainees with CIs. As language learning is inherently multimodal (Kuhl 2000), the identification training protocol incorporated explicit instructions on pitch patterns by showing visual cues (pedagogical pictures depicting F0 characteristics of height and contour) to draw distinctions between different tone types. HVPT with explicit instruction on visual F0 information has been proven beneficial to drawing native English learners' attention to F0 contours while learning Mandarin tones (Wiener et al. 2020).

4.2 Contributing Factors for Training-Induced Benefits in the CP of Lexical Tones

The overwhelming bulk of CI research has focused on training-induced benefits at the group level with few studies offering analysis and interpretation on the contributors to the individual differences in perceptual learning. The mixed-effects linear regression analyses in this study attempted to explore the potential factors predicting the magnitude of pre-post changes in lexical tone categorization among the pediatric trainees, and the data revealed two significant predictors, namely, baseline performance and duration of CI use. We found that pediatric CI individuals with relatively poorer performance in the CP of lexical tones in the pretest session tended to show greater pre-post improvements in their tonal CP outcome (i.e., more pre-post changes in identification boundary width) than those who started at a relatively higher level of performance. It is not surprising to see this type of negative relationship as the size for potential improvement is relative to the baseline condition (Zhang et al. 2020a). In addition, we also observed that an increase in duration of CI use contributed to the magnitude of benefits in lexical tone categorization following the HVPT protocol. Longer duration of CI experience in pediatric users would presumably result in better acclimation to their device. A recent study suggested that pediatric CI recipients showed a refined control of F0 fluctuations in lexical tone pronunciation along with longer experience of their CI device (Deroche et al. 2019a). It follows that a refinement of F0 information in more experienced CI children (i.e., pediatric CI users with longer duration of their devices) would be better at extracting pitch contour patterns from the highly variable input delivered by the intensive and repetitive identification training of lexical tones. In studying the developmental trajectory of CP of Mandarin tones in NH children from four- to seven-year-old (Chen et al. 2017), it has been shown that more categorical-like perception on lexical tones stems from the accumulating exposure to the tonal language. However, one should interpret the regression data with great caution, since longer duration of CI use may not necessarily correlate with greater reliance on F0 contours in lexical tone perception by children with CIs (Peng et al. 2017). Our regression result showed that none of the hypothesized factors could significantly predict the pre-post changes in the peakedness scores. This result echoed some earlier reports which showed that neither baseline performance nor demographic characteristics was significantly correlated with the magnitude of training benefits (Mishra et al. 2015; Mishra & Boddupally 2018). One possibility for the discrepant

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

findings between boundary width and peakedness score in this study might lie in the different complexity between identification and discrimination tasks. Identification and discrimination tasks poses different demands on sensory memory, long-term phonological memory, and working memory (Xu et al. 2006). For instance, in a discrimination trial with two stimuli, listeners need to transitorily memorize the two successive presented sounds for comparison. In other words, the discrimination process can be more demanding in terms of greater cognitive resources such as working memory to remember the acoustic and phonetic details of the two stimuli in each trial. It is not uncommon that discrimination scores are not fully predictable from identification scores in CP tests. In our study, the situation could be further influenced by the age factor as we tested kindergarten-aged participants. Further investigations with full consideration of cognitive factors in relation to participant age are warranted to gain a better understanding of the individual differences in training effects.

4.3 Clinical Implications and Future Directions

Findings of this study are of potential clinical significance for pediatric CI recipients who are native speakers of tonal languages. Consistent with the well-documented heterogeneity among children with CIs, individual results in our study revealed a wide range of outcomes for tonal CP in both the training and control groups. Despite the heterogeneity, all trained kindergarteners with CIs accrued training-induced benefits in lexical tone categorization, which were reflected in the relatively narrower boundary widths (i.e., steeper identification slopes) in posttest/follow-up test than in pretest. Our results validated the efficacy of HVPT in improving the pitch perception of Mandarin tones in native pediatric CI recipients with prelingual deafness, because effective reception of pitch information is a prerequisite for correct speech understanding in Mandarin (Huang et al. 2020). Our data add to the growing speech training literature

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

and lend support to the inclusion of high variability identification training of lexical tones in the rehabilitative regimens for children with CIs.

Several limitations need to be acknowledged. This HVPT study followed the seminal work of Lively et al. (1993) that originally aimed at overcoming phonetic difficulties in second language learners. In this line of work, higher variability in the language input is assumed to be more beneficial in phonetic learning and generalization although few studies have actually tested this assumption (Giannakopoulou et al. 2017). A handful of recent studies have directly contrasted the use of high and low variability training protocols, which revealed insignificant benefits for phonetic training with high variability (e.g., speech from four talkers) over low variability (e.g., speech from a single talker) in learning of non-native phonetic contrasts (Giannakopoulou et al. 2017; Dong et al. 2019; Wiener et al. 2020; Zhang et al. 2021b; Brekelmans et al. 2022). Some cross-linguistic studies reported an interaction between variability of training materials and perceptual aptitude of trainees (i.e., baseline ability for perceiving pitch) while learning non-native lexical tones (Perrachione et al. 2011; Sadakata & McQueen 2014; Dong et al. 2019; Qin et al. 2021). Learners with high aptitude tended to benefit from high-variability training whilst those with low aptitude gained more benefits from low-variability training. While nearly all HVPT studies used the identification training protocol, a recent study compared discrimination training with traditional identification training and demonstrated comparable effects for the two training approaches, as long as high variability was incorporated in the training (Shinohara & Iverson 2018). Further training research is needed to determine the optimal dosage of variability and test the feasibility of the discrimination training task for the CI population, which hold promise in promoting efficient and customizable intervention to take into account individual differences and facilitate speech learning on an individual basis. As each training session is rather short, future intervention studies can implement this computer-based training protocol and test its suitability for extended access to care via telepractice in auditory rehabilitation (Goehring et al. 2012; Galvan et al. 2014; Völter et al. 2021). In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

addition, the current study leaves open whether CI recipients' use of secondary acoustic cues in recognizing Mandarin tones could also be improved following the HVPT with training materials of natural speech that included duration and intensity cues. A recent study by Kim et al. (2021) has demonstrated that short-term intensive training could slightly but not significantly improve cue-weighting of F0 contour and amplitude envelope in NH adults recognizing CI simulations of Mandarin tones. In studies of second language learning and phonetic training, it is shown that non-native listeners may demonstrate enhanced sensitivity to the secondary acoustic cues (Zhang et al. 2009; Zhang et al. 2019b; Zhang et al. 2022). It is possible that the HVPT protocol might significantly increase the sensitivity of secondary acoustic cues of lexical tones among the pediatric trainees with CIs. Further investigations are needed to verify this possibility with test stimuli orthogonally manipulated F0 contour and amplitude envelope (c.f., Kim et al. 2021).

Although our training study revealed robust benefits of HVPT in lexical tone categorization for pediatric CI recipients, it remains unclear whether the HVPT protocol can be extended to the training of consonants and vowels that have entirely different acoustic cues. Moreover, assessments of training benefits need to take into account a variety of sensory, linguistic and cognitive abilities, but not just the abilities that are proximally related to the perceptual training protocol (Ingvalson & Wong 2013). Tests of transfer of learning can be extended to speech production and sentence recognition. In a recent study, Miller et al. (2016a) examined the neural correlates of the learning of two consonant contrasts (i.e., /ba/–/da/ and /wa/–/ja/) in postlingually deafened adults with CIs. Their results demonstrated significantly enhanced amplitudes of mismatch negativity (MMN) to the sound stimuli across different phonetic categories but not to the ones within the same category. It merits further investigations with electrophysiological approaches to illustrate the neural mechanisms of brain plasticity underlying lexical tone learning in children with CIs. Lastly, it should be acknowledged that the sample size is relatively small, especially for the regression analyses of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

potential contributors for the trained children's individual differences in training-related benefits. A larger subject pool and stronger statistical power should be considered in future studies to obtain more robust results.

5 Conclusions

The present study demonstrated that HVPT led to enhanced lexical tone categorization of synthesized T1 and T2 stimuli differing solely in pitch variations in Mandarin-speaking pediatric CI recipients with congenital deafness. Remarkably, the training-induced gains in Mandarin tone categorization sustained 10 weeks post training. These findings provide compelling evidence for improved pitch perception of lexical tones following the structured adaptive training protocol. Individual patient factors, including baseline performance and duration of CI use, were found to be significantly correlated with magnitudes of pre-post improvements. The results have important implications for further research to refine the computer-based training protocol for optimizing training benefits in pediatric CI users and test its feasibility for personalized aural rehabilitation via telepractice.

Conflict of Interest: The authors have declared that no competing interests existed at the time of publication.

References

Bates, D., Mächler, M., Bolker, B.M., et al. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67, 1–48.

Benjamini, Y., Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, 29, 1165–1188.

Boersma, P., Weenink, D. (2017). Praat: Doing phonetics by computer (Computer program, Version 6.0.33). Available at: http://www.praat.org.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

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

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., et al. (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.

Brekelmans, G., Lavan, N., Saito, H., et al. (2022). Does high talker variability improve the learning of non-native phoneme contrasts? A replication. *J. Mem. Lang.*, 126, 104352. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0749596X22000390 [Accessed August 20, 2022].

Chao, Y.R. (1948). *Mandarin primer: An intensive course in spoken Chinese*, Berkeley, CA: Harvard University Press.

Chen, F., Peng, G. (2021). Categorical perception of pitch contours and voice onset time in Mandarin-speaking adolescents with autism spectrum disorders. *J. Speech, Lang. Hear. Res.*, 64, 4468–4484.

Chen, F., Peng, G., Yan, N., et al. (2017). The development of categorical perception of Mandarin tones in four-to seven-year-old children. *J. Child Lang.*, 44, 1413–1434.

Chen, S., Yang, Y., Wayland, R. (2021). Categorical perception of Mandarin pitch directions by Cantonese-speaking musicians and non-musicians. *Front. Psychol.*, 12, 713949.

Chen, Y., Wong, L.L.N. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *Int. J. Audiol.*, 56, S7–S16.

Cheng, B., Zhang, X., Fan, S., et al. (2019). The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study. *Front. Psychol.*, 10, 1178.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Cheng, X., Liu, Y., Shu, Y., et al. (2018). Music training can improve music and speech perception in pediatric Mandarin-speaking cochlear implant users. *Trends Hear.*, 22, 2331216518759214.

Deroche, M.L.D., Lu, H.-P., Lin, Y.-S., et al. (2019a). Processing of acoustic information in lexical tone production and perception by pediatric cochlear implant recipients. *Front. Neurosci.*, 13, 639.

Deroche, M.L.D., Lu, H.P., Kulkarni, A.M., et al. (2019b). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Sci. Rep.*, 9, 109.

Dong, H., Clayards, M., Brown, H., et al. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, 7, e7191.

Drouin, J.R., Theodore, R.M. (2020). Leveraging interdisciplinary perspectives to optimize auditory training for cochlear implant users. *Lang. Linguist. Compass*, 14, e12394.

Feng, Y., Peng, G. (2022). Development of categorical speech perception in Mandarin-speaking children and adolescents. *Child Dev.*, 00, 1–16. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/cdev.13837 [Accessed August 17, 2022].

Finney, D.J. (1971). *Probit analysis* 3rd ed., Cambridge, UK: Cambridge University Press.

Fu, Q.-J., Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific J. Speech, Lang. Hear.*, 5, 45–57.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V, et al. (1998). Importance of tonal envelope cues in Chinese speech recognition. *J. Acoust. Soc. Am.*, 104, 505–510.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Fuhrmeister, P., Myers, E.B. (2020). Desirable and undesirable difficulties: Influences of variability, training schedule, and aptitude on nonnative phonetic learning. *Attention, Perception, Psychophys.*, 82, 2049–2065.

Fuller, C.D., Galvin III, J.J., Maat, B., et al. (2018). Comparison of two music training approaches on music and speech perception in cochlear implant users. *Trends Hear.*, 22, 1–22.

Galvan, C., Case, E., Todd Houston, K. (2014). Listening and learning: Using telepractice to serve children and adults with hearing loss. *Perspect. Telepractice*, 4, 11–22.

Gao, Q., Wong, L.L.N., Chen, F. (2021). A review of speech perception of Mandarin-speaking children with cochlear implantation. *Front. Neurosci.*, 15, 773694.

Gerrits, E., Schouten, M.E.H. (2004). Categorical perception depends on the discrimination task. *Percept. Psychophys.*, 66, 363–376.

Gfeller, K., Guthe, E., Driscoll, V., et al. (2015). A preliminary report of music-based training for adult cochlear implant users: Rationales and development. *Cochlear Implants Int.*, 16, S22–S31.

Giannakopoulou, A., Brown, H., Clayards, M., et al. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, 5, e3209.

Goehring, J.L., Hughes, M.L., Baudhuin, J.L. (2012). Evaluating the feasibility of using remote technology for cochlear implants. *Volta Rev.*, 112, 255–265.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.*, 105, 251–279.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Goldinger, S.D., Pisoni, D.B., Logan, J.S. (1991). On the nature of talker variability effects on recall of spoken word lists. *J. Exp. Psychol. Learn. Mem. Cogn.*, 17, 152–162.

Goldstone, R.L., Hendrickson, A.T. (2010). Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.*, 1, 69–78.

Good, A., Gordon, K.A., Papsin, B.C., et al. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear Hear.*, 38, 455–464.

Harnad, S. (2003). Categorical perception. In L. Nadel, ed. *Encyclopedia of Cognitive Science*. London, UK: Nature Publishing Group. Available at: https://eprints.soton.ac.uk/257719/1/catperc.html.

Hiskey, M.S. (1966). *Hiskey-Nebraska test of learning aptitude*, Cambridge, UK: Union College Press.

Huang, W., Wong, L.L.N., Chen, F., et al. (2020). Effects of fundamental frequency contours on sentence recognition in Mandarin-speaking children with cochlear implants. *J. Speech, Lang. Hear. Res.*, 63, 3855–3864.

Ingvalson, E.M., Wong, P. (2013). Training to improve language outcomes in cochlear implant recipients. *Front. Psychol.*, 4, 263.

Ingvalson, E.M., Wong, P.C.M. (2016). Auditory training: Predictors of success and optimal training paradigms. In N. M. Young & K. I. Kirk, eds. *Cochlear implants in children: Learning and the brain*. (*pp. 293–297*). Philadelphia, PA: Springer.

Iverson, P., 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Iverson, P., Hazan, V., 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.

Jiang, C., Hamm, J.P., Lim, V.K., et al. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Mem. Cognit.*, 40, 1109–1121.

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & J. W. Mullennix, eds. *Talker variability in speech peocessing*. (*pp. 145–165*). San Diego: Academic Press.

Kim, S., Chou, H.-H., Luo, X. (2021). Mandarin tone recognition training with cochlear implant simulation: Amplitude envelope enhancement and cue weighting. *J. Acoust. Soc. Am.*, 150, 1218–1230.

Koerner, T., Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sci.*, 7, 26.

Kral, A., Dorman, M.F., Wilson, B.S. (2019). Neuronal development of hearing and language: Cochlear implants and critical periods. *Annu. Rev. Neurosci.*, 42, 47–65.

Kral, A., Kronenberger, W.G., Pisoni, D.B., et al. (2016). Neurocognitive factors in sensory restoration of early deafness: A connectome model. *Lancet Neurol.*, 15, 610–621.

Kral, A., O'Donoghue, G.M. (2010). Profound deafness in childhood. *N. Engl. J. Med.*, 363, 1438–1450.

Kuhl, P.K. (2000). A new view of language acquisition. *Proc. Natl. Acad. Sci.*, 97, 11850–11857.

Kuhl, P.K. (2004). Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci.*, 5, 831–843.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Kuhl, P.K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.*, 50, 93–107.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., et al. (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.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.*, 82, 1–26.

Lakens, D., Caldwell, A.R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Adv. Methods Pract. Psychol. Sci.*, 4, 1–14.

Lenth, R., Singmann, H., Love, J., et al. (2018). emmeans: Estimated marginal means, aka least-squares means. *R Packag. version 1.3.0*. Available at: https://cran.r-project.org/package=emmeans.

Liberman, A.M., Harris, K.S., Hoffman, H.S., et al. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.*, 54, 358–368.

Lively, S.E., Logan, J.S., 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.

Lively, S.E., Pisoni, D.B., Yamada, R.A., et al. (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.

Lo, C.Y., Looi, V., Thompson, W.F., et al. (2020). Music training for children with sensorineural hearing loss improves speech-in-noise perception. *J. Speech, Lang. Hear. Res.*, 63, 1990–2015.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Lo, C.Y., McMahon, C.M., Looi, V., et al. (2015). Melodic contour training and its effect on speech in noise, consonant discrimination, and prosody perception for cochlear implant recipients. *Behav. Neurol.*, 2015, 352869.

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

Luo, X., Chang, Y., Lin, C.-Y., et al. (2014). Contribution of bimodal hearing to lexical tone normalization in Mandarin-speaking cochlear implant users. *Hear. Res.*, 312, 1–8.

Luo, X., Fu, Q.-J. (2004). Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants. *J. Acoust. Soc. Am.*, 116, 3659–3667.

Luo, X., Fu, Q.J., Galvin, J.J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif.*, 11, 301–315.

Luo, X., Fu, Q.J., Wei, C.G., et al. (2008). Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear Hear.*, 29, 957–970.

Luo, X., Fu, Q.J., Wu, H.P., et al. (2009). Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hear. Res.*, 256, 75–84.

Ma, J., Zhu, J., Yang, Y., et al. (2021). The development of categorical perception of segments and suprasegments in Mandarin-speaking preschoolers. *Front. Psychol.*, 12, 693366.

Macmillan, N.A., Creelman, C.D. (2005). *Detection theory: A user's guide* 2nd ed., Mahwah, NJ: Erlbaum.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

McMurray, B., Danelz, A., Rigler, H., et al. (2018). Speech categorization develops slowly through adolescence. *Dev. Psychol.*, 54, 1472–1491.

Miller, S., Zhang, Y., Nelson, P. (2016a). Neural correlates of phonetic learning in postlingually deafened cochlear implant listeners. *Ear Hear.*, 37, 514–528.

Miller, S., Zhang, Y., Nelson, P.B. (2016b). Efficacy of multiple-talker phonetic identification training in postlingually deafened cochlear implant listeners. *J. Speech, Lang. Hear. Res.*, 59, 90–98.

Mishra, S.K., Boddupally, S.P. (2018). Auditory cognitive training for pediatric cochlear implant recipients. *Ear Hear.*, 39, 48–59.

Mishra, S.K., Boddupally, S.P., Rayapati, D. (2015). Auditory learning in children with cochlear implants. *J. Speech, Lang. Hear. Res.*, 58, 1052–1060.

Moore, D.R., Shannon, R. V (2009). Beyond cochlear implants: Awakening the deafened brain. *Nat. Neurosci.*, 12, 686–691.

Nan, Y., Liu, L., Geiser, E., et al. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proc. Natl. Acad. Sci. U. S. A.*, 115, E6630–E6639.

Oxenham, A.J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends Amplif.*, 12, 316–331.

Peng, G., Zheng, H.-Y., Gong, T., et al. (2010). The influence of language experience on categorical perception of pitch contours. *J. Phon.*, 38, 616–624.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Peng, S.-C., Lu, H.-P., Lu, N., et al. (2017). Processing of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients. *J. Speech, Lang. Hear. Res.*, 60, 1223–1235.

Peng, S.-C., Tomblin, J.B., Cheung, H., et al. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear Hear.*, 25, 251–264.

Perrachione, T.K., Lee, J., Ha, L.Y.Y., et al. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *J. Acoust. Soc. Am.*, 130, 461–472.

Peterson, N.R., Pisoni, D.B., Miyamoto, R.T. (2010). Cochlear implants and spoken language processing abilities: Review and assessment of the literature. *Restor. Neurol. Neurosci.*, 28, 237–250.

Pisoni, D.B., Kronenberger, W.G., Harris, M.S., et al. (2017). Three challenges for future research on cochlear implants. *World J. Otorhinolaryngol. - Head Neck Surg.*, 3, 240–254.

Pisoni, D.B., Lively, S.E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange, ed. *Speech perception and linguistic experience: Issues in cross-language speech research*. (*pp. 433–459*). Baltimore, MD: York Press.

Qin, Z., Gong, M., Zhang, C. (2021). Neural responses in novice learners' perceptual learning and generalization of lexical tones: The effect of training variability. *Brain Lang.*, 223, 105029.

Qin, Z., Tremblay, A., Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *J. Phon.*, 73, 144–157.

Rayes, H., Al-Malky, G., Vickers, D. (2019). Systematic review of auditory training in pediatric cochlear implant recipients. *J. Speech, Lang. Hear. Res.*, 62, 1574–1593.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Sadakata, M., McQueen, J.M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Front. Psychol.*, 5, 1318.

Sakai, M., 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.

Shao, J., Lau, R.Y.M., Tang, P.O.C., et al. (2019). The effects of acoustic variation on the perception of lexical tone in Cantonese-speaking congenital amusics. *J. Speech, Lang. Hear. Res.*, 62, 190–205.

Shen, G., Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *J. Acoust. Soc. Am.*, 140, 4396–4403.

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

Stropahl, M., Besser, J., Launer, S. (2020). Auditory training supports auditory rehabilitation: A state-of-the-art review. *Ear Hear.*, Publish Ah, 697–704.

Tan, J., Dowell, R., Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *Am. J. Audiol.*, 25, 246–256.

Tao, D., Deng, R., Jiang, Y., et al. (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear Hear.*, 36, 102–110.

Torppa, R., Huotilainen, M. (2019). Why and how music can be used to rehabilitate and develop speech and language skills in hearing-impaired children. *Hear. Res.*, 380, 108–122.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Völter, C., Stöckmann, C., Schirmer, C., et al. (2021). Tablet-based telerehabilitation versus conventional face-to-face rehabilitation after cochlear implantation: Prospective intervention pilot study. *JMIR Rehabil. Assist. Technol.*, 8, e20405.

Wang, S., Liu, B., Dong, R., et al. (2012). Music and lexical tone perception in Chinese adult cochlear implant users. *Laryngoscope*, 122, 1353–1360.

Wang, W., Zhou, N., Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *Int. J. Audiol.*, 50, 270–278.

Wang, W.S.Y. (1973). The Chinese language. *Sci. Am.*, 228, 50–63.

Wang, Y., Jongman, A., Sereno, J.A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc. Am.*, 113, 1033–1043.

Wang, Y., Spence, M.M., Jongman, A., et al. (1999). Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am.*, 106, 3649–3658.

Whalen, D.H., Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25–47.

Wiener, S., Bradley, E.D. (2020). Harnessing the musician advantage: Short-term musical training affects non-native cue weighting of linguistic pitch. *Lang. Teach. Res.*

Wiener, S., Chan, M.K.M., Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions? *Mod. Lang. J.*, 104, 152–168.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

van Wieringen, A., Wouters, J. (2015). What can we expect of normally-developing children implanted at a young age with respect to their auditory, linguistic and cognitive skills? *Hear. Res.*, 322, 171–179.

Wong, P.C.M., Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.*, 28, 565–585.

Wu, H., Ma, X., Zhang, L., et al. (2015). Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Front. Psychol.*, 6, 436.

Wu, J.-L., Yang, H.-M., Lin, Y.-H., et al. (2007). Effects of computer-assisted speech training on Mandarin-speaking hearing-impaired children. *Audiol. Neurotol.*, 12, 307–312.

Xi, J., Zhang, L., Shu, H., et al. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170, 223–231.

Xu, Y., Gandour, J.T., 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.

Yang, X.J., Qu, C.Y., Sun, X.B., et al. (2011). Norm revision of H-NTLA for children from 3 to 7 years old in China. *Chinese J. Clin. Psychol.*, 19, 195–197.

Yu, K., Li, L., Chen, Y., et al. (2019). Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology*, 56, e13448.

Zhang, C., Chen, S. (2016). Toward an integrative model of talker normalization. *J. Exp. Psychol. Hum. Percept. Perform.*, 42, 1252–1268.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Zhang, C., Shao, J., Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PLoS One*, 12, e0183151.

Zhang, H., Chen, F., Yan, N., et al. (2016). The influence of language experience on the categorical perception of vowels: Evidence from Mandarin and Korean. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. (*pp. 873–877*).

Zhang, H., Ding, H., Zhang, Y. (2021a). High-variability phonetic training benefits lexical tone perception: An investigation on mandarin-speaking pediatric cochlear implant users. *J. Speech, Lang. Hear. Res.*, 64, 2070–2084.

Zhang, H., Wiener, S., Holt, L.L. (2022). Adjustment of cue weighting in speech by speakers and listeners: Evidence from amplitude and duration modifications of Mandarin Chinese tone. *J. Acoust. Soc. Am.*, 151, 992–1005.

Zhang, H., Zhang, J., Ding, H., et al. (2019a). Bimodal benefit in categorical perception of lexical tones for Mandarin-speaking children with cochlear implants. In *Proceedings of the International Congress of Phonetics Science*. Melbourne, Australia.

Zhang, H., Zhang, J., Ding, H., et al. (2020a). Bimodal benefits for lexical tone recognition: An investigation on Mandarin-speaking preschoolers with a cochlear implant and a contralateral hearing aid. *Brain Sci.*, 10, 238.

Zhang, H., Zhang, J., Ding, H., et al. (2020b). Efficacy of multi-talker phonetic training in Mandarin tone perception for native pediatric cochlear implant users. In *Proceedings of the 10th International Conference on Speech Prosody 2020*. (*pp. 819–823*). Tokyo, Japan.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Zhang, H., Zhang, J., Peng, G., et al. (2020c). Bimodal benefits revealed by categorical perception of lexical tones in Mandarin-speaking kindergarteners with a cochlear implant and a contralateral hearing aid. *J. Speech, Lang. Hear. Res.*, 63, 4238–4251.

Zhang, L., Wang, J., Hong, T., et al. (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.

Zhang, L., Xie, S., Li, Y., et al. (2020d). Perception of musical melody and rhythm as influenced by native language experience. *J. Acoust. Soc. Am.*, 147, EL385–EL390.

Zhang, X., Cheng, B., Qin, D., et al. (2021b). Is talker variability a critical component of effective phonetic training for nonnative speech? *J. Phon.*, 87, 101071.

Zhang, X., Cheng, B., Zhang, Y. (2021c). The role of talker variability in nonnative phonetic learning: A systematic review and meta-analysis. *J. Speech, Lang. Hear. Res.*, 64, 4802–4825.

Zhang, Y. (2016). Categorical perception. In R. Sybesma, W. Behr, Y. Gu, et al., eds. *Encyclopedia of Chinese language and linguistics*. Leiden, the Netherlands: Brill.

Zhang, Y., Kuhl, P., Imada, T., et al. (2019b). Neural commitment alters cue weighting of formant structure in speech perception: A cross-language MEG study. *J. Acoust. Soc. Am.*, 146, 2954. Available at: https://asa.scitation.org/doi/abs/10.1121/1.5137259 [Accessed August 29, 2022].

Zhang, Y., Kuhl, P.K., Imada, T., et al. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage*, 46, 226–240.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Zhang, Y., Zhang, L., Shu, H., et al. (2012). Universality of categorical perception deficit in developmental dyslexia: An investigation of Mandarin Chinese tones. *J. Child Psychol. Psychiatry*, 53, 874–882.

Zhu, J., Chen, X., Yang, Y. (2021). Effects of amateur musical experience on categorical perception of lexical tones by native Chinese adults: An ERP study. *Front. Psychol.*, 12, 611189.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Table 1. Demographic information of the child participants with CIs.

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

CA = chronological age; CI = cochlear implant; H-NTLA = Hiskey–Nebraska Test of Learning Aptitude; M = male; F = female; ACE = Advanced Combination Encoder; FS4-P = Fine Structure Processing Strategy; HiRes-Optima = High Resolution Optima; yrs = years.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Table 2. Demographic characteristics and available *p* values of independent-samples t-tests between training group and control group.

Characteristics	Training group	Control group	*p* value
CA (yrs)	4.91 (0.12)	5.07 (0.13)	$p = 0.41$
Age at CI (yrs)	1.44 (0.15)	1.7 (0.21)	$p = 0.32$
CI duration (yrs)	3.47 (0.17)	3.36 (0.17)	$p = 0.67$
H-NTLA	109 (2.33)	111 (2.32)	$p = 0.34$

CA = chronological age; CI = cochlear implant; yrs = years; H-NTLA = Hiskey–Nebraska Test of Learning Aptitude; data in brackets indicate mean standard errors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

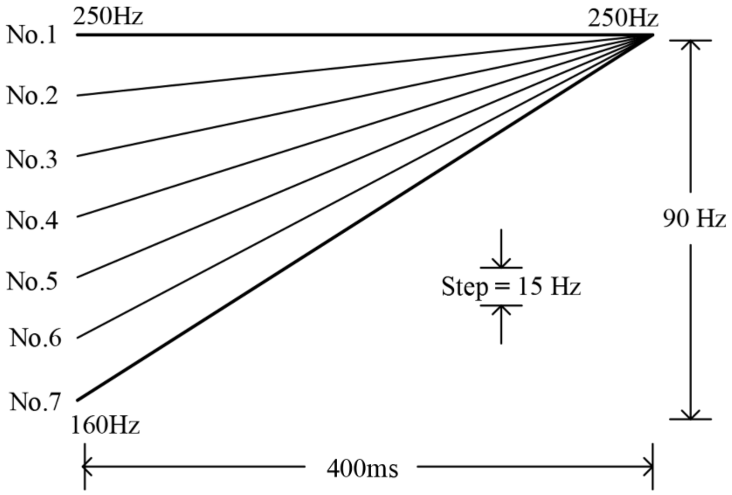
Table 3. Linear regression results indicating the relationship between duration of CI use / baseline performance and training-induced benefit magnitudes in terms of boundary width / peakedness score among the trained individuals with CIs.

Predictors	Boundary width				Peakedness score			
	β	SE	*t*	*p* value	β	SE	*t*	*p* value
Duration of CI use	-0.27	0.12	-2.27	0.04 *	-0.04	0.43	-0.1	0.92
Baseline performance	-0.89	0.11	-8.06	0.00001 ***	-0.75	0.49	-1.52	0.16

Three *asterisks* represent $p < .001$, one *asterisk* represents $p < .05$.

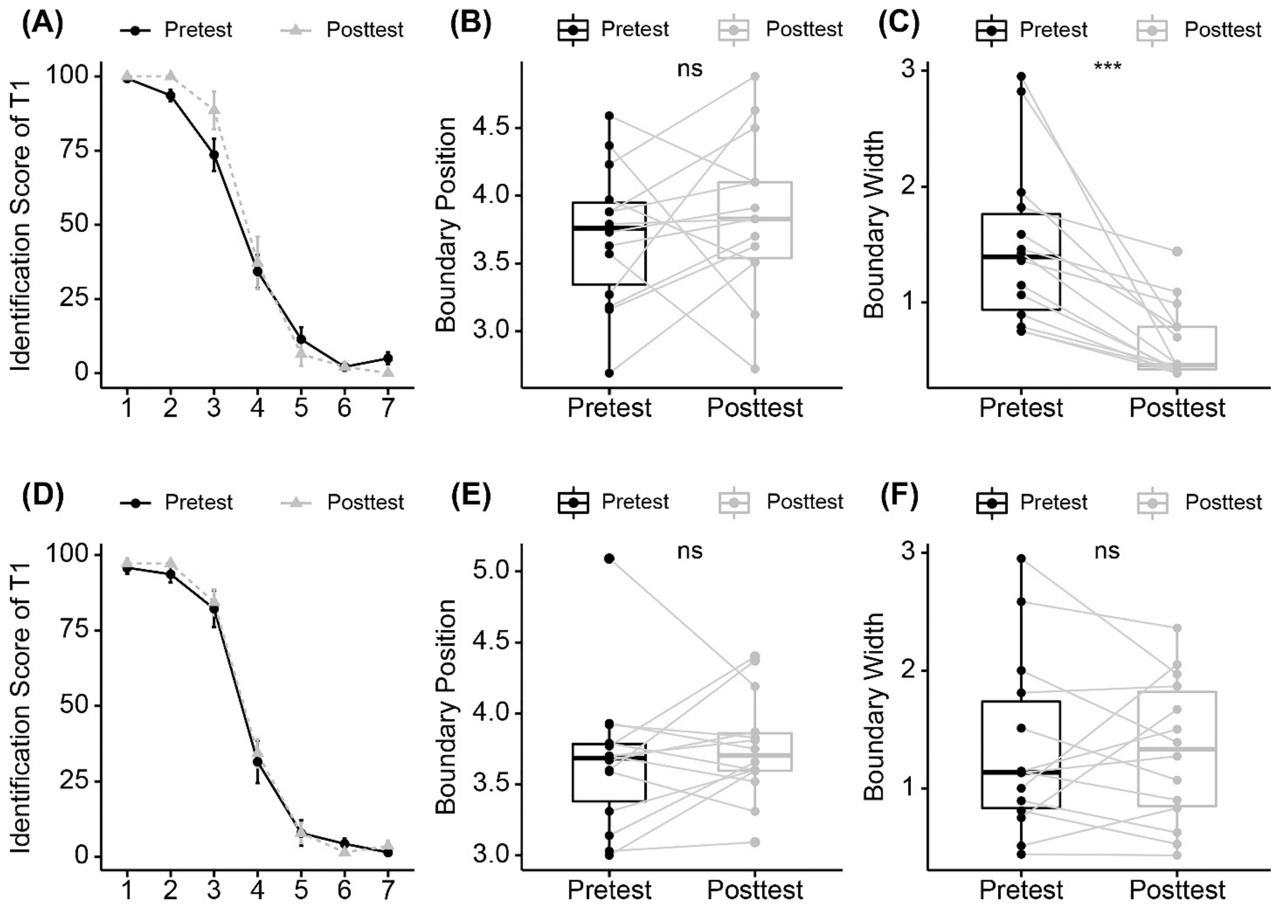
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Figure 1. Schematic diagram of F0 contours along the tonal continuum.



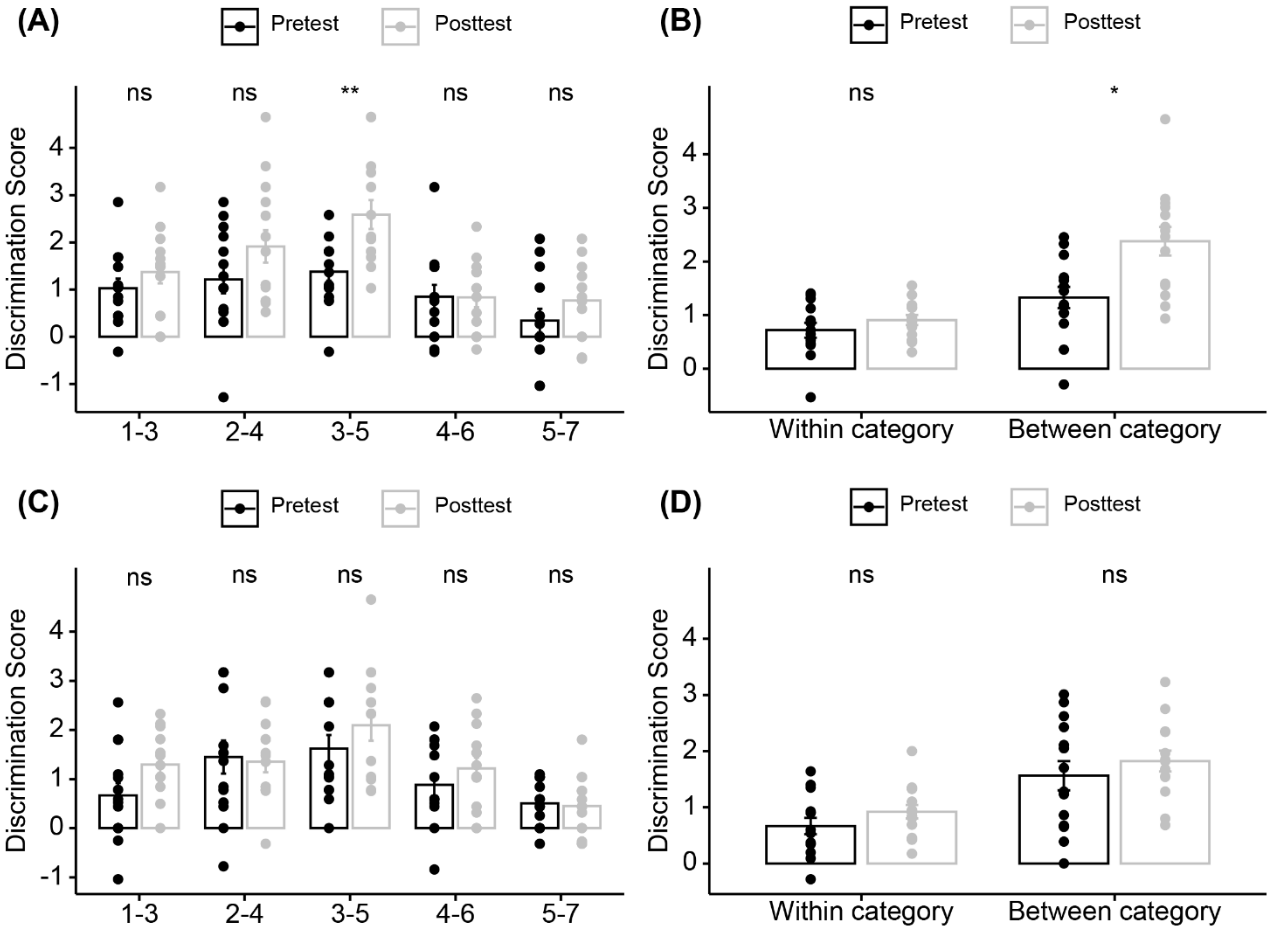
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Figure 2. Grand averaged identification data in pretest and posttest for the training group (A, B, and C) and the control group (D, E, and F). (A) Identification functions of T1 for the two test sessions of the training group; (B) Mean boundary positions for the two test sessions of the training group; (C) Mean boundary widths for the two test sessions of the training group; (D) Identification functions of T1 for the two test sessions of the control group; (E) Mean boundary positions for the two test sessions of the control group; (F) Mean boundary widths for the two test sessions of the control group. Dots in line plots (A and D) represent the grand mean discrimination score of T1 for each stimulus across all participants, whereas dots in box plots (B, C, E, and F) represent individual data of each participant in boundary position (B and E) or boundary width (C and F). Error bars represent the standard errors across all participants. Statistical significances are provided: three *asterisks* represent $p < 0.001$ and *ns* represents $p > 0.05$.



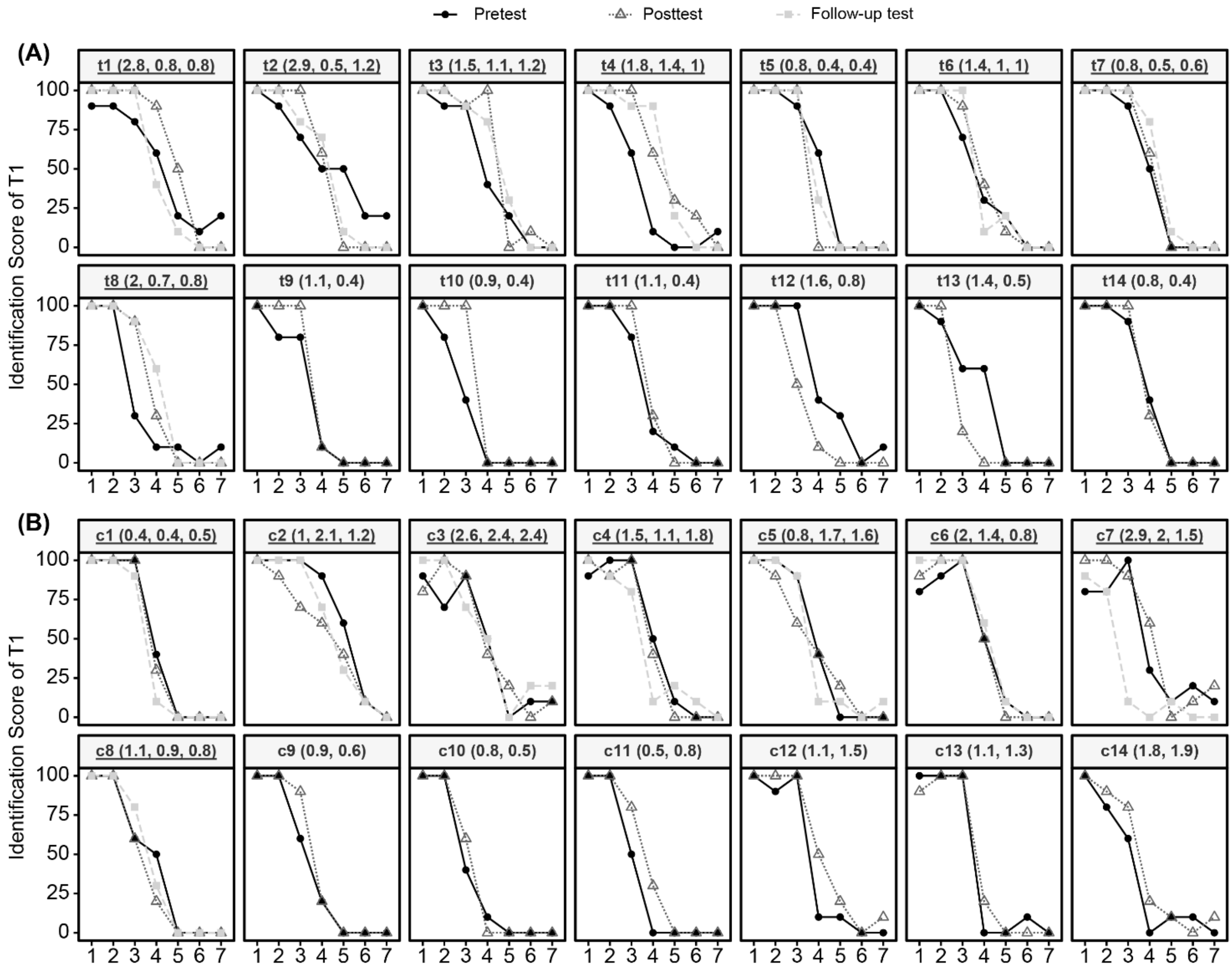
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Figure 3. Grand averaged discrimination data from pretest to posttest for the training group (A and B) and the control group (C and D). (A) Mean discrimination scores of different stimulus contrasts for the two test sessions of the training group; (B) Mean discrimination scores of between-category type and within-category type for the two test sessions of the training group; (C) Mean discrimination scores of different stimulus contrasts for the two test sessions of the control group; (D) Mean discrimination scores of between-category type and within-category type for the two test sessions of the control group. Dots in the histograms represent individual data of discrimination score of each stimulus pair (A and C) or of each discrimination type (B and D). Stimulus contrasts shown in the x-axis of (A) and (C) contain both forward order (e.g., 1-3, 5-7) and reverse order (e.g., 3-1, 7-5). Error bars represent the standard errors across all participants. Statistical significances are provided: one *asterisks* represents $p < 0.05$ and *ns* represents $p > 0.05$.



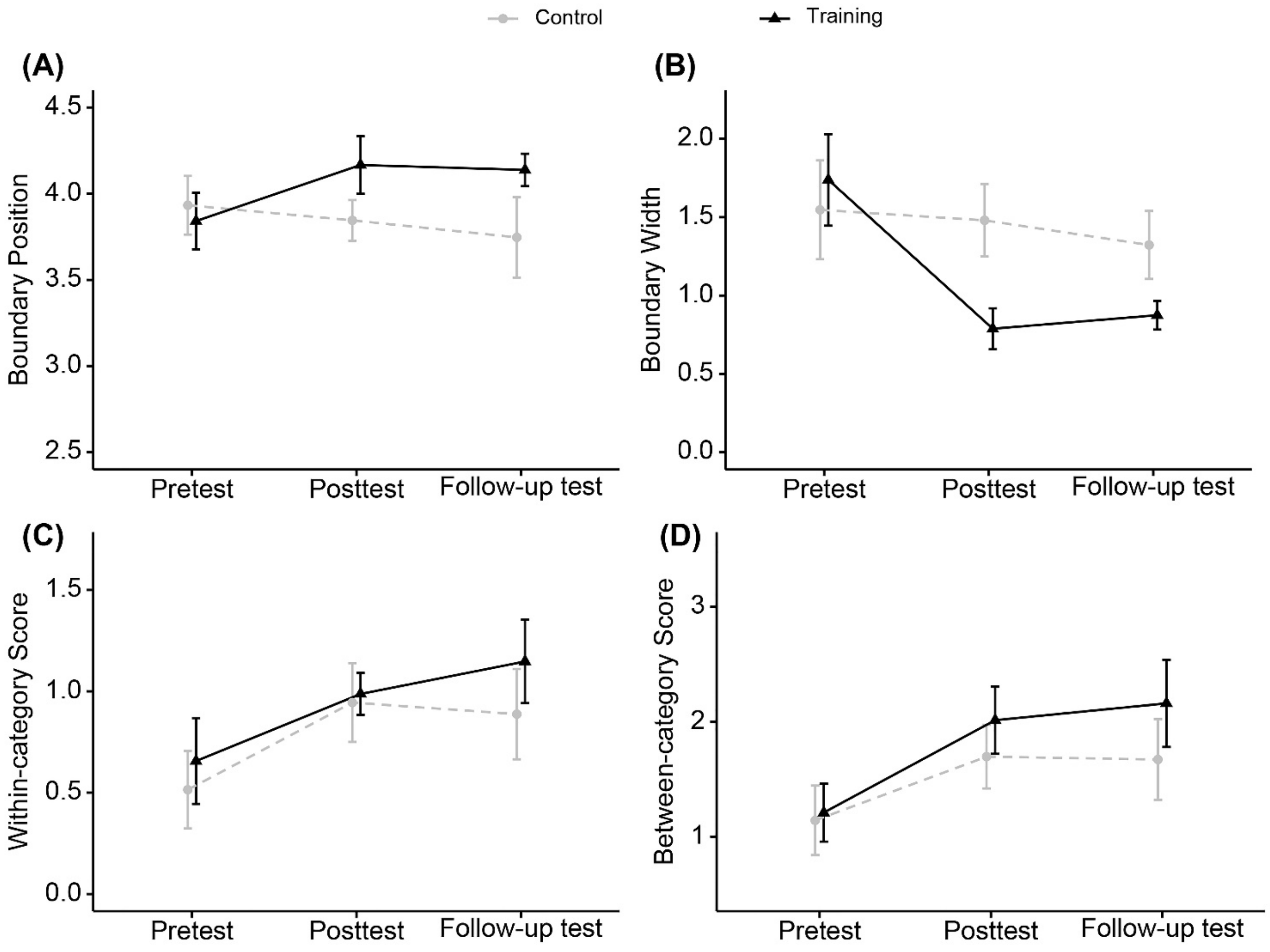
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Figure 4. Individual results of identification performance for training group and control group in different test sessions. (A) Identification functions of T1 for each trained child with CI; (B) Identification functions of T1 for each control child with CI. Data in brackets indicate boundary widths for pretest, posttest, and follow-up test (if follow-up test data are available). The underlined captions of individual plots indicate those who completed all three test sessions, including subjects t1 to t8 from the training group and c1 to c8 from the control group.



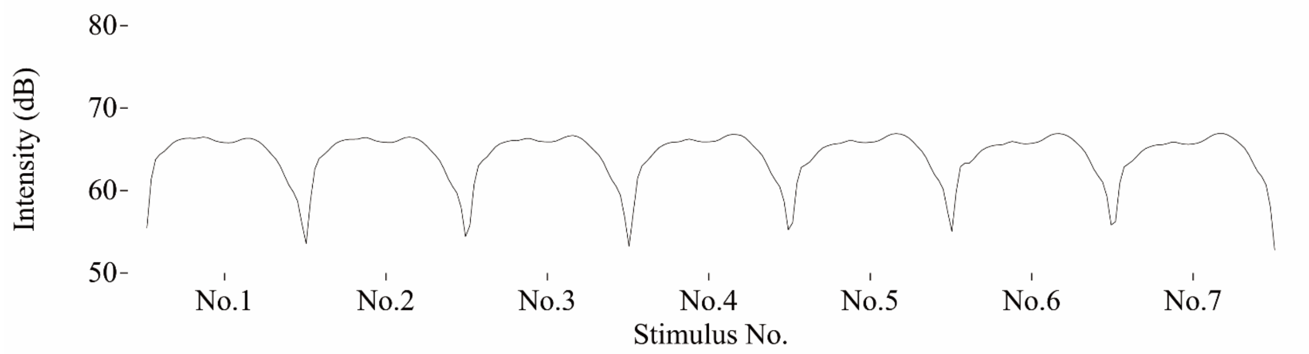
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Figure 5. Grand averaged identification and discrimination results for the 16 children with CIs completed all three test sessions. (A) Mean boundary positions for the trained and control children from pretest to posttest and follow-up test; (B) Mean boundary widths for the trained and control children from pretest to posttest and follow-up test; (C) Mean within-category scores for the trained and control children from pretest to posttest and follow-up test; (D) Mean between-category scores for the trained and control children from pretest to posttest and follow-up test. Error bars represent the standard errors across all participants.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Appendix A. Schematic diagram of amplitude envelope along the tonal continuum.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022

doi:10.20944/preprints202211.0007.v1

Appendix B. Actual amount of training for the trained children with CIs per training session, where the number "1" indicates the session was presented only once and "2" indicates the session was repeated once.

t1	1	2	1	1	1	6
t2	2	1	2	1	1	7
t3	2	2	2	2	1	9
t4	2	1	1	1	1	6
t5	2	1	1	2	1	7
t6	2	2	2	1	1	8
t7	1	1	1	1	1	5
t8	1	2	1	1	1	6
t9	2	1	1	1	1	6
t10	1	1	1	1	1	5
t11	2	1	1	1	1	6
t12	2	2	1	1	1	7
t13	1	1	1	1	1	5
t14	1	1	1	1	1	5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 November 2022 doi:10.20944/preprints202211.0007.v1

Appendix C. Linear regression results indicating the relationship between implanted age / duration of CI use and CP parameters of boundary width / peakedness score in pretest among all child participants with CIs.

Predictors	Boundary width				Peakedness score			
	β	SE	*t*	*p* value	β	SE	*t*	*p* value
Implanted age	-0.24	0.3	-0.8	0.43	0.01	0.3	0.03	0.97
Duration of CI use	-0.27	0.33	-0.82	0.42	-0.03	0.33	-0.1	0.92