
High Variability Phonetic Training Facilitates Transfer of Learning from Perception to Production in Children with Cochlear Implants

[Hao Zhang](#), Xuequn Dai, Wen Ma^{*}, [Hongwei Ding](#)^{*}, [Yang Zhang](#)^{*}

Posted Date: 26 September 2023

doi: 10.20944/preprints202309.1732.v1

Keywords: high variability phonetic training (HVPT); lexical tone; perception-production link; cochlear implant (CI); perceptual learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

High Variability Phonetic Training Facilitates Transfer of Learning from Perception to Production in Children with Cochlear Implants

Hao Zhang^a, Xuequn Dai^a, Wen Ma^{a,*}, Hongwei Ding^{b,*}, 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 Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, MN, USA

* Correspondence: Wen Ma: mawen@sdu.edu.cn (W.M.); hwding@sjtu.edu.cn (H.D.); zhanglab@umn.edu (Y.Z.)

Abstract: Purpose: This study builds upon an established effective training method to investigate the advantages of high variability phonetic identification training for enhancing lexical tone perception and production in Mandarin-speaking pediatric cochlear implant (CI) recipients, who typically face ongoing challenges in these areas. **Method:** Thirty two Mandarin-speaking children with CIs were quasi-randomly assigned into the training group (TG) and the control group (CG). The sixteen TG participants received five sessions of high variability phonetic training (HVPT) within a period of three weeks. The CG participants did not receive the training. Perception and production of Mandarin tones were administered before (pretest) and immediately after (posttest) the completion of HVPT via lexical tone recognition task and picture naming task. Both groups participated in the identical pretest and posttest with the same time frame between the two test sessions. **Results:** TG showed significant improvement from pretest to posttest in identifying Mandarin tones for both trained and untrained speech stimuli. Moreover, perceptual learning of HVPT facilitated trainees' lexical tone production as rated by a cohort of 10 Mandarin-speaking adults with normal hearing. In contrast, CG did not exhibit significant changes in either perception or production. **Conclusion:** The results represent initial evidence of HVPT-induced transfer of perceptual learning to lexical tone production in CI users, which supports the application of this speech training protocol to aural rehabilitation practice.

Keywords: high variability phonetic training (HVPT); lexical tone; perception-production link; cochlear implant (CI); perceptual learning

1. Introduction

Cochlear implant (CI) represents a great bioengineering achievement in the treatment of individuals with severe to profound sensorineural hearing loss (Tamati, Pisoni, & Moberly, 2022; B. S. Wilson, Dorman, Woldorff, & Tucci, 2011). Despite providing recipients with impressive access to sound, current CI devices transmit distorted acoustic signals via speech coding strategies with a limited number of electrodes, resulting in degraded spectral-temporal information (Moore & Shannon, 2009). The fine structure information is poorly resolved, with merely temporal envelope cues preserved, contributing to the compromised fundamental frequency (F0) and harmonics (Oxenham, 2008). As F0 is a primary acoustic correlate for pitch patterns in speech sounds (Whalen & Xu, 1992), pitch percept is generally weak and poses a unique challenge for CI users.

Due to impaired pitch perception in CI users, persistent challenges arise in acquiring lexical tones, particularly in those from tonal languages like Mandarin Chinese (Tan, Dowell, & Vogel, 2016 for a review). Mandarin uses four pitch variations (high-level, low-to-high rising, low-dipping, and high-falling) to represent its four tones (T1, T2, T3, T4) (Chao, 1948; W. S. Y. Wang, 1973). Prior studies have consistently documented significant deficits in both tone perception and production among

Mandarin-speaking CI children (Chen & Wong, 2017; Gao, Wong, & Chen, 2021; Tan et al., 2016 for reviews). These children typically achieve accuracy rates of about 67%-77% in tone recognition (Tao et al., 2015; H. Zhang, Zhang, Ding, & Zhang, 2020; Zhou, Huang, Chen, & Xu, 2013) and around 50% in tone production (Peng, Tomblin, Cheung, Lin, & Wang, 2004; Xu et al., 2011; Zhou & Xu, 2008). However, early acquisition of lexical tones is evident in typically developing children with normal hearing (NH), which stabilizes over the first 2-3 years of life (Singh & Fu, 2016 for a review). For instance, with a large subject pool (107 Mandarin-speaking children with CIs and 125 age mates with NH), Zhou et al. (2013) revealed that the group mean accuracies of tone perception and production were respectively 67.3% and 46.8% for CI children, whereas the accuracies were 98.7% and 94.8% for NH counterparts. This highlights significant room for improvement in Mandarin-speaking pediatric CI recipients' lexical tone acquisition.

Auditory training is known as a sound-based intervention for speech and hearing rehabilitation, which engages the central auditory system to make perceptual distinctions of sound contrasts through repetition and variation of sound stimuli together with adaptive listening and effective feedback (Cambridge, Taylor, Arnott, & Wilson, 2022; Rayes, Al-Malky, & Vickers, 2019 for reviews). Several reports have demonstrated the potential benefits of auditory training in lexical tone rehabilitation for CI users (X. Cheng et al., 2018; Kim, Chou, & Luo, 2021; Wu, Yang, Lin, & Fu, 2007; H. Zhang, Ding, & Zhang, 2021; H. Zhang, Ma, Ding, & Zhang, 2023). These studies employed different approaches, such as phonetic identification, tone recognition, and melodic contour identification training, highlighting improvements in lexical tone perception for trained CI recipients. Intensive music training also yielded perceptual gains, showing improved recognition of lexical tones and sentences (Cheng et al., 2018). Despite these positive findings, previous research mainly assessed training-induced gains in speech perception, overlooking cross-modal transfer to speech production. To address this gap, our study employs the widely recognized high variability phonetic training (HVPT) protocol. Our primary aim is to assess the potential success of HVPT for lexical tone acquisition in Mandarin-speaking CI children, focusing on robust generalization of perceptual training benefits to new stimuli and reliable far-transfer to lexical tone production.

As a well-established technique in second language learning, HVPT utilizes training materials with multiple talkers and varying phonetic contexts. HVPT is known for its benefits in robust generalization, long-term retention, and far-transfer of perceptual learning to production in nonnative speech acquisition (Barriuso & Hayes-Harb, 2018; Ingvalson & Wong, 2016; X. Zhang, Cheng, & Zhang, 2021 for reviews). In a seminal study, Logan et al. (1991) employed variable natural speech produced by multiple talkers to train native Japanese speakers in distinguishing English consonants /r/ and /l/. Results showed improved identification of the target speech sounds across talkers and stimuli, indicating robust generalization (Logan et al., 1991), and this generalization effect was attributed to talker variability (Lively, Logan, & Pisoni, 1993). Further research demonstrated long-term retention (at least three months) of perceptual gains (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994) and far-transfer to speech production (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). HVPT has also been applied to nonnative speakers learning lexical tones, showing transfer from perception to production (Dong, Clayards, Brown, & Wonnacott, 2019; Y. Wang, Jongman, & Sereno, 2003; Y. Wang, Spence, Jongman, & Sereno, 1999; Wiener, Chan, & Ito, 2020). For example, Wang et al. (1999, 2003) assessed HVPT's effectiveness in teaching Mandarin tones to native speakers of American English. Participants underwent an eight-session identification training over two weeks, resulting in improved tone identification, generalization to new words and untrained speakers, and a remarkable 18% increase in tone production accuracy rated by native Mandarin-speaking adults (Y. Wang et al., 2003).

Inspired by successes in second language learning, attempts to apply the HVPT protocol to CI users have yielded promising results. Miller and colleagues administered a two-week perceptual training program for postlingually deafened adults with CIs, significantly enhancing their recognition of phonetic contrasts (Miller, Zhang, & Nelson, 2016a, 2016b). Recent studies by Zhang and colleagues demonstrated that Mandarin-speaking CI children could obtain robust and lasting benefits in lexical tone perception through a five-session HVPT protocol (H. Zhang et al., 2021, 2023;

H. Zhang, Zhang, Ding, & Li, 2020). These gains extended to both familiar and unfamiliar talkers and lasted up to 10 weeks post-training. However, questions remain about whether these perceptual gains generalize to novel phonetic contexts, given that testing was conducted on the monosyllable /i/ that was also used in training. Furthermore, further research is needed to investigate cross-modal transfer of perceptual learning to lexical tone production in Mandarin-speaking pediatric CI users.

This study builds upon previous research to investigate the effectiveness of the HVPT protocol for Mandarin-speaking pediatric CI users' lexical tone rehabilitation. Three research questions are addressed: (a) whether training-induced gains generalize to recognizing lexical tones in novel phonetic contexts; (b) whether perceptual learning transfers to improved tone production during spontaneous speech; (c) whether a relationship exists between perception gains and production improvements. Based on prior research findings, we would expect the following results: (a) trained children with CIs would improve recognition of lexical tones for both trained and untrained stimuli, while control counterparts would show minimal pretest-posttest changes; (b) perceptual learning would transfer to improved tone production, with some tones benefiting more than others; (c) there would be a weak but positive relationship between gains in tone recognition and production. These findings can inform effective speech rehabilitation for tone language CI users.

2. Method

2.1. Participants

This study involved 32 Mandarin-speaking children with CIs (13 girls and 19 boys), recruited from the Shanghai Rehabilitation Center of the Deaf Children. They all had bilateral severe to profound sensorineural hearing loss, received unilateral CIs before 3.5 years of age, and primarily used spoken Mandarin in daily life. None had psychiatric disorders or brain injuries according to parental reports. Nonverbal intelligence was assessed using the Hiskey–Nebraska Test of Learning Aptitude (H-NTLA), with scores above the passing criteria of 84 (Hiskey, 1966; Yang, Qu, Sun, Zhu, & Wang, 2011). On average, their chronological age was 5.19 years, age at implantation was 1.85 years, CI experience was 3.32 years, and H-NTLA score was 108. Table 1 presents the demographic and CI device information. This study received ethics approval from the Ethics Committee of School of Foreign Languages, Shanghai Jiao Tong University, and parental consent was obtained before participation.

Table 1. Demographic information of the pediatric CI participants.

Subject (Sex)	Group	CA (yrs)	Speech processor	Speech strategy	CI side	Age at CI (yrs)	CI duration (yrs)	H- NTLA score
t1 (F)	TG	4.97	OPUS2	FS4-P	Right	1.58	3.38	105
t2 (F)	TG	5.03	OPUS2	FS4-P	Left	1.14	3.89	100
t3 (F)	TG	4.35	OPUS2	FS4-P	Right	1.04	3.3	126
t4 (M)	TG	4.23	OPUS2	FS4-P	Right	1	3.23	102
t5 (M)	TG	4.63	Nucleus5	ACE	Left	1.38	3.24	114
t6 (M)	TG	5	Nucleus6	ACE	Right	1.5	3.49	125
t7 (F)	TG	4.17	Naida	HiRes	Right	1.59	2.58	103
t8 (M)	TG	4.16	OPUS2	FS4-P	Right	1.57	2.59	105
t9 (M)	TG	4.97	Nucleus6	ACE	Right	2.32	2.65	108
t10 (F)	TG	6.39	Nucleus5	ACE	Right	1.88	4.51	98
t11 (F)	TG	6.39	Nucleus5	ACE	Right	1.88	4.51	96
t12 (F)	TG	6.33	Nucleus5	ACE	Right	1.61	4.72	112

t13 (M)	TG	4.78	OPUS2	FS4-P	Right	1.6	3.18	110
t14 (M)	TG	4.29	Nucleus6	ACE	Right	1.1	3.19	106
t15 (M)	TG	6.81	Freedom	ACE	Right	2.28	4.53	103
t16 (F)	TG	5.16	OPUS2	FS4-P	Right	3.46	1.7	108
c1 (F)	CG	4.36	Nucleus6	ACE	Right	1.41	2.95	96
c2 (M)	CG	5.17	OPUS2	FS4-P	Right	1.21	3.95	113
c3 (M)	CG	5.25	OPUS1	FS4-P	Right	1.71	3.54	125
c4 (F)	CG	4.89	OPUS2	FS4-P	Right	1.35	3.54	118
c5 (M)	CG	4.87	OPUS1	FS4-P	Right	1.13	3.74	108
c6 (F)	CG	5.64	OPUS1	FS4-P	Right	2.67	2.98	109
c7 (M)	CG	4.26	Nucleus5	ACE	Left	1.85	2.41	110
c8 (F)	CG	4.68	Naida	HiRes	Right	2.8	1.88	102
c9 (M)	CG	5.59	OPUS2	FS4-P	Right	1.72	3.87	105
c10 (M)	CG	5.5	Nucleus6	ACE	Right	1.97	3.53	108
c11 (F)	CG	6.7	Nucleus5	ACE	Right	1.94	4.76	112
c12 (M)	CG	4.59	Nucleus6	ACE	Right	2.99	1.6	98
c13 (M)	CG	5.38	Nucleus5	ACE	Right	2.38	3	106
c14 (M)	CG	6.36	Freedom	ACE	Left	2.35	4.01	108
c15 (M)	CG	4.35	Nucleus5	ACE	Right	1.55	2.8	102
c16 (M)	CG	6.8	Freedom	ACE	Right	3.38	3.42	114

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

Participants were quasi-randomly assigned into the training group (TG, $n = 16$) and the control group (CG, $n = 16$) to ensure similar demographics (c.f., Mishra, Boddupally, & Rayapati, 2015), including chronological age (TG $\text{mean} = 5.1$ years, CG $\text{mean} = 5.27$ years; $p = .57$), age at CI (TG $\text{mean} = 1.68$ years, CG $\text{mean} = 2.02$ years; $p = .14$), CI experience (TG $\text{mean} = 3.42$ years, CG $\text{mean} = 3.25$ years; $p = .57$), and nonverbal intelligence (TG $\text{mean} = 107$, CG $\text{mean} = 108$ years; $p = .77$). Table 2 provides an overview of subject profiles of the two groups.

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

Characteristics	TG	CG	t value	p value
CA (yrs)	5.1 (0.89)	5.27 (0.8)	0.57	.57
Age at CI (yrs)	1.68 (0.62)	2.02 (0.67)	1.50	.14
CI duration (yrs)	3.42 (0.84)	3.25 (0.82)	0.58	.57
H-NTLA	107 (8.49)	108 (7.29)	0.29	.77

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

2.2. Experimental Design

The study followed a pretest-training-posttest design, as depicted in Figure 1. The pretest established the child participants' baseline performance on lexical tone perception and production, while the posttest evaluated the training-induced gains. The posttest was performed with an interval

of three weeks following pretest. During this three-week period, TG participants received five sessions of HVPT, whereas CG participants did not receive formal training but underwent pretest and posttest assessments within the same time frames as the TG. All participants used their personal CI devices with their regular settings in a soundproof therapy room for the experiment.

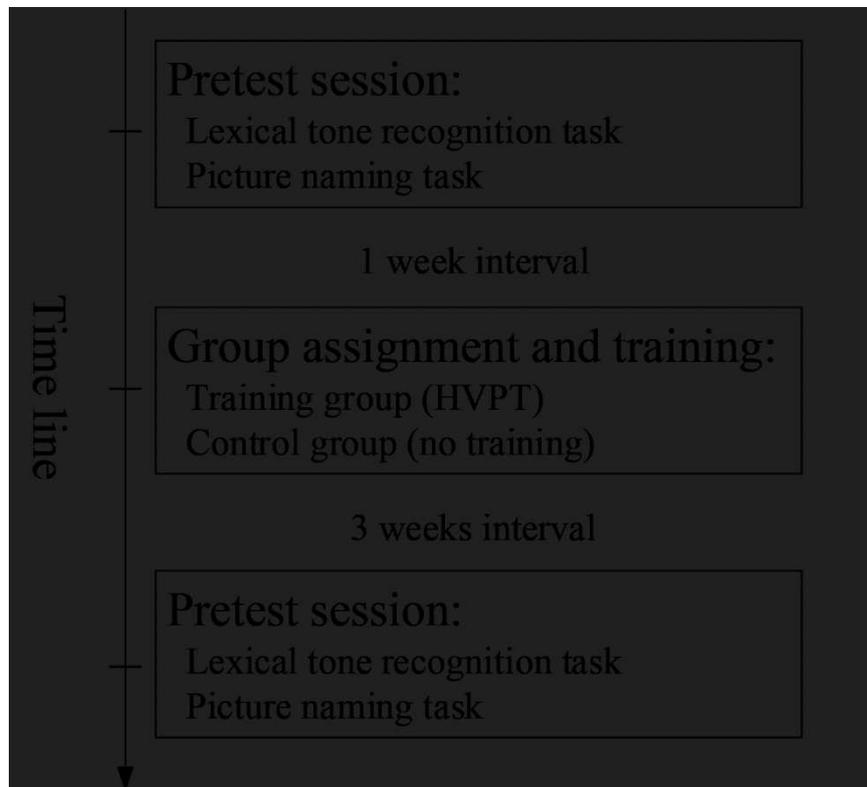


Figure 1. Schematic diagram of the experimental design. HVPT represents high variability phonetic training.

2.2.1. Test Procedures

For both pretest and posttest, lexical tone recognition and picture naming tasks were adopted to evaluate the perception and production performances of Mandarin tones. The two tasks were implemented in a random and counterbalanced order on two lab visits across the pediatric CI participants via E-Prime 2.0 program (Psychology Software Tools Inc., USA) on a laptop.

The lexical tone recognition task in this study followed procedures similar to our previous research (H. Zhang et al., 2021; H. Zhang, Zhang, Ding, & Li, 2020; H. Zhang, Zhang, Ding, & Zhang, 2020). Four native Mandarin-speaking adults (two males and two females) recorded the four Mandarin tones using the /i/ monosyllable (Pinyin Romanization “i”), producing five samples for each tone per speaker. Additionally, the same four speakers recorded the four tones using the /ɤ/ monosyllable (Pinyin Romanization “e”), with each speaker providing two samples for each tone. In total, 112 tonal stimuli were recorded (80 /i/ stimuli and 32 /ɤ/ stimuli) at a 44.1 kHz sampling rate and 16-bit depth in a sound-treated booth. Since /i/ monosyllables with the four tones were also treated as the training materials (see *Training Protocol* for details), the 80 /i/ stimuli and 32 /ɤ/ stimuli were employed to test perceptual gains of lexical tones on trained and untrained phonetic contexts. All stimuli were normalized to a root-mean-square (RMS) intensity level of 65 dB SPL, while preserving their natural duration. The stimuli were delivered individually through a loudspeaker (JBL CM220) that was placed approximately 1.2 m in front of the participant’s implanted ear. A four-alternative forced-choice (4 AFC) paradigm was adopted for tone recognition, with four pictures of cars representing the four Mandarin tones to simplify the task. These pictures matched the pitch contour of the target lexical tones: a car driving on a level road represented T1, a car driving on a rising road denoted T2, a car driving on a dipping road indicated T3, and a car driving on a falling

road signified T4. Participants were familiarized with the tone-picture matching process, underwent a practice session with trial-by-trial feedback, and then completed test sessions. The test included 80 /i/ stimuli in two blocks (trained phonetic condition) followed by 32 /x/ stimuli in one block (untrained phonetic condition). The entire lexical tone recognition task, including familiarization, practice, and test sessions, took approximately 25-35 minutes per participant.

The picture naming task involved 20 cartoon drawings of concrete objects, with five items corresponding to each type of Mandarin tone. These pictures were selected from the Mandarin Tone Identification Test (MTIT) (Zhu, Wong, & Chen, 2014a), specifically designed for Mandarin-speaking children with hearing impairment by researchers at the University of Hong Kong. Each picture could be named using a commonly used Chinese monosyllabic word typically known by children around the age of five (Zhu, Wong, & Chen, 2014b). To ensure that the child participants could correctly name the pictures with the intended lexical items, explicit instructions were provided before the formal test session. The formal testing only commenced after the pediatric CI participants had successfully learned the names of all 20 pictures. During the test, the picture stimuli were presented individually in random order, requiring children with CIs to name each presented picture three times, with at least a two-second interval between each pronunciation. All pronunciation samples were recorded using Praat software (Boersma & Weenink, 2017) on a laptop, utilizing an audio interface (Mbox Mini) paired with a head-mounted microphone (AKG C544L) at a sampling rate of 44.1 kHz (16 bit). The entire picture naming task, including the instruction phase and the test session, lasted approximately 15-25 minutes for each child participant. The stimuli used in the picture naming task can be found in Appendix A.

2.2.2. Training Protocol

The training stimuli and protocol of this experiment followed our prior studies (H. Zhang et al., 2021, 2023; H. Zhang, Zhang, Ding, & Li, 2020). The four Mandarin tones instantiated on three vowels (i.e., /i/, /a/, and /u/), resulting in 12 monosyllables as the training stimuli. A total of 600 training samples were obtained from 10 native Mandarin-speaking adults (five males and five females), with each monosyllable reproduced 5 times per speaker. The multiple talkers and diverse phonetic contexts contributed to the high-variable feature of the training stimuli. All training samples were recorded in a soundproof booth at a sampling rate of 44.1 kHz (16 bit), which were normalized to an RMS intensity level of 65 dB SPL but keep the natural duration. All training stimuli were delivered via a loudspeaker (JBL CM220) at a distance of around 1.2 m from the TG participant's implanted ear.

The training protocol consisted of five sessions of HVPT that were implemented through a computer-based training program (c.f., Miller et al., 2016a, 2016b). Phonetic identification task was adopted in the training program, because of its characteristics of more naturalistic and efficient relative to discrimination task (Lively et al., 1993; Pisoni & Lively, 1995). The phonetic identification training included a 4 AFC paradigm with the tone-picture matching requirement as the abovementioned *Test Procedures* of lexical tone recognition task. The child participants from TG were instructed to choose the matching picture of the presented training stimulus from the four options on the screen of a laptop. During the training sessions, trial-by-trial feedback was provided instantaneously; and thus, trial repetition was offered accordingly, with one repetition for trial with correct response while two repetitions for trial with incorrect response.

In addition, adaptive scaffolding and quick quizzes were incorporated to improve the efficacy of the training protocol (Y. Zhang et al., 2009). The adaptive scaffolding was prepared by controlling the extent of stimuli variability within each training session, since the blocked high-variability design was demonstrated to be beneficial for phonetic learning, especially among individuals with relatively lower perceptual aptitude (Fuhrmeister & Myers, 2020; Perrachione, Lee, Ha, & Wong, 2011; Sadakata & McQueen, 2014). The training schedule of this study began with the 120 training samples from two talkers (one male and one female) implemented in six blocks, with each block consisting of the realization of four Mandarin tones on one vowel from one specific talker. In other words, the acoustic-phonetic variability of the training materials was blocked by talkers and vowels. For the following training sessions, the training samples from two new talkers (one male and one female) were added

in the training blocks. Finally, training samples of all 10 talkers were involved in the last session. A quick quiz of lexical tone identification was provided for the trainees on 24 training samples from the two newly added talkers. The trained children were allowed to proceed to the next training session only when they exceeded 85% on the identification quiz. Otherwise, they were required to repeat the current training session; thereafter, the next training session would be provided. All participants from TG completed the five sessions of HVPT at their own pace over a period of three weeks.

2.3. Outcome Measures and Statistical Analyses

The performances of lexical tone perception and production were evaluated for the child participants of TG and CG in pretest and posttest sessions. For perception performance, we focused on the accuracy measure of the lexical tone recognition task. For production performance, we conducted perceptual judgements from a cohort of adult native speakers on child participants' Mandarin tones obtained in the picture naming task. It should be noted that a trained phonetician picked out the best exemplar from each child's three samples of each picture item, basing on the criterion of most accurate pronunciation in lexical tones (Yan, Chen, Gao, & Peng, 2021). Only the best sample for each lexical item was incorporated in the subsequent statistical analyses (Guo, Chen, Chang, & Yan, 2022; Guo, Chen, Yan, Gao, & Zhu, 2022). Therefore, a total of 1280 exemplar samples (32 child participants \times 20 lexical items \times 2 test sessions) were submitted to the evaluation of lexical tone production.

Statistical analyses were performed in R (R Core Team, 2020). Random intercepts and maximal slopes that would converge were treated as random effects while constructing statistical models (Barr, Levy, Scheepers, & Tily, 2013). The significance of fixed factors was estimated with α setting at .05, and the p values of fixed factors were obtained with the Satterthwaite's method in lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Multiple model comparisons were performed to determine the best-fit model with the lowest Akaike Information Criterion (AIC) using the anova function. Based on the best-fit model, post hoc pairwise comparisons were performed for the significant fixed factors using the emmeans package (Lenth, Singmann, Love, Buerkner, & Herve, 2018) with false discovery rate (FDR) correction (Benjamini & Yekutieli, 2001).

2.3.1. Accuracy Measure of Recognition Data

The accuracy rates for recognizing lexical tones, both on trained (/i/ stimuli) and untrained (/ʌ/ stimuli) phonetic conditions in the pretest and posttest sessions, were calculated for the participants. To account for saturation effects and ensure homoscedasticity, the accuracy percentages were transformed into rationalized arcsine units (RAU), a common practice in assessing speech intelligibility performance (Oleson, Brown, & McCreery, 2019; Studebaker, 1985; H. Zhang et al., 2021; H. Zhang, Zhang, Ding, & Zhang, 2020; L. Zhang et al., 2018). Linear mixed-effect (LME) models were constructed using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) to examine the impact of HVPT on lexical tone recognition. Given that all child participants ($n = 32$) completed the training in the trained phonetic context, while only a subset ($n = 20$, 10 from TG and 10 from CG) completed the untrained context, two separate LME models were created for each phonetic condition. The RAU score served as the dependent variable, and the models included fixed effects such as Test Session (pretest and posttest), Tone Type (T1, T2, T3, and T4), Group (TG and CG), and all possible interactions. Additionally, participant variation was considered as a random effect.

2.3.2. Perceptual Judgement of Production Data

Ten native Mandarin-speaking judges, comprising undergraduate and postgraduate students (five males and five females), with an average age of 21 years (ranging from 18 to 24 years), were recruited for the study. None of the judges had a history of speech, language, or hearing impairments. Prior to the perceptual judgment experiment, they underwent a familiarization phase with filtered speech and successfully passed a screening test, achieving over 90% accuracy in recognizing

Mandarin tones in filtered stimuli from native adults (Wong & Leung, 2018). Compensation was provided for their participation in the experiment.

We used the 1280 exemplar samples collected from all child participants (both TG and CG) during their pretest and posttest sessions as experimental stimuli. Building on the series of studies conducted by Wong and colleagues (e.g., Wong, 2012b; Wong & Leung, 2018; Wong, Schwartz, & Jenkins, 2005), these samples were low-pass filtered at 500 Hz, a procedure aimed at preserving pitch information while removing lexical information. The filtered stimuli were then normalized for intensity and grouped into blocks based on the child speakers who produced them, using E-prime 2.0. In total, 32 blocks of child productions were created, with stimuli within each block presented randomly. Judges conducted tone rating assessments at their own pace within a sound-treated room. They listened to the perceptual stimuli through Sennheiser HD280 PRO headphones at a comfortable volume. Judges had the option to replay a stimulus as many times as needed until they confidently identified it as either T1, T2, T3, or T4, at which point they recorded their response using designated keyboard buttons. To familiarize judges with the experimental procedure, a set of six practice stimuli from the screening test was provided, complete with trial-by-trial feedback, at the start of each block. The entire perceptual judgment experiment took approximately 1.5 hours, spanning two laboratory visits for each judge.

To evaluate the pediatric CI participants' accuracy in lexical tone production, a generalized linear mixed-effects (GLME) model was created using the package of lme4 (Bates et al., 2015). The binary outcome to each experimental stimulus ("1" indicating a correct response while "0" indicating an incorrect response) was treated as the dependent variable. In constructing the GLME model, Test Session (pretest and posttest), Tone Type (T1, T2, T3, and T4), Group (TG and CG), and all possible interactions were entered as fixed effects. Meanwhile, judge, participant, and lexical item were included as random effects.

2.3.3. Correlational Analysis between Tone Perception and Production Accuracy

To assess the potential relationship between pediatric CI children's lexical tone perception and production, Spearman's correlation was employed. This analysis involved all child participants ($n = 32$) and examined their tone perception accuracy in the lexical tone recognition task against their tone production accuracy in the picture naming task during the pretest session. Prior to conducting the correlation analysis, the children's tone production accuracy, as indicated by perceptual judgments, was transformed into RAU scores to match their tone recognition accuracy. Additionally, to explore the association between the benefits derived from HVPT in perception and production, Spearman's correlation was conducted on the gains of trained children ($n = 16$). These gains were calculated as the accuracy differences between the pretest and posttest sessions for lexical tone perception and production, respectively. These correlational analyses were conducted separately for each tone type as well as for all tones combined.

3. Results

3.1. Perceptual Gains induced by High Variability Phonetic Training

For the trained stimuli, the overall mean accuracies of lexical tone recognition in RAU score for TG and CG in pretest and posttest are displayed in in Figure 2. For recognition accuracy analysis, the best-fit model formula was as follows: $\text{lmer}(\text{RAU Score} \sim \text{Group} * \text{Test Session} + \text{Tone Type} * \text{Test Session} + (1 | \text{Participant}), \text{REML} = \text{FALSE})$. The LME results revealed a significant interaction effect of Group by Test Session, $\chi^2(1) = 18.98, p < .001$, which showed a significant difference in pretest-posttest changes between the two groups. Multiple post-hoc comparisons showed that significantly improved recognition performance from the pretest to posttest sessions was observed in the TG, $\beta = 21.12, SE = 2.98, t = 7.08, p < .001$, but not in the CG, $\beta = 3.03, SE = 2.98, t = 1.02, p = .31$, indicating that the significant pretest-posttest improvements in recognition accuracy were observed only in the TG rather than in the CG. Meanwhile, the difference on lexical tone recognition between the two groups was significant in posttest session, $\beta = 16.98, SE = 4.28, t = 3.97, p < .001$, but not in pretest session, $\beta =$

-1.11, $SE = 4.28$, $t = -0.26$, $p = .8$, which indicated that the two groups shared comparable baseline performance in pretest. Collectively, the results indicate that the trained pediatric CI recipients experienced significant and robust improvements in lexical tone recognition within the trained phonetic condition. Furthermore, there was a nearly significant interaction effect between Tone Type and Test Session ($\chi^2(3) = 7.71$, $p = .05$), suggesting that changes in performance from pretest to posttest varied to some extent among the four tone types. Exploratory post-hoc tests showed the recognition accuracy improved significantly for T1, $\beta = 12.01$, $SE = 4.22$, $t = 2.85$, $p = .005$, T2, $\beta = 9.88$, $SE = 4.22$, $t = 2.34$, $p = .02$, and T3, $\beta = 18.3$, $SE = 4.22$, $t = 4.34$, $p < .001$, but not for T4, $p > .05$. Figure 3 shows the mean recognition accuracy of each tone type in the trained phonetic context for the two groups in pretest and posttest sessions.

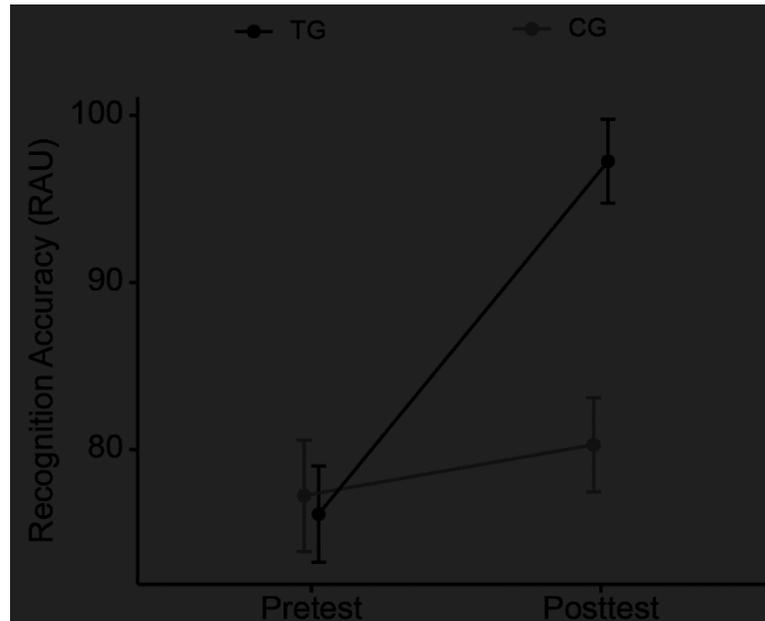


Figure 2. Overall mean recognition accuracy of lexical tones in rationalized arcsine unit (RAU) in pretest and posttest for the two groups. Error bars represent the standard errors. TG represents training group and CG represents control group.

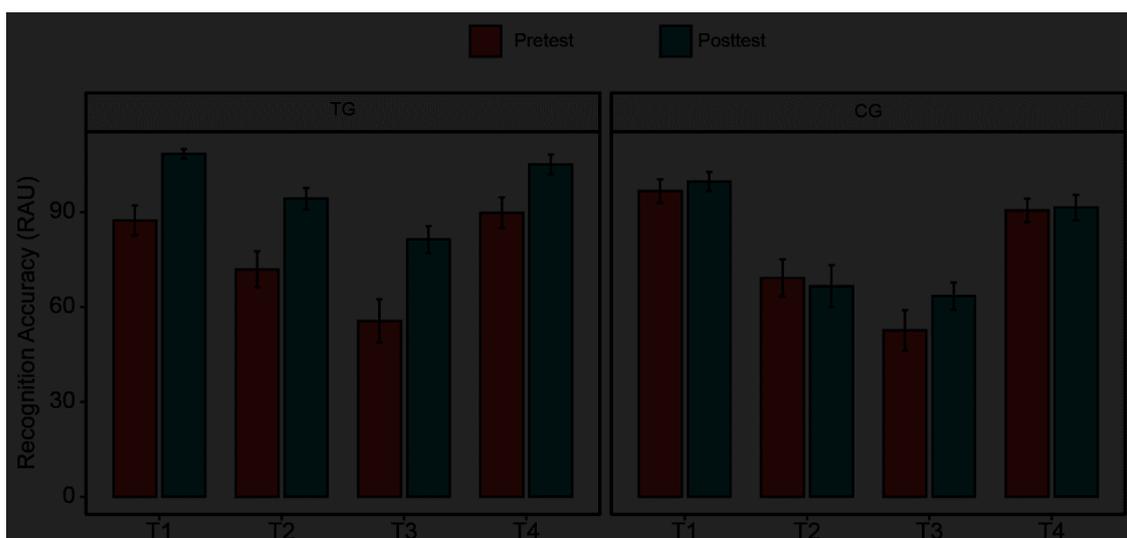


Figure 3. Mean recognition accuracy in rationalized arcsine unit (RAU) of each tone type for the trained phonetic context in the pretest and posttest for the training group (TG) and control group (CG). Error bars represent the standard errors.

To investigate whether the training induced gains on lexical tone recognition could generalize to the untrained phonetic environment, a similar LME model was built, with the best-fit model formula: $\text{lmer}(\text{RAU Score} \sim \text{Group} * \text{Test Session} + \text{Tone Type} + (1 | \text{Participant}), \text{REML} = \text{FALSE})$. The results showed significant main effects of Test Session, $\chi^2(1) = 13.1, p < .001$, and Tone Type, $\chi^2(3) = 60.2, p < .001$. Moreover, the interaction effect of Group by Test Session was also found significant, $\chi^2(1) = 18.98, p < .001$. Further post-hoc tests on the two-way interaction revealed that the difference between pretest and posttest sessions was significant for the TG, $\beta = 20.91, SE = 4.73, t = 4.42, p < .001$, but not the CG, $\beta = 2.27, SE = 4.73, t = 0.48, p = .63$, indicating that the significant pretest-posttest improvements on lexical tone recognition in the untrained phonetic condition were observed in the TG rather than in the CG. In addition, the difference between the two groups was significant in the posttest session, $\beta = 17.25, SE = 6.17, t = 2.8, p = .008$, but not in the pretest session, $\beta = -1.39, SE = 6.17, t = -0.23, p = .82$. These results suggest that the TG and CG had comparable baseline performance in lexical tone recognition during the pretest. Overall, the findings indicate that children who received HVPT demonstrated robust generalization of lexical tone recognition to untrained phonetic contexts. Figure 4 provides an overview of the recognition accuracy for each tone type within the untrained phonetic context.

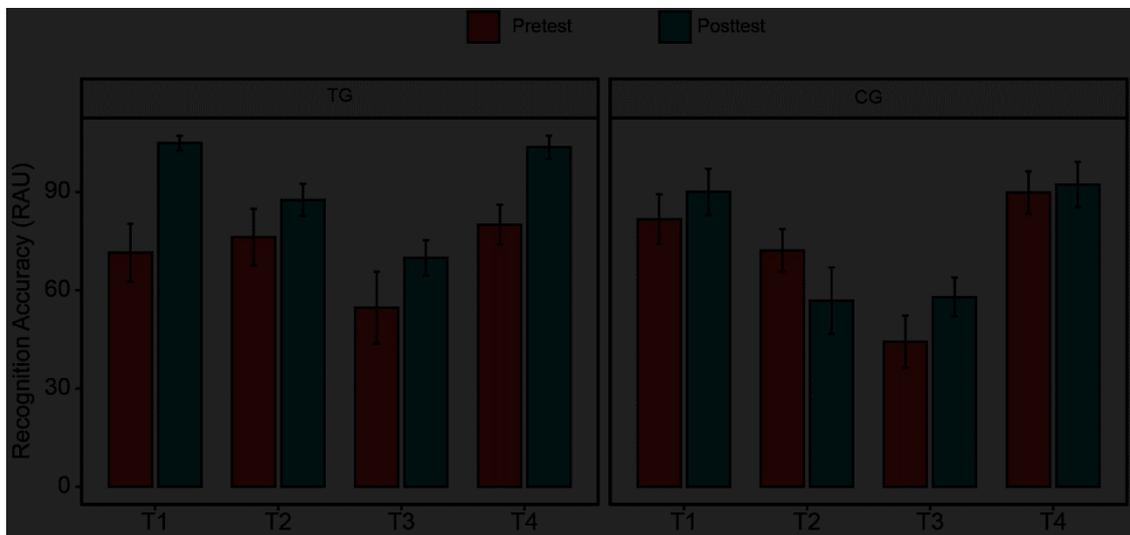


Figure 4. Mean recognition accuracy in rationalized arcsine unit (RAU) of each tone type for the untrained phonetic context in the pretest and posttest for the training group (TG) and control group (CG). Error bars represent the standard errors.

3.2. Far-transfer to Production Modality

Figure 5 displays the production accuracy of each tone for the TG and CG in different test sessions. The best fit model for analysis of perceptual judgement was: $\text{glmer}(\text{ACC} \sim \text{Group} * \text{Test Session} * \text{Tone Type} + (1 | \text{Judge}) + (1 | \text{Participant}) + (1 | \text{Lexical Item}), \text{REML} = \text{FALSE})$. The GLME model revealed significant main effects of Test Session, $\chi^2(1) = 3.93, p = .047$, and Tone Type, $\chi^2(3) = 39.52, p < .001$. Moreover, the interaction effects were also found significant, including the two-way interactions of Group by Test Session, $\chi^2(1) = 18.31, p < .001$, Group by Tone Type, $\chi^2(3) = 12.56, p = .006$, and Test Session by Tone Type, $\chi^2(3) = 18.37, p < .001$, as well as the three-way interaction of Group by Test Session by Tone Type, $\chi^2(3) = 8.79, p = .03$. Further post-hoc multiple comparisons were conducted on the three-way interaction effect. The post-hoc analysis showed that the production accuracy rate (%) of the trained children improved significantly from pretest to posttest in pronunciation of T1, $\beta = 0.53, SE = 0.25, z = 2.09, p = .036$, and T2, $\beta = 0.89, SE = 0.14, z = 6.27, p < .001$, but not T3, $\beta = 0.0005, SE = 0.12, z = 0, p = 1$, or T4, $\beta = 0.32, SE = 0.23, z = 1.39, p = .16$. The difference of production accuracy between pretest and posttest was insignificant for all of the four lexical tones in the control peers (all $ps > .2$). Meanwhile, there was no significant difference in production accuracy of each lexical tone in the pretest session (all $ps > .2$). Taken together, the results suggest that while

both the TG and CG participants displayed similar baseline performance during the pretest, only the CI children who underwent training showed significant enhancements in the accuracy of lexical tone production, particularly for T1 and T2, between the pretest and posttest sessions.

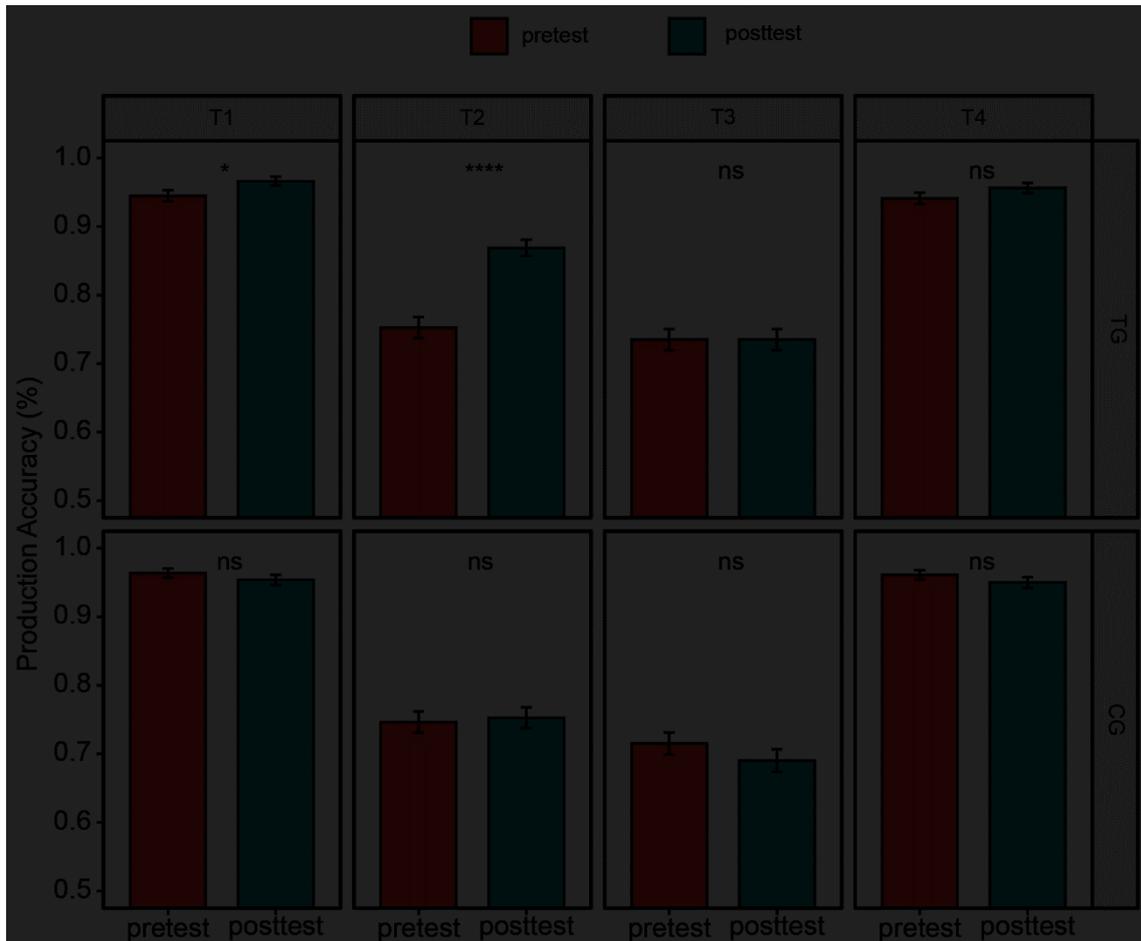


Figure 5. Mean production accuracy in percentage of each tone type for the untrained phonetic context in the pretest and posttest for the training group (TG) and control group (CG). Error bars represent the standard errors. **** indicates $p < .001$, * indicates $p < .05$, ns indicates $p > .05$.

3.3. Relationship between Perception and Production

Table 3 summarizes the correlational results between tone recognition and production accuracy for each tone type and with all tones combined among the pediatric CI participants. The analysis revealed that while the relationship between tone perception and production was not statistically significant for individual tones, there was a significant moderate relationship when considering all tones collectively. Furthermore, Table 4 outlines the correlation coefficients between training-induced improvements in perception and production, showing that the relationship between HVPT-induced perception gains and production gains did not reach statistical significance, whether analyzed for individual tones or when considering all tones together.

Table 3. Correlation between lexical tone perception and production for baseline performance ($n = 32$).

Tone Type	r -value	Effect Size	p -value
T1	-.023	small	.9
T2	-.18	small	.32
T3	.087	small	.64
T4	.32	medium	.078

All tones	.38	medium	<.01**
-----------	-----	--------	--------

** indicates $p < .01$.

Table 4. Correlation between lexical tone perception gains and production gains ($n = 16$).

Tone Type	r -value	Effect Size	p -value
T1	-.1	small	.7
T2	-.034	small	.9
T3	.057	small	.83
T4	.21	small	.43
All tones	.038	small	.76

4. Discussion

This study aimed to extend our prior work by assessing the effectiveness of HVPT for Mandarin-speaking pediatric CI users' lexical tone learning with additional tests of transfer of learning. Pretest and posttest assessments were conducted to evaluate tone recognition and production. The results aligned with our expectations based on nonnative phonetic training studies. Trained children improved recognition for both trained (i.e., /i/) and untrained (i.e., /ɤ/) syllables. Additionally, perceptual learning robustly enhanced tone production, even though the relationship between perception gains and production benefits was not significant. This represents a noteworthy discovery, potentially the first demonstrating a reliable perception-production connection in pediatric CI users via cross-modal transfer following auditory training.

4.1. Robust Generalization to Lexical Tone Recognition in Novel Phonetic Contexts

Following five sessions of HVPT, significant improvements in lexical tone recognition were observed for the TG child participants with CIs, both in trained and untrained phonetic contexts. In contrast, the CG children, who did not undergo formal training, showed no significant pretest-posttest changes in either context (see Figures 3 and 4). This underscores the substantial role of HVPT in eliciting perceptual gains, as both TG and CG had similar baseline performance. This finding replicated our previous research (H. Zhang et al., 2021, 2023) and further demonstrated that the perceptual gains could transfer to novel, untrained phonetic context by introducing high variability input. Robust generalization of perceptual learning is crucial in speech training (Lively et al., 1993; Logan et al., 1991; Pisoni & Lively, 1995). Our finding adds to the growing body of evidence that emphasizes the pivotal role of high variability input in successful learning of a variety of speech domains (Fuhrmeister & Myers, 2020; Lively et al., 1993; Shinohara & Iverson, 2018; X. Zhang, Cheng, Qin, & Zhang, 2021; X. Zhang, Cheng, & Zhang, 2021; X. Zhang, Cheng, Zou, Li, & Zhang, 2023). Moreover, examining a broader scientific literature spectrum, including visual perception, motor learning, language acquisition, computational modeling/deep learning, inductive reasoning, problem-solving, and education, it's evident that variability in input generally follows a pattern: while more variable input may initially hinder learning, it typically leads to robust generalization effects (Raviv, Lupyan, & Green, 2022).

The benefits of high variability input in perceptual training can be explained through various theoretical perspectives. One view, known as the abstractionist view, posits that listeners normalize phonetic variability by extracting generalized phonetic patterns across different talkers, allowing them to map these patterns to long-term memory representations (Ladefoged & Broadbent, 1957; Lieberman, 1973). Exposure to diverse acoustic cues during training necessitates a normalization process that emphasizes between-category phonetic differences while reducing sensitivity to within-category acoustic variations, ultimately enhancing perceptual sensitivity (Pisoni & Lively, 1995). This perspective aligns with computational models suggesting that generalization occurs when talker-specific idiosyncrasies (linguistically irrelevant cues) are dissociated from linguistically relevant cues (Ramscar & Baayen, 2013; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). In contrast, the exemplar

view contends that talker-specific idiosyncrasies are not excluded but rather encoded and stored alongside phonetically relevant information in long-term memory (Johnson, 1994, 1997). These talker-specific cues are thought to enhance learning by creating more associative connections, leading to more robust mental representations (Goldinger, 1998). This perspective aligns with recent training studies using the categorical perception paradigm, which observed improvements in within-category tone discrimination following HVPT for Mandarin-speaking pediatric CI users (H. Zhang et al., 2023). It is important to note that these two views are not necessarily contradictory; both emphasize the importance of introducing high variability input to establish more robust abstract representations. Future research, employing techniques such as electrophysiology (B. Cheng, Zhang, Fan, & Zhang, 2019) and eye-tracking (Qin, Tremblay, & Zhang, 2019), could further investigate how idiosyncratic information is used to form lexical tone representations in CI recipients undergoing formal auditory training.

4.2. Cross-modality Transfer of Perceptual Learning to Tone Production

The perceptual learning achieved through HVPT extended beyond lexical tone recognition and significantly improved the accuracy of lexical tone production in trained pediatric CI users. However, this improvement was not uniform across all four tone types, with notable gains observed for T1 and T2, while T3 and T4 showed no significant improvement (refer to Figure 5). This pattern aligns with previous research on the developmental trajectory of lexical tone production in Mandarin-speaking children aged 3-5 years, which suggests an acquisition order of the four Mandarin tones from T4 to T1 to T2 to T3 (Wong, 2012a, 2012b, 2013; Wong et al., 2005). T3, in particular, was found to be more challenging due to its complex motor sequencing and control demands, involving the activation of various laryngeal muscles (Wong, 2012b). The tangible benefits observed in T3 recognition (see Figure 2) may not have translated to significant improvements in T3 production, potentially due to immature motor control required for this complex tone. These findings support the hypotheses that perception precedes production in speech acquisition (Edwards, 1974; Greenlee, 1980; Kuhl et al., 2008) and underscore the importance of accurate tone perception as a precursor for good tone production in pediatric CI users (Xu et al., 2011; Zhou et al., 2013). Conversely, T4 production involves minimal motor control, primarily requiring the relaxation of the cricothyroid muscle (Wong, 2013). The lack of significant perceptual learning transfer to T4 production might be attributed to participants already having mastered this tone type to a ceiling-level degree. This asymmetric pattern of production improvement across the four tones aligns with recent research that suggests HVPT, combined with explicit instruction, improved nonnative beginners' productions of T1, T2, and T4 but had limited effectiveness in enhancing T3 production even after extensive classroom exposure (Wiener et al., 2020). The insignificant enhancement in T3 production of this study and nonnative learning research alike calls for future investigations exploring whether incorporating motor control exercises for laryngeal muscles into HVPT might improve T3 production accuracy for both native and nonnative learners.

In addition to the perceptual training-related benefits in production, we also evaluated correlations between lexical tone perception and production in terms of baseline performance (i.e., recognition and production accuracies in pretest) and magnitude of training-induced gains (i.e., difference between pretest and posttest) to uncover the perception-production link in Mandarin-speaking children with CIs. The significantly positive correlation between overall perception and production of lexical tones (i.e., with all four tones combined) replicated previous findings (e.g., Peng et al., 2004; Xu et al., 2011; Zhou et al., 2013) that pediatric CI users with exceptional performance in tone recognition tends to perform also well in tone production. Nevertheless, the relationship between perception gains and production gains were insignificant with astonishing small correlation coefficient ($r = .038$, $p = .76$). This observation was partially inconsistent with our prediction that was made on basis of perceptual training of second language phonemes. A recent meta-analysis presented a synthesis of 21 studies, which demonstrated a small to medium relationship between perception and production gains after training, although the relationship failed to reach significant ($r = .31$, $p = .18$) (Sakai & Moorman, 2018). The particularly small correlation coefficient of this study may be

partially due to the small number of trained participants with CIs ($n = 16$) and remarkable individual differences in the trainees' perception and production performances. More statistically robust and reliable correlation results could be found in future research recruiting more participants with relatively less heterogeneity. However, the insignificant findings echoed the meta-analysis results, suggesting that the contributing factors for production improvements may be partially or entirely independent of those for perception improvements (Sakai & Moorman, 2018).

The observed gains in production and the positive relationship between perception and production outcomes in Mandarin-speaking pediatric CI recipients undergoing perceptual training have important implications for the theoretical understanding of the interplay between perception and production in lexical tone rehabilitation. The motor theory (Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Lieberman & Mattingly, 1985) and the direct realist theory (Fowler, 1986, 1989) propose the existence of common speech representations shared by both perceptual and production modalities. Additionally, the native language magnet theory suggests that the strong connection between speech perception and production is developmental, forged through accumulated perceptual experience and learned mappings between the two modes (Kuhl et al., 2008; Kuhl & Meltzoff, 1996). Beyond shared mental representations, speech perception and production are believed to involve overlapping neural mechanisms. Neuroimaging studies have demonstrated co-activation of brain regions responsible for both speech perception and production tasks (e.g., Campbell et al., 2001; S. M. Wilson, Saygin, Sereno, & Iacoboni, 2004). Notably, the motor cortex (involved in speech production) is activated during passive speech listening, while the auditory cortex (associated with hearing speech sounds) is co-activated when participants view silent videos of mouthed speech. Hypothetically, the HVPT in this study extended the perceptual experience of lexical tones in trained pediatric CI recipients. This tuning of mental representations, coupled with potential changes in the underlying cortical processes due to neural plasticity, might have facilitated a more accurate representation of target lexical tones. These refined mental representations, in conjunction with established neural mechanisms and learned mappings between perception and production, likely contributed to improvements in lexical tone production. Future research should delve into the neural correlates of perceptual training-induced changes, employing a cognitive neuroscience approach. This would provide a comprehensive understanding of the perception-production link in the context of speech rehabilitation and learning for pediatric CI recipients.

4.3. Limitations and Future Directions

Several limitations of this study should be acknowledged. A notable limitation of this study concerns the evaluation of pediatric CI participants' lexical tone production, which relies solely on perceptual judgments by native Mandarin-speaking adults. Although the use of production accuracy measured on the same scale (RAU score) as recognition accuracy facilitates correlation analysis between perception and production, future investigations could benefit from incorporating alternative production measures, such as acoustic analyses, to capture more nuanced differences in perceptual training-induced gains in lexical tone production (Deroche, Lu, Lin, Chatterjee, & Peng, 2019; Mao, Chen, Xie, & Xu, 2020; Tang, Yuen, Xu Rattanasone, Gao, & Demuth, 2019, 2021; Tao, Liu, & Zhou, 2022).

Another limitation pertains to the absence of a long-term retention assessment. Evaluating long-term retention, which involves measuring the maintenance of training-induced benefits over an extended period, is crucial for assessing the stability of learning. Although long-term retention has received relatively limited attention in auditory training studies involving pediatric CI users (Rayes et al., 2019), it is an important indicator of the durability of learning outcomes. Future research should consider conducting follow-up assessments to determine whether the perceptual training-induced improvements in tone production persist over time for trained CI children.

Additionally, it is important to acknowledge the relatively small sample size in this study, which is a common challenge in training research due to the substantial time commitments required for both researchers and participants. Future studies should aim to increase the sample size substantially to enhance the robustness and generalizability of findings in the realm of aural interventions.

Despite these limitations, the findings of this study hold significant implications for rehabilitative strategies for pediatric CI users from tone languages. The study demonstrates tangible gains in recognizing lexical tones and highlights the cross-modality transfer of perceptual learning to tone production, affirming the effectiveness and efficiency of the HVPT protocol for lexical tone rehabilitation and learning in CI children. Given its cost-effectiveness and accessibility via the internet, the HVPT protocol has the potential to bridge the gap between laboratory research and clinical practice. However, future research endeavors should focus on confirming the ecological validity and therapeutic efficacy of HVPT for individuals with CIs in real-world clinical settings. Particular attention should be given to optimizing perceptual training methods for pronunciation improvements.

5. Conclusions

In summary, this study complements our previous investigations (H. Zhang et al., 2021, 2023) and further substantiates the efficacy of HVPT in the lexical tone rehabilitation and learning of Mandarin-speaking children with CIs. The results demonstrate perceptual improvements in recognizing lexical tones, encompassing both trained and untrained phonetic contexts. Crucially, these perceptual gains extend to the production domain, resulting in significant enhancements in lexical tone production accuracy. While the study reveals a clear perception-production link in terms of overall baseline performance, the correlation between perceptual gains and production gains remains statistically insignificant, indicating disproportionate amounts of gains in the two domains among the individuals. These findings hold both theoretical and methodological significance, offering valuable insights for future research to refine experimental designs and outcome measures for a more nuanced assessment of production benefits stemming from perceptual training in CI users.

Funding: This study was supported by grants from the National Social Science Foundation of China (18ZDA293, 21BYY020), and the Natural Science Foundation of Shandong Province (ZR2023QA117).

Acknowledgments: We thank the Shanghai Rehabilitation Center of the Deaf Children for the cooperation and assistance in implementing this study.

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

Appendix A. Stimuli of picture naming task in lexical tone production test.

Tone Type	Character	Syllable	Meaning
Tone 1	枪	/qiang1/	gun
	山	/shan1/	hill
	心	/xin1/	heart
	猫	/mao1/	cat
	车	/che1/	car
Tone 2	桥	/qiao2/	bridge
	河	/he2/	river
	梨	/li2/	pear
	门	/men2/	door
	羊	/yang2/	sheep
Tone 3	鼓	/gu3/	drum
	狗	/gou3/	dog
	火	/huo3/	fire
	伞	/san3/	umbrella
	草	/cao3/	grass

	信	/xin4/	letter
	肉	/rou4/	meat
Tone 4	饭	/fan4/	meal
	蛋	/dan4/	egg
	象	/xiang4/	elephant

References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *The CATESOL Journal*, 30(1), 177–194.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4), 1165–1188. <https://doi.org/10.1214/aos/1013699998>
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer (Computer program, Version 6.0.33). Retrieved from <http://www.praat.org>
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. <https://doi.org/10.3758/BF03206911>
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310. <https://doi.org/10.1121/1.418276>
- Cambridge, G., Taylor, T., Arnott, W., & Wilson, W. J. (2022). Auditory training for adults with cochlear implants: A systematic review. *International Journal of Audiology*. <https://doi.org/10.1080/14992027.2021.2014075>
- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., ... David, A. S. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research*, 12(2), 233–243. [https://doi.org/10.1016/S0926-6410\(01\)00054-4](https://doi.org/10.1016/S0926-6410(01)00054-4)
- Chao, Y. R. (1948). *Mandarin primer: An intensive course in spoken Chinese*. Berkeley, CA: Harvard University Press.
- Chen, Y., & Wong, L. L. N. (2017). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology*. <https://doi.org/10.1080/14992027.2017.1300694>
- Cheng, B., Zhang, X., Fan, S., & Zhang, Y. (2019). The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study. *Frontiers in Psychology*, 10, 1178. <https://doi.org/10.3389/fpsyg.2019.01178>
- Cheng, X., Liu, Y., Shu, Y., Tao, D.-D., Wang, B., Yuan, Y., ... Chen, B. (2018). Music training can improve music and speech perception in pediatric Mandarin-speaking cochlear implant users. *Trends in Hearing*, 22, 2331216518759214. <https://doi.org/10.1177/2331216518759214>
- Deroche, M. L. D., Lu, H.-P., Lin, Y.-S., Chatterjee, M., & Peng, S.-C. (2019). Processing of acoustic information in lexical tone production and perception by pediatric cochlear implant recipients. *Frontiers in Neuroscience*, 13, 639. <https://doi.org/10.3389/fnins.2019.00639>
- Dong, H., Clayards, M., Brown, H., & Wonnacott, E. (2019). The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones. *PeerJ*, 7, e7191. <https://doi.org/10.7717/peerj.7191>
- Edwards, M. L. (1974). Perception and production in child phonology: The testing of four hypotheses. *Journal of Child Language*, 1(2), 205–219. <https://doi.org/10.1017/S0305000900000659>
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, 14(1), 3–28. [https://doi.org/10.1016/s0095-4470\(19\)30607-2](https://doi.org/10.1016/s0095-4470(19)30607-2)
- Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology*, 1(2), 145–160. https://doi.org/10.1207/s15326969eco0102_3
- Fuhrmeister, P., & Myers, E. B. (2020). Desirable and undesirable difficulties: Influences of variability, training schedule, and aptitude on nonnative phonetic learning. *Attention, Perception, & Psychophysics*, 82, 2049–2065. <https://doi.org/10.3758/s13414-019-01925-y>
- Gao, Q., Wong, L. L. N., & Chen, F. (2021). A review of speech perception of Mandarin-speaking children with cochlear implantation. *Frontiers in Neuroscience*, 15, 773694. <https://doi.org/10.3389/fnins.2021.773694>

21. Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*. <https://doi.org/10.1037/0033-295X.105.2.251>
22. Greenlee, M. (1980). Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception. *Journal of Child Language*, 7(3), 459–468. <https://doi.org/10.1017/S0305000900002786>
23. Guo, C., Chen, F., Chang, Y., & Yan, J. (2022). Applying Random Forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production. *Biomedical Signal Processing and Control*, 77. <https://doi.org/10.1016/j.bspc.2022.103811>
24. Guo, C., Chen, F., Yan, J., Gao, X., & Zhu, M. (2022). Atypical prosodic realization by Mandarin-speaking autistic children: Evidence from tone sandhi and neutral tone. *Journal of Communication Disorders*, 100. <https://doi.org/10.1016/j.jcomdis.2022.106280>
25. Henshaw, H., & Ferguson, M. A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: A systematic review of the evidence. *PloS One*, 8(5), e62836. <https://doi.org/10.1371/journal.pone.0062836>
26. Hiskey, M. S. (1966). *Hiskey-Nebraska test of learning aptitude*. Cambridge, UK: Union College Press.
27. 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.
28. Johnson, K. (1994). Memory for vowel exemplars. *The Journal of the Acoustical Society of America*, 95(5), 2977–2977. <https://doi.org/10.1121/1.408940>
29. Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.
30. Kim, S., Chou, H.-H., & Luo, X. (2021). Mandarin tone recognition training with cochlear implant simulation: Amplitude envelope enhancement and cue weighting. *The Journal of the Acoustical Society of America*, 150(2), 1218–1230. <https://doi.org/10.1121/10.0005878>
31. Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979–1000. <https://doi.org/10.1098/rstb.2007.2154>
32. Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4), 2425–2438. <https://doi.org/10.1121/1.417951>
33. Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
34. Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America*, 29(1), 98–104. <https://doi.org/10.1121/1.1908694>
35. Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). emmeans: Estimated marginal means, aka least-squares means. *R Package Version 1.3.0*. Retrieved from <https://cran.r-project.org/package=emmeans>
36. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461. <https://doi.org/10.1037/h0020279>
37. Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
38. Liberman, P. (1973). On the evolution of language: A unified view. *Cognition*, 2(1), 59–94. [https://doi.org/10.1016/0010-0277\(72\)90030-3](https://doi.org/10.1016/0010-0277(72)90030-3)
39. 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. *The Journal of the Acoustical Society of America*, 94(3), 1242–1255. <https://doi.org/10.1121/1.408177>
40. Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, 96(4), 2076–2087. <https://doi.org/10.1121/1.410149>
41. Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
42. Mao, Y., Chen, H., Xie, S., & Xu, L. (2020). Acoustic assessment of tone production of prelingually-deafened Mandarin-speaking children with cochlear implants. *Frontiers in Neuroscience*, 14. <https://doi.org/10.3389/fnins.2020.592954>
43. Miller, S., Zhang, Y., & Nelson, P. (2016a). Neural correlates of phonetic learning in postlingually deafened cochlear implant listeners. *Ear and Hearing*, 37(5), 514–528. <https://doi.org/10.1097/AUD.0000000000000287>
44. Miller, S., Zhang, Y., & Nelson, P. B. (2016b). Efficacy of multiple-talker phonetic identification training in postlingually deafened cochlear implant listeners. *Journal of Speech, Language, and Hearing Research*, 59(1), 90–98. https://doi.org/10.1044/2015_JSLHR-H-15-0154

45. Mishra, S. K., Boddupally, S. P., & Rayapati, D. (2015). Auditory learning in children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 58(3), 1052–1060. https://doi.org/10.1044/2015_JSLHR-H-14-0340
46. Moore, D. R., & Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nature Neuroscience*, 12(6), 686–691. <https://doi.org/10.1038/nn.2326>
47. Oleson, J. J., Brown, G. D., & McCreery, R. (2019). The evolution of statistical methods in speech, language, and hearing sciences. *Journal of Speech, Language, and Hearing Research*, 62(3), 498–506. https://doi.org/10.1044/2018_JSLHR-H-ASTM-18-0378
48. Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, 12(4), 316–331. <https://doi.org/10.1177/1084713808325881>
49. Peng, S.-C., Tomblin, J. B., Cheung, H., Lin, Y.-S., & Wang, L.-S. (2004). Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. *Ear and Hearing*, 25(3), 251–264. <https://doi.org/10.1097/01.AUD.0000130797.73809.40>
50. Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461–472. <https://doi.org/10.1121/1.3593366>
51. 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.
52. 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. *Journal of Phonetics*, 73, 144–157. <https://doi.org/10.1016/j.wocn.2019.01.002>
53. R Core Team. (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from <https://www.r-project.org/>
54. Ramscar, M., & Baayen, H. (2013). Production, comprehension, and synthesis: A communicative perspective on language. *Frontiers in Psychology*, 4, 233. <https://doi.org/10.3389/fpsyg.2013.00233>
55. Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909–957. <https://doi.org/10.1111/j.1551-6709.2009.01092.x>
56. Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2022.03.007>
57. Rayes, H., Al-Malky, G., & Vickers, D. (2019). Systematic review of auditory training in pediatric cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*, 62(5), 1574–1593. https://doi.org/10.1044/2019_JSLHR-H-18-0252
58. Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, 5, 1318. <https://doi.org/10.3389/fpsyg.2014.01318>
59. 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. *Applied Psycholinguistics*, 39(1), 187–224. <https://doi.org/10.1017/S0142716417000418>
60. Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics*, 66, 242–251. <https://doi.org/10.1016/j.wocn.2017.11.002>
61. Singh, L., & Fu, C. S. L. (2016). A new view of language development: The acquisition of lexical tone. *Child Development*, 87(3), 834–854. <https://doi.org/10.1111/cdev.12512>
62. Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech, Language, and Hearing Research*, 28(3), 455–462. <https://doi.org/10.1044/jshr.2803.455>
63. Tamati, T. N., Pisoni, D. B., & Moberly, A. C. (2022). Speech and language outcomes in adults and children with cochlear implants. *Annual Review of Linguistics*, 8, 299–319. <https://doi.org/10.1146/annurev-linguistics-031220-011554>
64. Tan, J., Dowell, R., & Vogel, A. (2016). Mandarin lexical tone acquisition in cochlear implant users with prelingual deafness: A review. *American Journal of Audiology*, 25(3), 246–256. https://doi.org/10.1044/2016_AJA-15-0069
65. Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2019). The acquisition of Mandarin tonal processes by children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 62(5), 1309–1325. https://doi.org/10.1044/2018_JSLHR-S-18-0304
66. Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2021). Longer cochlear implant experience leads to better production of Mandarin tones for early implanted children. *Ear and Hearing*, 42(3), 1405–1411. <https://doi.org/10.1097/AUD.0000000000001036>
67. Tao, D., Deng, R., Jiang, Y., Galvin III, J. J., Fu, Q.-J., & Chen, B. (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear and Hearing*, 36(1), 102–110. <https://doi.org/10.1097/AUD.0000000000000086>

68. Tao, D., Liu, J.-S., & Zhou, N. (2022). Acoustic analysis of tone production in Mandarin-speaking bimodal cochlear implant users. *JASA Express Letters*, 2(5), 055201. <https://doi.org/10.1121/10.0010376>
69. Wang, W. S. Y. (1973). The Chinese language. *Scientific American*, 228(2), 50–63. <https://doi.org/10.1038/scientificamerican0273-50>
70. Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2), 1033–1043. <https://doi.org/10.1121/1.1531176>
71. Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649–3658. <https://doi.org/10.1121/1.428217>
72. Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25–47. <https://doi.org/10.1159/000261901>
73. Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions? *The Modern Language Journal*, 104(1), 152–168. <https://doi.org/10.1111/modl.12619>
74. Wilson, B. S., Dorman, M. F., Woldorff, M. G., & Tucci, D. L. (2011). Cochlear implants. Matching the prosthesis to the brain and facilitating desired plastic changes in brain function. In *Progress in Brain Research* (Vol. 194, pp. 117–129). <https://doi.org/10.1016/B978-0-444-53815-4.00012-1>
75. Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701–702. <https://doi.org/10.1038/nn1263>
76. Wong, P. (2012a). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, 40(1), 141–151. <https://doi.org/10.1016/j.wocn.2011.10.005>
77. Wong, P. (2012b). Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and in the United States: Interjudge reliability and perceptual results. *Journal of Speech, Language, and Hearing Research*, 55(5), 1423–1437. [https://doi.org/10.1044/1092-4388\(2012/11-0273\)](https://doi.org/10.1044/1092-4388(2012/11-0273))
78. Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America*, 133(1), 434–443. <https://doi.org/10.1121/1.4768883>
79. Wong, P., & Leung, C. T. T. (2018). Suprasegmental features are not acquired early: Perception and production of monosyllabic cantonese lexical tones in 4- to 6-year-old preschool children. *Journal of Speech, Language, and Hearing Research*, 61(5), 1070–1085. https://doi.org/10.1044/2018_JSLHR-S-17-0288
80. Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research*, 48, 1065–1079. [https://doi.org/10.1044/1092-4388\(2005/074\)](https://doi.org/10.1044/1092-4388(2005/074))
81. Wu, J.-L., Yang, H.-M., Lin, Y.-H., & Fu, Q.-J. (2007). Effects of computer-assisted speech training on Mandarin-speaking hearing-impaired children. *Audiology and Neurotology*, 12(5), 307–312. <https://doi.org/10.1159/000103211>
82. Xu, L., Chen, X., Lu, H., Zhou, N., Wang, S., Liu, Q., ... Han, D. (2011). Tone perception and production in pediatric cochlear implants users. *Acta Oto-Laryngologica*, 131(4), 395–398. <https://doi.org/10.3109/00016489.2010.536993>
83. Yan, J., Chen, F., Gao, X., & Peng, G. (2021). Auditory-motor mapping training facilitates speech and word learning in tone language-speaking children with autism: An early efficacy study. *Journal of Speech, Language, and Hearing Research*, 64(12), 4664–4681. https://doi.org/10.1044/2021_JSLHR-21-00029
84. Yang, X. J., Qu, C. Y., Sun, X. B., Zhu, F., & Wang, J. P. (2011). Norm revision of H-NTLA for children from 3 to 7 years old in China. *Chinese Journal of Clinical Psychology*, 19(2), 195–197.
85. Zhang, H., Ding, H., & Zhang, Y. (2021). High-variability phonetic training benefits lexical tone perception: An investigation on mandarin-speaking pediatric cochlear implant users. *Journal of Speech, Language, and Hearing Research*, 64(6), 2070–2084. https://doi.org/10.1044/2021_JSLHR-20-00631
86. Zhang, H., Ma, W., Ding, H., & Zhang, Y. (2023). Sustainable benefits of high variability phonetic training in Mandarin-speaking kindergarteners with cochlear implants: Evidence from categorical perception of lexical tones. *Ear & Hearing*, 44(5), 990–1006. <https://doi.org/10.1097/aud.0000000000001341>
87. Zhang, H., Zhang, J., Ding, H., & Li, Y. (2020). 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.
88. Zhang, H., Zhang, J., Ding, H., & Zhang, Y. (2020). Bimodal benefits for lexical tone recognition: An investigation on Mandarin-speaking preschoolers with a cochlear implant and a contralateral hearing aid. *Brain Sciences*, 10(4), 238. <https://doi.org/10.3390/brainsci10040238>
89. Zhang, L., Wang, J., Hong, T., Li, Y., Zhang, Y., & Shu, H. (2018). Mandarin-speaking, kindergarten-aged children with cochlear implants benefit from natural F0 patterns in the use of semantic context during speech recognition. *Journal of Speech, Language, and Hearing Research*, 61(8), 2146–2152. https://doi.org/10.1044/2018_JSLHR-H-17-0327

90. Zhang, X., Cheng, B., Qin, D., & Zhang, Y. (2021). Is talker variability a critical component of effective phonetic training for nonnative speech? *Journal of Phonetics*, 87, 101071. <https://doi.org/10.1016/j.wocn.2021.101071>
91. Zhang, X., Cheng, B., & Zhang, Y. (2021). The role of talker variability in nonnative phonetic learning: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, 64(12), 4802–4825. https://doi.org/10.1044/2021_jslhr-21-00181
92. Zhang, X., Cheng, B., Zou, Y., Li, X., & Zhang, Y. (2023). Cognitive factors in nonnative phonetic learning: Impacts of inhibitory control and working memory on the benefits and costs of talker variability. *Journal of Phonetics*, 100, 101266. <https://doi.org/10.1016/j.wocn.2023.101266>
93. Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E. B., ... Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage*, 46(1), 226–240. <https://doi.org/10.1016/j.neuroimage.2009.01.028>
94. Zhou, N., Huang, J., Chen, X., & Xu, L. (2013). Relationship between tone perception and production in prelingually-deafened children with cochlear implants. *Otology & Neurotology*, 34(3), 499–509. <https://doi.org/10.1097/MAO.0b013e318287ca86>
95. Zhou, N., & Xu, L. (2008). Development and evaluation of methods for assessing tone production skills in Mandarin-speaking children with cochlear implants. *The Journal of the Acoustical Society of America*, 123(3), 1653–1664. <https://doi.org/10.1121/1.2832623>
96. Zhu, S., Wong, L. L. N., & Chen, F. (2014a). Development and validation of a new Mandarin tone identification test. *International Journal of Pediatric Otorhinolaryngology*, 78(12), 2174–2182. <https://doi.org/10.1016/j.ijporl.2014.10.004>
97. Zhu, S., Wong, L. L. N., & Chen, F. (2014b). Tone identification in Mandarin-speaking children with profound hearing impairment. *International Journal of Pediatric Otorhinolaryngology*, 78(12), 2292–2296. <https://doi.org/10.1016/j.ijporl.2014.10.036>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.