10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Efficacy of Multi-Talker Phonetic Training in Mandarin Tone Perception for Native Pediatric Cochlear Implant Users

Hao Zhang [1], *Jing Zhang* [1], *Hongwei Ding* [1*], *Yongqin Li* [2]

[1] Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
[2] Shanghai Rehabilitation Center of the Deaf Children, Shanghai Rehabilitation & Vocational Training Center for the Disabled, Shanghai, China

zhang.hao@sjtu.edu.cn, hwding@sjtu.edu.cn

Abstract

Previous studies have documented remarkable benefits of cochlear implants (CIs) in speech and language outcomes for pediatric recipients. However, pitch perception poses a unique challenge for CI users, especially for the native tonal language listeners. The present study evaluated the efficacy of multi-talker phonetic training in improving lexical tone perception for Mandarin-speaking preschoolers with CIs. Eight unilaterally implanted children with bilateral severe-to-profound hearing loss were recruited to participate in the training protocol which involved a five-phase tonal identification training. Another eight CI peers with comparable demographic characteristics served as a control group. Both trainees and controls completed identical pre- and post-training tone recognition and categorical perception (CP) tests. The results indicated significant improvement in lexical tone recognition for the trainees after the training terminated, and the training effect transferred to their tonal CP performances. Such training-induced outcomes were not observed in the control group. To conclude, the multi-talker phonetic training regimen was validated effective and feasible to enhance CI children's tonal categorization.

Index Terms: multi-talker phonetic training, Mandarin tone perception, cochlear implant, preschoolers

1. Introduction

Cochlear implantation is the most effective clinical treatment to restore auditory sensation for individuals with severe-to-profound sensorineural deafness [1]. Albeit cochlear implant (CI) provides significant benefits in speech and language performances, CI recipients are well demonstrated poorly in pitch perception [2]–[4]. Pitch variations draw meaningful distinctions between words in tonal languages, such as Mandarin, which places greater reliance on pitch information for tonal language listeners' speech recognition. Therefore, it is of great interest to develop efficient and feasible training paradigms to promote pitch perception in Mandarin-speaking CI users.

Music training regimes have been documented applicable in improving pitch-related perceptual performances for CI recipients [5]–[7]. Benefits of music training are well demonstrated for the discrimination of melodic contour and rhythm in deaf children with CIs, and the training-induced gains can transfer to the perception of emotional speech prosody [6]. Moreover, music training can potentially contribute to improvements in lexical tone perception, since pitch cues are important for both music and tones. In a recent study, an eight-week melodic contour identification (MCI) training was introduced to pediatric Mandarin-speaking CI users [7]. The result suggested that music training was able to benefit CI children's music and lexical tone perception. These results have considerable clinical implications, since they suggest that CI children are able to improve their pitch perception following science-based auditory training interventions. However, it is noteworthy that the auditory training paradigms mentioned above are relatively lengthy and whether more fine-grained pitch perceptual performance, such as tonal categorical perception (CP), can be promoted through formal training waiting for further investigation.

CP requires the fine-grained phonetic categorization ability, mapping the variable acoustic signal onto a finite set of phonemes. The well-developed phonetic categorization of native phonemes is demonstrated to have a strong correlation with young children's later developed speech perception and production skills [8], [9]. The basic phonetic categorization ability lays the fundamental basis for children's typical language development. In perspective of Mandarin tones, a recent study revealed that native children with normal hearing show tonal CP as early as four years of age [10]. As a result, improvement in tonal categorization should be taken into consideration while developing efficient auditory training paradigms for Mandarin-speaking pediatric CI users.

Ample evidence has shown that phonetic categorization can be promoted by high variability phonetic training, which mainly comes from cross-linguistic studies. For example, it can effectively improve the phonetic categorization of the non-native /r/-/l/ contrast in Japanese listeners [11]–[13]. In addition, the effectiveness of high variability training has been validated in improving the categorization of Mandarin tones for non-native learners [14]. High variability of training materials is vital for efficient training outcomes. It is assumed that high variability identification training can increase auditory sensitivity to acoustic variations in an intended way: enhancing the sensitivity to primary acoustic differences of between-category without raising sensitivity to that of within-category [15].

[*] Corresponding author

10.21437/SpeechProsody.2020-167

For CI individuals, the efficacy of high variability identification training has been evaluated in postlingually deafened adults [16]. In that study, a pretest-intervention-posttest design was implemented with four two-hour sessions of phonetic identification training over a period of two weeks. The results demonstrated that the multi-talker phonetic identification training promoted the CP of speech contrasts of /ba/–/da/ and /wa/–/ja/. Moreover, that study also investigated the neural correlates underpinning the improvement of phonetic categorization in terms of enhanced mismatch negativity (MMN) responses to sound stimuli across different phonetic categories relative to stimuli within the same category [16].

The aforementioned findings are encouraging because they suggest the plasticity and potential enhancement for phonetic categorization. However, both second language learners and postlingually deafened adults have the normal auditory hearing experience, which is crucially different from the congenitally deaf children with CIs. Therefore, the benefit of high variability training awaits validation via further empirical investigations. The primary aim of this study is to evaluate the efficacy of multi-talker phonetic training in improving tonal categorization for Mandarin-speaking CI children.

2. Methods

2.1. Participants

Sixteen Mandarin-speaking preschoolers with an age range of 4-6 years were recruited for this study. All child participants were born with bilateral severe-to-profound sensorineural hearing loss and were implanted unilaterally. They were divided into the training group and control group, with each group involved eight participants. The two groups did not differ significantly in terms of chronological age, age at implantation CI experience or nonverbal intelligence (see Table 1). This study was approved by the Ethics Committee of School of Foreign Languages, Shanghai Jiao Tong University. Informed consent was received from the caregiver of each child.

Table 1: *Demographic characteristics and available p values of independent-samples t-tests between the two groups.*

Characteristics	Training group	Control group	*p* value
Chronological age (yrs)	5.07	5.09	$p = 0.93$
Age at CI (yrs)	1.57	1.79	$p = 0.59$
CI duration (yrs)	3.5	3.3	$p = 0.63$
H-NTLA	107	112	$p = 0.24$

2.2. Test procedure

All preschool participants were instructed to complete two measures for pre- and posttest, including the natural tone recognition test and the tonal CP test, inside a sound-treated therapy room. Both tests were implemented with E-Prime 2.0 (Psychology Software Tools Inc., USA) on a Windows-based laptop. The sound stimuli were delivered to each participant via bilateral loudspeakers (JBL CM220) which were placed approximately 1.2 m from the listener.

2.2.1. Natural tone recognition test

Test materials were naturally produced four lexical tones with the syllable /i/ in Mandarin, which were obtained from four native Mandarin-speaking adults (two males and two females). Each tone was recorded five times per speaker, resulting in 80 sound stimuli in total. All stimuli were normalized equally for root-mean-square intensity level of 65 dB SPL.

A four-alternative, forced-choice (4 AFC) identification paradigm was adopted in this test. A procedural learning stage was provided for the child participants, which instructed the children to match T1, T2, T3, and T4 with the corresponding four pictures of driving cars that are popular materials for formal instructions of Mandarin tones in primary schools. Before the formal test, a training session was provided to guarantee that all participants could follow the requirements of the test. During the formal test, the participants' picture-pointing responses were logged by the experimenter. A total of 80 tonal stimuli were randomly presented to each child within two blocks. Participants could have breaks between blocks.

The percentage accuracy of four tones was computed for each participant. Afterwards, the percentage accuracy was transformed into rationalized arcsine unit (RAU) score [17] for statistical analysis.

2.2.2. Tonal categorical perception test

Mandarin syllable /i/ with T1 and T2 were selected as endpoints in this test for the synthesis of the tonal continuum. The original samples, /i/ with T1 and T2, were obtained from a female native adult, with each sample having a duration of 400 ms. Pitch tier transfer was implemented in Praat [18] to create the template tokens of the endpoints, so as to guarantee the identical acoustic features except for the pitch contours between the two templates. Based on the template tokens, a seven-stimulus tonal continuum was re-synthesized with equal pitch interval, using STRAIGHT [19]. The amplitude and duration were kept constant across all seven stimuli to 65 dB SPL and 400 ms, respectively. Figure 1 shows the schematic diagram of the pitch contour for each stimulus along the continuum.

Two classic tasks of speech CP, including the identification task and the discrimination task, were introduced to each child participant. A two-alternative forced choice (2 AFC) paradigm was adopted in the identification task. The participants were required to recognize each tonal stimulus as T1 or T2. Two blocks were prepared with the seven stimuli repeated five times in each block, resulting in a total of 70-stimulus random presentation in the identification task. An AX paradigm was used in the discrimination task. The preschoolers were instructed to decide whether the presented contrastive tonal pairs were the "same" or "different". Seven pairs were the same condition with a tonal stimulus pairing with itself (e.g., 1-1, 3-3), whereas 10 pairs were the different conditions with two steps separating the two tonal stimuli (e.g., 1-3, 3-1). All 17 constructed pairs were repeated five times. Totally, 85 tonal pairs were presented randomly in the discrimination task.

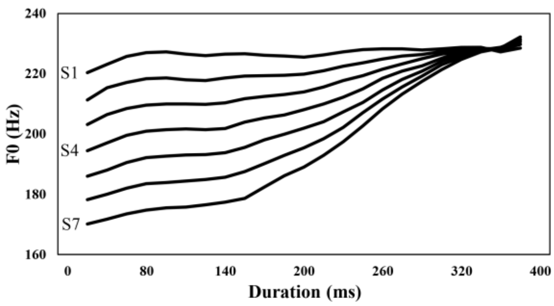

Figure 1: *Schematic diagram of the tonal stimuli.*

The identification scores and discrimination scores were estimated for each participant. Two key parameters of identification scores, boundary position (i.e., 50% cross-over point) and boundary width (i.e., the linear distance between 25% and 75%), were estimated via Probit analysis [20]. The discrimination scores were calculated according to the formula in Xu et al. [21], and were further divided into two types (i.e., the between-category accuracy and the within-category accuracy). Peakedness scores, which refer to different scores between the two discrimination types [22], were also calculated.

2.3. Training protocol

The phonetic training materials were three naturally produced Mandarin vowels (/i/, /a/, and /u/) with four lexical tones. The 12 training syllables were recorded from 10 native Mandarin-speaking adults (five males and five females), with each syllable reproduced five times. Thus, 600 training items with high variability were obtained in total.

A computer-based training program was developed according to the protocol described specifically in Miller et al. [23]. The trainees were introduced to complete five sessions of phonetic identification training. A 4 AFC identification task was involved in the training program, which required the trainees to determine the sound stimuli as T1, T2, T3, or T4. Feedback was provided for each training item. Meanwhile, one repetition was offered for items with correct identification responses, and two repetitions were offered for items with incorrect responses.

Training began with two unique talkers (one male and one female) in the first session, involving 120 training items implemented in six blocks. Blocked high variability design was employed to control the variability within each training block [24]. Afterwards, two additional talkers (one male and one female) were added in the following sessions, until all 10 talkers were involved in the last session. It should be noted that a tonal identification quiz of 24 training materials from the recently added two talkers was provided for the trainees after each session. The current training session terminated only after the trainees obtained an accuracy of over 85% for the quiz. Child participants completed all five training sessions at their own pace over a period of three weeks.

3. Results

3.1. Training effect in natural tone recognition

Mean RAU accuracy scores of natural tone recognition in pre- and posttest are shown in Figure 2 for both training and control CI preschoolers. The average identification accuracies across four tones for the trainees and controls were respectively 79.08 and 80.51 in pretest, while were respectively 99.67 and 80.65 in posttest. Training effect was assessed using a mixed-design of repeated measures ANOVA, with Group (training, control) as between-subject factor, Test Session (pretest, posttest) and Tone Type (T1, T2, T3, T4) as within-subject factor. When applicable, Bonferroni or Greenhouse–Geisser corrections were employed to adjust p values.

The result indicated a significant interaction of Group × Test Session (F (1, 14) = 13.51, p = 0.002). Post-hoc simple main effect showed significantly higher identification accuracy in posttest than in pretest for the training group (F (1, 14) = 27.38, p < 0.001), but not for the control group (p = 0.97).

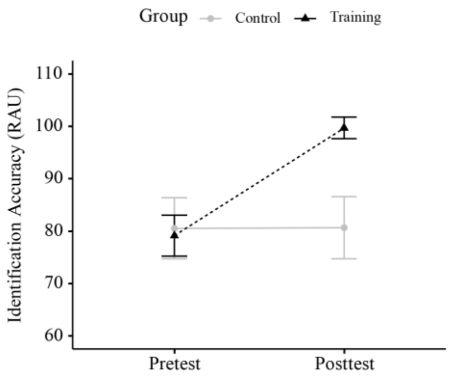

Figure 2: *Mean accuracy of natural tone recognition across pre- and posttest for training group and control group.*

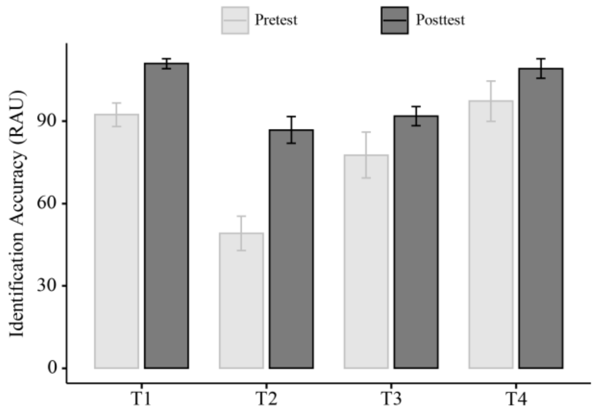

Figure 3: *Identification accuracy of four Mandarin tones across pre- and posttest for the training group.*

In addition, a significant interaction of Tone Type × Test Session (F (3, 42) = 3.57, p = 0.03) was observed. Simple main effect revealed that the training effect was significant for T2 (F (1, 14) = 32.27, p < 0.001).

3.2. Training effect in tonal categorical perception

The identification and discrimination curves of tonal CP in two test sessions for the training and control groups are depicted in Figure 4. Typical S shapes were shown for all identification curves. Meanwhile, prominent discrimination accuracy peaks were well aligned with the corresponding identification crossovers. Mixed-designs of ANOVA were performed to investigate whether the training effect could transfer to benefit the trained CI children's performances in tonal CP. In the ANOVA models, the key identification and discrimination parameters served as dependent variables, while Group and Test Session served as between-subject and within-subject variables, respectively.

For the boundary width analysis, a significant Group × Test Session (F (1, 14) = 4.9, p = 0.04) interaction was revealed. Simple main effect revealed that the boundary widths were significantly narrower in posttest than in pretest for the trainees (F (1, 14) = 11.34, p = 0.005), but not for the controls (p = 0.82). Additionally, a marginal significant Group × Test Session (F (1, 14) = 3.51, p = 0.08) interaction was found for the peakedness score analysis, with the trained children revealed significantly higher score in posttest over in pretest (F (1, 14) = 4.54, p = 0.05).

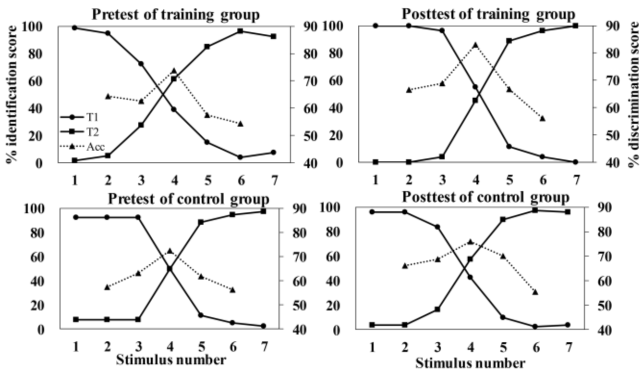

Figure 4: *Identification and discrimination curves of tonal CP across pre- and posttest for training group and control group.*

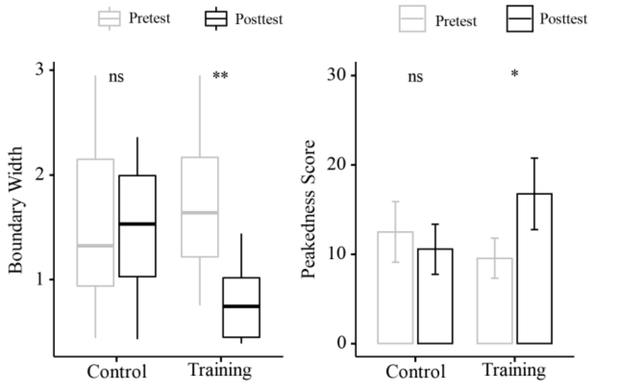

Figure 5: *Mean boundary width and peakedness score of tonal CP across pre- and posttest for training and control groups.*

4. Discussion

The present study aims to examine the effectiveness of multi-talker phonetic identification training in improving Mandarin-speaking CI children's lexical tone perception. The results indicated that the trained preschoolers received significant improvements in natural tone recognition. Moreover, a robust training effect transferred to their tonal CP performance.

Significant improvements in natural tone recognition were observed for the trained CI children by an average RAU score of 20.59 following the training paradigm in this study. This result was consistent with a recent study by Miller et al. [23] which demonstrated that multiple-talker phonetic training was able to improve adult CI users' perception of the /ba/-/da/ and /wa/-/ja/ speech contrasts. However, great variability existed in this study for the training-induced effects across different tone types, with T2 received more evident gains than the other three tones (see Figure 3). Several previous studies have documented a remarkable identification confusion between T2 and T3 for pediatric Mandarin-speaking children with CIs [25]–[27]. The confusion between T2 and T3 contributes to much poorer recognition of these two tones relative to T1 and T4, which in turn leaves more potential room for improvement in T2 and T3. Therefore, it is reasonable to propose that the extremely poor identification of T2 in the pre-training stage leads to the significant training benefits for T2 recognition. On the other hand, the great heterogeneity of T3 performance in the pretest might limit the benefits obtained on the group level. The data added to the extant literature in supporting the introduction of formal auditory training for CI users' speech rehabilitation. Moreover, the findings implicated that the training protocols should emphasize specific lexical tone contrasts (e.g., T2 and T3) to reach efficient and optimal outcomes.

Characteristics of CP were exhibited in the perception of T1 and T2 for the preschool participants with CIs in this study. The CP of lexical tones echoes prior studies that recruited Mandarin speakers of pediatric CI users with prelingual deafness [28] and adult participants with postlingual hearing loss [29]. Although both mentioned studies involved merely the identification task of CP paradigm, their results all revealed typical S-shaped functions for the tonal identification curves. Together with the current findings, all results converged to the conclusion that native listeners with sensorineural hearing loss are able to perceive Mandarin tones categorically, regardless of the well-known degraded spectral and temporal cues transmitted to these clinical populations.

Moreover, the robust benefits of multi-talker identification training could transfer to CP performance for the trained CI children. The boundary width of T1 and T2 identification functions became smaller, indicating a significant sharper categorical boundary for the T1-T2 contrast due to the training. In addition, the much higher peakedness score in posttest than in pretest demonstrated that the training significantly enhanced behavioral sensitivity to the between-category stimulus pairs over the within-category pairs of the T1-T2 tonal contrast. The current data support the notion that the phonetic identification training paradigm can alter perceptual sensitivity of phonetic contrasts selectively, with raising sensitivity to between-category phonetic differences whereas reducing sensitivity to within-category differences [15]. These results collectively suggest that the high variability training protocol adopted in this study can enhance the trainees' ability in tonal categorization. Although Mandarin tones can be categorized by the duration patterns and the concurrent amplitude envelope, a recent report demonstrated that tonal perceptual performance for pediatric CI recipients was positively correlated with their reliance on the pitch contour [25]. Noted that the stimuli used in the tonal CP test were manipulated only on the pitch cue while keeping all other acoustic cues constant. As a result, the mechanism underlying the efficacy of training protocol could be postulated that the formal training with high variability materials tends to enable the trained children to weight the pitch cue more heavily when categorizing lexical tones.

5. Conclusions

The present study demonstrated that high variability training with multiple talkers can improve lexical tone recognition for Mandarin-speaking children with CIs. Moreover, the training-induced benefits can transfer to the trained children's tonal CP performance, which was reflected by sharper identification boundary and higher peakedness score in posttest relative to in pretest. To conclude, this pattern of results provides evidence of the efficacy of multi-talker phonetic training in enhancing the tonal categorization ability for pediatric CI users.

6. Acknowledgement

This work was supported by the Major Programs of National Social Science Foundation of China (18ZDA293 and 13&ZD189). We thank the Shanghai Rehabilitation Center of the Deaf Children for the cooperation and assistance in implementing this study.

7. References

[1] A. A. Eshraghi, R. Nazarian, F. F. Telischi, S. M. Rajguru, E. Truy, and C. Gupta, “The cochlear implant: Historical aspects and future prospects,” *Anat. Rec. Adv. Integr. Anat. Evol. Biol.*, vol. 295, no. 11, pp. 1967–1980, 2012.

[2] S.-C. Peng, J. B. Tomblin, H. Cheung, Y.-S. Lin, and L.-S. Wang, “Perception and production of Mandarin tones in prelingually deaf children with cochlear implants,” *Ear Hear.*, vol. 25, no. 3, pp. 251–264, 2004.

[3] D. Tao, R. Deng, Y. Jiang, J. J. Galvin III, Q.-J. Fu, and B. Chen, “Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users,” *Ear Hear.*, vol. 36, no. 1, pp. 102-110, 2015.

[4] C. M. Holt, K. Demuth, and I. Yuen, “The use of prosodic cues in sentence processing by prelingually deaf users of cochlear implants,” *Ear Hear.*, vol. 37, no. 4, pp. e256–e262, 2016.

[5] K. Gfeller, E. Guthe, V. Driscoll, and C. J. Brown, “A preliminary report of music-based training for adult cochlear implant users: Rationales and development,” *Cochlear Implants Int.*, vol. 16, no. sup3, pp. S22–S31, 2015.

[6] A. Good *et al.*, “Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants,” *Ear Hear.*, vol. 38, no. 4, p. 455-464, 2017.

[7] X. Cheng *et al.*, “Music training can improve music and speech perception in pediatric Mandarin-speaking cochlear implant users,” *Trends Hear.*, vol. 22, pp. 1-12, 2018.

[8] P. K. Kuhl, B. T. Conboy, D. Padden, T. Nelson, and J. Pruitt, “Early speech perception and later language development: Implications for the "critical period",” *Lang. Learn. Dev.*, vol. 1, no. 3–4, pp. 237–264, 2005.

[9] P. Kuhl and M. Rivera-Gaxiola, “Neural substrates of language acquisition,” *Annu. Rev. Neurosci.*, vol. 31, pp. 511–534, 2008.

[10] F. Chen, G. Peng, N. Yan, and L. Wang, “The development of categorical perception of Mandarin tones in four-to seven-year-old children,” *J. Child Lang.*, vol. 44, no. 6, pp. 1413–1434, 2017.

[11] S. E. Lively, J. S. Logan, and D. B. Pisoni, “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.*, vol. 94, no. 3, pp. 1242–1255, 1993.

[12] S. E. Lively, D. B. Pisoni, R. A. Yamada, Y. Tohkura, and T. Yamada, “Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories,” *J. Acoust. Soc. Am.*, vol. 96, no. 4, pp. 2076–2087, 1994.

[13] Y. Zhang *et al.*, “Neural signatures of phonetic learning in adulthood: A magnetoencephalography study,” *Neuroimage*, vol. 46, no. 1, pp. 226–240, 2009.

[14] M. Sadakata and J. M. McQueen, “Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training,” *Front. Psychol.*, vol. 5, pp. 1318-1332, 2014.

[15] Y. Shinohara and P. Iverson, “High variability identification and discrimination training for Japanese speakers learning English/r/–/l,” *J. Phon.*, vol. 66, pp. 242–251, 2018.

[16] S. Miller, Y. Zhang, and P. Nelson, “Neural correlates of phonetic learning in postlingually deafened cochlear implant listeners,” *Ear Hear.*, vol. 37, no. 5, pp. 514–528, 2016.

[17] G. A. Studebaker, “A "rationalized" arcsine transform,” *J. Speech, Lang. Hear. Res.*, vol. 28, no. 3, pp. 455–462, 1985.

[18] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer [Computer software]. Version 6.0. 33.” 2017.

[19] H. Kawahara, I. Masuda-Katsuse, and A. De Cheveigne, “Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds,” *Speech Commun.*, vol. 27, no. 3–4, pp. 187–207, 1999.

[20] D. J. Finney, “*Probit analysis.*” New York: Cambridge University Press, 1971.

[21] Y. Xu, J. T. Gandour, and A. L. Francis, “Effects of language experience and stimulus complexity on the categorical perception of pitch direction,” *J. Acoust. Soc. Am.*, vol. 120, no. 2, pp. 1063–1074, 2006.

[22] C. Jiang, J. P. Hamm, V. K. Lim, I. J. Kirk, and Y. Yang, “Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics,” *Mem. Cognit.*, vol. 40, no. 7, pp. 1109–1121, 2012.

[23] S. E. Miller, Y. Zhang, and P. B. Nelson, “Efficacy of multiple-talker phonetic identification training in postlingually deafened cochlear implant listeners,” *J. Speech, Lang. Hear. Res.*, vol. 59, no. 1, pp. 90–98, 2016.

[24] T. K. Perrachione, J. Lee, L. Y. Y. Ha, and P. C. M. Wong, “Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design,” *J. Acoust. Soc. Am.*, vol. 130, no. 1, pp. 461–472, 2011.

[25] S.-C. Peng, H.-P. Lu, N. Lu, Y.-S. Lin, M. L. D. Deroche, and M. Chatterjee, “Processing of acoustic cues in lexical-tone identification by pediatric cochlear-implant recipients,” *J. Speech, Lang. Hear. Res.*, vol. 60, no. 5, pp. 1223–1235, 2017.

[26] Y. Chen, L. L. N. Wong, F. Chen, and X. Xi, “Tone and sentence perception in young Mandarin-speaking children with cochlear implants,” *Int. J. Pediatr. Otorhinolaryngol.*, vol. 78, no. 11, pp. 1923–1930, 2014.

[27] G. Li, S. D. Soli, and Y. Zheng, “Tone perception in Mandarin-speaking children with cochlear implants,” *Int. J. Audiol.*, vol. 56, no. sup2, pp. S49–S59, 2017.

[28] X. Luo, Y. Chang, C.-Y. Lin, and R. Y. Chang, “Contribution of bimodal hearing to lexical tone normalization in Mandarin-speaking cochlear implant users,” *Hear. Res.*, vol. 312, pp. 1–8, 2014.

[29] B. Qi, P. Liu, X. Gu, R. Dong, and B. Liu, “Categorical perception of lexical tones in native Mandarin-speaking listeners with sensorineural hearing loss,” *Acta Otolaryngol.*, vol. 138, no. 9, pp. 801–806, 2018.