ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

The Effect of Visual Articulatory Cues on the Identification of Mandarin Tones by Children With Cochlear Implants

Shanpeng Li,[a] Yinuo Wang,[b] Qianxi Yu,[a] Yan Feng,[a] and Ping Tang[a]

[a] MIIT Key Lab for Language Information Processing and Applications, School of Foreign Studies, Nanjing University of Science and Technology, China [b] Department of English, Linguistics and Theatre Studies, Faculty of Arts & Social Sciences, National University of Singapore

ARTICLE INFO

Article History:
Received September 12, 2023
Revision received January 24, 2024
Accepted March 7, 2024

Editor-in-Chief: Cara E. Stepp
Editor: Emily Beth Myers

https://doi.org/10.1044/2024_JSLHR-23-00559

ABSTRACT

Purpose: This study explored the facilitatory effect of visual articulatory cues on the identification of Mandarin lexical tones by children with cochlear implants (CIs) in both quiet and noisy environments. It also explored whether early implantation is associated with better use of visual cues in tonal identification.
Method: Participants included 106 children with CIs and 100 normal-hearing (NH) controls. A tonal identification task was employed using a two-alternative forced-choice picture-pointing paradigm. Participants' tonal identification accuracies were compared between audio-only (AO) and audiovisual (AV) modalities. Correlations between implantation ages and visual benefits (accuracy differences between AO and AV modalities) were also examined.
Results: Children with CIs demonstrated an improved identification accuracy from AO to AV modalities in the noisy environment. Additionally, earlier implantation was significantly correlated with a greater visual benefit in noise.
Conclusions: These findings indicated that children with CIs benefited from visual cues on tonal identification in noise, and early implantation enhanced the visual benefit. These results thus have practical implications on tonal perception interventions for Mandarin-speaking children with CIs.

Speech perception is a complex process that involves various perceptual, cognitive, and linguistic skills (Lalonde & McCreery, 2020). However, this process can be largely disrupted when acoustic signals are degraded such as in a noisy environment, especially for listeners with hearing impairments (Mattys et al., 2012). Despite this, accumulating evidence has shown that speech perception is also an audiovisual (AV) process where visual cues (such as speaker's articulatory movements) play an important role in speech perception (Richie & Kewley-Port, 2008; Rosenblum, 2008). When acoustic signals are degraded, listeners dynamically allocate more cognitive resources to the visual channel to assist their speech comprehension (De Gelder & Bertelson, 2003).

Correspondence to Shanpeng Li: shanpeng@njust.edu.cn. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Children can benefit from AV integration in speech perception, albeit to a lesser extent than adults (Ernst, 2008). It was recently revealed that the development of AV speech integration occurred as early as the ages of 5–6 years in Mandarin-speaking children (Weng et al., 2023). However, it is still not clear whether the ability of AV integration can aid children with hearing impairments in improving their performance in speech perception. This study, therefore, examined how children with hearing impairments, specifically those with cochlear implants (CIs), utilized visual articulatory cues to enhance their perception of an acoustically challenging speech unit, that is, lexical tones.

For listeners with profound hearing impairments, a CI is an electronic device that helps them restore hearing and receptive speech recognition ability. Nonetheless, CI users still face great challenges in pitch perception due to the limited resolution of fundamental frequency ($F0$) of CI devices (McDermott & Looi, 2004; Smith & Burnham, 2012). This would cause difficulties in word identification, especially in noisy environments, for CI users speaking tonal languages such as Mandarin Chinese, because these languages rely heavily on pitch to contrast word meanings. For example, the syllable /ma/ varies when carrying different tones: (a) "mother," with a high-level tone (Tone 1 or T1); (b) "hemp," with a rising tone (T2); (c) "horse," with a low-rising tone (T3); and (d) "scold" with a falling tone (T4). Therefore, the ability to perceive pitch information is critical for Mandarin Chinese listeners to identify word meanings.

Although lexical tone variations hinge on glottal and subglottal activities independent of vocal tract configurations, there is growing evidence demonstrating the existence of specific co-speech visual articulatory cues during tonal production (Attina et al., 2010; Garg et al., 2019). According to Garg et al. (2019), for instance, T1 is typically characterized by minimal head and eyebrow movement compared to other tones; T2 exhibits the most pronounced eyebrow lowering and raising and requires the longest duration to reach maximum head raising distance relative to other tones; T3 demonstrates the greatest amount of head raising and lowering movement, with the largest average distances traveled by the head and eyebrows; and T4 is associated with specific lip movements.

Relative to vowels and consonants, the visual–auditory associations of lexical tones are much less transparent. Therefore, there was still considerable controversy regarding whether these visual articulatory cues can aid listeners to perceive tones better. Some studies involving nonnative listeners reported visual benefits on the perception of Mandarin tones by native Dutch listeners (Y. Han et al., 2020; Wei et al., 2022) and on the perception of Cantonese tones by native Thai listeners (Burnham, Lau, et al., 2001). In contrast, for native listeners, previous studies only demonstrated a limited role of visual information on tone identification (Burnham, Ciocca, & Stokes, 2001; Burnham et al., 2015; Mixdorff et al., 2005).

Inconsistent results regarding the benefit of visual cues on tone identification were also found for listeners with hearing impairments. For example, although S.-Y. Liu et al. (2014) reported that visual cues did not enhance Mandarin tone perception by adult CI users, the visual benefit in tone perception was observed in Thai hearing aid users (Kasisopa et al., 2015) and Mandarin normal-hearing (NH) listeners based on CI-simulated speech (Smith & Burnham, 2012). The controversial findings regarding the role of visual input may stem from the different hearing restoration devices used in each study. Traditional hearing aids function by amplifying auditory signals for user perception, while CIs diverge by circumventing damaged hair cells to directly stimulate the auditory nerve by converting the mechanical sound energy into an electrical stimulus (K. Y. S. Lee et al., 2010; A. Li et al., 2014). To be specific, although the perceptual performance in CI simulation conditions outperformed actual CI users (Friesen et al., 2001), hearing aids did not significantly improve the speech intelligibility of tone perception among hearing-impaired listeners (K. Y. S. Lee et al., 2008; Gandour et al., 1984). This lack of improvement could lead to more reliance on the visual component of AV speech.

Given the fact that visual information only plays a complementary role in tonal identification when auditory information is insufficient (Hong et al., 2023), it is not surprising to see that NH adults do not consistently gain visual benefit during tone identification. However, whether and/or the extent to which visual articulatory cues can enhance tonal identification by children with CIs remains unclear.

Moreover, there has been extensive evidence showing that early implantation is typically assoicated with better language development, such as better word recognition performance (H. Liu et al., 2015), better phonemic awareness (Moberly et al., 2016), and stronger tonal production and perception abilities (Chen & Wong, 2016; Tang et al., 2019, 2021). The benefits of early implantation may be related to the fact that the central auditory system only keeps its maximal plasticity (in response to auditory stimulation) in early childhood (Gilley et al., 2006; Sharma & Dorman, 2006; Sharma et al., 2002). It is therefore critical to receive implantation within this period to prevent the degeneration of the central auditory system caused by early auditory deprivation (Finney et al., 2001; D. S. Lee et al., 2001). Therefore, it also raises the question of whether early implantation benefits the ability to better integrate visual cues into tone identification.

Therefore, this study examined the effect of visual information on the identification of Mandarin tones by children with CIs in both quiet and noisy environments. We first asked: (a) Do visual cues enhance children with CIs' tonal identification? If yes, we further asked: (b) Does earlier implantation facilitate better use of visual cues in tonal perception? We predicted that (a) children with CIs would show an improved tonal perception accuracy in AV as compared to audio-only (AO) settings and (b) children implanted earlier would exhibit greater benefits from visual cues in tonal perception.

Method

Participant

A total of 106 children with CIs were recruited from speech rehabilitation institutions from Northern China, specifically Beijing and Hebei Province. Fourteen children were excluded from the final data analysis due to failure to comprehend the task or voluntary withdrawal. The final data set included 94 children with CIs,[1] consisting of 50 boys and 44 girls. Their chronological age ranged from 3.08 to 7.33 years (M = 5.13 years, SD = 0.97 years) and implantation age ranged from 0.42 to 5 years (M = 2.17 years, SD = 1.07 years). According to written or oral reports from the rehabilitation institutions, none of these children had intellectual disorders or difficulties. In addition, 100 NH children (as controls) were recruited from a kindergarten in Northern China (Shandong Province). Their age ranged from 4 to 5.92 years (M = 5.07 years, SD = 0.52 years). All participants in this study were born and raised in Mandarin-speaking families. The study was conducted in accordance with the ethics protocol approved by Nanjing University of Science and Technology. Written informed consent was obtained from the children's parents.

Stimuli

A total of 36 monosyllabic Chinese words carrying four lexical tones (e.g., *shu1* "book"; 4 tones × 9 words) were used as stimuli. Monosyllabic words were used in this study to avoid any tonal coarticulation phenomena or tone sandhi process that might result in noncanonical tonal realizations. All words were picturable and fell within the top 50% of the most frequent monosyllabic words in the language input to NH children below 3 years, according to the Chang corpus (Chang, 1998) and the Tong corpus (Deng & Yip, 2018) from the CHILDES database (MacWhinney, 2000). Before recording, these materials were verified by two kindergarteners and tested (in a pilot experiment) with five NH children and five children with CIs to ensure their familiarity with young children. The target words were produced as naturally and clearly as possible by a 20-year-old female speaker in a sound-attenuated room. The audio and video of target words were recorded synchronously using a SONY HXR-NX 100 video camera with an external condenser microphone. The camera was positioned to film the upper body and the head of the speaker.

All video recordings for AV modality were edited into isolated clips, with a 720 × 480 pixel resolution, MPEG-4 coded format, and 30 frames per second in AVI format. These video clips were reedited as AO files without the video track for AO modality. In addition, following previous studies (Dohen & Lœvenbruck, 2009; Srinivasan & Massaro, 2003), masking the auditory signal is better for improvement when the visual modality is additionally available. Therefore, an eight-speaker babble noise was superimposed on the audio track in AO and AV conditions with signal-to-noise ratio (SNR) at 0 dB.

Procedure

Following D. Han et al. (2009), a two-alternative forced-choice (2AFC) picture-pointing paradigm was employed. The participant was instructed to discriminate between a pair of monosyllabic words that differ in tones (e.g., *shu1* "book" vs. *shu3* "mouse"). The four Mandarin tones were paired into six tone contrasts (i.e., T1–T2, T1–T3, T1–T4, T2–T3, T2–T4, T3–T4), with each contrast tested 6 times (2 words × 3 trials). A comprehensive list of the 36 words (2 words × 6 tone contrasts × 3 trials) is provided on the Open Science Framework (https://osf.io/sgcwb/).

The perceptual experiment was divided into two blocks: AO and AV. Each block consisted of 72 trials, among which 36 trials (6 trials × 6 tone contrasts) were presented in the quiet condition and the other 36 trials (6 trials × 6 tone contrasts) were presented in noise. To eliminate practice effects, the presentation of AO always preceded AV, and within each block, the trials were randomized. The order of the quiet and noise conditions was also counterbalanced.

All children were tested individually in quiet rooms. Before the formal identification test, each child underwent a familiarization phase to ensure they knew what the pictures represented. During the familiarization phase, the child was presented with all 36 pictures one by one using a 10.2-in. iPad and asked to name these pictures. In rare cases where the child failed to produce the target name, the experimenter would correct it and ask them to name it again. All children ultimately completed the familiarization successfully, with the entire phase typically lasting approximately 5–10 min.

Subsequently, during the formal identification test, the child was instructed to sit in front of a laptop equipped with two external loudspeakers. In each trial, either the audio stimulus (AO condition, where only the speech sound was available) or the audiovisual stimulus (AV condition, where both the speech sound and the video of the speaker during speech production were available) was first presented, followed by two pictures. Participants were instructed to choose the correct picture corresponding to the presented stimulus, using their fingers or a pen based on their preference to point to the pictures. Following their selection, the experimenter used an external numeric keypad to promptly input the response into the E-Prime software (Psychology Software Tools, Inc., 2016).

[1]Information regarding the brands of CI devices being used is available for 83 children: MED-EL (35 children), Cochlear (25 children), Advanced Bionics (17 children), and Nurotron (six children).

Data Coding and Statistical Analysis

Two types of analysis were conducted: the mixed-effect binomial logistic regression models and the correlation analysis. First, for each trial, children's responses were coded as 1 (*correct response*) or 0 (*incorrect response*). The mixed-effect binomial logistic regression models were employed to compare children's responses across different conditions. These models were implemented using the *glmmer* function in "lme4" package in R (R Core Team, 2023). The maximal random effect structure was used in model construction, which included random intercepts of Participant and Tone pairs, and random slopes of Modality, Environment, and Tone by Participant and Tone pairs (Barr et al., 2013). Then, all fixed factors were progressively added as random slopes using a forward-stepping approach until convergence failure occurred in model fitting. The R function *anova* was utilized to eliminate nonsignificant random slopes through model comparison based on Akaike information criterion. The final selected model was formulated as Response ~ Modality × Environment × Tone + (1 | Subject) + (1 | Tone pairs).

Statistical significance of each fixed factor was obtained using the *mixed* function in "afex" package (Singmann et al., 2021), providing omnibus effects of all fixed factors and interactions using F tests. When a significant effect (main and/or interaction effect) was observed, the *emmeans* function of the "emmeans" package (Lenth, 2022) was further executed to conduct Tukey's honestly significant difference (HSD) pairwise comparisons. In addition, a Pearson correlation analysis was performed using the *cor.test* function of the R package "stats" (R Core Team, 2023) to investigate the potential relationships between the implantation age, CI experiences and the observed effect of visual benefit (accuracy difference between AV and AO conditions: AV–AO) in tonal identification. All data and R scripts are available on the Open Science Framework (https://osf.io/sgcwb/).

Results

Tonal Identification Accuracy

Figure 1 illustrates the tonal identification accuracy (in percentage) of the NH group across modalities and environments. The results from mixed-effect logistic regression model were presented in Table 1, which showed that for NH children, neither the main effect of "Modality" ($p < .1$) nor its interaction with other effects was significant (Modality × Tone: $p = .165$; Modality × Environment: $p = .575$; Modality × Tone × Environment: $p = .568$). Therefore, there was no evidence supporting that NH children had shown any perception difference between AO and AV modalities in any environment.

Figure 2 illustrates the tonal identification accuracy of the CI group. The results of the mixed-effect logistic regression model reported a significant main effect of "Modality," $\chi^2(1) = 25.95$, $p < .001$, and a significant interaction between "Modality" and "Environment," $\chi^2(1) = 9.64$, $p = .002$. The Tukey's HSD post hoc test on this interaction suggested that children with CIs did not show any difference between AO and AV modalities in the quiet environment ($\beta = -.03$, $SE = .021$, $z = -1.45$, $p = .147$), while their accuracy significantly improved from AO to AV conditions in the noisy environment ($\beta = .155$, $SE = .026$, $z = 5.867$, $p < .0001$).

Correlation Analysis

Two Pearson correlation analyses (for quiet and noisy environments separately) were conducted to explore the relationship between the implantation age and the observed effect of visual benefit on tonal identification for the CI group. The results indicated a significant correlation between the two factors in the noisy environment

Figure 1. The identification accuracy of four tones (T1–T4) in audio-only and audiovisual modalities by NH children in quiet and noisy environments. AO = audio only; AV = audiovisual; NH = normal-hearing.

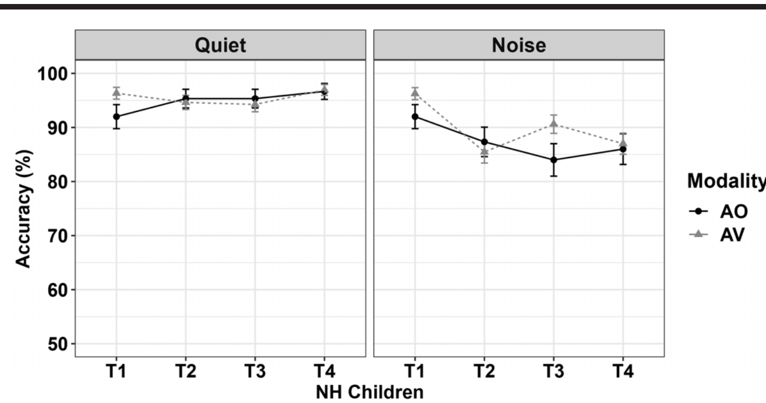


Table 1. The omnibus effect of each main and interaction fixed effect from the mixed-effect logistic regression model for children with cochlear implants (CI) and normal-hearing (NH) children.

Effect	*df*	NH		CI	
		χ^2	*p*	χ^2	*p*
Tone	3	3.65	.301	3.12	.374
Environment	1	42.96	< .001	65.03	< .001
Modality	1	2.7	.999	25.95	< .001
Tone × Environment	3	5.79	.122	5.75	.125
Tone × Modality	3	5.09	.165	3.06	.382
Environment × Modality	1	0.31	.575	9.64	.002
Tone × Environment × Modality	3	2.02	.568	5.12	.163

Note. *df* = degrees of freedom.

(r = −.3, t = −2.772, p = .007), but not in the quiet environment (r = .076, t = 0.675, p = .502). This implied that children implanted earlier demonstrated a larger degree of visual benefits in tone identification in noisy environments, as illustrated in Figure 3. Addtionally, we also examined the correlation between the length of CI experience and the visual benefit, while the results showed that they were not correlated in neither environment, that is, quiet environment, r = .02, t = 0.142, p = .887, and noisy environment, r = .095, t = 0.843, p = .402.

Discussion

The present study investigated the effect of visual articulatory cues on the perception of Mandarin tones by children with CIs. The results showed that, while NH children did not show any difference in tone identification between AO and AV modalities, children with CIs significantly improved their tonal identification accuracy from AO to AV modalities, although such improvement was observed in the noisy environment only. This suggested that visual cues enhanced the tone identification abilities of children with CIs, although only in the auditory-degraded noisy condition. Furthermore, our results also showed that early implantation helped children with CIs to better use visual cues to improve tone identification in noise.

It is worth noting that our results expand on S.-Y. Liu et al. (2014)'s findings, which showed that adult CI users did not gain visual benefits in Mandarin tonal perception on two factors. First, regarding the participants' age, our research involved children aged 3–7 years, while S.-Y. Liu et al. (2014) focused on adult CI users aged 18–56 years. It has been demonstrated that the ability of lipreading decreases with age (Başkent et al., 2014; Cienkowski & Carney, 2002; Shoop & Binnie, 1979). Consequently, adult listeners might experience reduced benefits from visual information as compared to those of children (Sommers et al., 2005).

Second, regarding the environment, the absence of visual benefit was noted in S.-Y. Liu et al. (2014) and the

Figure 2. The identification accuracy of four tones (T1–T4) in audio-only and audiovisual modalities by CI children in quiet and noisy environments. AO = audio only; AV = audiovisual; CI = cochlear implant.

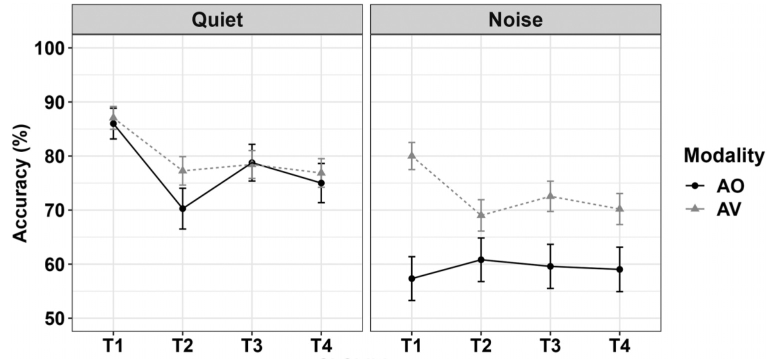
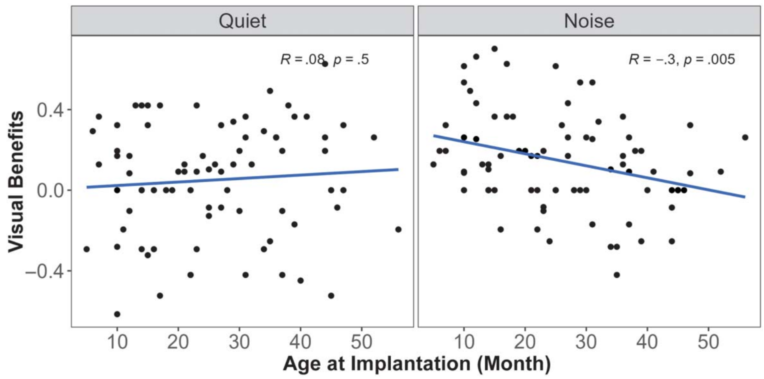


Figure 3. The scatter plots for the correlations between age at implanations and visual benefits (cacluated as the accuracy difference between audiovisual and audio-only conditions) on quiet and noise environment.



first block in the current study under quiet conditions. However, a significant visual benefit was observed in the second block in the current study, which involved SNR 0 dB babble noise added to the auditory modality. It is argued that the perception of Mandarin tones proves relatively straightforward for native speakers with normal hearing (C.-Y. Lee et al., 2013), even extending to children with 3 years of CI experience (Zheng et al., 2009). This potentially results in a ceiling effect and the impossibility of measuring a potentially significant improvement when the visual modality is additionally available. According to the fuzzy logical model of perception, both auditory and visual modalities influence the perception of speech, with the influence of one modality being greater when the other is ambiguous (Massaro & Cohen, 1983; Massaro & Oden, 1980). Therefore, noise in the auditory modality increases the manifestation of visual benefits for children with CIs.

Our findings on the CI children group, for the first time, demonstrated the benefits of visual articulatory cues on tone identification for this population, although the visual–acoustic associations of tones are much less transparent as compared to vowels and consonants. This provided support for the ability of children with CIs to integrate visual cues into speech perception. Furthermore, our correlation analysis also revealed an early implantation advantage for auditory and visual integration during tonal perception for children with CIs. Such benefits might be attributed to the increased cross-modal plasticity that occurs after early auditory deprivation, significantly improving these children's abilities in visual speech comprehension (Glick & Sharma, 2017). Early exposure to speech sounds enables children with CIs to better map the visual patterns of articulator movements to the corresponding phonological categories, thus leading to a more robust multisensory depiction of phonological categories compared to those who receive implants later.

The current finding on NH children was also consistent with previous results showing that NH adult listeners did not show any visual benefit on tone identification in Mandarin (M. Li et al., 2022), Cantonese (Wang et al., 2020), or Thai (Burnham et al., 2015). Two possible reasons might account for these results. First, the 2AFC task for tone identification might have been too easy for NH children, even under the 0 dB SNR babble noise condition. Therefore, the potential ceiling effect could have impeded the accurate measurement of the AV benefit, and it is possible that such benefits might occur under noisier conditions. This consideration is supported by Li et al. (2022), who observed an AV benefit in Mandarin participants aged 18–25 years when utilizing a −12 dB SNR. Second, these results might be related to the fact that visual cues only play a complementary role in tone identification (Hong et al., 2023), and listeners with NH predominantly use auditory information to perceive tones unless it is insufficient (Wang et al., 2008).

The present findings thus have important implications for clinical practice. CIs are widely used to treat severe-to-profound hearing loss in children. Although CIs provide significant benefits, especially when implanted at an early age, CI users still struggle with speech perception in challenging listening environments (e.g., noisy classrooms). Incorporating visual cues into CI treatment could help overcome these challenges and improve CI users' speech perception abilities. Clinicians working with CI users should consider incorporating visual cues into therapy to maximize outcomes.

There are several limitations in this study. First, the use of different CI devices with various signal-processing strategies might affect individual auditory performance (de Melo et al., 2013). Therefore, it is critical for future research to consider the impact of different strategies on the role of visual benefits during speech perception.

Second, our findings related to the benefits of early implantation are not directly indicative of milestones in the early stages, because the children in our study were not grouped by age. To bolster our findings, future research could group participants according to their age. Last, given the visual benefits observed, it would also be beneficial to investigate the relative contributions of different aspects of visual inputs. This suggests the value of utilizing eye-tracking studies in future research endeavors.

Conclusions

The present study provided evidence that visual cues facilitated Mandarin tone identification by children with CIs in noisy environments and that children implanted earlier benefited more from visual cues. These findings suggest the significance of incorporating visual cues into auditory training and underscore the importance of early implantation for children with severe hearing loss

Data Availability Statement

The data and code for this study are available at the Open Science Framework (https://osf.io/sgcwb/).

Acknowledgments

This research was supported by the National Social Sciences Foundation of China (20CYY012). We would like to thank Man Mei and Ji Li for data collection.

References

Attina, V., Gibert, G., Vatikiotis-Bateson, E., & Burnham, D. (2010). Production of Mandarin lexical tones: Auditory and visual components. *Proceedings of International Conference on Audio-Visual Speech Processing*.

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

Başkent, D., van Engelshoven, S., & Galvin, J. J., III. (2014). Susceptibility to interference by music and speech maskers in middle-aged adults. *The Journal of the Acoustical Society of America, 135*(3), EL147–EL153. https://doi.org/10.1121/1.4865261

Burnham, D., Ciocca, V., & Stokes, S. (2001). *Auditory-visual perception of lexical tone*. EUROSPEECH.

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. X., Schwarz, I., & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics, 36*(6), 1459–1491. https://doi.org/10.1017/S0142716414000496

Burnham, D., Lau, S., Tam, H., & Schoknecht, C. (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers and by non-tonal language speakers. *Proceedings of the AVSP 2001 International Conference on Auditory-Visual Speech Processing*.

Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese–speaking children's play narratives. *Narrative Inquiry, 8*(1), 77–111. https://doi.org/10.1075/ni.8.1.05cha

Chen, Y., & Wong, L. L. N. (2016). Speech perception in Mandarin-speaking children with cochlear implants: A systematic review. *International Journal of Audiology, 56*(Suppl. 2), S7–S16. https://doi.org/10.1080/14992027.2017.1300694

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing, 23*(5), 439–449. https://doi.org/10.1097/00003446-200210000-00006

De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7*(10), 460–467. https://doi.org/10.1016/j.tics.2003.08.014

de Melo, T. M., Bevilacqua, M. C., Costa, O. A., & Moret, A. L. M. (2013). Influence of signal processing strategy in auditory abilities. *Brazilian Journal of Otorhinolaryngology, 79*(5), 629–635. https://doi.org/10.5935/1808-8694.20130113

Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: The Tong corpus. *Journal of Chinese Linguistics, 46*(1), 69–92. https://doi.org/10.21415/T5PC7Q

Dohen, M., & Lœvenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech, 52*(2–3), 177–206. https://doi.org/10.1177/0023830909103166

Ernst, M. O. (2008). Multisensory integration: A late bloomer. *Current Biology, 18*(12), R519–R521. https://doi.org/10.1016/j.cub.2008.05.002

Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience, 4*(12), 1171–1173. https://doi.org/10.1038/nn763

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110*(2), 1150–1163. https://doi.org/10.1121/1.1381538

Gandour, J., Carney, A., Nimitbunnasarn, C., & Amatyakul, P. (1984). Tonal confusions in Thai patients with sensorineural hearing loss. *Journal of Speech and Hearing Research, 27*(1), 89–97. https://doi.org/10.1044/jshr.2701.89

Garg, S., Hamarneh, G., Jongman, A., Sereno, J. A., & Wang, Y. (2019). Computer-vision analysis reveals facial movements made during Mandarin tone production align with pitch trajectories. *Speech Communication, 113,* 47–62. https://doi.org/10.1016/j.specom.2019.08.003

Gilley, P. M., Sharma, A., Dorman, M., & Martin, K. (2006). Abnormalities in central auditory maturation in children with language-based learning problems. *Clinical Neurophysiology, 117*(9), 1949–1956. https://doi.org/10.1016/j.clinph.2006.05.015

Glick, H., & Sharma, A. (2017). Cross-modal plasticity in developmental and age-related hearing loss: Clinical implications. *Hearing Research, 343,* 191–201. https://doi.org/10.1016/j.heares.2016.08.012

Han, D., Liu, B., Zhou, N., Chen, X., Kong, Y., Liu, H., Zheng, Y., & Xu, L. (2009). Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users. *Ear and Hearing, 30*(2), 169–177. https://doi.org/10.1097/AUD.0b013e31819342cf

Han, Y., Goudbeek, M., Mos, M., & Swerts, M. (2020). Relative contribution of auditory and visual information to Mandarin Chinese tone identification by native and tone-naïve listeners. *Language and Speech, 63*(4), 856–876. https://doi.org/10.1177/0023830919889995

Hong, S., Wang, R., & Zeng, B. (2023). Incongruent visual cues affect the perception of Mandarin vowel but not tone. *Frontiers in Psychology, 13,* Article 971979. https://doi.org/10.3389/fpsyg.2022.971979

Kasisopa, B., Klangpornkun, N., & Burnham, D. (2015). Auditory-visual tone perception in hearing impaired Thai listeners. *Proceedings of INTERSPEECH*. https://doi.org/10.21437/Interspeech.2015-613

Lalonde, K., & McCreery, R. W. (2020). Audiovisual enhancement of speech perception in noise by school-age children who are hard of hearing. *Ear and Hearing, 41*(4), 705–719. https://doi.org/10.1097/AUD.0000000000000830

Lee, C.-Y., Tao, L., & Bond, Z. S. (2013). Effects of speaker variability and noise on Mandarin tone identification by native and non-native listeners. *Speech, Language and Hearing, 16*(1), 46–54. https://doi.org/10.1179/2050571X12Z.0000000003

Lee, D. S., Lee, J. S., Oh, S. H., Kim, S.-K., Kim, J.-W., Chung, J.-K., Lee, M. C., & Kim, C. S. (2001). Cross-modal plasticity and cochlear implants. *Nature, 409*(6817), 149–150. https://doi.org/10.1038/35051653

Lee, K. Y. S., van Hasselt, C. A., & Tong, M. C. (2008). Tone perception in Cantonese-speaking children with hearing aids. *Annals of Otology, Rhinology & Laryngology, 117*(4), 313–316. https://doi.org/10.1177/000348940811700413

Lee, K. Y. S., van Hasselt, C. A., & Tong, M. C. F. (2010). Lexical tone perception ability of profoundly hearing-impaired children: Performance of cochlear implant and hearing aid users. *Otology & Neurotology, 31*(7), 1079–1087. https://doi.org/10.1097/MAO.0b013e3181eb3242

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.5.5-1). Retrieved September 12, 2023, https://CRAN.R-project.org/package=emmeans

Li, A., Wang, N., Li, J., Zhang, J., & Liu, Z. (2014). Mandarin lexical tones identification among children with cochlear implants or hearing aids. *International Journal of Pediatric Otorhinolaryngology, 78*(11), 1945–1952. https://doi.org/10.1016/j.ijporl.2014.08.033

Li, M., Chen, X., Zhu, J., & Chen, F. (2022). Audiovisual Mandarin lexical tone perception in quiet and noisy contexts: The influence of visual cues and speech rate. *Journal of Speech, Language, and Hearing Research, 65*(11), 4385–4403. https://doi.org/10.1044/2022_JSLHR-22-00024

Liu, H., Liu, S., Kirk, K. I., Zhang, J., Ge, W., Zheng, J., Liu, Z., & Ni, X. (2015). Longitudinal performance of spoken word perception in Mandarin pediatric cochlear implant users. *International Journal of Pediatric Otorhinolaryngology, 79*(10), 1677–1682. https://doi.org/10.1016/j.ijporl.2015.07.023

Liu, S.-Y., Yu, G., Lee, L.-A., Liu, T.-C., Tsou, Y.-T., Lai, T.-J., & Wu, C.-M. (2014). Audiovisual speech perception at various presentation levels in Mandarin-speaking adults with cochlear implants. *PLOS ONE, 9*(9), Article e107252. https://doi.org/10.1371/journal.pone.0107252

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Erlbaum.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*(5), 753–771. https://doi.org/10.1037/0096-1523.9.5.753

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *The Journal of the Acoustical Society of America, 67*(3), 996–1013. https://doi.org/10.1121/1.383941

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

McDermott, H. J., & Looi, V. (2004). Perception of complex signals, including musical sounds, with cochlear implants. *International Congress Series, 1273,* 201–204. https://doi.org/10.1016/j.ics.2004.08.034

Mixdorff, H., Hu, Y., & Burnham, D. (2005). Visual cues in Mandarin tone perception. *Proceedings of INTERSPEECH.*

Moberly, A. C., Lowenstein, J. H., & Nittrouer, S. (2016). Early bimodal stimulation benefits language acquisition for children with cochlear implants. *Otology & Neurotology, 37*(1), 24–30. https://doi.org/10.1097/MAO.0000000000000871

Psychology Software Tools, Inc. (2016). *E-Prime 3.0* [Computer software]. https://support.pstnet.com/

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. Retrieved September 12, 2023, https://www.R-project.org

Richie, C., & Kewley-Port, D. (2008). The effects of auditory-visual vowel identification training on speech recognition under difficult listening conditions. *Journal of Speech, Language, and Hearing Research, 51*(6), 1607–1619. https://doi.org/10.1044/1092-4388(2008/07-0069)

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science, 17*(6), 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Sharma, A., & Dorman, M. F. (2006). Central auditory development in children with cochlear implants: Clinical implications. In A. R. Møller (Ed.), *Advances in oto-rhino-laryngology: Cochlear and brainstem implants* (Vol. 64, pp. 66–88). Karger. https://doi.org/10.1159/000094646

Sharma, A., Spahr, A., Dorman, M., & Todd, N. W. (2002). Early cochlear implantation in children allows normal development of central auditory pathways. *Annals of Otology, Rhinology & Laryngology, 111*(Suppl. 5), 38–41. https://doi.org/10.1177/00034894021110S508

Shoop, C., & Binnie, C. A. (1979). The effects of age upon the visual perception of speech. *Scandinavian Audiology, 8*(1), 3–8. https://doi.org/10.3109/01050397909076295

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *afex: Analysis of factorial experiments* (Version 0.28-1) [R package]. Retrieved September 12, 2023, https://CRAN.R-project.org/package=afex

Smith, D., & Burnham, D. (2012). Faciliation of Mandarin tone perception by visual speech in clear and degraded audio: Implications for cochlear implants. *The Journal of the Acoustical Society of America, 131*(2), 1480–1489. https://doi.org/10.1121/1.3672703

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing, 26*(3), 263–275. https://doi.org/10.1097/00003446-200506000-00003

Srinivasan, R. J., & Massaro, D. W. (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. *Language and Speech, 46*(1), 1–22. https://doi.org/10.1177/00238309030460010201

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

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*(5), 1405–1411. https://doi.org/10.1097/AUD.0000000000001036

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America, 124*(3), 1716–1726. https://doi.org/10.1121/1.2956483

Wang, Y., Sereno, J. A., & Jongman, A. (2020). Multi-modal perception of tone. In H.-M. Liu, F.-M. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition: Multidisciplinary approaches in Chinese languages* (pp. 159–173). Springer. https://doi.org/10.1007/978-981-15-7606-5_9

Wei, Y., Jia, L., Gao, F., & Wang, J. (2022). Visual-auditory integration and high-variability speech can facilitate Mandarin Chinese tone identification. *Journal of Speech, Language, and Hearing Research, 65*(11), 4096–4111. https://doi.org/10.1044/2022_JSLHR-21-00691

Weng, Y., Rong, Y., & Peng, G. (2023). The development of audiovisual speech perception in Mandarin-speaking children: Evidence from the McGurk paradigm. *Child Development*. Advance online publication. https://doi.org/10.1111/cdev.14022

Zheng, Y., Meng, Z., Wang, K., Tao, Y., Xu, K., & Soli, S. D. (2009). Development of the Mandarin Early Speech Perception Test: Children with normal hearing and the effects of dialect exposure. *Ear and Hearing, 30*(5), 600–612. https://doi.org/10.1097/AUD.0b013e3181b4aba8