Processing of fundamental frequency changes, emotional prosody and lexical tones by pediatric CI recipients

MONITA CHATTERJEE[1,*], MICKAEL L. D. DEROCHE[2], SHU-CHEN PENG[3], HUI-PING LU[4], NELSON LU[3], YUNG-SONG LIN[4,5], AND CHARLES J. LIMB[6]

[1] *Auditory Prostheses & Perception Laboratory, Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, USA*

[2] *Center for Research on Brain, Language and Music, McGill University, Montreal, QC, Canada*

[3] *Center for Devices and Radiological Health, United States Food and Drug Administration, Silver Spring, MD, USA*

[4] *Chi-Mei Medical Center, Department of Otolaryngology, Tainan, Taiwan*

[5] *Taipei Medical University, Taipei, Taiwan*

[6] *Department of Otolaryngology-Head & Neck Surgery, University of California San Francisco School of Medicine, San Francisco, CA, USA*

As cochlear implants (CIs) do not provide adequate representation of the harmonic structure of complex sounds, the perception of the voice fundamental frequency (F0) is severely limited in CI users. As F0 plays an important role in speech prosody and in lexical tones, this deficit has a negative impact on communication. Here we focus on the pediatric CI population, most of whom were prelingually deaf and were implanted before three years of age, within the most adaptive period of the brain's development. Our results suggest that, relative to their normally-hearing peers, school-age children with CIs have significant deficits in their sensitivity to both static and dynamic F0-changes. In addition, children with CIs also have deficits in their identification of emotional prosody and in lexical-tone recognition.

INTRODUCTION

Present-day cochlear implants (CIs) transmit acoustic-phonetic cues for speech recognition with sufficient fidelity for good-to-excellent speech recognition performance in quiet by the average patient, particularly when context cues are available and top-down reconstruction of the intended message can compensate for the degraded input. However, voice pitch, which conveys prosodic information and provides important cues for lexical tone recognition in tonal languages, is not transmitted appropriately by CIs. The voice fundamental frequency and its harmonics are lost in CI processing. Only the temporal envelopes extracted from the

*Corresponding author: monita.chatterjee@boystown.org

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

broader channel filters, which are modulated periodically when multiple harmonics of the fundamental frequency (F0) fall within the bandpass region of the filter, carry some voice pitch information. Unfortunately for CI patients, this envelope periodicity does not result in a salient enough percept to support music perception. For the recognition of prosody or lexical tones, however, the high level of pitch perception demanded by music may not be necessary. However, the extent to which the degraded pitch transmitted by CIs supports the recognition of these cues is not well understood.

For children who are implanted within the first months or years of life and are developing with the device in place, the implications of prosodic cue perception may be more profound than for post-lingually deaf adults, as language acquisition is thought to be mediated by prosodic information in infants and toddlers. These two populations provide an interesting comparison, bringing two different brains to the table. The post-lingually deaf adults developed auditory, cognitive and linguistic skills with good sensory and phonological representations. In contrast, pediatric CI recipients must develop all three with the degraded input of the device; Further, these very same cognition and language skills must then be harnessed in top-down repair of the degraded input. In studying this population, important questions regarding neural plasticity related factors (age at implantation, device experience, developmental factors) must be considered alongside other factors (sensitivity to the sensory input). For tone language speakers, we are additionally interested in the possible role of the tonal environment in the development of harmonic pitch sensitivity in pediatric CI recipients and in their normally hearing peers.

Here, we review recent work from our collaborative group on the recognition of emotional prosody and lexical tones by children with CIs in the context of the broader literature, and discuss implications for the development of CIs and rehabilitation of CI patients.

SENSITIVITY TO HARMONIC PITCH BY PEDIATRIC CI RECIPIENTS

Deroche *et al.* (2014) investigated harmonic pitch perception in pediatric CI users in the US and in Taiwan. They measured F0 difference limens in a 3-interval, 3-alternative forced-choice procedure, using the method of constant stimuli. The level of the stimuli was roved to avoid loudness cues. Psychometric functions were obtained, and thresholds for F0 discrimination were derived from these data. As expected, they reported large deficits in the children with CIs, both for a 100-Hz F0 reference and a 200-Hz F0 reference. No differences were observed between the children in the US and Taiwan, for both hearing status, suggesting that the tonal language environment did not influence static F0 discrimination in the normal-hearing (NH) or the CI children. There were small but significant age effects explaining about 10% of the variance across subjects, but no effects of age at implantation. In addition, they also reported that, in a task in which the participants had to discriminate between amplitude-modulation rate differences in sinusoidally amplitude-modulated broadband noise, the NH children's performance was the same as that of the CI children's (both being slightly worse than CI children's performance in the F0 discrimination task). Thus, when the NH children were compelled to use the same envelope periodicity cues available to CI children, their performance was not significantly different from that of the CI children's. This suggests further that the deficits in the CI children's performance may not be due to other differences between the populations (cognition, etc.), at least in these tasks. The CI children's performance in the F0 discrimination task was significantly correlated with that in the amplitude modulation (AM) rate discrimination task, supporting the idea that similar cues were used in the two tasks (despite small changes in the exact shape of the temporal envelopes and their coherence across channels). On the other hand, the NH children's thresholds in the two tasks were not correlated, supporting the idea that the NH children used very different cues in the two tasks (i.e., fine structure periodicity in the F0 discrimination task and temporal envelope periodicity in the AM rate discrimination task).

In a second study, Deroche *et al.* (2016) investigated the sensitivity to dynamic (swept F0) pitch in children with CIs in the US. They used two tasks, a direction labelling task and a direction discrimination task. In the first task, listeners heard a swept F0 harmonic complex and indicated whether it was rising or falling in a single-interval, 2-alternative forced-choice procedure. In the second task, they heard three swept F0 complexes: The first served as the reference, one of the remaining two intervals was identical to the reference in sweep range and direction while the other interval was identical in range but was swept in the opposite direction. The initial F0 was roved to avoid spectral-edge-related pitch cues, and level was roved to avoid any loudness cues. Results were similar to those obtained in the 2014 study, in that large deficits were observed in the CI children in both tasks. Significant age effects were also found (and corroborated by differences between children and adults in both populations), but no effects of age at implantation were observed.

We are presently conducting parallel studies in children in Taiwan in the swept-F0 tasks described above. In preliminary data, we observe a significant advantage in NH children in Taiwan over NH children in the US in both the direction labelling and the direction discrimination task (Fig. 1). This advantage is dominated by differences between the groups in the early developmental years, when the NH children in Taiwan appear to be already adult-like in both tasks, while the NH children in the US show a significant developmental effect, converging on the Taiwanese children's performance in their later teenage years. The data with child CI recipients shows no significant differences between the groups and no effects of age, although the difference between the US and Taiwanese CI users in the labelling task is marginally significant at p=0.051. Further data collection and analyses are ongoing to verify this trend.

To what extent are the F0 sweep rates in our studies relevant to those in lexical tones? He *et al.* (2016) attempted to relate the sensitivity to dynamic F0 sweeps to that of realistic F0 change rates in Mandarin tones. They adaptively modified the F0 range over which complex tones or recordings of lexical tones could be identified (as falling, rising, flat, dip, or peak). They found that NH adults could recognize 400-ms lexical tones with an F0 range only 40-cents wide, resulting in F0 change

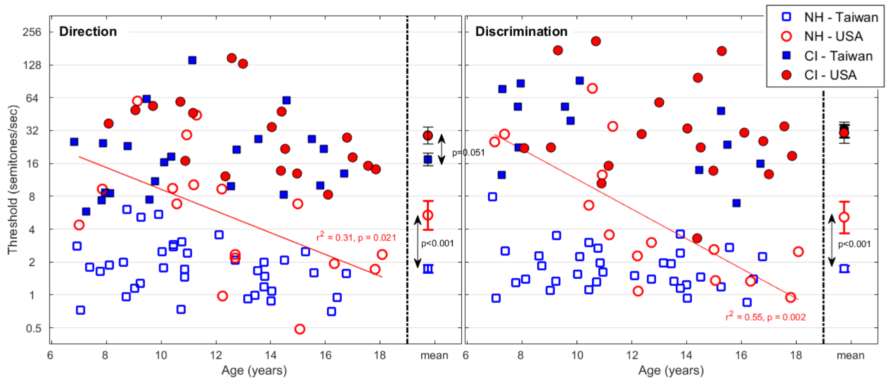

Fig. 1: Thresholds (in semitones/sec) obtained in an F0-sweep direction-labeling task (left-hand panel) and in an F0-sweep direction-discrimination task (right-hand panel) plotted against participants' age (in years). Blue and red symbols indicate results obtained in native speakers of American English and Mandarin, respectively. Open and filled symbols indicate results obtained in children with normal hearing and CIs, respectively. The right-hand side of each plot shows group means and standard deviations. (color online)

rates of only 1 semitone/s (or 2 semitones/s for dips). In contrast, CI adults needed F0 change rates of 2 octaves/s. These results and the size of the deficits exhibited by CI adults are very consistent with the dynamic F0 sensitivities reported by Deroche *et al.* (2016).

To summarize, our results indicate significant deficits in children with CIs (relative to their NH peers) in F0 discrimination of both static changes in F0 and swept-F0, dynamic changes. The data show evidence for an effect of linguistic environment, with the tone-language environment benefiting children with NH in both tasks, particularly in the early years of development. This finding is supported by other related studies in adults using both perceptual and physiological measures showing differences in the mechanisms of F0 coding and FM-sweep-coding in native speakers of Mandarin (Krishnan *et al.*, 2015; Krishnan *et al.*, 2011; Luo *et al.*, 2007a).

LEXICAL TONE RECOGNITION BY PEDIATRIC CI USERS

Lexical tones are relatively rapid changes in the F0 contour of syllables that signal the meaning of a word. In Mandarin, tones fall into four categories: Tone 1 consists of a high, flat F0, Tone 2 is predominantly rising in F0, Tone 3 is more complex, with a low initial F0, dipping to a minimum before rising again, and Tone 4 is a short, falling F0 contour. Similar to other studies, Peng *et al.* (2017) showed significant deficits in Taiwanese pediatric CI recipients' tone recognition (compared to NH peers) using a single-interval, 4-alternative forced-choice task in which participants pointed to the appropriate meaning depicted in a picture on the screen.

Performance was significantly predicted by the children's age at implantation but not by duration of device experience or by age at testing. This suggests that the ability to identify tones develops within a short period post-implant-activation, but does not change thereafter.

In a second task, Peng *et al.* (2017) asked whether CI children could utilize a secondary acoustic cue (duration) to their advantage in a tone recognition task. To do this, they used stimuli consisting of manipulated versions of an original utterance of the disyllabic word *"yanjing"*, in which the first syllable *yan* was kept constant but the second syllable *jing* was manipulated to have 8 steps of F0 contour, varying from Tone 1 (flat) to Tone 4 (falling). For each of these variations, six levels of duration of the second syllable were applied orthogonally. When the 2nd syllable is spoken in Tone 1, the word *yanjing* means "eyes", and when spoken in Tone 4, it means "eye glasses". The listeners heard each resynthesized utterance (in randomized order) and indicated which of the two meanings they thought it was associated with, in a single-interval, 2-alternative forced-choice task. The data were analysed using logistic regression to investigate the extent to which the children used F0 or duration cues to perform the task. It is to be noted that duration is not a strong cue for lexical tone recognition in Mandarin: It is a weak secondary cue at best.

Results (Fig. 2) showed that NH children used F0 cues extensively, but ignored the duration cue. On the other hand, the CI children made significant use of the F0 cue, but with less certainty than the NH children, while they used the duration cue significantly as well, much more so than the NH children. This suggests that even though CI children have significant deficits in F0 processing, they do rely on the cue in lexical tone recognition. Secondly, these results indicate that even though duration is not a strong acoustic cue for tones, the CI children implicitly develop the knowledge to use it appropriately to support their performance when it is explicitly provided. Age at implantation predicted their use of the duration cue, again suggesting an early development of tone-recognition skills.

Peng *et al.* (2017) also found that the use of the F0 cue in the yan jing task was a strong predictor of performance in lexical tone recognition with naturally uttered words. This is interesting, given the weakness of F0 sensitivity in this population, and underscores the dominance of F0 as a cue for lexical tones. The use of the duration cue, however, did not predict performance with the lexical tone recognition task, possibly because the cue is not reliable in natural Mandarin speech. However, the possibility remains that appropriate training with the secondary cue might help the children optimize their use of duration in natural speech recognition.

To summarize, children with CIs showed significant deficits in lexical tone recognition with naturally uttered words. They also showed a reliance on both F0 contour and duration cues in a Tone 1-Tone 4 cue-weighting task, while their NH peers only relied on the F0 contour. Lexical tone recognition was significantly predicted by the listeners' age at implantation, but not by their device experience.

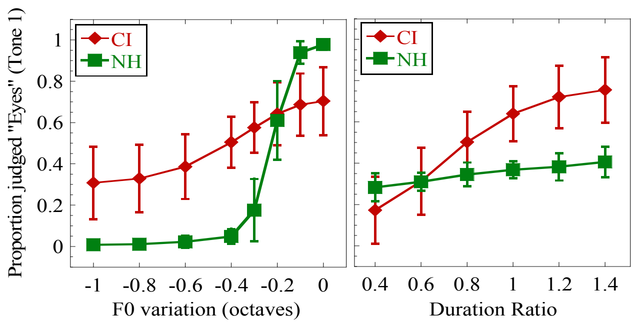

Fig. 2: The left hand panel shows the proportion of times the *yanjing* stimulus was associated with "Eyes" (Tone 1), plotted as a function of F0 variation of the second syllable by listeners with NH (squares) and with CIs (diamonds). The right hand panel shows the same, but plotted as a function of duration values of the stimuli. Error bars show +/- 1 s.d. from the mean. [Adapted from Peng et al. (2017)]

VOICE EMOTION RECOGNITION BY CHILDREN WITH CIS

Emotional prosody is critical for social communication. In children, appropriate emotional communication is important for the development of social interactions, both with their peers and their caregivers, and for social cognitive development in general. Luo *et al.* (2007b) showed that adult CI users had significant impairments in voice emotion perception. As many adult CI users are post-lingually deaf, it is reasonable to assume that they developed relatively normal understanding of emotions and emotional communication prior to their hearing loss and cochlear implantation. The literature on emotion understanding in children with CIs is relatively sparse, but suggests that while basic facial emotion recognition is developed in this population by the time they are school-aged, deficits in voice emotion recognition remain. We studied emotion recognition by school-age children with CIs using child-directed speech (i.e., exaggerated prosody), and compared performance with NH peers, adults with NH and adults with CIs (Chatterjee *et al.*, 2015). We found that, consistent with previous work in emotion recognition by CI recipients, the CI users showed significant deficits in voice emotion recognition. Interestingly, the post-lingually deaf adults and the pre-lingually deaf children with CIs showed very similar performance, suggesting a device-limitation in the task rather than factors related to cognition or language. However, this remains to be demonstrated. It is possible that similar performance can be achieved by different mechanisms.

Many of the participants in this study also participated in the studies on F0 processing by Deroche *et al.* (2014) and Deroche *et al.* (2016) described earlier. Based on correlational analyses, wee find that F0 sensitivity in all our tasks is a significant predictor of voice emotion recognition by CI listeners. This is noteworthy, given that CI patients show such large deficits in F0 processing, and also given that voice emotion is conveyed by numerous cues, including intensity, speaking rate, and timbre. This underscores the critical importance of F0 in the recognition of emotional prosody.

We also note that the stimuli in the Chatterjee *et al.* (2015) study contained exaggerated prosody, and that NH adults and children performed at ceiling in the task. This suggests that we underestimated the true deficit experienced by CI children in everyday communication, which consists of far more muted emotional cues.

Chatterjee *et al.* (2015) reported that the children with CIs showed a weak effect of device experience in their emotion recognition, but did not show a significant effect of developmental age. The children with NH showed a significant effect of age with full-spectrum stimuli, although their performance in this condition was near-ceiling, not significantly different from the adult NH listeners' performance, and significantly better than the performance of the children with CIs in the same condition. The children with NH also performed the task with 8-channel noise vocoded speech. While adults with NH showed performance similar to the mean CI performance in this condition, the children with NH showed significant deficits. Further analyses revealed a strong developmental effect (Fig. 3), with older children with NH showing significantly better performance than younger children, who clearly struggled to identify the emotions in the 8-channel vocoded condition. These findings underscore an important problem. The children with NH had experience

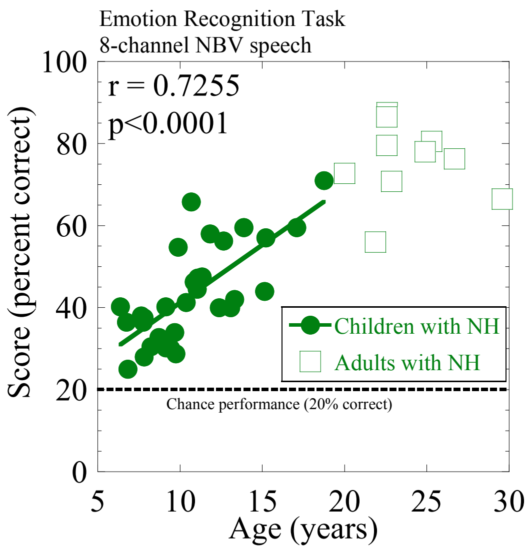

Fig. 3: Developmental effects in voice emotion recognition by children with NH when attending to 8-channel NBV speech. These children showed near-ceiling performance and smaller developmental effects in the same task with full-spectrum (unprocessed) speech. [Adapted from Chatterjee *et al.* (2015)]

with acoustic inputs, and different forms of degraded acoustic inputs (e.g., listening in reverberation or noise). They had presumably also developed language and cognition normally. Still, they struggled in this first encounter with CI-simulated speech. This speaks to the considerable difficulties faced by young children with CIs who do not have experience with acoustic inputs and have not developed the cognition or language skills that might help them to make sense of the new sensory signals. The advantage they have is the earlier implantation age and the greater plasticity of the brain within the first years of development. It seems critical, therefore, to ensure that rehabilitation efforts are maximized in the early months with the CI.

Unlike the lexical tone recognition data, the voice emotion recognition study did not show an effect of age at implantation on performance. Instead, duration of experience with the device (which was correlated with developmental age) was a significant predictor. This suggests that learning to read emotions continues as children develop, and that rehabilitation strategies may help by emphasizing prosodic cues even after the most sensitive period has passed.

SUMMARY AND CONCLUSIONS

To summarize, our results indicate that pediatric recipients of CIs must cope with large deficits in sensitivity to both static and dynamic changes in F0. In our studies, age at implantation is not predictive of performance in these psychophysical tasks, but the children with CIs show improvements in task performance as their duration of experience with the device (correlated with their chronological age at testing) increases. In preliminary work, we observe an advantage in dynamic-F0 sensitivity in native speakers of Mandarin over native speakers of American English. Sensitivity to voice emotion also shows deficits in children with CIs, and also shows age-related improvements. Although vocal emotions are represented by multiple acoustic cues, the dominance of F0 cues is demonstrated by the fact that F0 sensitivity is significantly correlated with performance in voice emotion recognition in CI users, even though this information is greatly degraded in this population. Native speakers of Mandarin developing with CIs showed significant deficits in lexical tone recognition relative to their NH peers, but also showed an ability to appropriately rely on a weak secondary acoustic cue, duration, in lexical tone contrast judgments. The age at implantation was a significant predictor of performance in lexical tone recognition, but increased duration of experience with the device did not correlate with performance in these tasks, suggesting a limit to adaptive advantages in sensitivity to tones early in life.

Overall, we conclude that the device-limitations placed on F0 coding in CIs are not sufficiently overcome by neural plasticity (age at implantation or duration of experience) as the child acquires experience with the device. Linguistic experience (tone language environment) may confer advantages to the NH population, but significant advantages have not been observed in our preliminary data with CI users as yet. The fact that the Mandarin-speaking children with CIs were able to utilize the secondary acoustic cue in tone identification suggests a role for training in tone recognition. We further conclude that F0-coding-limitations play a role in CI users' sensitivity to vocal emotion processing. These findings have implications for social communication, language acquisition, and rehabilitation efforts for the pediatric CI population and for device development strategies in the future.

ACKNOWLEGEMENTS

This work was supported by the following grants from the National Institutes of Health: R01 DC004786-08S1, R21 DC011905, R01 DC014233 (MC), and by support from the Chi-Mei Research Foundation and Taiwan National Science Council Grant 100-2314-B-038-037 (YSL).

REFERENCES

Chatterjee, M., Zion, D.J., Deroche, M.L., Burianek, B.A., Limb, C.J., Goren, A.P., Kulkarni, A.M., and Christensen, J.A. (**2014**) "Voice emotion recognition by cochlear-implanted children and their normally-hearing peers," Hear. Res., **322**, 151-162. doi: 10.1016/j.heares.2014.10.00

Deroche, M.L., Lu, H., Limb, C.J., Lin, Y. and Chatterjee, M. (**2014**). "Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin," Front. Neurosci. **8**, 282. doi: 10.3389/fnins.2014.00282

Deroche, M.L.D., Kulkarni, A.M., Christensen, J.A., Limb, C.J., and Chatterjee, M. (**2016**) "Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants," Front. Neurosci., **10**, 0007. doi:10.3389/ fnins.2016.00073

He, A., Deroche, M.L.D., Doong, J., Jiradejvong, P., and Limb, C.J. (**2016**). "Mandarin tone identification in cochlear implant users using exaggerated pitch contours," Otol. Neurotol., **37**, 324-331. doi: 10.1097/MAO.0000000000000980

Luo, H., Boemio, A., Gordon, M., and Poeppel, D. (**2007a**). "The perception of FM tones by Chinese and English listeners," Hear. Res., **224**, 75-83. doi: 10.1016/j.heares.2006.11.007

Luo, X., Fu, Q.J., and Galvin, J.J. 3rd. (**2007b**) "Vocal emotion recognition by normal-hearing listeners and cochlear implant users," Trends Amplif., **11**, 301-315. doi: 10.1177/1084713807305301

Krishnan, A., Gandour, J.T., Ananthakrishnan, S., Bidelman, G.M., and Smalt, C.J. (**2011**) "Linguistic status of timbre influences pitch encoding in the brainstem," Neuroreport, **22**, 801-803. doi: 10.1097/WNR.0b013e32834b2996

Krishnan, A., Gandour, J.T., Ananthakrishnan, S., and Vijayaraghavan, V. (**2015**). "Language experience enhances early cortical pitch-dependent responses," J. Neurolinguistics, **33**, 128-148. doi: 10.1016/j.jneuroling.2014.08.002

Peng, S.C., Lu, H.P., Lu, N., Lin, Y.S., Deroche, M.L., and Chatterjee, M. (**2017**) "Processing of acoustic cues in lexical tone identification by pediatric cochlear implant recipients," J. Speech Lang. Hear. Res., **60**, 1223-1235. doi: 10.1044/2016_JSLHR-S-16-0048