Children's Recognition of Emotional Prosody in Spectrally-Degraded Speech is Predicted by Their Age and Cognitive Status

Anna R Tinnemore[1], **Danielle J Zion**[2], **Aditya M Kulkarni**[3], and **Monita Chatterjee**[4]

[1]Department of Speech, Language and Hearing Sciences, University of Arizona, Tucson, AZ 85721; Present address: Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742

[2]Walter Reed National Military Medical Center, Bethesda, MD 20889

[3]Boys Town National Research Hospital, 555 N 30th St, Omaha, NE 68131

[4]Boys Town National Research Hospital, 555 N 30th St, Omaha, NE 68131

Abstract

Objective—It is known that school-aged children with cochlear implants show deficits in voice emotion recognition relative to normally-hearing peers. Little, however, is known about normally-hearing children's processing of emotional cues in cochlear-implant-simulated, spectrally degraded speech. The objective of this study was to investigate school-aged normally-hearing children's recognition of voice emotion, and the degree to which their performance could be predicted by their age, vocabulary and cognitive factors such as non-verbal intelligence and executive function.

Design—Normally-hearing children (6–19 years old) and young adults were tested on a voice-emotion recognition task under three different conditions of spectral degradation using cochlear-implant simulations (full-spectrum, 16-channel, and 8-channel noise-vocoded speech). Measures of vocabulary, non-verbal intelligence and executive function were obtained as well.

Results—Adults outperformed children on all tasks, and a strong developmental effect was observed. The children's age, the degree of spectral resolution and nonverbal IQ were predictors of performance, but vocabulary and executive functions were not, and no interactions were observed between age and spectral resolution.

Conclusions—These results indicate that cognitive function and age play important roles in children's ability to process emotional prosody in spectrally-degraded speech. The lack of an interaction between the degree of spectral resolution and children's age further suggests that younger and older children may benefit similarly from improvements in spectral resolution. The findings imply that younger and older children with cochlear implants may benefit similarly from technical advances that improve spectral resolution.

corresponding author: monita.chatterjee@boystown.org.

1. Introduction

Cochlear implants (CIs) transmit limited information about voice pitch and speech spectral contrasts to the listener. The spectro-temporally degraded signal heard by the patient, thus, does not allow for normal perception of prosodic cues in speech. In typically-developing children, attentional preference for speech with exaggerated prosodic contours (a hallmark of infant-directed or child-directed speech) has been observed (Fernald, 1985). These cues, including voice pitch contours, timbral changes, intensity contours and duration patterns, are thought to be important in infants' acquisition of their native language (e.g., Kemler Nelson et al., 1989), but also convey important information about the talker, including information about their internal state, mood or emotion, and generally enhance the quality of spoken communication. As the development of emotional communication is critical for children to establish strong social networks and acquire social cognition skills, the perception of emotional prosody by children with CIs warrants investigation. School-age children with CIs (cCI) have been shown to have normal concepts of emotion as revealed by their performance in basic facial emotion recognition tasks (Hopyan-Misakyan et al., 2009), but deficits have been identified in more comprehensive tests of emotion comprehension in cCI aged 4–11 years (Mancini et al., 2016) and in very young children with CIs aged 2.5–5 years (Wiefferink et al., 2013). As with post-lingually deaf adults with CIs (Luo et al., 2007), prelingually deaf cCI exhibit significant deficits in voice emotion recognition (Hopyan-Misakyan et al., 2009; Chatterjee et al., 2015; Geers et al., 2013).

Today's CIs utilize between 12 and 24 electrodes to transmit the acoustic information extracted from distinct frequency bands. However, factors such as spread of electric current within the cochlea to neural regions distant from the target electrode and patchy neural survival along the cochlea, limit the number of independent channels of information available to the listener. CI simulations are acoustic models of the resulting spectral degradation. Children with normal hearing (cNH) who are subjected to CI-simulated speech provide an important and informative comparison to cCI attending to undistorted (unprocessed) speech. Although CI simulations have various limitations in the degree to which they represent the electrical signal received by CI listeners, they have proved remarkably useful as a tool to investigate the processing and reconstruction of degraded auditory inputs by speech and language systems of the brain. A widely-used CI simulation is noiseband vocoded (NBV) speech, in which bands of noise model channels of information transmitted by the CI. In recent work, Chatterjee et al. (2015) reported that cNH aged 6–18 years (who showed near-ceiling performance in voice emotion recognition tasks using unprocessed speech) showed significant deficits in the task relative to adults with normal hearing (aNH) when both groups listened to 8-channel noiseband-vocoded (NBV) speech. A significant developmental effect was observed in the children attending to 8-channel NBV speech, explaining 52.64 % of the variance in the data. The authors speculated that the effect of age may be related to the development of cognitive and linguistic skills, improving children's ability to reconstruct the degraded input and recover the intended cues. The present study directly addresses this speculation. In an interesting contrast, Chatterjee et al. (2015) also found that cCI fared better on average with unprocessed speech than cNH did with 8-channel NBV speech. The cCI also showed strong variability in outcomes, but this

was only weakly accounted for by age. Chatterjee et al. (2015) theorized that the difference could be attributed to the fact that cCI had considerable experience with their devices, while the cNH, who had no prior experience with spectrally-degraded speech, had greater difficulty extracting the intended emotion from the degraded input. The sounds of NBV speech and electrically stimulated speech through a cochlear implant are likely quite different, limiting direct comparisons between the two groups of children. However, the adults with NH and the children with NH likely perceived NBV speech sounds similarly. Using the adults with NH as a reference, it seems reasonable to conclude that the younger cNH, who showed significant deficits compared to both aNH listening to NBV speech and cCI listening to full-spectrum speech, were missing some ability to extract the emotion information from the NBV stimuli. In previous work, Eisenberg et al. (2000) reported similar findings in a speech recognition task by cNH attending to NBV speech, and were able to link their performance to cognitive measures (such as working memory).

Voice emotion comprises a complex set of acoustic cues (e.g., Murray and Arnott, 1993; Banse and Scherer, 1996). Pitch cues are dominant, alongside spectral envelope, intensity, and duration cues. In an audio-only scenario (e.g., if the speaker is not directly facing the listener, or if the speaker's face is obscured by limited lighting) and when the voice pitch is degraded, as in CIs or in NBV speech, the listener would have to rely on appropriate use of the remaining cues for voice emotion recognition. Older children and adults may have an advantage here over younger children, who have limited experience with speech perception overall, have limited cognitive and linguistic resources, and may not have built up strong internal representations of the complex and variable acoustic structures associated with different vocal emotions. Social cognition is linked to executive function, and both continue to develop in children through puberty (for review, see Blakemore & Choudhury, 2006). As vocal emotion perception, the focus of the present study, represents an aspect of social cognition, performance in a vocal emotion recognition task may be additionally related to executive function in developing children. Thus, we hypothesize that developmental age, cognition, language skills and executive function, will all play a role in predicting children's performance in a voice emotion recognition task with degraded speech inputs.

Eight-channel NBV speech, as heard by aNH listeners, is thought to approximate the average speech information received by high-performing adult, postlingually deaf, CI listeners (e.g., Friesen et al., 2001). However, recent work suggests that with newer devices, CI users may benefit from increasing the number of channels available to carry speech information (Schvartz-Leyzac et al., 2017). Using child-directed speech materials (i.e., exaggerated prosody), Chatterjee et al. (2015) found that in voice emotion recognition, aNH listeners had mean performance levels with 8-channel NBV speech that most closely approximated the mean performance of both children and adults with CIs with full-spectrum speech. As expected, the aNH listeners in that study showed improvements with increasing numbers of channels from 8 to 16 (child NH participants only completed the full-spectrum and 8-channel conditions). Findings of Luo et al (2007), using adult-directed materials in a study of voice emotion recognition with adult NH and CI users, indicate that the CI users' mean performance with full-spectrum speech approximates NH listeners' mean performance with 4-channel sine-vocoded speech. The NH listeners in their study showed some improvements in performance when the number of channels was increased from 8 to 16.

This is consistent with previous studies showing that aNH listeners show continuing improvements in NBV speech recognition, particularly in background noise, as the number of vocoded channels increases from 8 to 16 (Friesen et al., 2001; Fu & Nogaki, 2005). Adult listeners also report an improvement in the subjective quality of speech with an increase in number of channels from 8 to 16 (Loizou, et al., 1999).

Thus, the previous literature shows that adults with NH benefit from increasing the spectral resolution in CI-simulated speech from 8 to 16 channels. This suggests that future modifications to CI technology that result in improved spectral resolution are likely to benefit post-lingually deaf adult CI users in both speech comprehension and emotion recognition tasks. The extent to which children might benefit from such improvements in spectral resolution in an emotion recognition task, however, remains unknown. Factors that might determine children's benefit include developmental age, cognitive status, and language abilities. At present, it is not feasible to change spectral resolution in listeners with CIs in a controlled manner: therefore, such studies must be limited to listeners with NH attending to degraded inputs. A caveat to this approach is that CI simulations are essentially limited in their representation of an impaired auditory system's response to electrical stimulation. Despite this acknowledged limitation, simulation studies have provided extraordinary insight into CI listeners' performance/constraints over the last few decades. If, for instance, we find that younger children benefit less from increased spectral resolution than older children do, and that the amount of benefit is related to their cognitive status or language abilities, then in the future, we could make more informed decisions for children with CIs receiving advanced technologies (e.g., focused stimulation which attempts to increase the number of channels available to transmit information). Such decisions might influence follow-up rehabilitation services provided to children with CIs receiving new modifications to their devices or processors at different developmental ages.

In the present study, we investigated the degree of benefit received by children when spectral resolution was increased from 8 to 16 channels of NBV speech. We hypothesized that younger children (6–10 years of age) might show reduced ability to benefit from the enhanced speech information than older children (older than 10). We note that cNH achieve near-ceiling performance in speech recognition with 16-channel NBV speech (Eisenberg et al., 2000). However, even while providing sufficient spectral resolution to support high levels of speech recognition, 16-channel NBV speech still presents limited voice pitch cues. In children with NH and in children with CIs, sensitivity to the fundamental frequency has been shown to be significantly correlated with performance in voice emotion tasks (Deroche et al., 2016). Thus, it is reasonable to hypothesize that voice emotion recognition might still be difficult for younger children in the 16-channel condition. The benefit children obtain from improved spectral resolution may also depend on their cognitive abilities. In previous studies, cCI's nonverbal cognitive abilities have been correlated with their performance in speech perception and indexical information processing tasks (e.g., Geers and Sedey, 2011; Geers et al., 2013). In the present study, we measured nonverbal intelligence, receptive vocabulary and executive function in the child participants. Although the focus of the study was on children and their performance with 8- and 16-channel NBV speech, we also obtained data with unprocessed (clean) versions of the stimuli. In addition, data were

obtained in a group of young NH adults (aNH) in all three conditions of spectral resolution, as a useful indicator of performance at the endpoint of the developmental trajectory.

2. Methods

2.1 Stimuli

The stimuli consisted of the female-talker sentences taken from the full corpus used by Chatterjee et al. (2015). Briefly, twelve emotion-neutral sentences from the Hearing In Noise Test (HINT) lists (e.g., *The cup is on the table*) were recorded by a 26-year-old female native English speaker in a child-directed manner with 5 emotions (*angry, happy, neutral, sad, scared*) for a total of 60 sentences. NBV versions (16-channel and 8-channel) were created using AngelSim (Emily Shannon Fu Foundation, tigerspeech.com) using standard settings (Chatterjee et al., 2015).

2.2 Tasks

Voice emotion recognition was measured using a single-interval, 5-alternative, forced-choice task. Stimuli were presented in randomized order within blocks (60 sentences in a particular condition of spectral resolution within a block). The different conditions were also presented in randomized order. Before each block of testing, listeners were provided with passive training using two sentences (processed to have the spectral resolution for that block) that were not included in the test, spoken in each of the five emotions. After each sentence was presented, the correct choice (emotion) was highlighted on the computer screen. This provided the listener with a sense of the speaking style of the talker in each emotion and condition (8 channel, 16 channel, etc.).

The Block Design and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence– Second Edition (WASI-II) were administered to child participants to obtain a composite measure of nonverbal intelligence (NVIQ; Wechsler, 1999). During the Block Design subtest of the WASI-II, the children were asked to replicate various patterns from the stimulus booklet using patterned blocks. When completing the Matrix Reasoning subtest, the children were asked to look at a series of pictures and asked to select a picture from an assortment of options to fill the missing space that best completed the series pattern. The overall Performance IQ (age-normed) resulting from the WASI-II was used in analyses.

The Peabody Picture Vocabulary Test – Form B (PPVT) provided a measure of receptive vocabulary (Dunn & Dunn, 2007). During this portion of testing, the children were asked to look at four images. The experimenter would then verbally present a word and ask the children to point to the picture on the page that best represented the word. The percentile score from the PPVT (age-normed) was used in analyses. Although the stimuli in this case were not calibrated, experimenters were trained in the procedure by clinically-trained colleagues at our institution, every effort was made to ensure that all of the children heard the presented words in the same way, were seated at the same distance from the experimenter, and were tested under the same conditions.

Parents completed the Behavior Rating Inventory of Executive Functioning (BRIEF) questionnaire (Gioia et al., 2000). The BRIEF includes measures of behavioral regulation

(including questions that address the child's ability to exercise inhibitory control over impulses, to switch attention from one task to another and to regulate emotions) and metacognition (including the child's ability to initiate activities, working memory, planning and organizing future events, goal-setting, organizing study or play materials, etc.). The Global Executive Composite score derived from the BRIEF (age-normed), which combines these measures, was used in our analyses.

The tests of cognition, language, and executive function were also presented in randomized order, typically during breaks in between the voice emotion recognition tasks.

2.3 Participants

Thirty children with normal hearing (thresholds 20 dB HL, 500–4000 Hz), aged 6.31 to 18.09 years (mean = 11.11 years, s.d. = 3.85 years) and 22 adults with normal hearing as defined above ranging in age from 20.01 to 33.2 years (mean = 24.79 years, s.d. = 4.00 years), participated. All participants were tested in the Full-spectrum, 16-channel NBV and 8-channel NBV conditions. Data from two of the children were excluded from analyses, as their performance with Full-spectrum speech fell 2.5 and 4.0 standard deviations below the mean, respectively. Of the 28 children whose data were included in analyses, 12 who were younger than 10 years of age were categorized as younger children (YCH), and the remaining 16 children who were older than 10 years of age were categorized as older children (OCH). Data were also obtained on NVIQ, receptive vocabulary (PPVT) and executive function (BRIEF) tasks in the children.

Owing to time constraints, not all of the children could complete all of the tasks. When participants had such time limitations, the voice emotion recognition tasks received priority, and the cognitive/linguistic tasks were completed as time permitted. For six of the children, data on one or more of the nonverbal intelligence, vocabulary, and executive function measures could not be obtained.

2.4 Analyses

Preliminary analyses were conducted in SPSS version 22 (IBM Corp.) and in R.3.1.2 (R Core Team, 2014) using the *ggplot2* package (Wickham, 2009). The linear mixed-effects models were generated in R 3.1.2, using the *nlme* package (Pinheiro et al., 2017).

3. Results

Figure 1 shows a scatterplot of the participants' performance (% correct) as a function of age. Chance performance was 20% correct (one out of five emotions correct). The regression lines were plotted only through the children's data (data points to the left of the 19-year point on the abscissa), and show significant correlations between performance in all conditions and age. As in Chatterjee et al. (2015), the Full-spectrum condition yielded near-ceiling performance in the children (indicating that the stimuli conveyed the emotions adequately). The age-effect in the Full-spectrum condition should be viewed with caution because of the obvious and expected ceiling effect. The main purpose of including this condition was to confirm that children could achieve excellent performance in the basic task.

The results were analyzed using a linear mixed-effects modeling approach. The percent correct scores were converted to arcsine-transformed units (Studebaker, 1981) in an attempt to normalize the distributions and to homogenize the variance. The transformation did not restore normality or homoscedasticity of the data (Shapiro Wilk test, Levene's test) for the Full-spectrum condition. For the 16-channel and 8-channel conditions, the data were homoscedastic before and after transformation (Levene's test). The percent correct data for the 16- and 8-channel conditions did not pass the tests of normality (Kolmogorov-Smirnov test, Shapiro-Wilk test), but the transformed data did pass the Kolmogorov-Smirnov test for normality (only data obtained in the 8-channel condition passed the Shapiro-Wilk test for normality). Based on these considerations, a decision was made to use the transformed dataset in analyses. We note that linear mixed effects models, which were the primary analysis tool used in this study, do not assume either homoscedasticity or normality of the data. However, they do assume that the model residuals are normally distributed, and this was checked by visual inspection of residual plots.

First, the overall effects of age groups and condition were analyzed. The subjects were divided into three groups: Younger Children (<10 years of age, YCH, N=12), Older Children (>10 years of age, OCH, N=16) and Adults. Figure 2 shows boxplots of percent correct scores obtained by each group, as a function of the degree of spectral resolution (abscissa). The LME included Group (YCH, OCH and Adults) and Condition (8-channel, 16-channel NBV and Full-spectrum speech) as fixed effects, and subject-based random intercepts. Initial analyses indicated that the inclusion of subject-based random slopes would not significantly improve the model fit. Results showed significant effects of Group ($t(40)=6.62$, $p<0.001$), and Condition ($t(106)=-3.46$, $p=0.0008$), and no significant interactions. The fitted model data were significantly correlated with the observed values ($r=0.53$, $p<0.0001$) and visual inspection of the model residuals suggested a reasonable fit.

The Full-spectrum condition was only included to confirm that children were able to identify the emotions correctly, and the ceiling-level performance in this condition confirmed this. For the remainder of the analyses, the Full-spectrum condition was excluded, and questions focused on the 8- and 16-channel conditions.

First, the analysis was conducted using Age (as a continuous variable), Condition (8 and 16 channel conditions only), Vocabulary, Executive Function (EF) and NVIQ as fixed effects, and subject-based random intercepts. Results showed significant effects of Age ($t(17)=4.6$, $p=0.0003$) and NVIQ ($t(17)=3.54$, $p=0.0025$), and no effects of Condition, Vocabulary or EF, and no significant interactions. Figure 3 illustrates the relation between NVIQ and individual children's scores obtained in the 8-channel and 16-channel conditions. As there was no interaction between Condition and NVIQ, the mean score obtained in the two conditions was calculated and plotted against the NVIQ scores to visualize the relation between performance and NVIQ. The black symbols show the mean of the scores obtained by each child in the two conditions, and the black line shows the results of a linear regression ($r=0.47$, $p=0.01$).

Next, the analysis was repeated with Age categorized into the two groups used in the first analysis (YCH and OCH). Fixed effects were Group, Condition, Vocabulary, EF, and NVIQ,

and subject-based random intercepts were included. Results showed significant effects of Group (t(17)=3.76, p=0.0016), Condition (t(19)= −2.30, p=0.0331), and NVIQ(t(17)=3.03, p=0.0076), no significant effects of Vocabulary or EF, and no significant interactions. The individual predictors were not significantly correlated with each other. Visual inspection of the residuals suggested a reasonable fit, and the predicted values generated by the model were significantly correlated with the observed values (r=0.95, p<0.0001).

4. Discussion

These results confirm a consistent and strong effect of children's age, indicate a further main effect of the degree of spectral resolution, and reveal a strong contribution of NVIQ in children's ability to identify emotions (i.e. emotional prosody) in spectrally degraded speech. The effect of spectral resolution became weaker or non-existent when the Full Spectrum condition was eliminated from analyses, suggesting that children had difficulty performing the task even with mild spectral degradation.

It is somewhat surprising that children continue to show large deficits in voice emotion recognition with 16-channel NBV speech compared to Full-spectrum speech, given that the 16-channel condition would be expected to yield near-perfect speech recognition scores in this population in quiet (Eisenberg et al., 2000). The lack of an interaction between Age and Condition also suggests that the ability to benefit from increased spectral resolution is independent of age and unrelated to age-related deficits in the task. In other words, the psychometric function is the same: it simply shifts up the vertical axis as children get older. The ceiling effects with Full-spectrum speech do not allow the observation of the vertical shift, but the present results suggest that improving spectral resolution beyond 16 channels would continue to benefit younger and older age groups to the same extent.

In addition to age, it seems that non-verbal IQ accounts for a significant proportion of the variability in the dataset. Geers et al. (2003) suggested a role for NVIQ in language development in cCI more than a decade ago. The present results place NVIQ as an important predictor of cNH's ability to detect voice emotion in CI-simulated speech. Collectively, these results suggest that the processing of spectrally-degraded information in speech (both indexical and linguistic) by children with CIs and children with NH is predicted by their general cognition (e.g., working memory measures in Eisenberg et al. and NVIQ in the present study). It is possible that mechanisms shared by indexical and linguistic processing are mediated by, or related to, general cognition. Geers et al. (2013) have suggested links between cCIs' performance in tasks involving indexical information processing and their language development. The present study did not find language (vocabulary) to be a predictor in the cNHs' performance, suggesting that mechanisms in cCI may be somewhat different from those in cNH. Additionally, the results showed no predictive value of the BRIEF measures for emotion recognition in our tasks. It is possible that this specific measure did not capture the aspects of EF that are most relevant to voice emotion recognition, but further investigation along these lines is clearly needed.

What causes younger children to require more spectral resolution to achieve the same performance in voice emotion recognition as older children? Possibly the answer lies in a

combination of factors. The specific task of voice emotion recognition requires salient pitch cues which remain degraded even with 16-channel NBV speech. Performance with degraded speech would require listeners to draw on their experience with vocal communication and rely on secondary acoustic cues associated with each emotion: for instance, happy vocalizations involve larger voice pitch fluctuations, but are also louder and faster, than sad vocalizations. Younger children have limited experience with spoken communication in general than older children, and their still-developing language systems might additionally increase their vulnerability to degradation of the input signal. Their ability to draw on these secondary cues may be more limited by these factors than older children's. Finally, it is likely that top-down reconstruction mechanisms and overall cognitive resources are generally more limited in younger children. This is suggested by the significant contribution of the nonverbal intelligence scores to children's performance with the NBV stimuli in the present dataset.

Previous results obtained in adult listeners with NH or with CIs suggest that in laboratory conditions in which specific phonetic contrasts or speech intonation patterns are explicitly provided at different levels of salience of the primary and secondary acoustic cues, both groups can use the secondary cue when the primary cue is degraded, albeit stronger intersubject variability tends to be observed in listeners with CIs (Moberly et al., 2014; Winn et al, 2012; Winn et al., 2013; Peng et al., 2009; Peng et al., 2012). The literature on children's ability to use secondary acoustic cues to identify speech or speech prosody under conditions of spectral degradation, however, remains limited. Nittrouer et al (2014) showed that 8-year old children with CIs showed cue-weighting patterns which were related to their auditory sensitivity to those cues, their phonemic awareness and word recognition abilities. However, developmental effects in these areas remain unknown. When specific cues are available to children with CIs, they show patterns similar to those observed in children with NH, although their usage of these cues is less efficient (Nittrouer et al., 2014; Giezen et al., 2010). Cue-weighting patterns, however, may not always be related to performance in natural speech perception. In Mandarin-speaking pediatric CI recipients performing a cue-weighting task in lexical tone identification, Peng et al. (2017) showed greater reliance on the secondary cue (duration) than on the primary cue (F0 change) in children with CIs than their NH peers. However it was their reliance on the primary cue, rather than reliance on the secondary cue, that was predictive of their real-world tone recognition performance.

Although such studies are informative about specific acoustic cues utilized by children with CIs in speech perception tasks, many of them have focused on a single age group. Thus, relatively little is known about developmental effects. In addition, even less is known about the development of cue-weighting patterns in children with NH attending to degraded speech cues.

Studies in NH pre-school children and toddlers suggest that the ability to process spectrally sparse speech (sinewave speech or NBV speech with 8 channels) is in place at early stages of development (Nittrouer et al., 2009, Newman & Chatterjee, 2013; Newman et al., 2015). However, Newman & Chatterjee (2013) and Newman et al (2015) also suggest that the performance of young children in such tasks depends strongly on the degree of spectral resolution. This is consistent with findings of Eisenberg et al. (2000). While these studies

focused on speech recognition, the present results, and those of Chatterjee et al (2015), also suggest a developmental trajectory in NH children's perceptions of voice emotion in NBV speech. What still remains unclear is which specific skills are improving to help children achieve better performance at older ages.

Factors such as age at implantation or experience with cochlear implants cannot be simulated in NH listeners. Both are important factors in the language development trajectory of children with CIs (e.g., Niparko et al., 2010). Thus, the reader is cautioned against drawing direct inferences about children with CIs from the present dataset. Taking these limitations into account, however, what the present results do indicate is that younger and older children, and adults, are able to take similar advantage of improvements in the bottom-up stream of information in their processing of emotional prosody. This bodes well for future advancements in CIs, particularly in the area of spectral coding.

Acknowledgments

References

Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. J Person Soc Psychol. 1996; 70(3):614–636.

Blakemore S-J, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. J Ch Psychol Psych. 2006; 47(3–4):296–312.

Chatterjee M, Zion DJ, Deroche ML, Burianek BA, Limb CJ, Goren AP, Kulkarni AM, Christensen JA. Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. Hear Res. 2014; 322:151–162. [PubMed: 25448167]

Deroche MLD, Kulkarni AM, Christensen JA, Limb CJ, Chatterjee M. Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. Front Neurosc. 2016; 10:0007.doi: 10.3389/fnins.2016.00073

Eisenberg LS, Shannon RV, Martinez AS, Wygonski J, Boothroyd A. Speech recognition with reduced spectral cues as a function of age. J Acoust Soc Am. 2000; 107(5):2704–2710. [PubMed: 10830392]

Fernald A. Four-month-old infants prefer to listen to motherese. Infant Behav Dev. 1985; 8(2):181–195.

Friesen LM, Shannon RV, Baskent D, Wang X. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J Acoust Soc Am. 2001; 110:1150–1163. [PubMed: 11519582]

Fu QJ, Nogaki G. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J Assoc Res Otolaryngol. 2005; 6(1):19–27. [PubMed: 15735937]

Geers AE, Davidson LS, Uchanski RM, Nicholas JG. Interdependence of linguistic and indexical speech perception skills in school-age children with early cochlear implantation. Ear Hear. 2013; 34:562–574. [PubMed: 23652814]

Geers A, Sedey AL. Language and verbal reasoning skills in adolescents with 10 or more years of cochlear implant experience. Ear Hear. 2011; 32:39S–48S. [PubMed: 21832889]

Geers A, Brenner C, Davidson L. Factors associated with development of speech perception skills in children implanted by age five. Ear Hear. 2003; 24:24S–35S. [PubMed: 12612478]

Giezen MR, Escudero P, Baker A. Use of acoustic cues by children with cochlear implants. J Speech Lang Hear Res. 2010; 53:1440–1457. [PubMed: 20689031]

Gioia GA, Isquith PK, Guy SC, Kenworthy L. Behavior rating inventory of executive function. Child Neuropsychol. 2000; 6:235–238. [PubMed: 11419452]

Hopyan-Misakyan TM, Gordon KA, Dennis M, Papsin BC. Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. Child Neuropsychol. 2009; 15:136–146. [PubMed: 18828045]

IBM Corp. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp; Released 2013

Kemler Nelson DG, Hirsh-Pasek K, Jusczyk PW, Cassidy KW. How the prosodic cues in motherese might assist language learning. J Child Lang. 1989; 16(1):55–68. [PubMed: 2925815]

Loizou PC, Dorman M, Tu Z. On the number of channels needed to understand speech. The Journal of the Acoustical Society of America. 1999; 106(4):2097–2103. [PubMed: 10530032]

Moberly AC, Lowenstein JH, Tarr E, Caldwell-Tarr A, Welling DB, Shahin AJ, Nittrouer S. Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? J Speech Lang Hear Res. 2014; 57(2):566–582. [PubMed: 24686722]

Murray IR, Arnott JL. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. J Acoust Soc Am. 1993; 93(2):1097–1108. [PubMed: 8445120]

Newman RS, Chatterjee M, Morini G**, Remez R. Toddlers' comprehension of degraded signals: noise-vocoded vs. sine-wave analogs. J Acoust Soc Am. 2015; 138:EL.311–317.

Newman R, Chatterjee M. Toddlers' recognition of noise-vocoded speech. J Acoust Soc Am. 2013; 133(1):483–494. [PubMed: 23297920]

Niparko JK1, Tobey EA, Thal DJ, Eisenberg LS, Wang NY, Quittner AL, Fink NE. CDaCI Investigative Team. Spoken language development in children following cochlear implantation. J Am Med Assoc. 2010; 303(15):1498–1506.

Nittrouer S, Lowenstein JH, Packer RR. Children discover the spectral skeletons in their native language before the amplitude envelopes. J Exp Psychol Hum Percept Perform. 2009; 35:1245–1253. DOI: 10.1037/a0015020 [PubMed: 19653762]

Nittrouer S, Caldwell-Tarr A, Moberly AC, Lowenstein JH. Perceptual weighting strategies of children with cochlear implants and normal hearing. J Commun Disord. 2014; 52:111–133. [PubMed: 25307477]

Peng SC, Lu N, Chatterjee M. Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. Audiol & Neurotol. 2009; 14(5):327–337.

Peng SC, Chatterjee M, Lu N. Acoustic cue integration in speech intonation recognition with cochlear implants. Trends in Amp. 2012; 16(2):67–82.

Pinheiro J, Bates D, DebRoy S, Sarkar D. R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131. 2017. https://CRAN.R-project.org/package=nlme

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. URL http://www.R-project.org/

Schvartz-Leyzac KC, Zwolan TA, Pfingst BE. Effects of electrode deactivation on speech recognition in multichannel cochlear implant recipients. Cochlear Implants International. 2017; 8(6):324–334.

Wechsler D. Wechsler Abbreviated Scale of Intelligence (WASI). Psychological Corporation; San Antonio: 1999.

Wiefferink CH, Rieffe C, Ketelaar L, De Reeve L, Frijns JH. Emotion understanding in deaf children with cochlear implants. J Deaf Stud Deaf Educ. 2013; 18(2):175–186. [PubMed: 23232770]

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag; New York: 2009.

Winn MB, Rhone AE, Chatterjee M, Idsardi WJ. The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. Frontiers Psychol. 2013; 4:824.doi: 10.3389/fpsyg.2013.00824

Winn MB, Chatterjee M, Idsardi WJ. The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. J Acoust Soc Am. 2012; 131(2):1465–1479. [PubMed: 22352517]
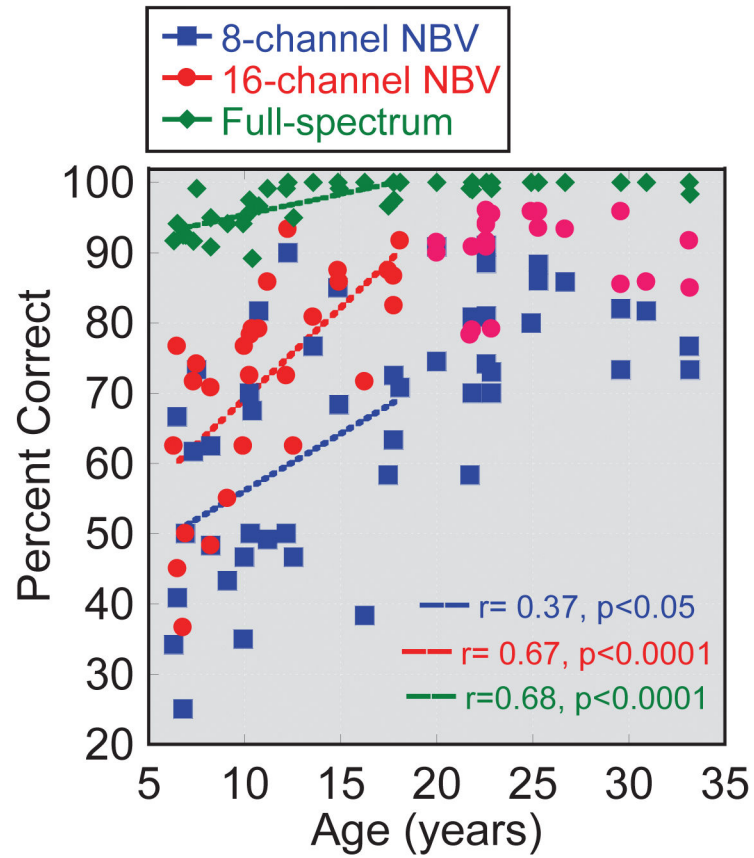
Fig. 1.
Percent correct scores obtained by all participants as a function of age (abscissa). The degree of spectral resolution is shown by the different symbols. Lines indicate regression through the children's data only.
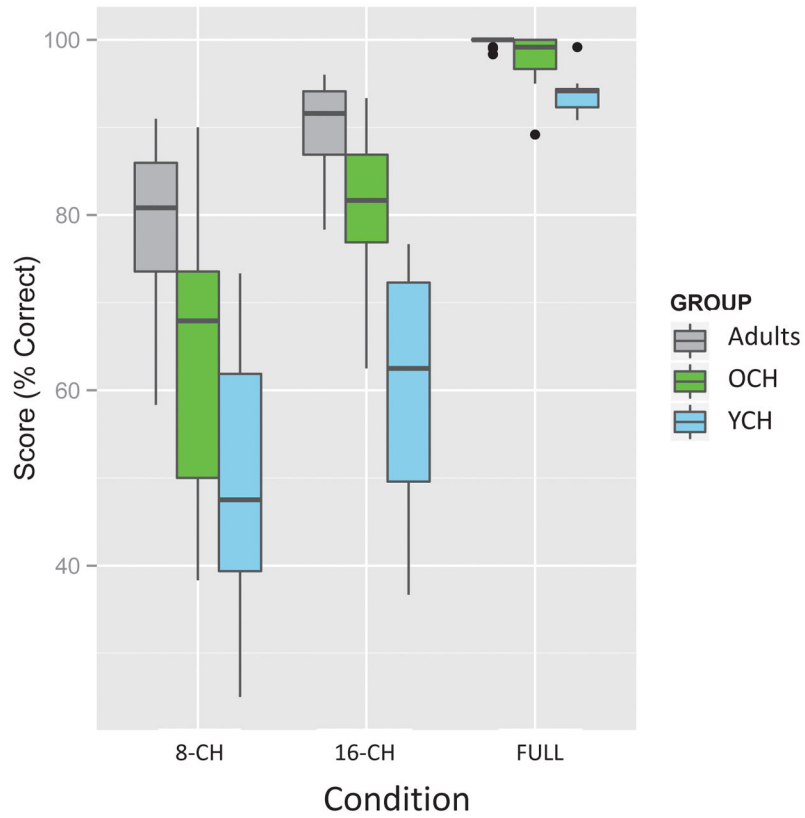
Fig. 2.
Boxplots of percent correct scores obtained by the younger and older children, and adults, for the different conditions of spectral resolution (8-channel NBV, 16-channel NBV, and Full-spectrum).
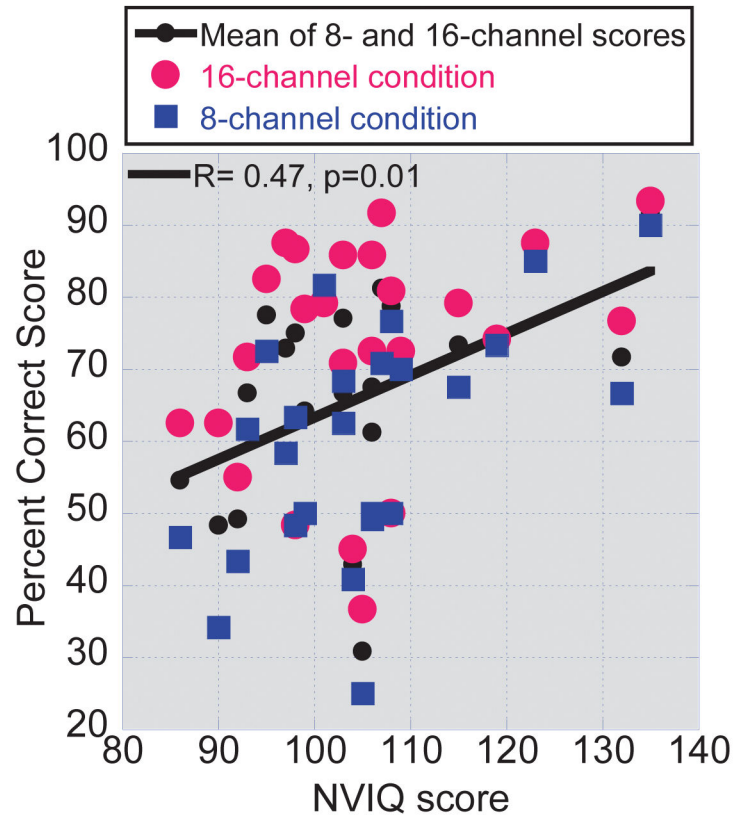
Fig. 3.
Scatterplot of percent correct scores obtained by all children in the 8-channel (blue) and 16-channel conditions, as a function of their NVIQ scores. The mean of the two scores is shown in black symbols, and black line represents a linear regression.