

HHS Public Access

Author manuscript *Hear Res.* Author manuscript; available in PMC 2016 April 01.

Published in final edited form as:

Hear Res. 2015 April; 322: 151-162. doi:10.1016/j.heares.2014.10.003.

Voice emotion recognition by cochlear-implanted children and their normally-hearing peers

Monita Chatterjee^{1,*}, Danielle Zion², Mickael L. Deroche³, Brooke Burianek¹, Charles Limb³, Alison Goren^{1,2}, Aditya M. Kulkarni¹, and Julie A. Christensen¹

¹Auditory Prostheses & Perception Lab, Boys Town National Research Hospital, 555 N 30th St, Omaha, NE 68131

²Department of Hearing & Speech Sciences, University of Maryland, 0100 LeFrak Hall, College Park, MD 20742

³Department of Otolaryngology, Johns Hopkins University School of Medicine, 818 Ross Research Building, 720 Rutland Avenue, Baltimore, MD

Abstract

Despite their remarkable success in bringing spoken language to hearing impaired listeners, the signal transmitted through cochlear implants (CIs) remains impoverished in spectro-temporal fine structure. As a consequence, pitch-dominant information such as voice emotion, is diminished. For young children, the ability to correctly identify the mood/intent of the speaker (which may not always be visible in their facial expression) is an important aspect of social and linguistic development. Previous work in the field has shown that children with cochlear implants (cCI) have significant deficits in voice emotion recognition relative to their normally hearing peers (cNH). Here, we report on voice emotion recognition by a cohort of 36 school-aged cCI. Additionally, we provide for the first time, a comparison of their performance to that of cNH and NH adults (aNH) listening to CI simulations of the same stimuli. We also provide comparisons to the performance of adult listeners with CIs (aCI), most of whom learned language primarily through normal acoustic hearing. Results indicate that, despite strong variability, on average, cCI perform similarly to their adult counterparts; that both groups' mean performance is similar to aNHs' performance with 8-channel noise-vocoded speech; that cNH achieve excellent scores in voice emotion recognition with full-spectrum speech, but on average, show significantly poorer scores than aNH with 8-channel noise-vocoded speech. A strong developmental effect was observed in the cNH with noise-vocoded speech in this task. These results point to the considerable benefit obtained by cochlear-implanted children from their devices, but also underscore the need for further research and development in this important and neglected area.

^{*}Correspondence: Monita Chatterjee, Auditory Prostheses & Perception Laboratory, Boys Town National Research Hospital, 555 N 30th St, Omaha, NE 68131, monita.chatterjee@boystown.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cochlear implants; voice emotion recognition; prosody; pitch

1. INTRODUCTION

Cochlear implants (CIs) today achieve remarkable success in delivering speech information to severely hearing-impaired or profoundly deaf listeners. More than 300,000 patients world-wide now use CIs (NIH Publication no. 11-4798, http://www.nidcd.nih.gov/health/ hearing/pages/coch.aspx), and many of them are children who were born deaf or lost their hearing within the first few years of life. The spectro-temporally sparse signal transmitted through the device supports high levels of speech understanding in quiet by many postlingually deaf adult CI patients who learned spoken language normally as children, a testament to the remarkable robustness of the pattern recognition ability of the human speech recognition system. The top-down "filling in" of degraded speech does not, however, compensate for all of the missing information. For example, the spectro-temporal detail that is needed to support the perception of harmonic pitch is lost, and as a result, pitch-dominant aspects of speech such as question-statement contrasts, vocal emotion, and lexical tone recognition are weakly transmitted by CIs (Shannon, 1983; Zeng, 2002; Chatterjee & Peng, 2008; Kong et al., 2009; Luo et al., 2007; Peng et al., 2004; Luo & Fu, 2004; Ciocca et al., 2002; Wei et al., 2004). This limitation also has a strong impact on the perception and production of melody by CI patients (Kong et al., 2004; Xu et al., 2009). However, music demands a level of accuracy in pitch that may not be necessary to understand or produce speech prosody. Although voice pitch is a primary acoustic cue for prosody, other cues such as intensity and duration also change to inform the listener of a question/statement contrast, or a heightened emotion. The listener might be able to make use of the accompanying cues to decode the speaker's intent/mood when pitch cues are degraded. In fact, work from our research group (e.g., Peng et al., 2009, 2012) has shown that adult listeners with CIs (aCIs), along with normally hearing adults (aNHs) attending to CI-simulated speech, do shift their attention to such co-varying cues in a relatively simple question/statement identification task. Thus, recognition of speech prosody may in fact be an *achievable goal* for patients with CIs. These aspects of spoken communication are critical for the listener to fully understand the communicative intent and mood of the speaker. Deficits in this regard can influence perceived quality of life, social interactions and, particularly for children, social development. Schorr et al. (2009) reported that, in children with CIs (cCI), perceived quality of life was predicted by their performance on a voice emotion recognition test (positive, negative or neutral), but not by their word recognition scores. Poor social function and psychopathological symptoms have also been tied to deficits in emotion understanding in children and adults (Eisenberg et al., 2010). In a recent study, Geers et al. (2013) reported links between voice emotion recognition and language skills in children with CIs, although it is unclear if there is a causal relationship between the two.

Many school-aged children with normal hearing have adult-like pitch discrimination (Deroche et al., 2012) and therefore strong access to the dominant cue for voice emotion. Voice emotion recognition is well established by age 5 in normally hearing children (cNH),

but continues to develop over time (Sauter et al., 2013; Tonks et al., 2007). Children with hearing impairment, however, have deficits in emotion recognition, some of which can be alleviated by training (Dyck et al., 2004; Dyck & Denver, 2003). Peng et al. (2008) reported deficits in the perception and production of speech intonation in question/statement contrasts by English-speaking cCI. Mandarin-speaking cCI show significant deficits in both perception and production of lexical tones (e.g., Peng et al., 2004; Ciocca et al., 2002). Facial emotion recognition seems to be delayed in pre-school cCI (Wiefferink et al., 2013), but is well established in school-age cCI (Hopyan-Misakyan et al., 2009). This suggests that children listening through CIs are able to establish firm concepts of emotion fairly early. However, studies to date indicate that voice emotion recognition by cCI remains significantly poorer than their normally hearing peers (Most & Aviner, 2009; Wang et al., 2013; Ketelaar et al., 2012; Volkova et al., 2013). Nakata et al. (2012) found that imitative voice emotion production scores of Japanese cCI aged 5–13 years were significantly poorer than cNHs' and correlated with their voice emotion perception scores. Similarly, acoustic analyses of 5 to 7-year-old "star" cCI and cNHs' imitations of happy and sad utterances (spoken by a normally hearing child model) showed significant deficits in the cCIs' imitations (Wang et al., 2013).

A major form of spectral degradation in CIs occurs through excessive spread of electrical current away from the electrode (e.g., Shannon, 1983; Kral et al., 1998; Chatterjee & Shannon, 1998). The spectral smearing of CIs is most commonly simulated in acoustic hearing by noise-band-vocoding (NBV), in which the broadly spreading electric field is represented by broad bands of noise centered at specific frequencies and temporally modulated by envelopes extracted from corresponding bands of the original speech spectrum (Shannon et al., 1995). Studies comparing post-lingually deaf adult CI patients' (aCIs') performance in phoneme and sentence recognition tasks in quiet and in noise, with the performance of adult NH listeners (aNH) in similar tasks, report equivalent performance when aNH listeners are attending to 4–8 spectral channels of NBV speech (e.g., Friesen et al, 2001), with the best aCIs' performance being equivalent to aNH listeners' performance with 8-channel NBV.

One aspect of the picture that still remains obscure is a comparison of how cNH process voice emotion information in spectrally degraded speech (of the sort heard through CIs), and cCIs' processing of voice emotion in natural speech. Such a comparison would tell us how much of the deficit in voice emotion recognition by cCI is due to difficulties with processing degraded speech faced by typically developing children, and how much of the deficit stems from other differences between the two populations. For instance, cNH have been shown to have more difficulty in recognizing degraded speech, particularly noise-vocoded speech, which transmits speech information in a manner similar to CI processors, than adults (e.g., Eisenberg et al., 2001; Nittrouer et al., 2009; Lowenstein et al., 2012). It is possible, therefore, that children faced with the dual load of perceiving degraded speech as well as decoding its associated emotion, may have greater difficulty than adults. Would cCI face the same difficulties as cNH who are relatively naïve to spectral degradations? Or would their experience with degraded inputs benefit their performance in voice emotion recognition?

A secondary question addressed by the present study relates to the performance of pre/peri/ postlingually deaf aCI as opposed to that of pre/perilingually deaf cCI. The post-lingually deaf aCI learned oral language with normal or only moderately impaired hearing as children, while the latter have been primarily exposed to electric hearing during development. How would this difference in early inputs to the system, influence their performance in voice emotion recognition? Postlingually deaf aCI might have three important advantages over the cCI: first, they would have greater knowledge of the acoustic properties of English and the modifications associated with different emotions. Second, they would have greater proficiency in their native language. Third, it is known that cNH have greater difficulties with spectrally degraded inputs than aNH (Eisenberg et al., 2001; Lowenstein et al., 2009; Nittrouer and Lowenstein, 2014). If cCI have similar difficulties with the spectrally degraded input provided by their device relative to aCI, then cCI might achieve poorer performance than aCI in tasks involving speech processing, in general. On the other hand, cCI might have the advantage of greater neuroplasticity and more efficient languagelearning during the sensitive period (Tomblin et al., 2007) with their device. Finally, many aCI who are not classified as prelingually deaf, had progressive hearing loss that began relatively early in life, resulting in a lack of consistent input in their developing years. Although many of these patients excel with their devices, the relatively sparse auditory input in the early years may impact their performance in many speech perception tasks (e.g., Koehlinger et al., 2013). These patients, and also those with prelingual deafness, might have developed various coping strategies over the years which the cCI have not as yet had sufficient time to fully acquire. Coping strategies such as altered patterns of cue-weighting (e.g., Peng et al., 2009, 2012; Winn et al., 2012; Winn et al., 2013) might be different for listeners dealing with spectral degradation vs. bandwidth reduction. It is still unclear how different challenges to hearing early in life (i.e., spectral degradation for cCI vs. reduction in the range of audible frequencies for aCI who had early hearing loss) alters adults' and childrens' listening strategies with CIs. Although the limited scope of the present study did not allow the investigation of these issues in full, the interpretation of the results might be enriched by keeping them in mind.

In the present study, we measured voice emotion recognition by school-aged cNH and cCI (6 – 18 years old). The cNH performed the task with both original (full-spectrum) speech and spectrally degraded, 8-channel NBV speech. Smaller numbers of aNH and aCI also participated. The aCI group comprised eight post-lingually deaf and one clearly prelingually deaf patients; most of the post-lingually deaf patients had some hearing impairment starting relatively early in life (with or without proper intervention). As it was expected that the cCI would have difficulty in the task, the stimuli were recorded in a child-directed manner, so the cues were exaggerated relative to adult-directed speech.

2. METHODS

2.1. Participants

1. Recording Task—One male and one female talker recorded the stimuli used in this study. They were both 26 years old, native speakers of American English, and spoke with a general American dialect.

2. Listening Task—Child participants included 31 cNH (15 boys, 16 girls, age range: 6.38 – 18.76 years, mean age 10.76 years, median age 10.37 years, standard deviation (s.d.): 3.056 years) and 36 cCI (15 boys, 21 girls, 19 users of the Advanced Bionics[™] CI and 17 users of Cochlear[™] CIs; age range: 6.83 – 18.44 years, mean age 12.15 years, median age 11.73 years, s.d.: 3.49 years, mean duration of device use 8.76 years). Adult participants included 10 aNH (3 men, 7 women, mean age 23.90 years, s.d. 2.76 years) and nine aCI (5 men, 4 women, age range: 27.34 – 69.81 years, mean age 52.16 years, s.d.: 13.22 years, mean duration of device use 8.00 years, 2 users of AB CIs and 7 users of Cochlear CIs). Table 1 lists further specifics of the aCI group. Note that the results of the one prelingually deaf adult participant fell within the range of the others', and therefore were combined with the remaining aCIs' data in analyses. Twenty-five of the 36 cCI were recruited and tested at the Johns Hopkins University School of Medicine in Baltimore, MD. The remaining subjects (all of the NH listeners and aCI listeners, and 11 of the cCI) were tested at Boys Town National Research Hospital in Omaha, NE. No significant differences were found between the cCIs' results obtained at the two sites (independent samples t-tests), and the data were combined for further analyses. Those CI users who had bimodal hearing were tested with only the CI, and those who had bilateral CIs were tested only on the side implanted first.

2.2 Stimuli

Twelve sentences were selected from the HINT database (Table 2), and spoken by one male and one female talker with five emotions (angry, happy, neutral, sad, and scared) in a childdirected manner. Participants for the recording task were seated in a sound-treated booth and were positioned 12 inches in front of a SHURE SM63 microphone with Marantz PMD661 solid state recorder and produced the sentences, three times each, according to the five target emotions. The sentences were selected based on their semantically emotion-neutral content according to the investigators' judgment. The sentences were selected in this manner to minimize any biasing effect of context. For example, the sentence "Big dogs can be dangerous" may lead to bias towards the emotion scared, while the sentence "The picture came from a book" carries much less, if any, emotional bias. After recording was completed, original audio files (44.1 kHz sampling rate, 16 bit) were edited using Adobe Audition version 1.5 software. Editing involved selecting the best of the three productions for each sentence and saving each as an individual audio file. Recording sessions were divided into five blocks, one for each emotion, and breaks were provided in between each block. Before beginning each recording block, participants were instructed to picture a personal or imaginary scenario in order to elicit a child-directed style of speech for the target emotion. The two talkers were selected from pilot work with four talkers (two male and two female), as the male and female resulting in the best emotion recognition by a group of NH children and adults. Noise-vocoded versions of the same stimuli were created using AngelSimTM (Emily Shannon Fu Foundation, www.tigerspeech.com). The method for noise vocoding paralleled that described by Shannon et al. (1995). For an n-channel vocoder, the speech signal was bandpass filtered into n logarithmically spaced bands (24 dB/octave) following the Greenwood frequency-place map. The time-varying speech envelope from each band was extracted using half-wave rectification and low-pass filtering (24 dB/oct filter, 160 Hz cutoff: the 160 Hz cutoff frequency was chosen to approximate the envelope discrimination

abilities of the average CI listener (e.g., Chatterjee & Peng, 2008; Chatterjee & Oberzut, 2011)). The envelope derived from each band was used to modulate a band-pass filtered white noise with the same filter parameters, center frequency and bandwidth. The different modulated noise bands were summed to create the final vocoded output. All stimuli were presented via an Edirol UA soundcard and a single loudspeaker located approximately 2 feet from the listeners, at an average level of 65 dB SPL.

2.3. Task

Participants in the listening task heard one presentation of the sentence, and indicated from a closed set which emotion he/she thought was associated with it. Sentences and emotions were fully randomized within each condition. Four conditions were available for testing in all: Full-spectrum speech, 16-channel NBV speech, 8-channel NBV speech, and 4-channel NBV speech. For each condition, stimuli included sentences spoken by the male and the female talker. Children and adults with CIs heard only full-spectrum speech. NH adults heard all four conditions. NH children heard full-spectrum and 8-channel NBV speech. Sentences were presented in blocks consisting of a given talker (male or female) and condition. Prior to listening to each block, the participant was given passive training with two sentences spoken by that talker (these sentences were not used in testing) in each of the five emotions. Each sentence/emotion would be presented, and the correct emotion button would light up on the screen. The purpose of this exercise was to familiarize the listener to the talker's speaking style and also what the talker sounded like in each condition. Blocks consisted of 60 trials (12 sentences, 5 emotions). Participants were encouraged to take breaks between blocks. No feedback was provided during the formal test.

3. RESULTS

3.1 Acoustic analyses of full-spectrum stimuli

The stimuli were analyzed using Praat v. 5.3.56 (Boersma, 2001; Boersma & Weenink, 2014) for mean F0 height (Hz), F0 range (ratio of maximum to minimum F0), mean intensity (dB SPL), overall duration (s), and the range of intensity (max – min in dB). Fig. 1 shows the results of the acoustic analyses. Each point represents the mean of all 12 sentences for each talker, and error bars represent standard deviations. Repeated measures ANOVAs (RMANOVAs) were conducted on the data shown in each panel, with Talker (male or female) and Emotion (five levels) as the factors:

For *average F0 height*, there were significant main effects of Talker (F(1,11) = 994.75, p<0.001) and Emotion (F(2.32, 25.56) = 893.11, p<0.001; Greenhouse Geisser correction for sphericity), and a significant interaction between the two (F(2.29, 25.188) = 28.25, p<0.001; Greenhouse Geisser correction for sphericity). Analyses of simple main effects showed the smallest difference between talkers for the *scared* emotion (the male talker raised his voice pitch considerably for this emotion: see Fig. 1), but as expected, significant differences were observed between the talkers (p<0.01 or better) at all levels, with the female talker's F0 being higher than the male's.

For *F0 range*, there was no main effect of Talker, but a significant main effect of Emotion (F(4, 44)=16.49, p<0.001) and a significant interaction (F(4, 44)=19.89, p<0.001). Analyses

of simple main effects showed the greatest difference between talkers for the *happy* sentences (F(1,11) = 25.262. p<0.001) and no significant differences for the *angry* sentences, but moderately significant differences for the remaining emotions (p<0.01 or better), with the male talker's F0 range being greater than the female's.

For *duration*, there were significant main effects of Talker (F(1,11) = 5.32, p =0.042) and Emotion (F(4, 44) = 342.68, p <0.001), and a significant interaction (F(1.9, 20.902) = 11.157, p=0.001; Greenhouse Geisser correction for sphericity). Analyses of simple main effects showed no significant differences between talkers for *angry* and *sad* sentences, but significantly longer duration for the male talker's *neutral* (F(1,11)=15.521, p =0.002) and *scared* sentences (F(1,11) = 70.991, p <0.001), and significantly longer duration for the female talker's *happy* sentences (F(1,11) = 14.286, p=0.003).

For *mean dB SPL*, there were significant main effects of Talker (F(1,11) = 47.23, p < 0.001) and Emotion (F(4, 44) = 120.76, p<0.001) and a significant interaction (F(2.51, 27.60) = 32.665, p<0.001; Greenhouse Geisser correction for sphericity). Analyses of simple main effects showed no significant differences between talkers for *neutral*, *angry* and *happy* sentences, but significantly higher mean intensities for the male talker for *sad* (F(1,11) = 130.41, p<0.001) and *scared* (F(1,11) = 163.70, p<0.001) sentences.

For *intensity range*, there was no significant main effect of Talker, but a significant effect of Emotion (F(2.28, 25.06)= 31.84, p<0.001; Greenhouse Geisser correction for sphericity) and a significant interaction (F(4, 44) = 9.334, p<0.001). Analyses of simple main effects showed no significant differences between talkers for *neutral* and *angry* sentences, but a greater intensity range of the female talker's *happy* sentences (marginally significant, F(1,11) = 4.87, p=0.05), and a greater intensity range of the male talker's *sad* (F(1,11) = 18.10, p=0.001) and *scared* (F(1,11) = 12.99, p = 0.004) sentences.

A deeper understanding of the data might be gained by considering the *discriminability* of the stimuli for different pairs of emotions. A discriminability measure (d') was defined for each cue and each pair of emotions, as the absolute value of the difference between the mean values of the cue (taken across the 12 sentences) for the two emotions, divided by their average standard deviation. A d' matrix of the pairwise discriminability of the five emotions, was constructed for each cue (see Table 3 for examples). The sum of all of the d's within the matrix for each cue was computed as a measure of the net discriminability provided by that cue. Fig. 2 plots this index for each of the four cues, for each of the two talkers. It seems that F0-based cues carry the greater weight of discriminability, at least in the limited set of stimuli in the present study. In addition, it is apparent that the male and female talkers' vocal emotions were most different in discriminability based on F0 range (female talker more distinctive), with smaller differences in intensity range (female talker more distinctive) and average intensity (male talker more distinctive) but very similar in the discriminability based on the other two cues, duration and F0 height. Thus, if the listeners achieve better performance with the male talker's utterances, this might suggest that average intensity is an important cue for voice emotion recognition. On the other hand, if the listeners achieve better performance with the female talker's utterances, this might suggest that the F0 range and intensity range contain more important information.

3.2 Acoustic analyses of 8-channel NBV stimuli

The signal processing involved in generating the NBV stimuli is expected to remove F0 cues and fine spectral detail, but retain amplitude envelope cues within the broad frequency channels. The intensity cues might change somewhat because of bandpass filtering of the original signal as well as lowpass filtering of the temporal envelope. Duration cues should not change significantly. Preliminary analyses of the NBV stimuli confirmed the absence of any discernible F0 cues. Duration cues, as also expected, remained unchanged compared to the full-spectrum stimuli.

The 8-channel NBV stimuli were analyzed in detail and compared with the full spectrum stimuli. A RMANOVA of the *duration* cue showed no significant effects of processing, as expected, and no interactions of processing with emotion or talker. Figure 3 overlays plots for mean dB SPL (upper plot) and intensity range (lower plot) for the NBV (circles) and full-spectrum (diamonds) versions of the stimuli recorded by the male (orange) and female (blue) talkers. Although the differences appear small, there were statistically significant changes in these two variables as a result of the processing. The NBV stimuli were slightly reduced in mean dB SPL relative to the original stimuli. The average reduction was 1.72 dB for the female talker (s.d. 0.33 dB) and 2.44 dB for the male talker (s.d. 1.11 dB). A RMANOVA showed significant main effects of processing (F(1,11) = 64.92, p<0.001), talker (F(1,11) = 61.28, p<0.001) and emotion (F(2.22, 24.43) = 143.83, p<0.001), and significant interactions between processing and emotion (F(2.06,22.69)=6.77, p=0.005; Greenhouse Geisser correction for sphericity) and between talker and emotion (F(2.35,25.88)=36.30, p<0.001). Analyses of simple main effects showed that for the female talker, all emotions showed significant effects of processing, with the greatest effects observed for the Sad stimuli, followed by Angry, Neutral, Happy, and Scared (ordered by F-ratio). For the male talker, *Scared* and *Happy* stimuli did not show a significant effect of processing; ordered by F-ratio, the greatest effects were observed for Neutral, followed by Angry and Sad. Although there were some statistically significant changes in the pattern, we doubt that they were related to real perceptual outcomes in the free-field stimulation used here; dB SPL values might change more than +/-4 dB with small head movements.

An RMANOVA on the *intensity range* data showed significant effects of processing (F(1,11) = 44.31, p<0.001) and emotion (F(4,44) = 30.43, p<0.001), no main effect of talker, and a significant interaction between processing and emotion (F(4,44) = 7.91, p<0.001). Analyses of simple main effects on the pooled data from the two talkers' sentences showed significant effects of processing at each emotion except for *Scared*, with the F-ratio being greatest for *Neutral*, followed by *Angry, Sad*, and *Happy*, in that order.

To summarize, the NBV stimuli showed small but significant differences in mean intensity and intensity range due to processing, more so for some emotions than for others, as reflected in the significant interactions between processing and emotion.

As with the full-spectrum stimuli, discriminability matrices were computed and the summed d' was derived for each cue. A comparison of the discriminability index for full-spectrum and NBV stimuli is shown in Fig. 4. The ordinate is on the same scale as in Fig. 2, for ease of comparison. The patterns are very similar for full-spectrum and NBV stimuli, but the

discriminability indices for both *mean intensity* and *intensity range* appear to be reduced for the female talker's utterances as a result of the processing.

3.3 Group mean scores with full-spectrum speech

Figure 5 shows mean emotion recognition scores (% correct: note that chance is at 20% correct) for full-spectrum speech obtained by each group of participants with the sentences recorded by the male and female talkers. Generally, the male talker's vocal emotions were harder to recognize; this difference was most apparent for the aCI group. A repeated-measures mixed ANOVA showed a significant main effect of talker (F(1,82) = 39.35, p<0.001) and a significant interaction between talker and subject group (F(3, 82) = 6.851, p<0.001). Pairwise comparisons showed that the cCIs' scores were significantly poorer than cNHs' (p<0.001) and aNHs' (p<0.001), but not different from aCIs' scores. No significant differences were found between cNHs' and aNHs' scores. The interaction effect should be interpreted with caution because of the ceiling effects obtained with NH listeners. The pattern of findings was not altered after transformation to rationalized arcsine units (RAUs) in an attempt to linearize the space (Studebaker, 1985).

Fig. 6 shows results obtained by aNH and cNH under full-spectrum as well as NBV conditions (cNH attended only to 8-channel NBV speech), as well as the CI listeners' mean scores for full-spectrum speech. Not surprisingly, the aNHs' performance declined as the spectral resolution worsened. The cNHs' mean score with 8-channel NBV speech was much lower than that of the aNHs' mean scores in the same condition. A mixed ANOVA on the 8-channel NBV data with talker as the within-subjects factor and subject group (cNH and aNH) as the between-subjects factor showed that the female talker's emotions were more recognizable (F(1,39) = 37.08, p<0.001) and confirmed that the aNHs' scores were significantly higher than the cNHs' scores (F(1,39) = 60.37, p<0.001); no significant interactions were found between talker and subject group.

The average cCIs' and aCIs' scores with full-spectrum speech were close to the aNHs' scores with 8-channel NBV speech (and considerably higher than cNHs' scores with 8-channel NBV speech). A mixed ANOVA with talker as the within-subjects factor and device type as the between-subjects factor, was conducted on the data obtained from the CI users to investigate any effects of device type (Cochlear Corp. or Advanced Bionics Corp.). The main effect of talker remained significant (F(1, 41) = 24.21, p< 0.001), but there was no effect of device/manufacturer and no interactions.

3.4 Effects of age, time in sound and age of implantation

Preliminary graphing indicated no clear differences in the relation between the data obtained using the male and female talkers' stimuli and the age variables, so the data were averaged across talkers prior to further analyses. The cCIs' data passed the Shapiro-Wilk normality test as well as the constant variance test, but were not significantly correlated with chronological age or age of implantation. A weak correlation (r = 0.37, p = 0.029) was observed between cCIs' scores and duration of experience with the device (i.e., the difference between the chronological age and age of implantation), with age at implantation

partialled out as a control variable. Note, however, that the significance level would not survive the Bonferroni correction for multiple comparisons (criterion α =0.025).

The cNHs' percent correct scores with full-spectrum speech failed the normality test but passed both the normality and constant variance tests after RAU transformation. The RAU-transformed data showed a significant age effect (r = 0.57, p = 0.0009). Fig. 7 shows the original (% correct) and RAU-transformed data in the full-spectrum condition, for the cNH and aNH, plotted as a function of age. It is apparent that the ceiling effect places a strong constraint on these data. The aNHs' data, as expected, showed no effects of age and was similar to the older cNHs' scores.

The cNHs' data with 8-channel NBV speech passed both normality and constant variance tests, and showed a significant correlation with age (r = 0.73, p < 0.0001). Fig. 8 shows the time course over which the gap closes between the cNHs' scores and the aNHs' scores in the 8-channel NBV condition. Percent correct (averaged across the data obtained using the two talkers' sentences) scores are plotted against chronological age of the participant. Filled circles show data obtained from the cNH, while open squares show data obtained from the aNH. The effect of age is strong and survives the Bonferroni correction for multiple correlations, suggesting that younger cNH who are naïve to NBV speech have the greatest difficulty in decoding voice emotion from NBV speech, and that this difficulty is gradually alleviated as the child develops into adulthood. Again, as expected, the aNHs' scores do not show age effects, and are similar to the scores obtained by the older cNH in this task.

3.5. Individual variability in cCIs' data

It may be of use to consider the individual cCIs' scores in light of equivalent scores obtained by aNH under different levels of spectral degradation. Fig. 9 shows the individual cCIs' scores (blue circles), plotted against the aCIs' scores (red squares). The individual data are plotted in no particular order along the abscissa. The solid horizontal lines show the mean scores obtained by aNH listeners under the four different levels of spectral degradation (indicated on the right hand ordinate). Five of the cCI scored around or below the 4-channel NBV scores obtained by aNH, while quite a number of cCI scored around or above the 16channel NBV scores obtained by aNH. Note that the aCIs' scores span a similar range of performance. This view of the data underscores the considerable range of difficulty experienced by CI listeners in voice emotion recognition, even with the exaggerated prosody of child-directed speech, and parallels the range of performance observed in speech recognition tasks (e.g., Friesen et al., 2001).

3.6 Beyond simple measures of accuracy: Confusion matrices and d's

The present study has focused on percent correct scores, which provide an overall sense of accuracy but do not allow for deeper investigation into error patterns. Future reports from our laboratory will include larger sample sizes and will be heavily focused on perceptual confusion matrices and their analyses, but we would like to leave the reader with some indication of the perceptual confusions made by the listeners in the present study. Figure 10 shows the mean confusion matrices, and under each condition of spectral resolution tested. The

cells are color-coded to represent the strength of the numerical values, but the actual values are also indicated. Note that the possible values of the cells range from 0 to 12 (12 sentences in each emotion for each talker). A visual inspection of the patterns reveals that for aNH (top four rows), the matrices become more and more diagonally dominant as spectral clarity increases. A similar qualitative change occurs for cNH, but their off-diagonal cells are much more populated for 8-channel NBV speech than observed in aNHs' confusion matrices. Consistent with their similar accuracy scores (reported in previous sections), aCIs' and cCIs' confusion matrices share commonalities in the patterns.

These qualitative observations are quantitatively supported in Fig. 11, which shows d' values based on hit rates and false alarm rates derived from the mean confusion matrices of Fig 10. Left and right panels correspond to the female and male talkers respectively. Within each panel, d' is plotted for each emotion (abscissa) and for each of the age groups and conditions. The upward pointing arrows correspond to the conditions (always with full-spectrum speech) in which the d' was infinite. The aCIs' and cCIs' mean d' values cluster around the aNHs' d' values for 8 and 16 channel NBV speech, again consistent with the patterns revealed by accuracy scores reported earlier. The cNHs' data with full-spectrum speech is excellent, but more prone to error than the aNHs' data in this condition. Future studies will focus on specific differences between the emotions, particularly with regard to the benefits achieved by increasing spectral resolution, and investigate the relationship between the perceptual confusion matrices and the confusion matrices based on acoustic features of the stimuli.

4.0 DISCUSSION

The excellent emotion recognition scores by the cNH and aNH in the full-spectrum condition confirmed that the stimuli conveyed each emotion with sufficient salience. This was confirmed by both the percent correct scores and the confusion matrices. Acoustic analyses were broadly consistent with results of Luo et al. (2007), although there were at least two points of difference between the stimuli used in their studies and the present one: first, the stimuli in Luo et al.'s study included questions which might have introduced a different pattern of acoustic cues than the statements and second, the stimuli in the present study were uttered in a child-directed manner while those in Luo et al.'s study varied more across the emotions than in Luo et al.'s study. However, the overall intensity patterns are similar, with *Anxious, Happy, and Angry* spoken loudest in Luo et al.'s study. Similarly, *Happy* was spoken with the greatest F0 range and mean F0 height in both studies.

Across all subject groups, voice emotion recognition scores obtained with the female talker's utterances were significantly better than with the male talker's utterances, both in the full-spectrum and NBV conditions. As the discriminability measure based on the acoustic analyses of full-spectrum stimuli showed that the female talker's sentences contained more information in the F0 range and the intensity range while the male talker's sentences contained more information in the mean intensity patterns, we inferred that F0 range and intensity range (which may co-vary to some extent) are likely to be the more

important cues for voice emotion in the full-spectrum condition. Acoustic analyses of the 8channel NBV stimuli showed no remaining F0 cue, no changes to duration as a result of processing and small but statistically significant changes to mean intensity and intensity range after processing. The intensity range discriminability index for the female talker actually dropped below the male talker's, but the female talker's voice emotion was still significantly more recognizable than the male talker's. This result leads us to wonder whether the pitch cue available to the listener in the temporal envelope of NBV speech still provides important information to support voice emotion recognition. The lowpass filter of the temporal envelope cut off at 160 Hz, so it would be surprising if sufficient envelope cues remained to support the task; however, note that Fu et al (2004) reported significant improvements in NH listeners' gender identification scores of 8-channel NBV speech when the envelope lowpass filter cutoff was increased from 40 Hz to 160 Hz. Schvartz and Chatterjee (2012) found similar improvements, although older NH listeners showed smaller benefits from changing temporal envelope filter cutoffs. Given that the present analyses also suggest that the female talker's intensity cues did not exceed the male talker's for the NBV stimuli, we tentatively conclude that the acoustic analyses presented here do not account for all of the perceptual data.

The cCIs' mean score in the task was as good as the aCIs', and the mean scores of both groups approximated the aNHs' scores with 8-channel NBV speech. The best-performing individuals in both groups, thus, exceeded the 8-channel NBV performance by aNH. This is illustrated in Fig. 9. It is apparent that the best-performing cCI can equal or exceed the aNHs' average performance with 16-channel NBV speech, and that some of these best-performing children are among the youngest. This finding speaks to the considerable benefit received by cCI from their devices. The similarity between the average performance of the cCI and the post-lingually deaf aCI in this task also suggests that, as far as the task in the present study is concerned, the cCI have, on average, been able to overcome any limitations imposed by their severe hearing loss in early childhood. The large variability in cCIs' data was similar to that observed in the aCIs' data, with the poorest performers, while still scoring above chance, fell below the mean aNHs' scores with 4-channel NBV speech. The scope of the present study was unfortunately too limited to allow a full understanding of the issues underlying this variability. The considerable range of performance underscores the need for more effective intervention in this population.

The cNHs' results highlight two important points. First, these data confirm that younger children have greater difficulty processing spectrally degraded speech. This has been shown in various studies of speech perception under conditions of noise-vocoding, sinewave-speech, reverberation, and background noise. The present study adds voice emotion recognition to this growing body of literature. We speculate that the task in the present study invokes mechanisms that extract speech information as well as those that extract voice emotion from the degraded sensory input. Thus, even though the task did not require the children to understand the sentences, we speculate that obligatory speech perception mechanisms enter into play whenever stimuli are speech-like, and that these processes take away significant cognitive resources from the task of emotion recognition, particularly when the input lacks some of the salient features that are critical for both tasks and has lost some of its redundancy. The still-developing auditory system and brain of younger listeners may

have greater difficulty with the task under these conditions. The second point highlighted by these data is that the cCI appear to be able to perform remarkably well in comparison, again speaking to the considerable benefit rendered by their device.

One issue to consider in comparing results obtained with CI simulations with results obtained with actual CI users, is that simulations have inherent limitations. For instance, the NBV simulations in the present study did not simulate the basally shifted spectral patterns normally presented to CI listeners. The literature suggests strong adaptation to these shifts by CI patients; such adaptation might alter speech processing mechanisms by the auditory system, particularly pitch mechanisms (Reiss et al., 2014), and would be difficult to replicate in the laboratory with NH listeners. Specific details of actual speech processing strategies used by patients' devices, were also ignored in the simulations, which provided a generic version of the "continuous interleaved sampling" strategy first described by Wilson et al (1991). The purpose of the simulation was only to provide a broad comparison to CI patients' performance and to results of previous studies. It is to be further noted that NBV simulations have proved remarkably successful in predicting both levels and patterns of performance by CI patients, both in speech perception tasks and in voice-pitch-related tasks (Friesen et al., 2001; Fu and Shannon, 1999; Fu et al., 2004; Baskent and Shannon, 2004; Peng et al., 2009; Peng et al., 2012). It will be of interest to observe the results of future attempts at more closely simulating specifics of processing strategies and the information transfer at the electrode-neuron interface.

As indicated by the acoustic analyses of the stimuli, voice emotion contrasts include changes in multiple dimensions of the input, and the listener's task is to detect that unique combination of the multi-dimensional patterns that signal the target emotion. The input, therefore, contains considerable redundancy, but much of this is lost when the signal is spectrally degraded. This study suggests that, with the rather exaggerated prosody of childdirected speech, children with cochlear implants can achieve remarkable success in voice emotion recognition; however, a significant proportion did not fare so well, with scores falling somewhere between the levels of performance achieved by aNH listening to 4- and 8-channel NBV speech. Five of the cCI achieved scores even lower than the mean 4-channel NBV scores obtained by aNH. Natural speech signals in everyday life present greater challenges, occurring in noisy and reverberant environments, often spoken rapidly or with reduced prosodic cues, and with different dialects and accents. Although facial expressions provide useful information in difficult listening conditions, the voice emotion content of speakers who are not directly facing the listener plays an important role in social communication and incidental learning, particularly in the developing years. Thus, greater attention clearly needs to be paid to improving the transmission and perception of emotional prosody, both in rehabilitative efforts and in device/processor development.

Finally, we note three caveats regarding the present study. First, the acoustic analyses presented here did not consider other candidate cues for vocal emotion, such as the spectral centroid. Second, the stimulus set was limited, including only two talkers and only the exaggerated prosody of child-directed speech. Finally, neither the cNH nor the aNH were given active training or practice with the task and stimuli, thus precluding the study of training effects. Planned, future studies in our laboratory will include more comprehensive

acoustic analyses, a larger database of stimuli, including multiple talkers and both adult- and child-directed speech materials, and investigations of training effects in both children and adults.

Acknowledgments

This work was funded by NIH grants R21 DC011905 and R01 DC014233 (PI: MC), T35 DC008757 (PI: Michael Gorga), and P30 DC 004662 (PI: Michael Gorga). We thank Dr. Qian-Jie Fu and the Emily Shannon Fu Foundation for the software used to control the voice emotion recognition test, and Dr. Sandra Gordon-Salant for the laboratory facilities at the University of Maryland used to record the stimuli. We are grateful to the children who participated in these experiments, and their families, for their support of this research.

References

- Baskent D, Shannon RV. Frequency-place compression and expansion in cochlear implant listeners. J Acoust Soc Am. 2004; 116(5):3130–3140. [PubMed: 15603158]
- Boersma P. Praat, a system for doing phonetics by computer. Glot International. 2001; 5(9/10):341–345.
- Boersma, P.; Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 5.3.56. 2014. retrieved 20 September 2013 from http://www.praat.org/
- Chatterjee M, Oberzut C. Detection and rate discrimination of amplitude modulation in electric hearing. J Acoust Soc Am. 2011; 130(3):1567–1580. [PubMed: 21895095]
- Chatterjee M, Peng SC. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. Hearing Research. 2008; 235:143–156. [PubMed: 18093766]
- Chatterjee M, Shannon RV. Forward masked excitation patterns in multielectrode electrical stimulation. J Acoust Soc Am. 1998; 103(5):2565–2572. [PubMed: 9604350]
- Ciocca V, Francis AL, Aisha R, Wong L. The perception of Cantonese lexical tones by early-deafened cochlear implantees. J Acoust Soc Am. 2002; 111:2250–2256. [PubMed: 12051445]
- Deroche ML, Zion DJ, Schurman JR, Chatterjee M. Sensitivity of school-aged children to pitch-related cues. Journal of Acoustical Society of America. 2012; 131:2938–2947.
- Dyck MJ, Denver E. Can the emotion recognition ability of deaf children be enhanced? A pilot study. Journal of Deaf Studies and Deaf Education. 2003; 8:348–356. [PubMed: 15448058]
- Dyck MJ, Farrugia C, Shochet IM, Holmes-Brown M. Emotion recognition/understanding ability in hearing or vision-impaired children: do sounds, sights, or words make the difference? Journal of Child Psychology and Psychiatry. 2004; 45:789–800. [PubMed: 15056310]
- Eisenberg LS, Shannon RV, Martinez AS, Wygonski J, Boothroyd A. Speech recognition with reduced spectral cues as a function of age. Journal of Acoustical Society of America. 2001; 107:2704– 2710.
- Eisenberg N, Spinrad TL, Eggum NM. Emotion-related self regulation and its relation to children's maladjustment. Annual Review of Clinical Psychology. 2010; 6:495–525.
- 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. Journal of Acoustical Society of America. 2001; 110:1150–1163.
- Fu QJ, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. Journal of the Acoustical Society of America. 1999; 105(3):1889–1900. [PubMed: 10089611]
- Fu QJ, Chinchilla S, Galvin JJ. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. J Assoc Res Otolaryngol. 2004; 5(3):253– 260. [PubMed: 15492884]
- 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 and Hearing. 2013; 34:562–574. [PubMed: 23652814]

- 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 Neuropsychology. 2009; 15:136–146. [PubMed: 18828045]
- Ketelaar L, Rieffe C, Wiefferink CH, Frijns JHM. Social competence and empathy in young children with cochlear implants and normal hearing. Laryngoscope. 2012; 123:518–523. [PubMed: 22996702]
- Koehlinger K, Owen Van Horne AJ, Moeller MP. Grammatical outcomes of 3 & 6 year old children with mild to severe hearing loss. Journal of Speech, Language, and Hearing Research. 2013; 56:1701–1714.
- Kong YY, Cruz R, Jones JA, Zeng FG. Music perception with temporal cues in acoustic and electric hearing. Ear & Hearing. 2004; 25(2):173–185. [PubMed: 15064662]
- Kral A, Hartmann R, Mortazavi D, Klinke R. Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents. Hearing Research. 1998; 121:11–28. [PubMed: 9682804]
- Lowenstein JH, Nittrouer S, Tarr E. Children weight dynamic spectral structure more than adults: evidence from equivalent signals. J Acoust Soc Am. 2012; 132(6):EL443–449. [PubMed: 23231206]
- Luo X, Fu Q-J, Galvin J. Vocal emotion recognition by normal-hearing listeners and cochlear-implant users. Trends and Amplification. 2007; 11:301–315.
- Luo X, Fu Q-J. Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. J Acoust Soc Am. 2004; 116:3659–3667. [PubMed: 15658716]
- Most T, Aviner C. Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing AIDS, and normal hearing. Journal of Deaf Studies and Deaf Education. 2009; 14:449–464. [PubMed: 19398533]
- Nakata T, Trehub SE, Kanda Y. Effect of cochlear implants on children's perception and production of speech prosody. Journal of Acoustical Society of America. 2012; 131:1307–1314.
- Nittrouer S, Lowenstein JH. Separating the effects of acoustic and phonetic factors in linguistic processing with impoverished signals by adults and children. Applied Psycholinguistics. 2014; 35:333–370. [PubMed: 24729642]
- Nittrouer S, Lowenstein JH, Packer RR. Children discover the spectral skeletons in their native language before the amplitude envelopes. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35:1245–1253. [PubMed: 19653762]
- Peng SC, Chatterjee M, Lu N. Acoustic cue integration in speech intonation recognition with cochlear implants. Trends and Amplification. 2012; 16:67–82.
- Peng SC, Lu N, Chatterjee M. Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. Audiology Neuro-Otology. 2009; 14:327–337. [PubMed: 19372651]
- Peng SC, Tomblin JB, Turner CW. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. Ear and Hearing. 2008; 29:336– 351. [PubMed: 18344873]
- Peng SC, Tomblin JB, Cheung H, Lin YS, Wang LS. Perception and production of Mandarin tones in prelingually deaf children with cochlear implants. Ear and Hearing. 2004; 25:251–264. [PubMed: 15179116]
- Reiss LA, Turner CW, Karsten SA, Gantz BJ. Plasticity in human pitch perception induced by tonotopically mismatched electro-acoustic stimulation. Neuroscience. 2014; 256:43–52. [PubMed: 24157931]
- Sauter DA, Panattoni C, Happé F. Children's recognition of emotions from vocal cues. British Journal of Developmental Psychology. 2013; 31:97–113. [PubMed: 23331109]
- Schorr EA, Roth FP, Fox NA. Quality of life for children with cochlear implants: perceived benefits and problems and the perception of single words and emotional sounds. Journal of Speech, Language, and Hearing Research. 2009; 52:141–152.
- Schvartz KC, Chatterjee M. Gender identification in younger and older adults: use of spectral and temporal cues. Ear & Hearing. 2012; 33(3):411–420. [PubMed: 22237163]
- Shannon RV. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. Hearing Research. 1983; 11:157–189. [PubMed: 6619003]

- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]
- Studebaker GA. A "rationalized" arcsine transform. Journal of Speech Hearing Research. 1985; 28:455–462. [PubMed: 4046587]
- Tomblin JB, Barker BA, Hubbs S. Developmental constraints on language development in children with cochlear implants. International Journal of Audiology. 2007; 46:512–523. [PubMed: 17828667]
- Tonks J, Williams WH, Frampton I, Yates P, Slater A. Assessing emotion recognition in 9–15 year olds: Preliminary analysis of abilities in reading emotion from faces, voices, and eyes. Brain Injury. 2007; 21:623–629. [PubMed: 17577713]
- Volkova A, Trehub SE, Schellenberg EG, Papsin BC, Gordon KA. Children with bilateral cochlear implants identify emotion in speech and music. Cochlear Implants International. 2013; 14:80–91. [PubMed: 23453221]
- Wang DJ, Trehub SE, Volkova A, Van Lieshout P. Child implant users' imitation of happy- and sadsounding speech. Frontiers in Psychology. 2013; 4:351. [PubMed: 23801976]
- Wei CG, Cao K, Zeng FG. Mandarin tone recognition in cochlear-implant subjects. Hearing Res. 2004; 197:87–95.
- Wiefferink CH, Rieffe C, Ketelaar L, De Raeve L, Frijns JHM. Emotion understanding in deaf children with a cochlear implant. Journal of Deaf Studies and Deaf Education. 2013; 18:175–186. [PubMed: 23232770]
- Wang DJ, Trehub SE, Volkova A, van Lieshout P. Child implant users' imitation of happy- and sadsounding speech. Front Psychol. 2013; 4(351)10.3389/fpsyg.2013.00351
- Winn MB, Chatterjee M, Idsardi WJ. Roles of voice onset time and F0 in stop consonant voicing perception: effects of masking noise and low-pass filtering. Journal of Speech, Language, and Hearing Research. 2013; 56:1097–1107.
- Winn MB, Chatterjee M, Idsardi WJ. The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. Journal of Acoustical Society of America. 2012; 131:1465–1479.
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. Nature. 1991; 352:236–238. [PubMed: 1857418]
- Xu L, Zhou N, Chen X, Li Y, Schultz HM, Zhao X, Han D. Vocal singing by prelingually-deafened children with cochlear implants. Hearing Research. 2009; 255:129–134. [PubMed: 19560528]
- Zeng FG. Temporal pitch in electric hearing. Hearing Research. 2002; 174:101–106. [PubMed: 12433401]

- Voice emotion recognition was investigated in children with normal hearing, children with cochlear implants (CIs), adults with normal hearing and adults with CIs. Preliminary acoustic analyses of the stimuli and perceptual confusion matrices are presented alongside the percent correct scores.
- Children with cochlear implants (CIs) showed deficits in voice emotion recognition relative to their normally hearing peers, but they were no worse at the task than adult listeners with CIs.
- Some CI children achieved excellent performance in the task
- Children with normal hearing (NH) were as good at the task as NH adults
- However, younger NH children had difficulty in the task with CI-simulated speech





Results of acoustic analyses of the male (circles) and female (squares) talker's utterances, plotted for each of the five emotions (abscissa). Each panel corresponds to a different acoustic cue. Error bars show +/-1 s.d. from the mean.



Fig. 2.

Summed discriminability indices for the different cues (abscissa) plotted for the male (orange) and female (blue) talkers, and for full-spectrum stimuli. Results are shown for Mean Intensity (Int.), Duration (Dur.), F0 height (F0 ht.), F0 range (F0 rng.), and Intensity Range (Int. Rng.).





Results of acoustic analyses for the cues of Mean Intensity (dB SPL) (top) and Intensity Range (bottom), compared for NBV (circles) and full spectrum (diamonds) stimuli and for the male (orange) and female (blue) talker, respectively.



Fig. 4.

Summed discriminability indices for Mean Intensity (Int.), Duration (Dur.), and Intensity Range (Int. Rng.), for full spectrum (solid bars) and NBV (hatched bars) stimuli, and for the male and female talkers (orange and blue, respectively).



Fig. 5.

Mean emotion recognition scores with full spectrum stimuli for the four subject groups, for the male (orange) and female (blue) talkers, respectively. Error bars show +/-1 s.d.. The solid horizontal line shows chance performance.



Fig. 6.

Mean emotion recognition scores plotted against the spectral resolution condition, for the four subject groups. Note that aNH were tested under all conditions; cCI and aCI were tested only in the full-spectrum condition, and cNH were tested in full-spectrum and 8-channel NBV conditions. Error bars show +/-1 s.d. from the mean. Left and right hand panels show results obtained with the female and male talker, respectively.



Fig. 7.

RAU transformed scores (filled symbols) and percent correct scores (open symbols) plotted against age, for cNH (circles) and aNH (squares) and listening to full-spectrum stimuli. The solid line shows the regression line through the RAU-transformed data for the cNH only (r and p values are also indicated).



Fig. 8.

Percent correct scores plotted against age, for cNH (filled symbols) and aNH (open symbols) listening to 8-channel NBV stimuli. The regression line was plotted through the data obtained from cNH only (r and p values indicated).

		Ma	le Tal	lker				Fer	nale 1	alker			
		ANGRY	HAPPY	NEUTRAL	SAD	SCARED		ANGRY	HAPPY	NEUTRAL	SAD		SCARED
	ANGRY	3.27	1.23	5.32	2.0	0 0.18	ANGRY	7.32	1.18	2.09		1.18	0.23
aNH	HAPPY	2.14	5.23	1.59	0.3	2 2.73	HAPPY	3.86	6.14	0.95		0.18	0.86
4 611	NEUTRAL	0.09	0.05	8.64	3.0	0.18	NEUTRAL	0.27	0.95	8.82		1.55	0.41
4-CH	SAD	2.18	0.82	2.50	6.2	.7 0.23	SAD	0.45	0.41	3.27		7.86	0.00
	SCARED	4.50	1.59	0.23	0.0	5.68	SCARED	1.68	0.95	1.14		0.36	7.86
	111001	ANGRY	HAPPY	NEUTRAL	SAD	SCARED	111001	ANGRY	НАРРҮ	NEUTRAL	SAD		SCARED
aNH	ANGRY	5.50	0.32	4.09	1.9	0.18	ANGRY	9.36	0.32	1.//		0.41	0.14
anni	HAPPY	0.55	8.73	10.73	0.0	1.95	HAPPY	0.91	10.73	0.14		1.69	0.23
8-CH	SAD	2.77	0.05	1 77	6.0	0.09	CAD	0.00	0.50	3.77		1.00	0.10
	SCARED	2.77	2 73	0.23	0.9	6.86	SCARED	1.05	0.09	1 18		0.04	9.23
	JUNILU	6.6.5	2.75	0.23	0.0	0.00	JCHILD	1.05	0.55	1.10		0.00	5.25
		ANGRY	HAPPY	NEUTRAL	SAD	SCARED		ANGRY	HAPPY	NEUTRAL	SAD		SCARED
	ANGRY	6.68	0.00	4.14	1.0	9 0.09	ANGRY	9.91	0.14	1.41		0.50	0.05
aNH	HAPPY	0.09	10.00	0.14	0.0	0 1.77	HAPPY	0.09	11.73	0.05		0.00	0.14
16 CH	NEUTRAL	0.00	0.00	10.73	1.2	7 0.00	NEUTRAL	0.05	0.05	10.73		1.14	0.05
10-CH	SAD	2.00	0.36	2.82	6.7	7 0.05	SAD	0.64	0.00	1.59		9.73	0.05
	SCARED	2.05	1.86	0.27	0.0	0 7.82	SCARED	0.55	0.32	0.55		0.00	10.59
		ANICOV	LIADOV		CAD	CCADED		ANCOV	LIADDV	NEUTOAL	CAD		COADED
	ANGRY	11 64	0.00	0.26	340	JCARED	ANGRY	12.00	0.00	0.00	SAD	0.00	0.00
aNH	LADDY	0.00	11.72	0.30	0.0	0.00	LADDY	0.00	12.00	0.00		0.00	0.00
	NELITRAL	0.00	0.00	11.95	0.0	5 0.00	NELITRAL	0.05	0.00	11.91	1	0.05	0.00
Full	SAD	0.00	0.00	0.00	12.0	0.00	SAD	0.00	0.00	0.00		12.00	0.00
	SCARED	0.05	0.91	0.05	0.0	11.00	SCARED	0.00	0.09	0.05	_	0.00	11.86
												0.00001	
		ANGRY	HAPPY	NEUTRAL	SAD	SCARED		ANGRY	HAPPY	NEUTRAL	SAD		SCARED
CNH	ANGRY	2.52	1.29	5.03	2.4	2 0.74	ANGRY	6.56	1.45	2.52		0.98	0.48
CIVIT	HAPPY	3.18	3.06	2.15	0.6	3 2.94	HAPPY	3.65	3.90	1.55		0.52	2.39
8-CH	NEUTRAL	0.47	0.94	6.98	3.2	0.32	NEUTRAL	0.53	1.65	6.52		2.79	0.52
0 011	SAD	2.26	1.24	3.27	4.3	5 0.87	SAD	1.15	0.90	3.79		5.48	0.68
	SUARED	4.61	1.60	0.65	0.2	4.92	SCARED	1.26	1.24	1.48		0.35	7.00
		ANGRY	HAPPY	NEUTRAL	SAD	SCARED		ANGRY	HAPPY	NEUTRAL	SAD		SCARED
	ANGRY	11.45	0.00	0.48	0.0	3 0.03	ANGRY	11.84	0.00	0.11		0.02	0.03
CNH	HAPPY	0.02	11.50	0.32	0.0	0.15	HAPPY	0.00	11.95	0.03		0.02	0.00
Eull	NEUTRAL	0.24	0.02	11.71	0.0	0.00	NEUTRAL	0.03	0.03	11.53		0.39	0.02
Full	SAD	0.02	0.02	0.15	11.4	4 0.39	SAD	0.00	0.02	0.24		11.55	0.19
	SCARED	0.02	1.03	0.15	0.2	9 10.52	SCARED	0.03	0.29	0.19		0.15	11.34
		ANCOV	LIADDV	NEUTOAL	CAD	CARED		ANCOV	LIADDY	NEUTOAL	CAD		CARED
	ANGRY	7.00	0.61	2 20	340	a 0.22	ANCRY	ANGRT	0.17	1 20	340	0.50	O 11
aCl	HADDY	1 11	6.92	0.22	0.0	2 2 20	HADDY	0.72	0.17	0.11		0.00	1 20
	NEUTRAL	0.06	0.56	10.83	0.5	0 0.06	NELITRAL	0.28	0.33	10.22		0.89	0.28
Full	SAD	1.22	0.39	1.72	8.6	0.06	SAD	0.28	0.22	1.33		10.06	0.11
	SCARED	1.56	2.00	0.22	0.1	1 8.11	SCARED	0.67	0.83	0.50	_	0.00	10.00
	a al article									0.00			10100
		ANGRY	HAPPY	NEUTRAL	SAD	SCARED		ANGRY	HAPPY	NEUTRAL	SAD		SCARED
cCl	ANGRY	7.50	0.38	3.35	0.4	9 0.29	ANGRY	9.15	0.31	1.44		0.81	0.29
CCI	HAPPY	0.31	8.31	0.88	0.4	2 2.10	HAPPY	0.22	10.90	0.28		0.11	0.49
Full	NEUTRAL	1.31	0.33	8.99	1.1	5 0.22	NEUTRAL	0.21	0.33	8.36		2.88	0.22
i un	SAD	0.86	0.32	1.78	8.6	0.39	SAD	0.28	0.13	1.38		9.92	0.31
	SCARED	0.72	2.35	0.46	0.7	8 7.69	SCARED	0.15	1.38	1.01		0.56	8.90

Fig. 9.

Percent correct scores plotted against age, for cCI (blue circles) and aCI (red squares) listening to full-spectrum stimuli. The individual data are not plotted in any particular order along the abscissa. Solid horizontal lines indicate aNHs' mean scores under different conditions of spectral resolution (no. of channels, shown on the right hand ordinate), for comparison.



Fig. 10.

Mean confusion matrices obtained with stimuli recorded by the male (left panel) and female (right panel) talkers, and for the different listener groups and different conditions of spectral resolution (top to bottom). Confusion matrices are presented with the stimuli organized vertically and the response categories organized horizontally. Each cell shows the number of responses for that particular stimulus and response combination: the range is from 0 (white) to 12 (darkest green).



Fig. 11.

Values of d' calculated for each of the confusion matrices shown in Fig. 10, plotted against the corresponding emotion. Left and right panels show results obtained with sentences recorded by the male and female talker, respectively. Within each panel, the different symbols show different levels of spectral resolution (e.g., squares represent the full spectrum condition). The different colors show results obtained with different subject groups.

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 1

oss).
ng l
earii
ц =
HL
aids;
aring
: he
∃S =
~
(H/
group. (H ₁
ICI group. (H/
he aCI group. (H/
bout the aCI group. (H,
on about the aCI group. (H ₁
mation about the aCI group. (H ₁
information about the aCI group. (H ₁
led information about the aCI group. (H ₁
etailed information about the aCI group. (H ₁

Participant	Age at testing (years)	Gender	Age of Implantation (years)	Duration of device use (years)	Manufacturer/Device	Pre/Postlingual deafness	Notes
BT_N3	27.34	Male	18.10	9.24	Cochlear Ltd./Nucleus 24	Prelingual (bilateral hearing aids until implantation)	Diagnosed at 8 mos: bilateral HAs until implantation
BT_N4	51.79	Female	40.64	11.15	Cochlear Ltd./Nucleus 24R(CS)	Postlingual: early onset hearing loss	HL started at age 2, possibly ototoxicity- related; HAs from age 2.5 years until age 38 years.
BT_N5	53.53	Female	50.75	2.78	Cochlear Ltd./Nucleus CI512	Postlingual: adult onset hearing loss	Sudden HL at age 49 years. Unknown cause.
BT_N6	55.09	Male	44.11	10.98	Cochlear Ltd./Nucleus 24R(CS)	Postlingual: adult onset hearing loss	Progressive HL of unknown cause started age 29. Possibly noise exposure
BT_N7	58.61	Female	48.76	9.86	Cochlear Ltd./Nucleus Freedom	Postlingual: possibly early onset hearing loss	Diagnosed with HL at age 19, suspected undiagnosed HL earlier in life.
BT_N11	50.65	Male	42.51	8.14	Cochlear Ltd./Freedom	Postlingual: early onset hearing loss	Diagnosed with HL at 18 months, but no intervention possible; HAs at age 26.
BT_N12	69.82	Male	66.01	3.81	Cochlear Ltd/CI 512	Postlingual: possibly early onset hearing loss	Possibly some undiagnosed HL as a child; progressive; HAs for 20 years prior to implantation
BT_C_01	36.94	Female	30.97	5.97	Advanced Bionics/Clarion 90K 1J	Postlingual: early onset hearing loss	HL onset at age 9 (meningitis); residual hearing in RE. Bilateral HAs until implantation.
$BT_{-}C_{-}03$	65.75	Male	55.62	10.13	Advanced Bionics/Clarion C II	Postlingual	Progressive HL onset after age 20. Suspected noise exposure/family history.

Chatterjee et al.

TABLE 2

List of sentences

Item #	Sentence (6 syllables each)
1	Her coat is on the chair.
2	The road goes up the hill.
3	They're going out tonight.
4	He wore his yellow shirt.
5	They took some food outside.
6	The truck drove up the road.
7	The tall man tied his shoes.
8	The mailman shut the gate.
9	The lady wore a coat.
10	The chicken laid some eggs.
11	A fish swam in the pond.
12	Snow falls in the winter.

Table 3

Examples of discriminability matrices for mean dB SPL (top) and F0 range (bottom), computed based on analyses of the female talker's utterances.

	ď c	discriminab	ility matrix: m	ean dB SP	L, female tall	ker	
	ANGRY	HAPPY	NEUTRAL	SAD	SCARED		SUM (ABS)
ANGRY	0	2.1555	2.3086	1.8752	0.7385		7.0778
НАРРҮ		0	4.0699	3.6100	1.2309		8.9108
NEUTRAL			0	0.3046	2.7046		3.0092
SAD				0	2.3086		2.3086
SCARED					0		0
						TOTAL	21.3064
	•	d' discrimin.	ability matrix:	F0 range,	female talkeı	L	

		ł' discrimin	ability matrix:	F0 range,	female talkeı		
	ANGRY	Addyh	NEUTRAL	SAD	SCARED		(SAA) MUS
ANGRY	0	8.7219	1.6826	1.4301	0.5845		12.4191
Addah		0	10.9195	7.9976	9.0907		28.0078
NEUTRAL			0	3.5465	0.3966		3.9431
GAD				0	2.0331		2.0331
SCARED					0		0
						TOTAL	46.4031