Title
Perception of Child-Directed Versus Adult-Directed Emotional Speech in Pediatric Cochlear Implant Users.

Journal
Ear & Hearing, 41(5)

ISSN
0196-0202

Authors
Barrett, Karen Chan
Chatterjee, Monita
Caldwell, Meredith T
et al.

Publication Date
2020

DOI
10.1097/aud.0000000000000862

Peer reviewed

HHS Public Access
Author manuscript
Ear Hear. Author manuscript; available in PMC 2021 September 01.

Published in final edited form as:
Ear Hear. 2020 ; 41(5): 1372–1382. doi:10.1097/AUD.0000000000000862.

Perception of Child-Directed vs. Adult-Directed Emotional Speech in Pediatric Cochlear Implant Users

Karen Chan Barrett[1], **Monita Chatterjee**[2], **Meredith Caldwell**[3], **Mickael L. D. Deroche**[4], **Patpong Jiradejvong**[1], **Aditya M. Kulkarni**[2], **Charles J. Limb**[1]

[1]Department of Otolaryngology–Head & Neck Surgery, University of California, San Francisco, San Francisco, CA, USA

[2]Auditory Prostheses and Perception Lab, Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, USA

[3]Vanderbilt University Medical Center, Nashville, TN, USA

[4]Department of Psychology, Concordia University, Montreal, Quebec, Canada

Abstract

Objectives: Cochlear implants (CIs) are remarkable in allowing individuals with severe-to-profound hearing loss to perceive speech. Despite these gains in speech understanding, however, CI users often struggle to perceive elements such as vocal emotion and prosody, as CIs are unable to transmit the spectro-temporal detail needed to decode affective cues. This issue becomes particularly important for children with CIs, but little is known about their emotional development. In a previous study, pediatric CI users showed deficits in voice emotion recognition with child-directed stimuli featuring exaggerated prosody. However, the large intersubject variability and differential developmental trajectory known in this population incited us to question the extent to which exaggerated prosody would facilitate performance in this task. Thus, we revisited the question with both adult-directed and child-directed stimuli.

Design: Vocal emotion recognition was measured using both child-directed (CDS) and adult-directed (ADS) speech conditions. Pediatric CI users, aged 7-19 years old, with no cognitive or visual impairments and who communicated through oral communication with English as the primary language participated in the experiment (n = 27). Stimuli comprised twelve sentences selected from the HINT database. The sentences were spoken by male and female talkers in a CDS or ADS manner, in each of the five target emotions (*happy, sad, neutral, scared, angry*). The chosen sentences were semantically emotion-neutral. Percent correct emotion recognition scores were analyzed for each participant in each condition (CDS vs. ADS). Children also completed cognitive tests of nonverbal IQ and receptive vocabulary, while parents completed questionnaires of CI and hearing history. It was predicted that the reduced prosodic variations found in the ADS condition would result in lower vocal emotion recognition scores compared to the CDS condition.

Author Contribution Statement: KCB tested participants, analyzed the data, and co-wrote the manuscript; MC co-designed the experiment and co-analyzed the data; MC co-designed the experiments and tested participants; PJ helped to set up the user interfaces for data collection at UCSF; MD analyzed dynamic pitch threshold data used in multiple regression analyses, AK set up the initial user interfaces for testing; CJL oversaw the entire project, helped with recruitment, and co-wrote the manuscript. All authors discussed the results and implications and commented on the manuscript.

Moreover, it was hypothesized that cognitive factors, perceptual sensitivity to complex pitch changes, and elements of each child's hearing history may serve as predictors of performance on vocal emotion recognition.

Results: Consistent with our hypothesis, pediatric CI users scored higher on CDS compared to ADS speech stimuli, suggesting that speaking with an exaggerated prosody – akin to "motherese" – may be a viable way to convey emotional content. Significant talker effects were also observed in that higher scores were found for the female talker for both conditions. Multiple regression analysis showed that nonverbal IQ was a significant predictor of CDS emotion recognition scores while Years using CI was a significant predictor of ADS scores. Confusion matrix analyses revealed a dependence of results on specific emotions; for the CDS condition's female talker, participants had high sensitivity (*d'* scores) to *happy* and low sensitivity to the *neutral* sentences while for the ADS condition, low sensitivity was found for the *scared* sentences.

Conclusions: In general, participants had higher vocal emotion recognition to the CDS condition which also had more variability in pitch and intensity and thus more exaggerated prosody, in comparison to the ADS condition. Results suggest that pediatric CI users struggle with vocal emotion perception in general, particularly to adult-directed speech. We believe these results have broad implications for understanding how CI users perceive emotions both from an auditory communication standpoint and a socio-developmental perspective.

INTRODUCTION

Cochlear Implants (CI) are surgically-inserted prosthetic devices that provide sound and enable speech perception in deaf individuals with severe hearing loss. Despite the impressive ability of CI users to understand speech in quiet environments, CI users struggle with complex acoustic stimuli and environments due to the limited spectral-temporal detail provided by the implants. For example, in comparison to normal hearing (NH) individuals, adult CI users struggle to perceive pitch-dominant aspects of speech such as prosody (e.g. question/statement contrasts, nonverbal vocal expressions), vocal emotion, and lexical tone recognition (Chatterjee & Peng, 2008; Chatterjee et al., 2015; Ciocca et al., 2002; Luo & Fu, 2004; Luo et al., 2007; Most & Aviner, 2009; Paquette et al., 2018; Peng et al., 2004; Wei et al., 2004). CI user deficits are also seen in aspects of music comprehension such as melody and pitch perception *(*Gfeller et al., 2002; Kong et al., 2004; Limb & Roy, 2014; McDermott, 2004), and aspects of musical emotions at the neurophysiological level (Deroche et al., 2019)

The ability to comprehend emotion, in particular, is vital for human communication as it informs a listener about a speaker's intent and affect. Emotion can be conveyed by both visual and auditory cues. For vocal emotion, acoustic elements like tempo, pitch (e.g. the fundamental frequency or F0 of a speaker), and intensity are powerful cues of emotion (Banse & Scherer, 1996; Jiam et al., 2017; Murray & Arnott, 1993). In NH children, accurate emotion perception and self-regulation are important for positive social development (Eisenberg et al., 2010). In child CI users, research has found that a lower self-reported quality of life is related to poorer performance in vocal emotion recognition (Schorr et al., 2009). Likewise, researchers have found that high emotion recognition scores (i.e. better perception of indexical components of speech) in elementary-aged CI children are associated with higher language levels and better developed social skills, although it is unclear whether there is a causal relationship between the two (Geers et al., 2013). Vocal emotion is therefore a critical aspect of social cognition.

Several studies have shown that vocal emotion production and perception in pediatric CI users are generally poorer than their hearing peers. These abilities are typically assessed by asking experimental participants to speak utterances while expressing a particular emotion (i.e. production) or by listening to semantically neutral sentences spoken with a particular target emotion and having participants identify the emotion expressed (i.e. perception). In studies of CI children compared to matched NH controls, CI children have shown both poorer voice emotion production and discrimination (Nakata et al., 2012; Wang et al., 2013). Other experiments have likewise found that pediatric CI users perform worse than matched NH individuals on emotion recognition in both speech and music (Volkova et al., 2013). When presented with emotion perception tests presented in various modalities such as with auditory-only, visual-only, or auditory-visual cues, CI participants performed worse on emotion perception than the NH controls (Most & Aviner, 2009; Wiefferink et al. 2013). In one study, it was found that school-aged CI users performed worse than matched NH controls on emotion recognition in speech, but performed similarly on identifying emotion in faces (i.e. with visual cues), suggesting that CI children do not struggle to identify emotion overall, but have specific difficulties with affective prosody in speech (Hopyan-Misakyan et al., 2009). Moreover, deficits in CI users are also seen in more sophisticated assessments of emotion, such as understanding ambivalent emotional responses to a given situation, understanding moral rules, and knowing how and when to use psychological emotion regulation strategies (Mancini et al., 2016).

It has been posited that speech featuring more exaggerated prosody and larger acoustic variations may be easier for children with hearing loss to process (Cannon & Chatterjee, 2019). Indeed, this child-directed speech (CDS) has been seen to be more effective in maintaining the attention of young listeners (Dominey & Dodane, 2004). The greater variation of acoustic cues in CDS highlight linguistic and prosodic elements of speech. Earlier work from our group found that CI users performed worse on emotion recognition of CDS compared to both adult and child NH listeners, with individual data showing a considerable range of difficulty on this task, even with the inflated prosody of CDS stimuli (Chatterjee et al., 2015). However, listeners are not always confronted with highly exaggerated prosodic cues in many listening situations, instead having to rely on more subtle prosodic changes such as those found in adult-directed speech (ADS). Older children in particular, are likely addressed with the more normal prosody of adult-directed speech. Therefore, it is important to understand the degree to which the deficit observed in other studies with the exaggerated prosody of CDS stimuli may be magnified when the stimuli consist of ADS stimuli.

Additionally, previous studies have found that different factors can predict participants' emotion recognition abilities. Previous studies found that when NH children listened to degraded CI simulated-speech in an emotion recognition task, a child's age and nonverbal intelligence were predictors of performance (Tinnemore et al., 2018). In children with mild to moderate hearing loss, vocabulary skills have been linked to performance on emotion recognition tasks, (Cannon & Chatterjee, 2019). These findings are consistent with a body of literature indicating that younger children struggle more with the identification of degraded speech or speech in noise than older children, and that linguistic and cognitive status accounts for intersubject variation in these outcomes in children with hearing loss and children with normal hearing attending to degraded speech (e.g. Eisenberg et al., 2000; McCreery et al., 2019). In adults with cochlear implants, cognitive status has been shown to be a predictor of speech recognition outcomes in some studies (e.g. Schvartz, Chatterjee and Gordon-Salant, 2008; O'Neill, Kreft and Oxenham, 2019) but not in others (e.g. Moberly et al., 2017). Children with cochlear implants are at risk for impaired executive function, visual and verbal working memory (e.g. Nittrouer, Caldwell-Tarr and Lowenstein, 2013; Kronenberger et al., 2014; AuBuchon, Pisoni and Kronenberger, 2015, 2019). Cognitive and linguistic skills are likely involved in the outcomes of many speech perception tasks. In school-age children with cochlear implants, the development of these skills may depend on a number of factors such as their access to speech information and intervention, parental support, etc.. We therefore hypothesized that children with CIs' performance in vocal emotion recognition would be predicted by their cognitive status, particularly when the task is more challenging (as with ADS stimuli).

In this study, we investigated both 1) how voice emotion recognition varies in pediatric CI users when listening to child-directed vs. adult-directed speech as well as 2) factors that predict vocal emotion recognition. In previous experiments by our group, emotion recognition was tested in hearing loss individuals in both a CDS and ADS condition (Cannon & Chatterjee, 2018). Acoustic analysis affirmed that across emotions, the CDS material showed greater variation in pitch and intensity than the ADS condition (Cannon & Chatterjee, 2019). Because of the greater prosodic variation observed in the CDS condition, we predicted that pediatric CI users would demonstrate better emotion recognition to the CDS condition than the ADS condition. Additionally, we explored factors that may predict performance on emotion identification. Participants completed cognitive tests and questionnaires of their hearing history, giving information about years using CI, age of implantation, and their oral communication experience. We hypothesized that perceptual sensitivity to complex pitch changes, cognitive factors such as receptive language and nonverbal IQ, as well as elements of participants' hearing history may serve as predictors of vocal emotion recognition performance.

MATERIALS AND METHODS

Participants

Twenty-seven pediatric CI users (20 females, 7 males) ages 7-19 years old (average age = 11.5 years old ± 3.6) participated in the study. For inclusion in this experiment, participants had to have significant hearing loss in both ears, have at least one CI, communicate orally, use English as the primary language, and have no visual or cognitive impairments. Participants with unilateral hearing loss were not included. Parents completed a demographic questionnaire on behalf of the pediatric participants which included information about the gender and age of the participant. Parents also completed a CI Case History questionnaire, which included information about parental education levels (later used as a metric of socioeconomic status), family history of hearing loss, and detailed information about the age of CI implantation and daily usage of the CI by the participant. The average age of implantation was at 2.76 years old ± 2.6. The mean duration of CI use was 9.15 years ± 3.6. The CI users utilized a variety of devices and processing strategies (see Table 1 in Supplementary Materials).

This study was approved by the Institutional Review Board at the University of California, San Francisco (UCSF). Subjects were recruited via flyers posted in CI centers and clinics, through referral from pediatric CI audiologists in hospitals and private clinics, as well as through hearing support groups throughout the Northern California area. Participants came to the UCSF Sound and Music Perception lab where informed consent and parental assent was obtained for all participants. All research protocols followed approved IRB guidelines. Participants were compensated for their participation in the experiment and parking costs at UCSF were reimbursed.

Test Battery

Following informed consent, participants completed a behavioral and perceptual test battery as detailed below. Participants always commenced with the cognitive tests (i.e. Peabody Picture Vocabulary test followed by the Wechsler Abbreviated Scale of Intelligence-II test) so that participants were alert and energized for the cognitive tests. Following this, participants completed the Emotion Recognition Test. The total test battery took roughly 2-3 hours with most participants completing the testing in one all-day session with multiple breaks.

Peabody Picture Vocabulary Test: The Peabody Picture Vocabulary Test (PPVT, 4th edition, Form B, Pearson, Bloomington, Indiana, USA) is a test of receptive language (Dunn & Dunn, 2007). Participants were asked to look at four images. The experimenter verbally presented a word and asked the participant to point to the picture on the page that best represented the word. Two training sentences were included prior to testing. The percentile score from the PPVT (age-normed) was used in analyses. Although the stimuli for the PPVT test (which were presented live) were not calibrated, the PPVT was always performed by the first and third authors so that 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.

Wechsler Abbreviated Scale of Intelligence-II: The Wechsler-Abbreviated Scale of Intelligence-II (WASI-II, 2nd edition, Pearson, Bloomington, IN, USA) consists of several subtests. For this experiment, the Block Design and Matrix Reasoning Subtests of the WASI II were both administered to obtain a composite measure of non-verbal IQ (NVIQ) (Wechsler & Others, 1999). For the Block Design subtest, participants were asked to replicate various patterns from the stimulus booklet using the patterned blocks. For the Matrix Reasoning subtest, the children were presented with a series of pictures and then asked to select one of five options to fill the missing space to best complete the series pattern. The overall performance intelligence (age-normed) NVIQ score resulting from these two subtests was used in analyses.

Voice Emotion Recognition: Voice emotion recognition was measured using a single interval, five-alternative, forced-choice task using a custom Matlab-based software program used in our group's previous experiments (Cannon & Chatterjee, 2019; Tinnemore et al., 2018). Twelve semantically neutral sentences taken from the HINT database (ex. "her coat is on the chair") were spoken by different male and female talkers in either a child-directed manner (CDS condition) or an adult-directed manner (ADS condition). During preparation of these stimuli, talkers were asked to think of a scenario that would elicit one of the five target emotions (*happy, sad, angry, scared, or neutral*) as if they were talking to a child as young as age 6 for the CDS condition and as if they were speaking to an adult for the ADS condition. The ADS stimuli were selected from recordings completed by a number of male and female talkers. The two talkers for the ADS and CDS condition were the male and female talker from a cohort of 10 whose recordings resulted in the highest performance on the voice emotion recognition task by a group of NH adults. In total, each speaker produced 60 stimuli (12 sentences x 5 emotions). In addition, two different sentences were recorded for the five target emotions that were used as practice materials.

The experimental software program presented the stimuli via a Microsoft Surface Pro tablet (10.6 inch ClearType HD Display) located inside a soundproof booth. The protocol was similar to that described in previous studies (Chatterjee et al., 2015). The sentences were presented to the participants in soundfield, with sound coming from a single loudspeaker (Sony SS-MB150H) located approximately 1 meter from the participants at a mean level of 65 db SPL. Participants heard one presentation of the sentence, and indicated on the screen, from a closed set, the emotion that they thought was expressed. No feedback was given during the test and no repetition of the sentences was allowed. One run consisted of 60 stimuli with target emotions and sentences being fully randomized. Prior to listening to each run, 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. For the training, each sentence/emotion would be presented and the correct emotion button would light up on the screen. Thus, training served to familiarize the listener to the talker's particular speaking style and what that talker sounded like in each condition. Within the ADS condition, participants performed two runs with the male talker's stimuli and two runs with the female talker's stimuli (a total of 4 runs). The same was done for the CDS condition and participants were encouraged to take breaks between runs as needed. Across the study participants, the test order for performing the ADS or CDS condition was counterbalanced. Moreover, within a condition, the gender of the talker (female or male) was randomized. The mean percent correct across a talker and condition was calculated to obtain the final accuracy score for each run. Additionally, confusion matrices were recorded for each run (see Supplementary Materials Table 2). For this emotion test, participants used their self-reported better ear or earlier-implanted ear while the contralateral ear was plugged.

Data Analysis

Statistical analyses were completed in SPSS 24 (IBM) to compare percent correct scores for the ADS vs. CDS conditions as well as further analysis comparing scores across talkers. Moreover, multiple regression analyses was performed to analyze the impact of various predictors (ex. age of CI implantation, years using CI, NVIQ scores) on accuracy scores.

Finally, *d'* scores (Macmillan & Douglas Creelman, 2005) were calculated for participants' performance on each target emotion and statistical analyses were performed to explore differences in performance for different emotions. Normality of data was confirmed by P-P plots and Kolmogorov-Smirov tests prior to analysis. Acoustic analyses were conducted using the software program Praat (Boersma, 2001).

RESULTS

The data were first analyzed to confirm that no differences existed due to the varied hearing profiles (see Supplementary Materials Table 1) of the CI users. Percent correct emotion recognition scores were analyzed for any differences according to CI manufacturer, given that all three CI manufacturers (Advanced Bionics, Med-El, and Cochlear) were represented in our participant sample. One-way analysis of variance (ANOVA) revealed no significant difference according to CI manufacturer for either the CDS [$F(2, 24) = 0.44$, $p = 0.65$] or ADS [$F(2, 24) = 0.42$, $p = 0.66$] percent correct scores. One-way ANOVAs also revealed no significant differences in performance depending on whether a child was bilaterally or unilaterally implanted for CDS [$F(1, 25) = 3.09$, $p = 0.09$] or ADS [$F(1, 25) = 0.56$, $p = 0.46$] percent correct scores. Finally, one-way ANOVAs revealed no significant differences according to whether the left or right ear was tested for either the CDS [$F(1, 25) = 0.07$, $p = 0.80$] or ADS [$F(1, 25) = 0.68$ $p = 0.42$] percent correct scores. These results suggest that experimental differences (see results below) are likely due to differences in the stimulus conditions rather than because of the hearing profile of participants or because of the CI manufacturer chosen by a patient.

Acoustic characteristics of ADS and CDS stimuli.

Figure 1 shows boxplots of the acoustic features of the stimuli. For each emotion, values are shown relative to the "neutral" emotion. It is evident that the pitch variation across emotions (top row) and the intensity variation across emotions (bottom row) are considerably smaller for the ADS than for the CDS stimuli. The pattern of duration differences (middle row) is less different between the talkers and from ADS to CDS. We note that both the difference between mean values of an acoustic feature being used to contrast two emotions and the variance should be considered in determining how well that particular acoustic feature distinguishes the emotions (e.g., Chatterjee et al., 2015).

Within ADS and CDS types of speech, the female and male talkers also show differences in their patterns, underscoring the importance of considering talker variability in emotional productions. These differences have been quantified in previous publications that used the same stimuli (Chatterjee et al., 2015; Cannon & Chatterjee, 2019). For instance, in the ADS stimuli, the female talker produced more obvious intensity differences (bottom row) across emotions than the male talker. In the CDS stimuli, the female talker produced less variable pitch contrasts (smaller whiskers in the boxplots), a larger range of durations, and a larger range of intensities across emotions than the male talker.

Differences in performance in ADS vs. CDS condition.

Figure 2 compares the overall percent correct scores in the ADS condition compared to the CDS condition, with higher percent correct scores indicating better performance (note that chance performance is at 20% correct). Paired samples *t*-tests revealed that scores were significantly higher for the CDS condition [$t(26)=11.53$, $p < 0.001$] compared to the ADS condition (see Figure 2). When scores were further separated by talker (i.e. male or female), a significant talker effect was also found (see Figure 3). Participants had better performance for the female talker in both the CDS condition [$t(26)= 6.88$, $p < 0.001$, figure 3A] and the ADS condition [$t(26) = 5.86$, $p < 0.001$, figure 3B]. Individual data depicted in panels 3C and 3D.

Analyses of Predictors of Performance.

To examine whether performance on the administered cognitive assessments or aspects of a participant's hearing profile or development (e.g. years of experience using a CI, SES) were predictors of accuracy, we performed multiple regression analyses. Examined factors included: age of CI implantation, years of experience using the CI, standard scores taken from PPVT (an index of receptive vocabulary), nonverbal IQ (NVIQ, the WASI composite perceptual reasoning index score), and socioeconomic status (SES). SES was estimated through caretaker education levels as obtained from the questionnaire. SES scores for the primary caretaker and the secondary caretaker were highly correlated (Pearson correlation coefficient, $r = 0.50$, $p < 0.05$), and therefore averaged to produce an SES score for multiple regression analyses. Participants' chronological age was highly correlated with years of experience using a CI (presumably because the age of implantation of participants was quite young with many implanted under the age of 1) and was therefore not included in regression analysis.

Standard multiple regression analyses were used to assess the ability of these five factors to predict CDS and ADS percent correct scores. Preliminary analyses showed that there was no violation of linearity, multicollinearity, or homoscedasticity. For the CDS condition, the results of the regression indicated that the model was able to explain 40.9% of the variance (i.e. $r^2 = 0.409$) in CDS percent correct scores, $F(5,21) = 2.90$, $p < 0.05$. Only NVIQ made a significant contribution to the model ($\beta = 0.47$, $p < 0.05$). For the ADS condition, the results of the regression indicated that the model was able to explain 39.8% of the variance (i.e. $r^2 = 0.398$) in ADS percent correct scores, $F(5, 21) = 2.77$, $p < 0.05$. Only years using CI made a significant contribution to the model ($\beta = .538$, $p < 0.01$). See Table 1 for summary of multiple regression analyses.

Thus, for the CDS condition, nonverbal IQ performance predicted emotion recognition performance and for the ADS condition, years of experience using a CI predicted emotion recognition performance. Follow-up stepwise hierarchical regression analyses confirmed these findings as well. For the CDS condition, SES was entered in step 1, explaining 5.4% of the variance in the model. After entry of years of implantation and age of implantation were entered on step 2, the model explained 22.9% of the variance. After entry of NVIQ in step 3, the total variance explained by the model was 40.9%, $F(4, 22) = 3.80$, $p < 0.05$. NVIQ explained an additional 18% of the variance in CDS percent correct scores, after controlling for SES, years of implantation, and age of implantation, R square change = 0.18, F change = (1, 22) = 6.69, $p < 0.05$. In the final model where PPVT scores were entered on step 4, the model still explained 40.9% of the variance, and only NVIQ was a statistically significant predictor (β = 0.61, $p < 0.05$). For the ADS condition, SES was entered in Step 1 with the model explaining 2.7% of the variance. Age of implantation was entered in Step 2, with the model explaining 0.5% of the model. PPVT was entered in Step 3, explaining 14.9% of the variance, and NVIQ was entered in step 4 explaining 16.4% of the variance. In the final step, years of experience using CI was entered and the model explained 39.8% of the variance (i.e. R squared = 0.25, $F(5, 21) = 2.77$, $p < 0.05$). Years using CI explained an additional 23.4% of the variance after the other factors were controlled for, R square change = 0.234, F change(1, 21) = 8.15, $p < 0.01$, and was the only significant contributing predictor (β = 0.54, $p < 0.01$). Thus regression analyses suggest that NVIQ is a predictor of performance with CDS while years using CI is a significant predictor of performance with ADS.

Predictors of performance incorporating F0 Threshold:

In addition to the paradigms described in the Methods section above, participants from this project had also completed a test of dynamic pitch sensitivity whose data was presented in a previous paper (Deroche et al., 2019). The Deroche et al. 2019 paper examined whether speaking a tonal language aided in dynamic pitch perception; NH and CI children ages 6-19 from Taiwan and the US were tested on both a dynamic pitch labeling task (label whether F0 was rising or falling) as well as a discrimination task (pick the odd one out when one F0 was rising and the other two were falling, at a constant sweep rate). They measured performance for a range of sweep rates— very shallow to very steep—which allowed the fit of a psychometric function from which thresholds could be extracted a fixed level of discriminability (d' of 0.77). They found that a tonal language benefit existed for NH children in both tasks, and for the CI children in the labeling task (even though all CI children struggled overall in both tasks).

For the experiment detailed here, we used these dynamic F0 thresholds from Deroche et al. 2019 to investigate whether dynamic pitch abilities served as predictors of vocal emotion recognition performance, positing that abilities to encode F0 sweeps may aid in vocal prosody perception. Dynamic pitch thresholds correlated both with CDS percent correct scores ($r = -0.69$, $p < 0.01$) and ADS percent correct scores ($r = -0.66$, $p < 0.001$), suggesting that F0 thresholds might predict performance and therefore new, separate multiple regression analyses were run.

To avoid the inclusion of too many predictors, which could lead to non-significant models, predictors were carefully selected to only include Age of Implantation, Years of CI experience, NVIQ, and F0 threshold. SES was not included in this multiple regression analysis because preliminary analysis suggested it was highly correlated with, and therefore operating through, age of implantation. PPVT scores were highly correlated with NVIQ so PPVT was not included in this particular model either; NVIQ was chosen given previous research suggesting that NVIQ may be a predictor of vocal emotion recognition for NH children listening to vocoded speech as a CI simulation (Tinnemore et al., 2018). For the CDS condition, standard multiple regression analyses were used to assess the ability of these four factors to predict CDS percent correct scores. One subject's data (BT22) was found to be an outlying case as assessed by the Cook's distance max value and was excluded from analysis. With the removal of this influential point, it was found that the model was able to explain 45.4% of the variance (i.e. $r^2 = 0.45$) in CDS percent correct scores $F(4, 15) = 3.12$, $p < 0.05$, however none of the predictors made a significant contribution to the model (see Table 2). Hierarchical regression analyses confirmed this as well. When age of implantation was entered in step 1, the model explained 2.6% of the variance. After entry of years using CI in step 2, the model then explained 15.6% of the variance. After entry of NVIQ in step 3, the total variance explained by the model was 43.5%, $F(3, 16)= 4.11$, $p < 0.05$. NVIQ explained an additional 27.9% of the variance after controlling for age of implantation and years using CI, R square change = 27.9%, F Change (1, 16) = 0.01. In the final model where F0 threshold was entered on step 4, the model still explained 45.4% of the variance $F(4, 15) = 3.12$, $p < 0.05$ and none of the predictors made a significant contribution to the model. Thus, it appears that F0 threshold does not appear to contribute significantly to the model in predicting performance with CDS stimuli (see Table 2).

For the ADS condition, standard multiple regression analysis was performed. The results of the regression indicated that the model was able to explain 50.0% of the variance (i.e. $r^2 = 0.50$) in ADS percent correct scores $F(4, 16) = 4.01$, $p < 0.05$. In this case, F0 threshold made a trending significant contribution to the model ($\beta = -0.81$, $p = 0.055$, see Table 4). Hierarchical regression analyses was performed to explore this in more detail. When age of implantation was entered in step 1, the model explained 1.7% of the variance. After entry of years using CI in step 2, the model then explained 22.2% of the variance. After the entry of NVIQ in step 3, the total variance explained by the model was 36.7%, $F(3, 17) = 3.29$, $p < 0.05$, with years of CI contributing significantly to the model ($p < 0.05$) and NVIQ giving a trending significant contribution ($p = 0.065$). NVIQ explained an additional 14.5% of the variance after controlling for age of implantation and years using CI, R square change = 0.15, F Change (1, 17) = 0.065. In the final model where F0 threshold was entered on step 4, the model now explained 50.0% of the variance $F(4, 16) = 4.01$, $p < 0.05$ with F0 threshold providing a trending significant contribution ($p = 0.055$). F0 threshold explained an additional 13.3 % of the variance after controlling for the other 3 predictors, R square change = .133, F change (1, 16) = 0.055 (see Table 2).

In summary, for accuracy in the CDS condition, dynamic pitch F0 threshold was not a significant predictor. For the ADS condition, the standard and hierarchical multiple regression analyses suggest that F0 thresholds were marginally significant predictors of performance. It appears, therefore, that it is only in the harder condition—the ADS condition —that F0 threshold may help with performance.

Sensitivity to Different Target Emotions:

The presented results have focused on percent correct scores which provide an overall sense of accuracy, but do not allow for deeper investigation into error patterns. To further explore sensitivity to different emotions, *d'* scores were analyzed based on confusion matrices (see Supplementary Materials Table 2 for an example confusion matrix). Hit rates and false alarm rates were analyzed separately for each talker (i.e. male and female) and condition (ADS vs.

CDS). Any hit rates equal to 1 were adjusted to 0.9999 while false alarm rates equal to 0 were adjusted to 0.0001.The *d'* scores were calculated as the difference between z-scores derived from the hit rates and false alarm rates.

A two-way ANOVA examined the effect of talker (CDS Female, CDS Male, ADS Female, ADS Male) and emotion (happy, scared, neutral, sad, angry) on *d'* scores. Results revealed a significant interaction between the effects of talker and emotion on *d'* scores $F(12, 520) = 4.39$, $p < 0.001$ as well as significant main effects of talker [$F(3, 520) = 74.2$, $p < 0.001$] and emotion [$F(4, 520) = 7.23$, $p < 0.001$]. Analysis was also performed on the subcomponents of the *d'* score— namely hit rates and false alarm rates. A two-way ANOVA of hit rates revealed a significant interaction between the effects of talker and emotion on hit rates $F(12, 520) = 4.39$, $p < 0.001$ and significant main effects of talker [$F(3, 520) = 49.6$, $p < 0.001$] and emotion [$F(4, 520) = 3.92$, $p < 0.001$]. A two-way ANOVA of false alarm rates also revealed a significant interaction between the effects of talker and emotion on false alarm rates $F(12, 520) = 2.463$, $p < 0.01$ and significant main effects of talker [$F(3, 520) = 32.3$, $p <0.001$] and emotion [$F(4, 520) = 27.3$, $p < .001$]. *Post-hoc* analysis was performed to explore these differences in more detail (see below, analysis separated by condition).

CDS d' Score Analysis by Emotion: A one-way ANOVA was conducted to explore the impact of emotion on sensitivity (i.e. *d'* scores) for the CDS female talker (see Figure 4). There was a significant difference in sensitivity across emotions $F(4, 64)= 9.50$, $p < 0.001$ and *post–hoc* pairwise comparisons (Tukey's) revealed: 1) *happy d'* scores were significantly higher than *scared* ($p < 0.05$), *neutral* ($p < 0.001$), and *sad* ($p < 0.01$), 2) *neutral d'* scores were significantly lower than *angry* ($p < 0.001$), and 3) *sad d'* scores were trending towards being lower than *angry* ($p = 0.07$). Unlike for the CDS female talker, a one-way ANOVA for the CDS male talker showed no significant difference in *d'* scores according to emotion $F(4, 130) = 1.16$, $p = 0.33$ (See Figure 4, top). Differences according to emotion are thus only seen for the female talker with participants generally demonstrating high sensitivity for the *happy* emotion and low sensitivity for the *neutral* emotion.

Statistical analysis of the hit rates (an index of accuracy) and false alarm rates (see Figure 4, bottom panels) showed a significant difference in hit rates according to emotion for the CDS female talker, $F(4, 64) = 3.75$, $p < 0.01$. Tukey's *Post-hoc* comparisons revealed that *happy* hit rates were significantly higher than *neutral* hit rates ($p < 0.01$). For the male talker, a one-way ANOVA of the CDS hit rates revealed no significant difference according to emotion $F(4, 130) = 0.70$, $p = 0.59$. Thus, as with *d'* scores, differences in accuracy for particular emotions was only seen with the female talker, with participants having high hit rates for *happy* sentences. For false alarm rates, a one-way ANOVA of the CDS female talker revealed a significant difference according to emotion $F(4, 130)= 3.37$, $p < 0.01$, with Tukey's *post-hoc* comparisons revealing that the *sad* emotion elicited a higher false alarm rate than *angry* ($p < 0.01$, see Figure 4, bottom right). For the CDS male talker, a one-way ANOVA revealed significant differences in false alarm rates according to emotion $F(4, 130) = 3.17$, $p < 0.05$, in that participants had higher false alarm rates to the *neutral* emotion compared to the *angry* emotion ($p < 0.01$).

ADS d' score analysis by Emotion: A one-way ANOVA explored the impact of emotion on sensitivity for the ADS female talker and showed a significant difference in *d'* scores according to emotion $F(4, 130) = 8.31$, $p < 0.001$. Tukey's *post-hoc* comparisons revealed that 1) *d'* scores to *happy* were higher than to *scared* ($p < 0.001$), 2) *d'* scores to *scared* were lower than for *neutral* ($p < 0.05$), *sad* ($p < 0.001$), and *angry* ($p < 0.001$). For the ADS male talker, a one-way ANOVA revealed no significant difference in *d'* scores according to emotion $F(4, 64) = 1.74$, $p = 0.15$. Thus, as with the CDS condition, *d'* scores differed according to emotion only for the female talker, with participants showing low sensitivity detection to *scared* (see Figure 5).

Statistical analysis on the hit rate and false alarm rates revealed differences according to emotion as well (see Figure 5, bottom panels). For the ADS female talker, a one-way ANOVA revealed a significant difference in hit rates according to emotion $F(4, 130) = 12.36$, $p < 0.001$, where 1) hit rates for *happy* were higher than hit rates for *scared* ($p < 0.001$, as revealed through Tukey's *post-hoc* tests) and 2) Hit rates for *scared* were lower than hit rates for *neutral* ($p < 0.001$), *sad* ($p < 0.001$), and *angry* ($p < 0.001$). Thus for the female talker, in general, accuracy was low for the *scared* emotion. For the ADS male talker, a one-way ANOVA revealed only a trending difference among hit rates according to emotion $F(4, 130) = 2.41$, $p = 0.052$. Analysis of the false alarm rate data, revealed a significant difference in false alarm rates according to emotion for the ADS female talker, $F(4, 64) = 10.035$, $p < 0.001$. Tukey's *post-hoc* comparisons showed that 1) *happy* had lower false alarm rates than *neutral* ($p < 0.001$) or *sad* ($p < 0.01$), 2) *neutral* had higher false alarm rates than *sad* ($p < 0.05$) and *angry* ($p < 0.001$) and 3) *sad* had higher false alarm rates than *angry* ($p < 0.001$). Thus, participants appeared to respond with *neutral* or *sad* (i.e. high false alarm rates) when unsure about the target emotion. For the ADS male talker, a one-way ANOVA revealed that false alarm rates varied significantly according to the target emotion $F(4, 130) = 18.57$, p < 0.001; Tukey's *post-hoc* comparisons revealed that 1) *happy* had lower false alarm rates than *neutral* ($p < 0.001$) and *sad* ($p < 0.001$), 2) *neutral* had higher false alarm rates than *scared* ($p < 0.001$), *sad* ($p < 0.01$) and *angry* ($p < 0.001$), and finally that 3) *scared* had higher false alarm rates than *angry* ($p < 0.01$). As with the female talker, participants tended to respond with *neutral* or *sad* (hence the high false alarm rates) when unsure about the target emotion.

DISCUSSION

In this experiment, pediatric CI users were tested on their vocal emotion recognition performance for both child-directed (CDS) and adult-directed (ADS) speech. As hypothesized, participants had better emotion perception for the CDS condition compared to the ADS condition. These results mirror what has been reported in NH children who also had poorer performance with adult directed materials compared to child-directed materials (Cannon & Chatterjee, 2019). It is likely that the larger variations in intensity and pitch in the CDS condition—the inflated prosody—made it an easier condition for pediatric CI participants. A talker effect was found as well in that in both ADS and CDS conditions, participants had higher emotion recognition scores with the female talker's stimuli, mimicking previous study results with NH and children with hearing loss who likewise showed this talker effect (Cannon & Chatterjee, 2019; Chatterjee et al., 2015). Different male and female talkers were used for the ADS and CDS material, and, given that there were only two talkers for each set, these results cannot be interpreted as a sex/gender effect. However, this does point to the fact that variability exists in how speakers communicate emotions; previous acoustic analyses of the CDS and ADS stimuli found differences between talkers in patterns of acoustic cues across emotions (Chatterjee et al., 2015; Cannon & Chatterjee, 2019). It is possible that in this test, the different female speakers used for the CDS and ADS conditions were simply more animated in their conveyance of the target emotions compared to the two male speakers.

Multiple regression analyses showed that nonverbal intelligence (NVIQ) performance was a significant predictor of CDS performance, while years of CI experience was a significant predictor of ADS performance. NVIQ is a predictor of emotion recognition by NH children listening to vocoded speech (Tinnemore et al., 2018). Research on child CI users have found that NVIQ, an element of general cognition, is a predictor of language outcomes and development (Geers & Nicholas, 2013; Geers & Sedey, 2011). Moreover, it has been posited that general cognition skills such as working memory and NVIQ mediate both linguistic and indexical speech (including prosody and emotion perception) processing (Tobey et al., 2003; Geers et al., 2013; Tinnemore et al., 2018). Language supports emotional exchanges, helping to label and conceptualize emotions, and is therefore vital for learning to discriminate emotions, which may explain why NVIQ served as predictor of CDS emotion recognition performance here.

It is interesting to note that different factors were predictors for the ADS and CDS conditions. Participants' nonverbal IQ was a significant predictor in the easier emotion recognition condition (i.e. CDS), but for the harder emotion recognition condition (i.e. ADS), increased experience using a CI predicted emotion perception. We speculate that with CDS stimuli, cognitive and linguistic ability is a fundamental requisite to be able to associate the perceived prosodic information with the correct emotional label. On the other hand, with ADS stimuli, the benefit of experience (as in years of experience with the device and developmental age) seems far more important in a child's ability to associate the reduced acoustic cues with the correct emotional label. With increased experience, children are exposed to many more linguistic styles and inputs, particularly with adult speakers, than children with only one or two years of experience with sound, who typically have limited exposure to adult conversation. It is possible that cognitive and linguistic skills still matter in shaping emotion recognition outcomes with ADS stimuli, but it seems likely that their contribution is more limited than that of extensive experience. Currently, the role of early CI implantation, and hence years of CI experience within this pediatric population, in the development of speech emotion skills is not well understood. While some data refute the idea that longer duration of CI use provides an advantage in emotion recognition (Hopyan-Misakyan et al., 2009), other studies show positive correlations between duration of implant use with complex emotion comprehension (Mancini et al., 2016) and with speech perception (Artières et al., 2009; Kiefer et al., 1996). Discrepancies in findings may be due to the variability in amount of CI experience in the different experimental cohorts (short vs. long durations of CI experience, as well as chronological age of pediatric CI participants) as well as variations in stimuli and in the way emotion recognition is being tested across experiments.

For another set of multiple regression analyses, pitch thresholds from a dynamic pitch sensitivity task (see Deroche et al., 2019) were included to explore their impact on emotion recognition. It was found that perceptual sensitivity to complex pitches was not significant predictors for the CDS condition, but were trending towards a significant contribution towards ADS condition performance. Dynamic pitch thresholds have been linked to other aspects of sound processing such as static pitch sensitivity and emotion recognition in child CI users (Deroche et al., 2016). In this experiment, results suggest that for more difficult emotion perception tasks, participants may need to draw more heavily upon experience and hearing acuity, such as dynamic pitch thresholds, to aid in distinguishing between emotions.

Finally, confusion matrix analysis showed that participants sometimes showed differential sensitivity to particular emotions. For the CDS condition, sensitivity (i.e. *d'* scores) was particularly high for the *happy* emotion and low for the *neutral* emotion but only for the female talker. Previous acoustic analysis of CDS materials had found that *scared, happy*, and *angry* were spoken the loudest and that *happy*, in particular, had the greatest F0 range (Chatterjee et al., 2015), which may explain the high sensitivity to *happy* sentences in this experiment. For the ADS condition, sensitivity to the *scared* condition was quite low, and again only for the female talker. An exhaustive acoustic analysis and discussion of the different ways in which emotions are vocally expressed is beyond the scope of this study, but our findings confirm previous observations based on acoustic analyses of vocal emotions (Banse & Scherer, 1996). Differences in acoustic characteristics such as speech rate, pitch range, pitch changes, and intensity characterize the expression of different emotions, leading to differences in ease of perception of these different emotions by pediatric CI users (Mildner & Koska, 2014). Indeed, research has shown that certain emotions are more difficult to perceive than others in both adult (Luo et al., 2007) and child (Geers et al., 2013) CI users. Such emotion-specific sensitivity may be further modulated by variability in speaking style and across talkers, as well as by individual CI patients' sensitivity to specific cues for emotion. Generally, CI patients may be more sensitive to differences in intensity (e.g., Luo et al., 2007) or speaking rate than in differences in voice pitch, while normally hearing listeners use voice pitch as a dominant cue to vocal emotions.

Given the difficulty that pediatric cochlear implant users face in identifying emotions in vocal speech and its importance for social development, identifying ways to improve emotion recognition is of great importance for pediatric CI users. Our results here suggest that emotion identification is easier when inflated pitch and intensity cues (i.e. exaggerated prosody) are used. Other research has posited that since pitch cues are fundamental but poorly represented auditory cues of vocal emotion identification, training regimens that focus on pitch perception rehabilitation and sound processing may lead to improvement in emotion recognition performance in CI users (Jiam et al., 2017; Mildner & Koska, 2014). Indeed a pilot study of an 11-session psychoeducational program focused on improving emotional understanding in deaf children lead to improvements in emotion vocabulary and emotion comprehension (Dyck & Denver, 2003). More recently, it was found that music training, as opposed to art training, lead to improvements not only on musical discrimination of melodic contour and rhythm, but also on speech prosody perception in child CI users (Good et al., 2017). Additional research is necessary, but these music and emotion-based training programs may show promise for improving emotion recognition in pediatric CI users. Improved emotion identification, in turn, may have a significant impact on socio-developmental skills for pediatric CI users.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We'd like to thank members of the UCSF Sound and Music Perception lab (Melanie Gilbert and Lauren Jacobs) for their *ad-hoc* assistance on this project. This work was funded by NIH/NIDCD grant number R01DC014233. This work was first presented as a poster at the Association for Research in Otolaryngology (ARO) 2018 Midwinter Meeting in San Diego, California. For questions regarding this paper, please contact Karen Chan Barrett at Karen.Barrett@ucsf.edu.

Disclosures:

C.L. has received research support from Advanced Bionics (medical advisory board, consultant), Med-El (advisory board), Oticon (consultant), and Spiral Therapeutics (chief medical officer) for unrelated work. All other authors declare no competing interest.

WORKS CITED

A. Macmillan N, & Douglas Creelman C. (2005). Detection theory: a user's guide. 492.

Artières F, Vieu A, Mondain M, Uziel A, & Venail F (2009). Impact of early cochlear implantation on the linguistic development of the deaf child. Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology, 30(6), 736–742.

AuBuchon AM, Pisoni DB and Kronenberger WG (2015) 'Short-Term and Working Memory Impairments in Early-Implanted, Long-Term Cochlear Implant Users Are Independent of Audibility and Speech Production', Ear and Hearing. doi: 10.1097/AUD.0000000000000189.

AuBuchon AM, Pisoni DB and Kronenberger WG (2019) 'Evaluating pediatric cochlear implant users' encoding, storage, and retrieval strategies in verbal working memory', Journal of Speech, Language, and Hearing Research. doi: 10.1044/2018_JSLHR-H-18-0201.

Banse R, & Scherer KR (1996). Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology, 70(3), 614–636. [PubMed: 8851745]

Boersma Paul & Weenink David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.06, retrieved 8 November 2019 from http://www.praat.org/.

Cannon SA, & Chatterjee M (2019). Voice Emotion Recognition by Children With Mild-to-Moderate Hearing Loss. Ear and Hearing. 10.1097/AUD.0000000000000637

Chatterjee M, & Peng S-C (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. Hearing Research, Vol. 235, pp. 143–156. 10.1016/j.heares.2007.11.004 [PubMed: 18093766]

Chatterjee M, Zion DJ, Deroche ML, Burianek BA, Limb CJ, Goren AP, … Christensen JA (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. Hearing Research, 322, 151–162. [PubMed: 25448167]

Ciocca V, Francis AL, Aisha R, & Wong L (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. The Journal of the Acoustical Society of America, 111(5 Pt 1), 2250–2256. [PubMed: 12051445]

Deroche MLD, Kulkarni AM, Christensen JA, Limb CJ, & Chatterjee M (2016). Deficits in the Sensitivity to Pitch Sweeps by School-Aged Children Wearing Cochlear Implants. Frontiers in Neuroscience, 10, 73. [PubMed: 26973451]

Deroche MLD, Lu H-P, Kulkarni AM, Caldwell M, Barrett KC, Peng S-C, … Chatterjee M (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. Scientific Reports, 9(1), 109. [PubMed: 30643156]

Dominey PF, & Dodane C (2004). Indeterminacy in language acquisition: the role of child directed speech and joint attention. Journal of Neurolinguistics, 17(2), 121–145.

Dunn LM, & Dunn DM (2007). PPVT-4: Peabody picture vocabulary test. Pearson Assessments.

Dyck MJ, & Denver E (2003). Can the Emotion Recognition Ability of Deaf Children be Enhanced? A Pilot Study. Journal of Deaf Studies and Deaf Education, 8(3), 348–356. [PubMed: 15448058]

Eisenberg N, Spinrad TL, & Eggum ND (2010). Emotion-Related Self-Regulation and Its Relation to Children’s Maladjustment. Annual Review of Clinical Psychology, Vol. 6, pp. 495–525. 10.1146/annurev.clinpsy.121208.131208

Eisenberg LS et al. (2000) ‘Speech recognition with reduced spectral cues as a function of age’, The Journal of the Acoustical Society of America. doi: 10.1121/1.428656.

Geers AE, Davidson LS, Uchanski RM, & Nicholas JG (2013). Interdependence of linguistic and indexical speech perception skills in school-age children with early cochlear implantation. Ear and Hearing, 34(5), 562–574. [PubMed: 23652814]

Geers AE, & Nicholas JG (2013). Enduring advantages of early cochlear implantation for spoken language development. Journal of Speech, Language, and Hearing Research: JSLHR, 56(2), 643–655. [PubMed: 23275406]

Geers AE, & Sedey AL (2011). Language and verbal reasoning skills in adolescents with 10 or more years of cochlear implant experience. Ear and Hearing, 32(1 Suppl), 39S – 48S. [PubMed: 21832889]

Gfeller K, Turner C, Mehr M, Woodworth G, Fearn R, Knutson JF, … Stordahl J (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. Cochlear Implants International, 3(1), 29–53. [PubMed: 18792110]

Good A, Gordon KA, Papsin BC, Nespoli G, Hopyan T, Peretz I, & Russo FA (2017). Benefits of Music Training for Perception of Emotional Speech Prosody in Deaf Children With Cochlear Implants. Ear and Hearing, 38(4), 455–464. [PubMed: 28085739]

Hopyan-Misakyan TM, Gordon KA, Dennis M, & Papsin BC (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 15(2), 136–146. [PubMed: 18828045]

Jiam NT, Caldwell M, Deroche ML, Chatterjee M, & Limb CJ (2017). Voice emotion perception and production in cochlear implant users. Hearing Research, 352, 30–39. [PubMed: 28088500]

Kiefer J, Gall V, Desloovere C, Knecht R, Mikowski A, & von Ilberg C (1996). A follow-up study of long-term results after cochlear implantation in children and adolescents. European Archives of Oto-Rhino-Laryngology: Official Journal of the European Federation of Oto-Rhino-Laryngological Societies : Affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery, 253(3), 158–166.

Kong Y-Y, Cruz R, Jones JA, & Zeng F-G (2004). Music perception with temporal cues in acoustic and electric hearing. Ear and Hearing, 25(2), 173–185. [PubMed: 15064662]

Kronenberger WG et al. (2014) ‘Neurocognitive risk in children with cochlear implants’, JAMA Otolaryngology - Head and Neck Surgery. doi: 10.1001/jamaoto.2014.757.

Limb CJ, & Roy AT (2014). Technological, biological, and acoustical constraints to music perception in cochlear implant users. Hearing Research, 308, 13–26. [PubMed: 23665130]

Luo X, & Fu Q-J (2004). Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. The Journal of the Acoustical Society of America, 116(6), 3659–3667. [PubMed: 15658716]

Luo X, Fu Q-J, & Galvin JJ (2007). Cochlear Implants Special Issue Article: Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users. Trends in Amplification, 11(4), 301–315. [PubMed: 18003871]

Mancini P, Giallini I, Prosperini L, D’alessandro HD, Guerzoni L, Murri A, … Nicastri M (2016). Level of emotion comprehension in children with mid to long term cochlear implant use: How basic and more complex emotion recognition relates to language and age at implantation. International Journal of Pediatric Otorhinolaryngology, 87, 219–232. [PubMed: 27368475]

McCreery R et al. (2019) ‘Auditory, Cognitive, and Linguistic Factors Predict Speech Recognition in Adverse Listening Conditions for Children with Hearing Loss’, Front. Neurosci, 13(1093).

McDermott HJ (2004). Music perception with cochlear implants: a review. Trends in Amplification, 8(2), 49–82. [PubMed: 15497033]

Mildner V, & Koska T (2014). Recognition and production of emotions in children with cochlear implants. Clinical Linguistics & Phonetics, 28(7-8), 543–554. [PubMed: 25000377]

Moberly AC et al. (2017) 'Speech recognition in adults with cochlear implants: The effects of working memory, phonological sensitivity, and aging', Journal of Speech, Language, and Hearing Research. doi: 10.1044/2016_JSLHR-H-16-0119.

Most T, & Aviner C (2009). 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. Retrieved from https://academic.oup.com/jdsde/article-abstract/14/4/449/488373

Murray IR, & Arnott JL (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. The Journal of the Acoustical Society of America, 93(2), 1097–1108. [PubMed: 8445120]

Nakata T, Trehub SE, & Kanda Y (2012). Effect of cochlear implants on children's perception and production of speech prosody. The Journal of the Acoustical Society of America, 131(2), 1307–1314. [PubMed: 22352504]

Nittrouer S, Caldwell-Tarr A and Lowenstein JH (2013) 'Working memory in children with cochlear implants: Problems are in storage, not processing', International Journal of Pediatric Otorhinolaryngology. doi: 10.1016/j.ijporl.2013.09.001.

O'Neill ER, Kreft HA and Oxenham AJ (2019) 'Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions', The Journal of the Acoustical Society of America. doi: 10.1121/1.5116009.

Paquette S, Ahmed GD, Goffi-Gomez MV, Hoshino ACH, Peretz I, & Lehmann A (2018). Musical and vocal emotion perception for cochlear implants users. Hearing Research, 370, 272–282. [PubMed: 30181063]

Peng S-C, Tomblin JB, Cheung H, Lin Y-S, & Wang L-S (2004). Perception and production of mandarin tones in prelingually deaf children with cochlear implants. Ear and Hearing, 25(3), 251–264. [PubMed: 15179116]

Schorr EA, Roth FP, & Fox NA (2009). 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: JSLHR, 52(1), 141–152. [PubMed: 18664684]

Schvartz KC, Chatterjee M and Gordon-Salant S (2008) 'Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners', The Journal of the Acoustical Society of America. doi: 10.1121/1.2997434.

Tinnemore AR, Zion DJ, Kulkarni AM, & Chatterjee M (2018). Children's Recognition of Emotional Prosody in Spectrally Degraded Speech Is Predicted by Their Age and Cognitive Status. Ear and Hearing, 39(5), 874–880. [PubMed: 29337761]

Tobey EA et al. (2003) 'Factors Associated with Development of Speech Production Skills in Children Implanted by Age Five', Ear and Hearing. doi: 10.1097/01.aud.0000051688.48224.a6.

Volkova A, Trehub SE, Schellenberg EG, Papsin BC, & Gordon KA (2013). Children with bilateral cochlear implants identify emotion in speech and music. Cochlear Implants International, 14(2), 80–91. [PubMed: 23453221]

Wang DJ, Trehub SE, Volkova A, & van Lieshout P (2013). Child implant users' imitation of happy- and sad-sounding speech. Frontiers in Psychology, 4, 351. [PubMed: 23801976]

Wechsler D, & Others. (1999). Manual for the Wechsler abbreviated scale of intelligence. San Antonio, TX: Psychological Corporation.

Wei C-G, Cao K, & Zeng F-G (2004). Mandarin tone recognition in cochlear-implant subjects. Hearing Research, 197(1-2), 87–95. [PubMed: 15504607]

Wiefferink CH, Rieffe C, Ketelaar L, De Raeve L, & Frijns JHM (2013). Emotion understanding in deaf children with a cochlear implant. Journal of Deaf Studies and Deaf Education, 18(2), 175–186. [PubMed: 23232770]

Wiefferink CH, Rieffe C, Ketelaar L, De Raeve L, & Frijns JHM (2013). Emotion understanding in deaf children with a cochlear implant. Journal of Deaf Studies and Deaf Education, 18(2), 175–186. [PubMed: 23232770]

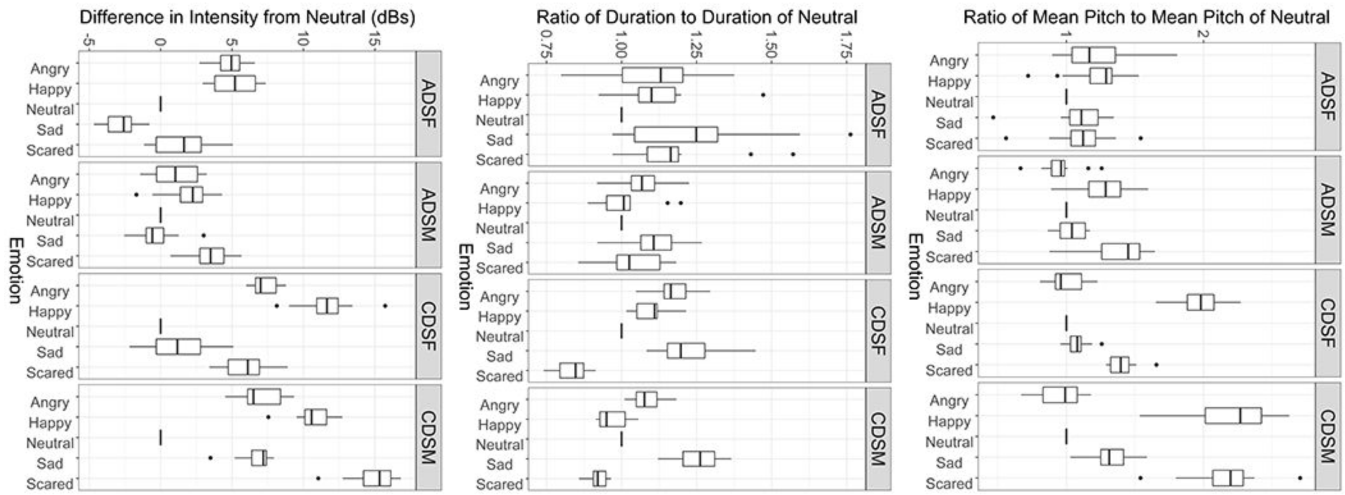

Figure 1. Acoustic characteristics of the adult- and child-directed stimuli for each emotion, relative to the neutral emotion.

In each row, ADSF and ADSM indicate adult-directed speech by the female and male talkers respectively, and CDSF and CDSM indicate child-directed speech by the female and male talkers respectively. Within each panel, the abscissa shows the specific emotion. The ordinates in the top, middle and bottom rows show boxplots of the ratio of the mean pitch (fundamental frequency) of the recorded sentences spoken in each emotion to the mean pitch of the same sentences spoken in the neutral emotion, the ratio of the duration of the recorded sentences spoken in each emotion to that of the same sentences spoken in the neutral emotion, and the intensity difference between the recorded sentences in each emotion and the same sentences in the neutral emotion, respectively.

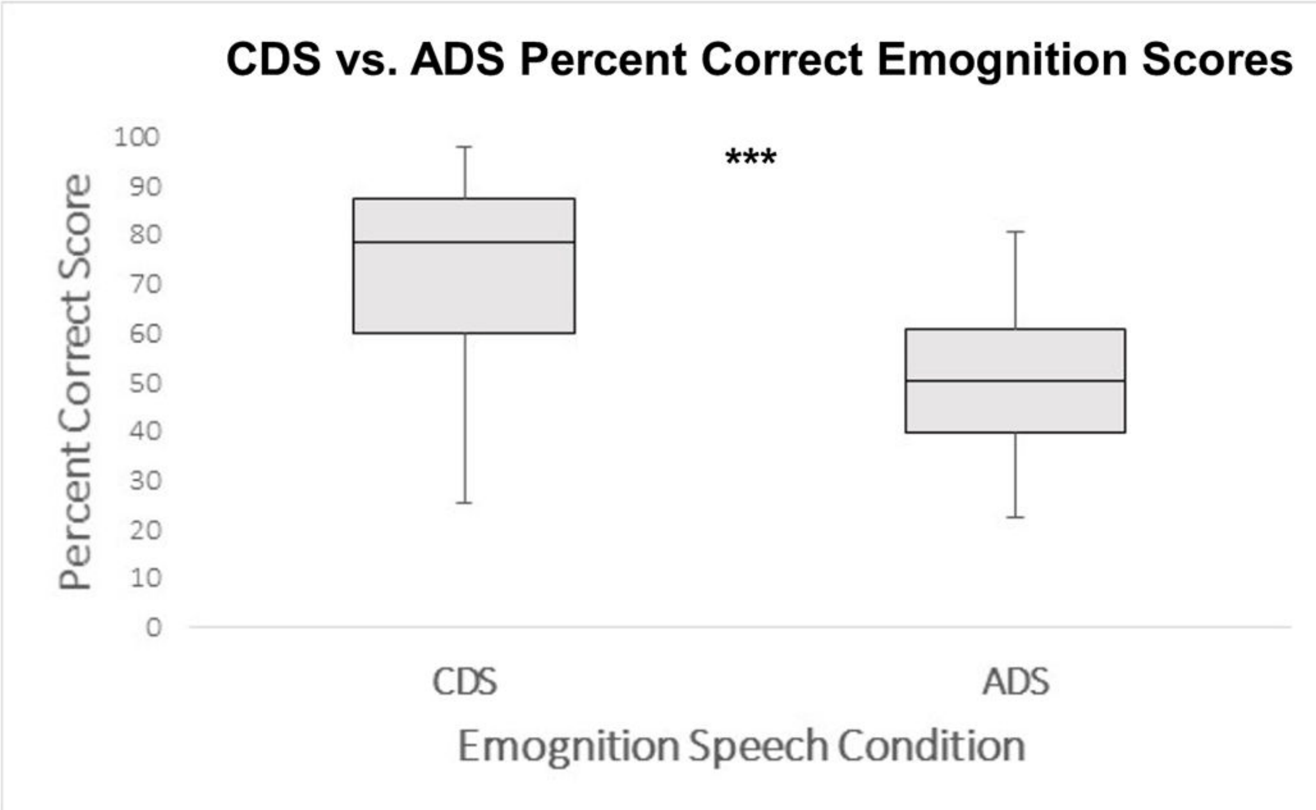

Figure 2. Percent Correct scores across the Conditions.

Comparison of Percent Correct Scores for CDS vs. ADS conditions reveal that performance is significantly higher for CDS compared to ADS condition [$t(26)=11.53$, $p < 0.001$]. Mean CDS Score: 71.7% +/− SD 20.3, Mean ADS Score: 50.0% +/− SD 14.4.

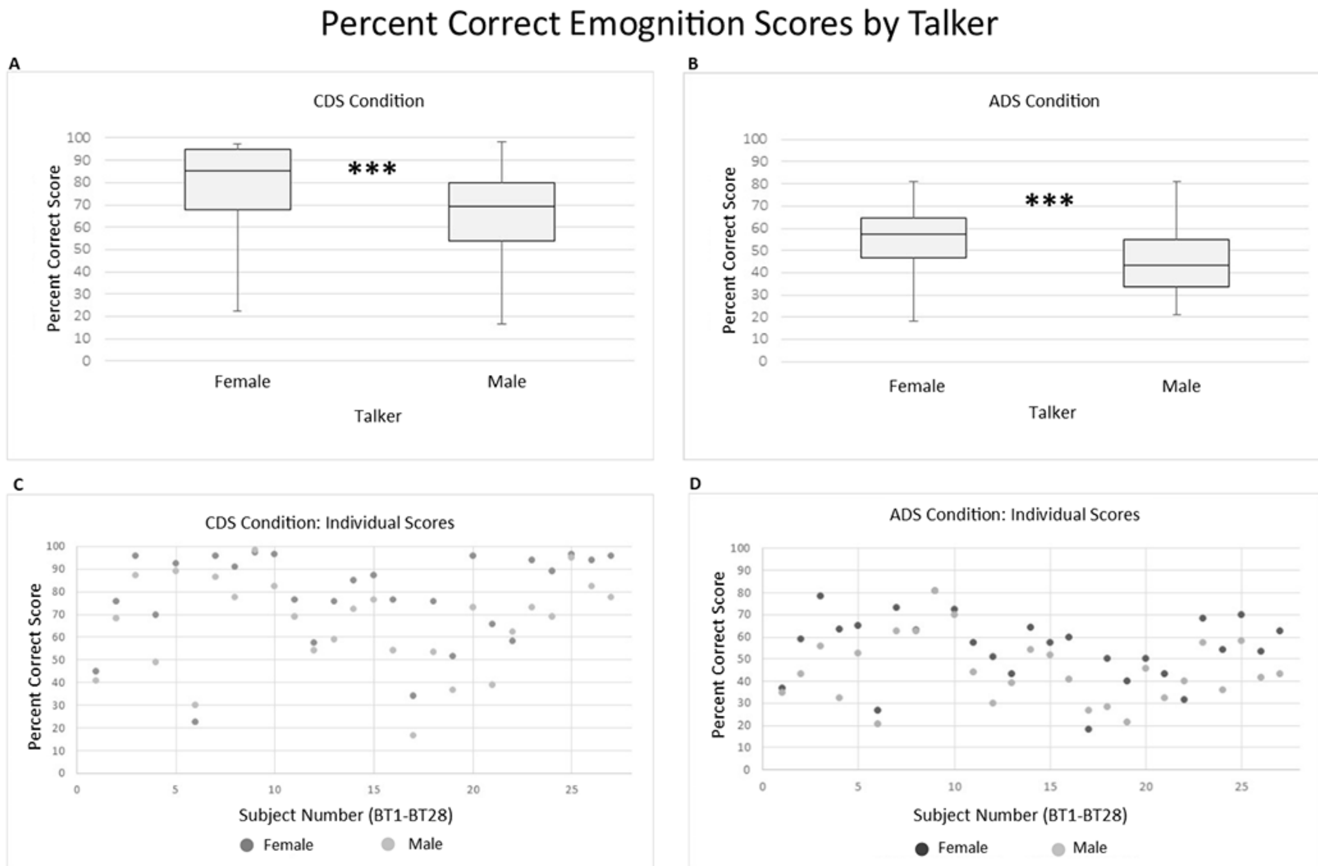

Figure 3. Percent Correct Scores separated by Talker.

Panels A and B show group average percent correct scores while Panels C and D depict individual scores. Participants showed higher percent correct score for the female talker in both the CDS (Female Mean Score: 77.5% +/− SD 20.8; Male Mean Score: 65.8% +/− SD 20.7, see panel 3A) and ADS condition (Female Mean Score: 55.3% +/− SD 15.5; Male Mean Score: 44.7% +/− SD 14.8, see panel 3B). Individual data separated by talker (female vs. male) is depicted for the CDS condition (Panel 3C, bottom left) and for the ADS condition (Panel 3B, bottom right).

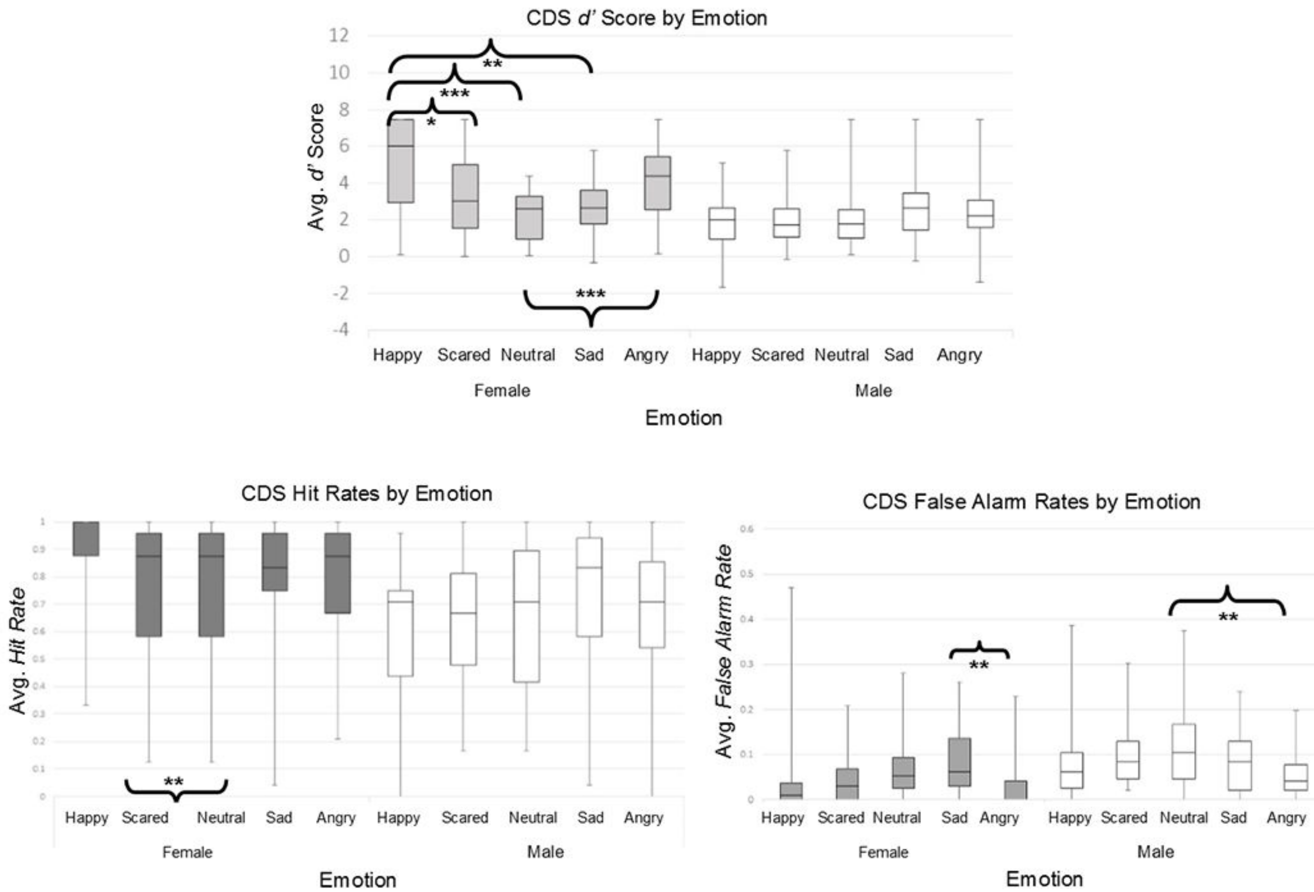

Figure 4. CDS condition confusion matrix data analysis evaluating responses to specific target emotion.

Top: Boxplots represent average *d'* score for each emotion separated out by talker. Left, bottom: Boxplot represents average hit rates (subcomponent of the *d'* score) for each emotion. Right, bottom: Boxplot represents average false alarm rates (subcomponent of the *d'* score) for each emotion.

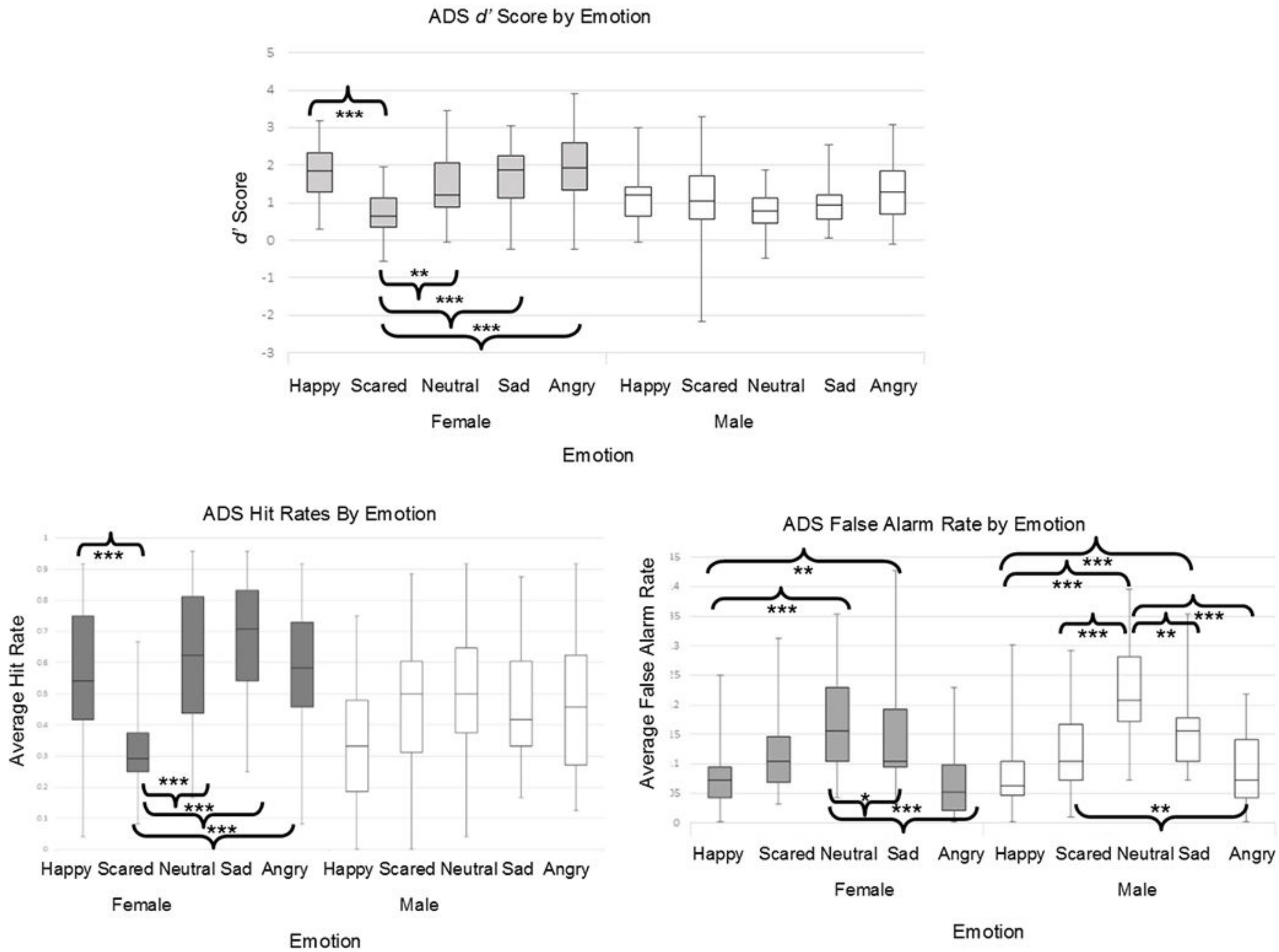

Figure 5. ADS condition confusion matrix data analysis evaluating responses to specific target emotion.

Top: Boxplots represent average *d’* score for each emotion separated out by talker. Left, bottom: Boxplot represents average hit rates (subcomponent of the *d’* score) for each emotion. Right, bottom: Boxplot represents average false alarm rates (subcomponent of the *d’* score) for each emotion.

TABLE 1.

Summary of Results of Standard Multiple Regression Analyses

Condition	Predictor Variable	Beta (β)	Significance
CDS	Age of implantation	0.32	0.16
	Years of experience using CI	0.29	0.14
	SES	0.28	0.21
	NVIQ	0.47	*0.04
	PPVT	0.00	1.00
ADS	Age of implantation	0.28	0.22
	Years of experience using CI	0.54	**0.01
	SES	0.08	0.74
	NVIQ	0.27	0.23
	PPVT	0.19	0.39

* significance at $p < 0.05$.

** significance at $p < 0.01$.

ADS indicates adult-directed speech; CDS, child-directed speech; CI, Cochlear implant; NVIQ, nonverbal intelligence; PPVT, Peabody Picture Vocabulary Test; SES socioeconomic status.

TABLE 2.

Standard Multiple Regression Analyses Including Dynamic Pitch F0 Thresholds From Deroche et al. (2019b) as a predictor variable

Condition	Predictor Variable	Beta (β)	Significance
CDS	Age of implantation	−0.01	0.98
	Years of experience using CI	0.28	0.50
	NVIQ	0.42	0.15
	Dynamic pitch F0 threshold	−0.31	0.48
ADS	Age of implantation	−0.28	0.37
	Years of experience using CI	−0.12	0.75
	NVIQ	−0.16	0.49
	Dynamic pitch F0 threshold	−0.81	[~*]0.06

In the CDS condition, although the model was significant, none of the predictors made a significant contribution. In the ADS condition, the model was significant with F0 threshold making a trending contribution.

[~*]trending significance at $p < 0.05$. CI indicates cochlear implants; NVIQ, nonverbal intelligence.