



Published in final edited form as:

*Ear Hear.* 2021 ; 42(6): 1727–1740. doi:10.1097/AUD.0000000000001057.

## Weighting of prosodic and lexical-semantic cues for emotion identification in spectrally-degraded speech and with cochlear implants

Margaret E. Richter<sup>1</sup>, Monita Chatterjee<sup>2</sup>

<sup>1</sup>Division of Speech and Hearing Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

<sup>2</sup>Boys Town National Research Hospital, Omaha, NE, USA

### Abstract

**Objectives:** Normally hearing (NH) listeners rely more on prosodic cues than on lexical-semantic cues for emotion perception in speech. In everyday spoken communication, the ability to decipher conflicting information between prosodic and lexical-semantic cues to emotion can be important: e.g., in identifying sarcasm or irony. Speech degradation in cochlear implants (CIs) can be sufficiently overcome to identify lexical-semantic cues, but the distortion of voice pitch cues makes it particularly challenging to hear prosody with CIs. The purpose of this study was to examine changes in relative reliance on prosodic and lexical-semantic cues in NH adults listening to spectrally degraded speech and adult CI users. We hypothesized that, compared to NH counterparts, CI users would show increased reliance on lexical-semantic cues and reduced reliance on prosodic cues for emotion perception. We predicted that NH listeners would show a similar pattern when listening to CI-simulated versions of emotional speech.

**Design:** Sixteen NH adults and eight postlingually deafened adult CI users participated in the study. Sentences were created to convey five lexical-semantic emotions (Angry, Happy, Neutral, Sad, and Scared), with five sentences expressing each category of emotion. Each of these 25 sentences was then recorded with the five (Angry, Happy, Neutral, Sad, and Scared) prosodic emotions by two adult female talkers. The resulting stimulus set included 125 recordings (25 sentences x five prosodic emotions) per talker, of which 25 were congruent (consistent lexical-semantic and prosodic cues to emotion) and the remaining 100 were incongruent (conflicting lexical-semantic and prosodic cues to emotion). The recordings were processed to have three levels of spectral degradation: full-spectrum, CI-simulated (noise-vocoded) to have eight channels and 16 channels of spectral information respectively. Twenty-five recordings (one sentence per lexical-semantic emotion recorded in all five prosodies) were used for a practice run in the full-spectrum condition. The remaining 100 recordings were used as test stimuli. For each talker and condition of spectral degradation, listeners indicated the emotion associated with each recording in a single-interval, five-alternative forced-choice task. The responses were scored as proportion

---

**All Correspondence should be addressed to:** Monita Chatterjee, Boys Town National Research Hospital, 555 N 30<sup>th</sup> St, Omaha, NE 68131, Voice 531 355 5069 Monita.Chatterjee@boystown.org.

There are no conflicts of interest, financial, or otherwise.

correct, where “correct” responses corresponded to the lexical-semantic emotion. CI users heard only the full-spectrum condition.

**Results:** The results showed a significant interaction between Hearing Status (NH, CI) and Congruency in identifying the lexical-semantic emotion associated with the stimuli. This interaction was as predicted – i.e., CI users showed increased reliance on lexical-semantic cues in the Incongruent conditions, while NH listeners showed increased reliance on the prosodic cues in the Incongruent conditions. As predicted, NH listeners showed increased reliance on lexical-semantic cues to emotion when the stimuli were spectrally degraded.

**Conclusions:** The current study confirmed previous findings of prosodic dominance for emotion perception by NH listeners in the full-spectrum condition. Further, novel findings with CI patients and NH listeners in the CI-simulated conditions showed reduced reliance on prosodic cues and increased reliance on lexical-semantic cues to emotion. These results have implications for CI listeners’ ability to perceive conflicts between prosodic and lexical-semantic cues, with repercussions for their identification of sarcasm and humor. Understanding instances of sarcasm or humor can impact a person’s ability to develop relationships, follow conversation, understand vocal emotion and intended message of a speaker, following jokes, and everyday communication in general.

---

## INTRODUCTION

The ability to identify our conversational partner’s intended emotion is essential for successful social interactions. Emotions in speech are conveyed in two concurrent ways: first, via prosodic cues, including changes in vocal pitch, timbre, loudness, speaking rate etc., and second, via lexical-semantic cues, i.e., the meanings of the words carrying the prosody. The extant literature suggests that when the two cues are in conflict (e.g., *This is just great!* spoken in a positive, happy tone versus in an angry tone), normal-hearing listeners rely primarily on the prosodic cues to determine the talker’s intended emotion, with a small age-related shift toward lexical-semantic cues (Ben David et al., 2016; Ben David et al., 2019).

In individuals with severe-to-profound hearing loss who use cochlear implants (CIs), harmonic pitch perception is severely impaired (e.g., Guerts & Wouters, 2001; Chatterjee & Peng, 2008; Oxenham, 2008; Milczynski et al., 2009; Luo et al., 2012; Crew & Galvin, 2012; Wang et al., 2011; Deroche et al., 2014; Tao et al., 2015; Deroche et al., 2016; Deroche et al., 2019). Acoustic cues to emotional prosody include voice pitch and its changes, intensity, duration or speaking rate, and vocal timbre, but voice pitch is dominant among these cues (e.g., Banse & Scherer, 1996; Chatterjee et al., 2015). Therefore, the loss of voice pitch information results in significant deficits in CI users’ perception of vocal emotions (Luo et al. 2007; Hopyan-Misakyan et al. 2009; Chatterjee et al. 2015; Tinnemore et al. 2018; Barrett et al., 2020). Consistent with findings in CI listeners, identification of emotional prosody is also impaired in normally hearing listeners subjected to CI-simulated, or vocoded, speech (Shannon et al., 1995; Chatterjee et al., 2015; Tinnemore et al., 2018; Ritter and Vongpaisal 2018; Everhardt et al., 2020). Despite CI users’ limitations in pitch perception and in identification of prosodic cues, the average CI user shows excellent sentence recognition with high-context materials in quiet environments (e.g., James et al.,

2019). Thus, their access to lexical-semantic cues to emotion may exceed their access to prosodic cues to spoken emotions.

Depending on the complexity and informational richness of the lexical-semantic content, the process of reconstructing the meaning of words and sentences may be facilitated by top-down cognitive and linguistic resources available to the listener (Stickney & Assmann 2001; Moberly et al. 2017; Bosen & Luckasen 2019; Kirby et al. 2019). However, the reconstruction and understanding of prosodic information from a degraded input is more challenging without information about the talker's facial expression (Ritter and Vongpaisal, 2018) or the social context in which the conversation is occurring. We reason that under these conditions, in the relative paucity of prosodic cues, the listener's natural reliance on prosody for emotion identification is likely to be reduced in favor of a greater reliance on lexical-semantic cues. It is unclear, however, whether CI users, who have long experience with degraded speech, might show a different pattern of reliance than normally hearing listeners, who are accustomed to hearing high-fidelity acoustic representation of prosody and a strong reliance on prosodic cues for everyday emotional communication.

The specific condition of interest in this context is a condition when prosodic and lexical-semantic cues conflict with each other. Such incongruent conditions may be encountered when the speaker is communicating irony or sarcasm. In this condition, the tone of the voice often communicates the true feeling, while the lexical-semantic cue indicates the opposite. While a number of cues are associated with ironic or sarcastic delivery, one of these is a contrast between the intonational cues in an utterance and the expected intonation based on the meaning of the spoken words (e.g., Attardo et al., 2003). If a listener is unable to decipher the emotion in the tone and must rely on the lexical-semantic content only, they might lose the entire point of the communication. Other than deliberate incongruence between lexical-semantic and prosodic cues as when expressing sarcasm, misalignment between the two may occur in other situations. For instance, individuals may unintentionally reveal an inner emotion in the tone of voice when their spoken words are not intended to do so. An inability to understand or detect such misalignment could have a negative impact on their ability develop relationships, follow conversation, understand vocal emotion and intended message of a speaker, follow jokes, and everyday communication in general. Additionally, individual talkers may also vary in the degree to which these cues are coherently present in their speech, which can further impact understanding. Knowledge about how individuals perceive spoken emotions in these situations is likely to deepen our understanding of communication difficulties encountered by CI users. For patients and their communication partners, such knowledge may help improve and customize their communication strategies to better accommodate the limitations of listening with the device.

We use multiple sensory inputs to decipher emotions, and incongruence may occur between sensory modalities as well. Previous studies of audio-visual integration of emotional cues (facial expressions combined with speech prosody) have shown that emotion identification improves when two sources of information are available over the case when only one source of information is provided (e.g., Takagi et al., 2015; Massaro and Egan, 1996; Pell, 2002). In such studies, facial cues to emotion are consistently dominant over vocal cues, with a pattern that depends on the individual emotion. For instance, happiness shows greater

facial dominance than negative emotions such as sadness or anger. This dominance is observed in the incongruent conditions, where the emotional expression on the face conflicts with the vocal prosody in the audio signal. CI recipients' sensitivity to facial cues for emotion perception has been reported to be no different from NH observers, but compared to NH counterparts, CI users show a stronger effect of incongruency in emotion expression between auditory and visual modalities (Fengler et al., 2017). This may reflect a greater reliance on the visual modality for emotion in CI users compared to NH peers.

Studies of audiovisual integration of emotional cues, however, generally focus on vocal prosody in the audio signal by using emotion-neutral speech materials that are spoken with different audio-visual expressions. Relatively few studies have focused on the integration of two sources of emotional information within the same modality, the focus of the present study. Further, little is known about how different sources of information about emotion are weighted when one source is degraded relative to the other, a key focus of the present study.

Here, we tested the hypothesis that prosodic dominance in speech emotion recognition is altered when voice pitch cues are degraded. Specifically, we predicted that when prosodic and lexical-semantic cues to emotion were in conflict with each other, CI users would show greater reliance on lexical-semantic cues than normally hearing listeners. Similarly, we predicted that normally hearing listeners would show a parallel shift away from prosodic cues to lexical-semantic cues when attending to CI-simulated speech. A question of interest was whether the normally hearing participants would show a greater reliance on prosody than CI users even with CI-simulated speech, owing to the aforementioned differences between the two groups in their listening experiences. We designed experiments in which listeners were presented with spoken sentences that conveyed emotional meaning (lexical-semantic emotion), but were spoken with varying emotional prosody. In some cases, the lexical-semantic and prosodic cues to emotion were congruent. In other cases, they were incongruent. Listeners were asked to take both the words and the tone of the speaker into account and decide which emotion the talker intended to convey. Their responses were scored as correct or incorrect based on the lexical-semantic category of emotion. Thus, if the sentence "My father is the best" was spoken with sad prosody and the listener perceived it as "happy", the response was scored correct. On the other hand, if the listener perceived it as "sad", the response was scored incorrect. We predicted that listeners would show strong reliance on prosodic cues when signals were undistorted. In this case, they should show near-ceiling performance in the Congruent condition (when the prosodic cue would align with the lexical-semantic emotion category), and near-floor performance in the Incongruent condition (when the prosodic cue would be misleading). We also predicted that listeners would show reduced reliance on the prosodic cue when the signal was degraded, or when listening through a cochlear implant. In this case, they would show reduced performance in the Congruent condition (as the prosodic cue and the lexical-semantic cue would both be degraded). In the Incongruent condition, they would show improved performance, as they would rely more on the lexical-semantic cue to emotion.

## MATERIALS AND METHODS

### Participants

The participants in this study included 16 adults with normal hearing (11 women, five men, age range: 19.14 to 34.79 years, mean age: 26.04 years) and eight postlingually deafened adult CI users (three women, five men, age range: 34.00 to 71.76 years, mean age: 60.06 years). Relevant information about the participants with CI(s) is shown in Table 1. All participants were recruited and tested at Boys Town National Research Hospital (BTNRH) in Omaha, NE. Informed consent was obtained from each participant and protocols were approved by BTNRH's Institutional Review Board under protocol #11–24. All participants were compensated for their participation in the study and CI users were additionally compensated for any travel costs incurred.

### Stimuli

Stimuli were created for this project to convey emotion with both semantic and prosodic cues. A total of 50 sentences were initially composed with 10 sentences in each semantic emotion category (Angry, Happy, Neutral, Sad, and Scared). Frequency of word use and phonological neighborhood were calculated for each sentence using the Washington University St. Louis English Lexicon Database (Balota et al. 2007). The average of word frequency and phonological neighborhood for the keywords of the sentences were calculated. Word frequency calculations were completed because the frequency of a word, how common or easily identifiable the word is, impacts how quickly a word can be recognized (Luce and Pisoni, 1998). Similarly, phonological neighborhood density, or how many words in the language differ by one phoneme to the target word, also impacts how quickly a word can be recognized (Luce and Pisoni, 1998). The calculated values are shown in Figure 1 for the different emotions. A one-way ANOVA with  $\log_{10}(\text{Keyword Frequency Average})$  as the dependent variable and Semantic Emotion as the independent variable showed no significant effects of Semantic Emotion category. A similar analysis with Keyword Phonological Neighborhood Density as the dependent variable also showed no significant effect of Semantic Emotion category. The number of samples is relatively low in each case, so this result should be treated with caution.

These 50 sentences were informally surveyed for perceptual accuracy of the intended emotion based on lexical-semantic cues. A list of the sentences (in random order) was created and the five possible emotions were indicated next to each sentence. Fourteen adults, all colleagues at BTNRH, volunteered to read each sentence and indicated the emotion that seemed to be best communicated by it. The five sentences with the highest mean accuracy scores were chosen for each emotion to be audio- recorded. These sentences are listed in Table 2. Each of the 25 sentences were recorded with five prosodic emotions (Angry, Happy, Neutral, Sad, and Scared), resulting in 125 recorded sentences. Sentences were recorded using an AKG C 2000B recording microphone, the output of which was routed through an external A/D converter (Edirol UA-25X) to the PC. Recordings were made at 44,100 Hz and with 16-bit resolution. Adobe Audition (v. 3.0 and v. 6.0) were used to record and process the sound files. Each sentence was repeated three times to ensure a good quality sentence was produced. All sentences were recorded by two female members of the laboratory who

were native speakers of American English. Files were high-pass filtered (75 Hz corner frequency, 24 dB/octave roll off) to reduce any background noise. Of the three repetitions, the highest quality recording was selected to be used in the stimulus set. The selection was made informally by the researchers.

Sentences were noise band vocoded (NBV) using AngelSim software ([tigerspeech.com](http://tigerspeech.com), Emily Shannon Fu Foundation) to have eight or 16 channels of spectral information following the method described by Shannon et al. (1995). The analysis frequency range was limited to 200–7000 Hz. The bandpass filters (fourth-order Butterworth, –24 dB/octave slopes and corner frequencies based on the Greenwood map (Greenwood 2001)). The temporal envelope was extracted from the original speech via half-wave rectification and lowpass filtering (fourth-order Butterworth, –24 dB/octave slope, and a corner frequency of 400 Hz)

One sentence per lexical-semantic emotion (recorded in all five prosodic emotional styles) was removed from the test stimulus set to create a practice condition. The final set of stimuli included: 25 full-spectrum sentences for the practice condition, with 100 full-spectrum sentences and 16- and eight-channel noise-vocoded versions of the same for the test conditions. This set was repeated for both female talkers. Normal-hearing participants completed the three test conditions, per talker, twice. Talker order was counterbalanced by participant and the six test blocks per talker were randomized. Table 3 shows an example of the order of testing for an example normally hearing participant. CI recipients only heard the full-spectrum condition and talker order was counterbalanced by participant. Two examples each of sentences 1 and 3 from Table 2, one in the Congruent and one in the Incongruent condition, are provided for the reader's benefit (see Supplemental Digital Content (SDC) audio files). SDC 1 consists of the sentence *I won the game* from Table 2 (the first sentence in the lexical-semantic emotion category Happy) spoken with Happy prosody (Congruent), and unprocessed (full-spectrum condition). SDC 2 consists of the same sentence, but spoken with Sad prosody (Incongruent), and also full-spectrum. SDC 3 consists of the sentence *He might kill her*, the third sentence in the lexical-semantic emotion category Scared, spoken with Scared prosody (Congruent) and in full-spectrum condition. SDC 4 consists of the same sentence, but spoken with Happy prosody (Incongruent), and in full-spectrum condition. SDC 5, SDC 6, SDC 7, and SDC 8 are the same recordings, but presented in the 16-channel noise-vocoded condition. Similarly, SDC 9, SDC 10, SDC 11, and SDC 12 are the same recordings, but presented in eight-channel noise-vocoded condition.

### Testing Protocol

Prior to completing the emotion recognition task, normally hearing participants received an air-conduction pure tone hearing evaluation from 250–8000 Hz using TDH headphones. The participants had a PTA (500–4000Hz) of 20 dB or better with no thresholds poorer than 25 dB in either ear. Both groups of subjects were read scripted instructions and the task was explained in detail. Participants then completed the emotion recognition task. To complete the task, participants listened to each recording and indicated which of the five emotions they thought it was most closely associated with. The experiment was controlled using a custom Matlab-based software program. The protocol has been described in detail

in previous publications (Cannon & Chatterjee, 2019; Christensen et al., 2019). Briefly, stimuli were played in randomized order. After each stimulus was played, the five choices of emotion appeared at the right-hand side of the screen, with cartoon images depicting each emotion and a text label (Happy, Angry, etc.) appearing below each image. The participant's task was to click on their selected emotion. Repetitions were not allowed, and no feedback was provided. No training was provided, but participants completed the practice condition described previously. Stimuli were presented at 65 dB (+/- 2dB) SPL and calibrated to a 1 kHz tone at the same root mean square level as the mean level of the stimuli within the block. Finally, participants completed an emotion evaluation survey in which they were presented with the sentences in written form and assessed the emotion associated with each sentence. This provided a sense of the degree to which each sentence successfully communicated the lexical-semantic emotion category. Participants were provided breaks as needed during testing and compensation for their participation in the study.

Both NH and CI users identified the intended semantic emotions with high levels of accuracy (Figure 2). A linear mixed effects modeling analysis of the NH participants' data with Semantic Emotion as a fixed effect and random slopes for the effect of Emotion by subject showed no effect of Semantic Emotion on the percent correct scores ( $\beta=0.313$ , s.e.=0.87,  $t(52.98)=0.355$ ,  $p=0.724$ ). Ceiling effects were clearly present, and this was reflected in a non-normal distribution of residuals. Therefore, a nonparametric Kruskal-Wallis test was also conducted and showed no effects of emotion category ( $\chi^2=3.90$ ,  $df=4$ ,  $p=0.42$ ). Parallel analyses of the CI users' data also showed no effect of semantic emotion in the LME analysis ( $\beta=-0.5958$ , s.e.=0.75,  $t(90.024)=-0.796$ ,  $p=0.428$ ). As in the NH data, ceiling effects skewed the distribution of model residuals. Therefore, the Kruskal-Wallis test was conducted on the CI listeners' data and also showed no effect of Semantic Emotion ( $\chi^2=7.09$ ,  $df=4$ ,  $p=0.1312$ ).

### Statistical Analyses

Statistical analyses were conducted in R v. 3.6.1 (R Core Team 2019). Plots were created using *ggplot2* (Wickham 2016). Linear mixed-effects models were constructed using *lme4* (Bates et al. 2015). Plots and histograms of model residuals, as well as normal quantile-quantile plots (qqnorm plots) were visually inspected to ensure normality and confirm model fit. The *lmerTest* package (Kuznetsova et al. 2017) was used in conjunction with *lme4* to obtain model results. The results reported include the estimated coefficient ( $\beta$ ) and its standard error (s.e.), and the p-value estimated by the *lmerTest* package which uses Satterthwaite's method to estimate degrees of freedom. When the fully complex model with random intercepts and slopes did not converge, the random effects structure was simplified until convergence was reached. Post-hoc multiple paired comparisons were completed using t-tests with the Holm (1979) correction; all p-values reported are corrected values. Prior to analyses, outliers were removed from the datasets. Tukey fences (Tukey, 1977) were used to determine outliers. This was done separately for datasets corresponding to NH and CI participants. Overall, 6.7% of the full data set were deemed outliers, of which 6.3%, 7.9%, 7.0% of the NH full-spectrum, 16-channel and eight-channel datasets were outliers respectively, and 4.3% of the data obtained with CI participants (full-spectrum only).

The model approach included two methods, one we refer to as “Linear Coding” (in which the levels within the factors Emotion and Spectral Condition were mapped to ordered integers) and the second as “Categorical Coding” (in which the levels were dummy-coded prior to analysis). The Linear Coding approach is appealing for its greater parsimony and simplicity (e.g., Lazic, 2008). On the other hand, this coding scheme makes the untestable assumption that the distances between levels within the category can legitimately be represented on an interval scale. The Categorical Coding scheme does not make such assumptions. For each predictor, the *anova* function in R was used to determine whether or not that predictor contributed significantly to the model. This function compares the goodness of fit of one model with a second version that includes a new predictor, based on chi-square analysis of likelihood ratios. The levels in the variables Spectral Condition and Emotion were coded as ordered integers and these predictors were treated as continuous variables in the “Linear Coding” approach. The mapping between the category levels and the integers was done such that assumptions of linearity were not violated (e.g., Pasta, 2009). This was done by visual inspection of both the data and of residuals after linear models were fitted to the data. For Spectral Condition, full-spectrum was coded as 0, 16-channel as 1, and eight-channel as 2. For Emotion, Happy was coded as 1, Scared as 2, Neutral as 3, Sad as 4, and Angry as 5. Effects of individual emotions or talkers were not of central interest, but they were informative about how robust the findings were to variations in talker and emotion. Post-hoc analyses guided interpretation of the findings. For Categorical Coding, we relied on dummy-coding of the levels in the categorical variables, the default in the *lme4* package in R. Model residual distributions confirmed goodness of fit of the final models in both approaches. To obtain greater insight into effects and interactions, post-hoc analyses were conducted, primarily comprising pairwise t-tests. Results obtained with the two kinds of analyses were consistent with one another, but some minor differences remained. To present the full picture, we report results obtained with both kinds of analyses below. Unless otherwise indicated, effects that were not found to be significant are not reported for the sake of brevity.

## RESULTS

### 1. Effects of Congruency and Spectral Degradation in Participants with NH

Figure 3 shows boxplots of the NH participants’ accuracy scores (proportion of lexical-semantic emotions correct) obtained with full-spectrum, 16-channel and eight-channel versions of the materials under Congruent and Incongruent conditions. The ordinate shows accuracy (percent correct scored on the lexical semantic emotion) for the two talkers’ recordings (top and bottom row) for each condition of spectral degradation (abscissa) and each lexical-semantic emotion category (columns), and for Congruent (grey) and Incongruent (black) conditions. The general pattern of results supports our hypotheses. When the signal is undistorted (full-spectrum), listeners showed best performance in the Congruent condition and worst performance in the Incongruent condition, suggesting greater reliance on prosodic cues overall. As the signal became more and more degraded, performance decreased in the Congruent condition and improved in the Incongruent condition, suggesting greater reliance on the lexical-semantic cues to emotion. Note that in the Incongruent condition, scores ranged from below chance in the full-spectrum condition



to at or above chance in the eight-channel condition (chance is at 20% correct in this task). This suggests that participants were misled by the salient prosodic cues in the full-spectrum, Incongruent condition.

A linear mixed effects analysis with Spectral Condition, Talker, Emotion, and Congruency as fixed effects was constructed. Results obtained with Linear and Categorical Coding are presented below.

Analysis with Linear Coding: The most complex model including random slopes for all factors did not converge, indicating that the sample size was too small to support the complex random effects structure. A model with a simpler random effects configuration -- subject-based random intercepts, and subject-based random slopes for the effect of Emotion -- was constructed and did converge. This was based on an expectation that the greatest intersubject variation would be based on emotion-based differences in how participants weigh the cues. Results showed significant effects of Spectral Condition (estimated coefficient ( $\beta$ )=-0.078, standard error (s.e.)=0.025;  $t(928)=-3.09$ ,  $p=0.002$ ), Emotion ( $\beta=-0.036$ , s.e.=0.011;  $t(244.84)=-3.393$ ,  $p=0.0008$ ) and Congruency ( $\beta=-0.921$ , s.e.=0.046;  $t(928)=-19.972$ ,  $p<0.0001$ ). The effect of Congruency is observed in Fig. 3 and Fig. 5A, with higher overall accuracy for congruent vs. incongruent cues. Most relevant to our research question a significant interaction between Spectral Condition and Congruency ( $\beta=0.162$ , s.e.=0.036;  $t(928)=4.529$ ,  $p<0.0001$ ) was observed. A significant interaction between Emotion and Congruency was also observed ( $\beta=0.046$ , s.e. = 0.013;  $t(928)=3.298$ ,  $p = 0.001$ ). Although a main effect of Talker was not found, a significant three-way interaction was observed between Talker, Emotion and Congruency.

Post-hoc analyses focused on obtaining deeper understanding of the interactions. The interaction between Spectral Condition and Congruency was followed up with post-hoc pairwise t-tests (Holm correction for multiple comparisons: Holm, 1979) on the effects of Spectral Condition conducted separately for Congruent and Incongruent conditions. In both Congruent and Incongruent conditions, significant differences were found between full-spectrum, 16-channel and eight-channel conditions. In the Congruent condition, accuracy in the full-spectrum condition was found to be significantly better than in the 16-channel ( $p=0.0011$ ) and eight-channel ( $p<0.00001$ ) conditions, and accuracy in the 16-channel condition was significantly better than in the eight-channel condition ( $p=0.003$ ). In the Incongruent condition, accuracy in the full-spectrum condition was significantly worse than in the 16-channel condition ( $p<0.00001$ ) and in the eight-channel condition ( $p<0.00001$ ) and accuracy in the 16-channel condition was significantly worse than in the eight-channel condition ( $p<0.0001$ ). These findings are consistent with the patterns observed in Fig 3, and confirm the hypothesis that listeners place greater weight on lexical-semantic cues and less weight on prosodic cues to emotion as the degree of spectral degradation increases.

The interaction between Emotion and Congruency was followed up by conducting separate pairwise comparisons between emotions for Incongruent and Congruent conditions. For the Congruent condition, pairwise t-tests with Holm correction showed significant differences between Angry and Happy ( $p<0.0001$ ), Angry and Neutral ( $p=0.019$ ), and Happy and Sad ( $p<0.0001$ ) conditions. For the Incongruent condition, significant differences were observed

between Angry and Neutral ( $p=0.012$ ), Happy and Scared ( $p=0.049$ ), and Neutral and Scared ( $p<0.0001$ ) conditions. The remaining comparisons were not significant.

The effect of Talker modified the Emotion x Congruency interaction, as revealed in the three-way interaction. This was investigated by repeating the pairwise t-tests above for the two talkers separately. For Talker 1, in the Congruent condition, a number of differences were observed between the emotions. Thus, significant differences were observed between Angry and Happy ( $p<0.0001$ ), Angry and Neutral ( $p<0.0001$ ), Angry and Sad ( $p=0.0004$ ), and Angry and Scared ( $p=0.0315$ ) emotions. The remaining differences were not significant. In the Incongruent condition, Talker 1's materials only produced significant differences between Angry and Scared ( $p=0.007$ ) and between Neutral and Scared ( $p=0.0046$ ). The remaining differences were not significant.

On the other hand, for Talker 2, in the Congruent condition, significant differences were observed between Happy and Neutral ( $p=0.0423$ ), Scared and Neutral ( $p=0.001$ ), Angry and Sad ( $p=0.0051$ ) and Happy and Sad ( $p<0.0001$ ). The remaining differences were not significant. In the Incongruent condition for this talker, significant differences were only observed between Angry and Neutral ( $p<0.0001$ ) and Scared and Neutral ( $p=0.0029$ ). The remaining differences were not significant.

Analysis with Categorical Coding: Results were generally consistent with those found with Linear Coding. Note that as dummy coding was used, the effects of the different emotions are only reported with reference to the Angry emotion. Significant effects of Spectral Condition (full spectrum vs. 16-channel:  $\beta=0.086$ ,  $s.e.=0.039$ ,  $t(896.04)=2.344$ ,  $p=0.019$ ), Emotion (Happy vs. Angry:  $\beta=0.087$ ,  $s.e.=0.041$ ,  $t(225.46)=2.135$ ,  $p=0.034$ ; Sad vs. Angry:  $\beta=-0.1214$ ,  $s.e.=0.045$ ,  $t(110.27)=-2.7$ ,  $p=0.008$ ), Congruency (lower scores for Incongruent than for Congruent stimuli,  $\beta=-0.59$ ,  $s.e.=0.035$ ,  $t(896.04)=-16.727$ ,  $p<0.00001$ ), and Talker ( $\beta=-0.123$ ,  $s.e.=0.035$ ,  $t(896.04)=-3.491$ ,  $p<0.001$ ) were observed. The most important interaction, between Spectral Condition and Congruency, was also observed in this analysis: the effect of Congruency differed between 16-channel and eight-channel conditions ( $\beta=0.181$ ,  $s.e.=0.045$ ,  $t(896.04)=4.005$ ,  $p<0.0001$ ), and also between 16-channel and full-spectrum conditions ( $\beta=-0.225$ ,  $s.e.=0.045$ ,  $t(896.04)=-4.996$ ,  $p<0.00001$ ). The effect of Emotion (only the difference between Angry and Happy) was modified by Congruency ( $\beta=0.137$ ,  $s.e.=0.047$ ,  $t(896.04)=-2.875$ ,  $p=0.004$ ). The effect of Spectral Condition was also different for the two talkers (full-spectrum vs. eight-channel condition,  $\beta=-0.114$ ,  $s.e.=0.045$ ,  $t(896.04)=-2.533$ ,  $p=0.011$ ). The effect of Emotion interacted with Talker (Angry vs. Happy,  $\beta=0.098$ ,  $s.e.=0.048$ ,  $t(896.04)=2.054$ ,  $p=0.04$ ; Angry vs. Neutral,  $\beta=0.186$ ,  $s.e.=0.048$ ,  $t(896.04)=3.909$ ,  $p<0.0001$ ; Angry vs. Sad,  $\beta=0.243$ ,  $s.e.=0.048$ ,  $t(896.04)=5.107$ ,  $p<0.00001$ ). Three way interactions were also observed. Thus, the difference between Angry and Sad emotions interacted with Congruency and Spectral Condition (difference between full-spectrum and eight channel conditions,  $\beta=-0.21$ ,  $s.e.=0.058$ ,  $t(896.04)=-3.622$ ,  $p=0.0003$ ). A three way interaction was also observed between Talker, Condition and Emotion (the difference between Angry and Neutral emotions interacted with the difference between full-spectrum and eight channels and with Talkers ( $\beta=0.131$ ,  $s.e.=0.058$ ,  $t(896.04)=2.247$ ,  $p=0.025$ ), and the difference

between Angry and Scared emotions in the eight-channel condition was also modified by Talker ( $\beta=0.136$ ,  $s.e.=0.058$ ,  $t(896.04)=2.347$ ,  $p=0.019$ ).

In summary, these analyses showed significant interactions between Congruency and Spectral Condition, in the direction supporting the hypothesis that listeners place greater weight on lexical-semantic cues and reduced weight on prosodic cues as the signal becomes increasingly degraded. Talker-based differences between the individual emotions were observed that were different for Congruent and Incongruent conditions, with more differences between emotions being observed in Congruent than in Incongruent conditions. In the Congruent conditions, the prosodic cues, which align with the lexical semantic cues, are likely to play more of a role than in Incongruent conditions. The lexical-semantic cues were not changing between talkers, but the prosodic cues were likely to vary from emotion to emotion and talker to talker. Thus, it is not surprising that talker-based differences between emotions are more apparent in the Congruent than in the Incongruent condition. Results were generally consistent and parallel between the two forms of coding of the variables. Both supported the crucial element of our hypothesis – the predicted interaction between Spectral Condition and Congruency. Talker-based variations were evident in both forms of analyses. Interactions between Emotion and Congruency were also found in both analyses. Details of the interactions were different, as expected because of the differences in coding and in the forms of analyses.

**Analyses excluding the Neutral emotion**—The Neutral semantic emotion may be thought to be somewhat different from the rest in this experiment: in this case, the Incongruent condition may not be thought of as truly incongruent in the sense of presenting two emotions that are as contrasting as, for instance, Happy and Sad. The results in the Neutral condition, however, seemed consistent with those in the other emotion conditions (Fig. 3).

A re-analysis of the data excluding the Neutral condition showed no essential differences in the findings. For the sake of brevity, we only report the results obtained with Categorical Coding. As with previous analyses, we obtained significant effects of Emotion (lower scores with Sad relative to Angry,  $\beta=-0.177$ ,  $s.e.=0.04144$ ;  $t(752)=-4.352$ ,  $p<0.0001$ ) Spectral Condition (lower scores with 16-channels than with full-spectrum,  $\beta=-0.093$ ,  $s.e.=0.04$ ,  $t(752)=-2.323$ ,  $p=0.02$ ; lower scores with eight-channel than with full-spectrum,  $\beta=-0.111$ ,  $s.e.=0.04$ ,  $t(752)=-2.752$ ,  $p=0.006$ ), Congruency (lower scores with Incongruent than with Congruent stimuli,  $\beta=-0.819$ ,  $s.e.=0.0394$ ;  $t(752)=-21.224$ ,  $p<0.0001$ ), and Talker ( $\beta=-0.081$ ,  $s.e.=0.039$ ,  $t(752)=-2.105$ ,  $p=0.036$ ). Crucially, and consistent with previous analyses, a significant interaction between Spectral Condition and Congruency was observed (higher accuracy for the 16-channel condition than the full-spectrum condition in the Incongruent condition relative to the Congruent condition,  $\beta=0.244$ ,  $s.e.=0.05$ ;  $t(752)=4.804$ ,  $p<0.00001$ ; similarly, for the eight-channel condition re: full-spectrum,  $\beta=0.405$ ,  $s.e.=0.05$ ,  $t(752)=8.12$ ,  $p<0.00001$ ). A significant interaction between Emotion and Congruency was also observed (specifically for the difference between Angry and Sad emotions,  $\beta=0.18$ ,  $s.e.=0.05$ ;  $t(752)=3.503$ ,  $p=0.00053$ ). A two-way interaction between Emotion and Condition was also observed (the difference between full-spectrum and eight-channel conditions interacted with the Angry-Sad difference,  $\beta=0.124$ ,  $s.e.=0.05$ ,  $t(752)=-2.271$ ,

$p=0.023$ ). Other significant two-way interactions were observed between Talker and Condition (difference between Talkers interacted with the full-spectrum – eight-channel difference,  $\beta=-0.162$ ,  $s.e.=0.05$ ,  $t(752)=-3.242$ ,  $p=0.001$ ), Emotion and Talker (the Angry - Sad difference interacted with Talker differences,  $\beta=0.245$ ,  $s.e.=0.05$ ,  $t(752)=4.754$ ,  $p<0.00001$ ). Three-way interactions between Condition, Emotion and Congruency were found (full-spectrum – eight-channel difference interacted with the Angry-Happy difference for the effect of Congruency,  $\beta=-0.144$ ,  $s.e.=0.063$ ,  $t(752)=-2.292$ ,  $p=0.022$ ; the same interaction was found for Emotion changing from Angry to Sad,  $\beta=-0.305$ ,  $s.e.=0.063$ ,  $t(752)=-4.832$ ,  $p<0.00001$ ). A three-way interaction between Spectral Condition, Emotion, and Talker was also significant (full-spectrum – eight-channel difference interacted with the Angry-Happy difference for the effect of Talker,  $\beta=0.205$ ,  $s.e.=0.063$ ,  $t(752)=3.252$ ,  $p=0.0012$ ). Finally, a three-way interaction was observed between Emotion, Congruency, and Talker (interaction between Congruency and Talker effects when Emotion changed from Angry to Sad,  $\beta=-0.148$ ,  $s.e.=0.051$ ,  $t(752)=-2.87$ ,  $p=0.0042$ ).

## 2. Comparison of results obtained in NH and CI listeners

Figure 4 compares the results obtained with NH and CI listeners in the full-spectrum condition. A linear mixed effects analysis was conducted with Talker, Congruency, Hearing Status (NH, CI) and Emotion as fixed effects, and subject-based random intercepts (model attempts with more complex random effects structures did not converge, likely because of the limited sample size). Again, analyses with both Linear Coding and Categorical Coding are reported below.

Results with Linear Coding: Results showed significant effects of Talker ( $\beta=-0.167$ ,  $s.e.=0.083$ ;  $t(189.57)=2.021$ ,  $p=0.046$ ), Congruency ( $\beta=-0.41$ ,  $s.e.=0.082$ ;  $t(189.57)=-5.033$ ,  $p<0.0001$ ), a significant interaction between Emotion and Talker ( $\beta=0.0578$ ,  $s.e.=0.025$ ;  $t(109.14)=2.345$ ,  $p=0.021$ ), and most relevant to the present study, a significant interaction between Congruency and Hearing Status ( $\beta=-0.278$ ,  $s.e.=0.090$ ;  $t(129.34)=-3.101$ ,  $p=0.0024$ ). No other significant interactions were observed. The interaction of greatest interest, that between Hearing Status and Congruency, can be observed in Fig. 4, and indicates that CI users rely less on prosodic cues in the Congruent condition than NH listeners and more on semantic cues in the Incongruent condition than the NH listeners. This finding, which supports our hypothesis, was confirmed by conducting follow-up one-way ANOVAs (with Accuracy as the dependent variable and Hearing Status as the independent variable) on the results obtained in the Congruent and Incongruent conditions separately. In the Congruent condition, the ANOVA showed a significant effect of Hearing Status ( $F(1, 238)= 58.96$ ,  $p<0.0001$ ). In the Incongruent condition, the greater reliance on CI users on the lexical-semantic cues was also confirmed by the ANOVA, which showed a significant effect of Hearing Status on Accuracy ( $F(1, 238)=133$ ,  $p<0.0001$ ).

The Emotion x Talker interaction was pursued by conducting a pairwise t-test for the effect of Emotion for the two talkers separately. No significant differences were found across emotions for either talker, and the interaction could not be supported: this was possibly due to a lack of statistical power.

Analyses with Categorical (Dummy) Coding: Results were largely consistent with those obtained with Linear Coding. Results showed significant effects of Talker ( $\beta=-0.201$ ,  $s.e.=0.06$ ,  $t(1096)=-3.406$ ,  $p=0.0006$ ) and Congruency ( $\beta=-0.416$ ,  $s.e.=0.06$ ,  $t(1096)=-7.038$ ,  $p<0.00001$ ). Significant interactions between Emotion and Talker, Congruency and Talker, and again, the crucially significant interaction between Congruency and Hearing Status, were observed. The Emotion x Talker interaction was reflected in significant differences between Talkers 1 and 2 for differences between Angry and all other emotions (Happy:  $\beta=0.192$ ,  $s.e.=0.078$ ,  $t(1096)=2.458$ ,  $p=0.014$ ; Neutral:  $\beta=0.239$ ,  $s.e.=0.078$ ,  $t(1096)=0.002$ ; Sad:  $\beta=0.244$ ,  $s.e.=0.078$ ,  $t(1096)=3.123$ ,  $p=0.002$ ; Scared:  $\beta=0.302$ ,  $s.e.=0.078$ ,  $t(1096)=3.867$ ,  $p=0.0001$ ). The Congruency x Talker interaction was observed in a significantly different effect of Talker on the Congruency effect ( $\beta=0.157$ ,  $s.e.=0.058$ ,  $t(1096)=2.711$ ,  $p=0.007$ ). The Congruency x Hearing Status interaction was as expected, with NH listeners showing significantly lower scores than CI listeners in the Incongruent condition ( $\beta=-0.194$ ,  $s.e.=0.061$ ,  $t(1096)=-3.178$ ,  $p=0.002$ ). Two three-way interactions were observed. The first showed that the Congruency x Hearing Status interaction was weakly but significantly different for the Sad emotion relative to the Angry emotion ( $\beta=0.155$ ,  $s.e.=0.079$ ,  $t(1096)=1.97$ ,  $p=0.049$ ). The second showed that the Congruency x Talker interaction was significantly different for the Angry vs. the Sad emotion ( $\beta=-0.155$ ,  $s.e.=0.06$ ,  $t(1096)=-2.813$ ,  $p=0.005$ ).

The analysis was repeated excluding the Neutral condition. For the sake of brevity, only the results of the analysis with Categorical Coding are reported here. Significant effects of Hearing Status (higher accuracy for NH than CI participants,  $\beta=0.125$ ,  $se=0.056$ ,  $t(318.9)=2.242$ ,  $p=0.026$ ), Emotion (lower accuracy for Scared than Angry,  $\beta=-0.14$ ,  $se=0.06$ ,  $t(360)=-2.324$ ,  $p=0.021$ ), Talker ( $\beta=-0.234$ ,  $se=0.06$ ,  $t(360)=-3.7$ ,  $p=0.0001$ ), and Congruency (lower accuracy for Incongruent than Congruent stimuli,  $\beta=-0.449$ ,  $se=0.06$ ,  $t(360)=-7.423$ ,  $p<0.00001$ ) were found. The crucial two-way interaction between Congruency and Hearing Status was observed here, in the same direction as in previous analyses ( $\beta=-0.344$ ,  $se=0.074$ ,  $t(360)=-4.638$ ,  $p<0.00001$ ). Significant interactions between Emotion and Hearing Status (larger difference between Angry and Scared emotions for the NH participants than the CI participants,  $\beta=0.219$ ,  $se=0.074$ ,  $t(360)=2.951$ ,  $p=0.0034$ ; smaller difference between Angry and Sad emotions for NH than for CI participants,  $\beta=-0.156$ ,  $se=0.074$ ,  $t(360)=-2.108$ ,  $p=0.036$ ) were observed. Two way interactions involving Talkers were observed with Hearing Status, Emotion, and Congruency. Thus, the difference between Talkers 1 and 2 was larger for NH than CI listeners ( $\beta=-0.18$ ,  $se=0.074$ ,  $t(360)=2.424$ ,  $p=0.016$ ), and larger for Happy ( $\beta=0.234$ ,  $se=0.086$ ,  $t(360)=2.739$ ,  $p=0.0065$ ), Sad ( $\beta=0.266$ ,  $se=0.086$ ,  $t(360)=3.104$ ,  $p=0.0021$ ), and Scared ( $\beta=0.375$ ,  $se=0.086$ ,  $t(360)=4.382$ ,  $p<0.0001$ ) emotions than for Angry; and for Congruent than Incongruent stimuli ( $\beta=0.23$ ,  $se=0.086$ ,  $t(360)=2.602$ ,  $p=0.01$ ). A three-way interaction was observed between Hearing Status, Emotion, and Congruency (Hearing Status modified the interaction between the Angry-Sad difference and Congruency,  $\beta=0.283$ ,  $se=0.105$ ,  $t(360)=2.702$ ,  $p=0.007$ ). Three-way interactions were also observed between Hearing Status, Emotion, and Congruency (Hearing Status modified the interaction between the Angry-Scared difference and Congruency,  $\beta=-0.492$ ,  $se=0.105$ ,  $t(360)=-4.696$ ,  $p<0.00001$ ), and between Hearing Status, Congruency, and Talker ( $\beta=-0.263$ ,  $se=0.105$ ,  $t(360)=-2.255$ ,

$p=0.025$ ). The latter was further modified by Emotion in a four-way interaction when Emotion changed from Angry to Scared ( $\beta=0.385$ ,  $se=0.148$ ,  $t(360)=2.596$ ,  $p=0.01$ ).

Figures 5A and 5B show the primary effects and interactions of interest in the accuracy data. Figure 5A shows the data averaged across semantic emotion categories for the NH listeners, plotted against the degree of spectral degradation, and separated by Talker (rows) and Congruency (columns). Figure 5B compares the full-spectrum data from the NH listeners with the corresponding data obtained in CI listeners. Individual data points are overlaid on the boxplots in each case. These summary plots show clearly that NH listeners show greater reliance on the lexical-semantic cues and reduced reliance on prosodic cues to emotion as the signal becomes increasingly degraded. The side by side comparison in Fig. 5B also shows that a similar effect is observed in CI listeners compared to NH listeners, where the CI listeners show increased reliance on lexical semantic cues and reduced reliance on prosodic cues than the NH listeners. The CI listeners' performance appears to align most closely with the NH listeners' performance with eight-channel vocoding.

## Analysis of Reaction Times

### **1. Effects of Congruency and Spectral Degradation on Reaction Times in**

**Participants with NH:** While the focus of this study was on the influence of spectral degradation on emotion recognition, reaction time data were also analyzed to interrogate effects of spectral degradation and hearing status. Analyzing these data is of interest as reaction time data can show an influence of listening effort, cognitive load, and intelligibility on the ability to complete the task (Pals et al., 2015). The reaction times were analyzed for effects of Congruency, Spectral Condition, Emotion and Talker for the NH and for CI participants. Based on outlier analysis, reaction times longer than 2.839s were excluded from analyses. For the NH listeners' data, initial LME analyses produced residuals with a slight rightward skew and a mildly concave qqnorm plot, suggesting a slight departure from normality in the distribution. Log-transforming the reaction times successfully addressed this issue. LME analysis on the log-transformed reaction times with fixed effects of Spectral Condition, Congruency, Emotion and Talker including random subject-based intercepts and slopes was conducted on the NH listeners' data.

The fully complex model with random slopes did not converge, so a simpler model with only subject-based random intercepts was implemented. No effects of Talker were observed. The final model showed significant effects of Congruency ( $\beta=0.052$ ,  $s.e.=0.005$ ,  $t(9428)=10.105$ ,  $p<0.0001$ ), Spectral Condition ( $\beta=0.018$ ,  $s.e.=0.003$ ,  $t(9428)=7.102$ ,  $p<0.0001$ ) and Emotion ( $\beta=-0.004$ ,  $s.e.=0.0015$ ,  $t(9428)=-2.818$ ,  $p=0.005$ ). Reaction times were slightly longer for Incongruent than for Congruent stimuli, which can be observed in Figure 6A. A post-hoc pairwise t-test (Holm correction for multiple comparisons) showed that the reaction times were significantly shorter for the full-spectrum condition than for the 16-channel and the eight-channel conditions ( $p<0.00001$  in each case), but no significant differences were found between the 16-channel and the eight-channel conditions. Pairwise t-tests following up on the effect of Emotion (Holm correction applied) showed only significant differences between Neutral and Sad ( $p=0.001$ ) and between Neutral and Scared ( $p=0.0009$ ) emotions.

Results with Categorical Coding: As with linear coding, the fully complex model did not converge, and the final model included fixed effects of Spectral Condition, Talker, Emotion, and Congruency, with subject-based random intercepts. The results in this case showed a significant effect of Emotion (reaction times were shorter for Sad than for Angry,  $\beta=-0.07$ ,  $s.e.=0.03$ ,  $t(9398)=-2.332$ ,  $p=0.02$ ), which is modified in a two-way interaction with Congruency ( $\beta=0.079$ ,  $s.e.=0.032$ ,  $t(9398)=2.484$ ,  $p=0.013$ ). The effect of congruency also differed with Talker in a two-way interaction ( $\beta=0.063$ ,  $se=0.027$ ,  $t(9396)=2.373$ ,  $p=0.018$ ). This was further modified in a 3-way interaction between Talker, Congruency, and Emotion, with a reduction in the effect for the Sad emotion than for Angry ( $\beta=-0.087$ ,  $se=0.032$ ,  $t(9396)=-2.732$ ,  $p=0.006$ ). A significant two-way interaction was also observed between Emotion and Talker, in which the difference in reaction time between the Angry and Scared emotions was different between talkers ( $\beta=0.077$ ,  $se=0.034$ ,  $t(9396)=2.284$ ,  $p=0.022$ ). Finally, the Congruency x Talker interaction was modified in a three-way interaction with Emotion, specifically the difference between Angry and Sad emotions ( $-\beta=0.087$ ,  $se=0.032$ ,  $t(9396)=-2.732$ ,  $p=0.006$ ).

## **2. Comparison of NH and CI Participants' Reaction Times in the Full-Spectrum**

**Condition:** Based on outlier analysis of the CI participants' dataset, reaction times longer than 7.51 s were excluded from analysis. Initial LME analyses of the NH and CI participants' reaction times in the full-spectrum condition also showed residuals with a rightward skewed distribution and concave qqnorm plot. Again, log-transforming the reaction times addressed the issue. An LME analysis was conducted on the log-transformed reaction times with Hearing Status, Talker, Emotion and Congruency as fixed effects. A fully complex model with subject-based random slopes did not converge. The most complex model that converged had subject-based random slopes for the effects of Congruency and Emotion, including their interaction. Results showed significant effects of Hearing Status ( $\beta=-0.519$ ,  $s.e.=0.073$ ,  $t(37.38)=-7.148$ ,  $p<0.00001$ ) and Talker ( $\beta=-0.174$ ,  $s.e.=0.051$ ,  $t(4699)=-3.412$ ,  $p=0.0006$ ), but not Emotion or Congruency. Significant two-way interactions between Talker and Hearing Status ( $\beta=0.063$ ,  $s.e.=0.027$ ,  $t(4699.81)=2.337$ ,  $p=0.0195$ ), Talker and Congruency ( $\beta=0.169$ ,  $s.e.=0.057$ ,  $t(4699)=2.961$ ,  $p=0.003$ ), and Talker and Emotion ( $\beta=0.421$ ,  $s.e.=0.015$ ,  $t(4699)=2.73$ ,  $p=0.0064$ ), were observed. Three way interactions between Talker, Congruency and Hearing Status ( $\beta=-0.233$ ,  $s.e.=0.07$ ,  $t(4699)=-3.331$ ,  $p=0.0009$ ), between Talker, Congruency and Emotion ( $\beta=-0.041$ ,  $s.e.=0.017$ ,  $t(4699)=-2.4$ ,  $p=0.0164$ ), and between Talker, Hearing Status and Emotion ( $\beta=-0.048$ ,  $s.e.=0.019$ ,  $t(4699)=-2.514$ ,  $p=0.012$ ) were also observed. Finally, a four-way interaction between Talker, Congruency, Hearing Status, and Emotion was observed ( $\beta=0.0568$ ,  $s.e.=0.021$ ,  $t(4699)=2.687$ ,  $p=0.0072$ ).

As all of the interactions observed involved Talker effects, a follow-up LME analysis (using Linear Coding) was conducted for each Talker separately.

For Talker 1's materials, a significant effect of Congruency ( $\beta=0.054$ ,  $s.e.=0.01$ ,  $t(2325)=5.52$ ,  $p<0.00001$ ) was observed, with incongruent stimuli producing longer reaction times than congruent ones. A significant effect of Hearing Status was also observed ( $\beta=-0.469$ ,  $s.e.=0.053$ ,  $t(24)=-8.9$ ,  $p<0.00001$ ) with longer reaction times obtained with CI users than NH listeners, along with a weak effect of Emotion ( $\beta=-0.007$ ,  $s.e.=0.004$ ,

$t(24)=-2.091, p=0.0473$ ). A follow-up pairwise  $t$ -test of the effect of Emotion did not show significant differences between the reaction times obtained for the different semantic emotion categories. No significant interactions were observed.

A similar LME conducted on the results obtained with Talker 2's materials showed similar results, with a significant effect of Congruency, with longer reaction times for incongruent stimuli ( $\beta=0.057, s.e.=0.01, t(2347)=5.747, p<0.00001$ ); a significant effect of Hearing Status ( $\beta=-0.536, s.e.=0.043, t(24.59)=-12.356, p<0.00001$ ). One difference from Talker 1's materials was that no effect of Emotion was observed with Talker 2's materials. Similar to Talker 1's materials, no interactions were observed with Talker 2's materials.

Analyses with the Categorical coding scheme yielded similar results. Significant overall effects of Hearing Status (shorter reaction times for NH participants;  $\beta=-0.511, se=0.063, t(57.91)=-8.12, p<0.00001$ ) and Talker ( $\beta=-0.153, se=0.05, t(494.6)=-3.094, p=0.002$ ) were found, but not of Congruency or Emotion. Two way interactions between Talker and Congruency ( $\beta=0.147, se=0.055, t(11010)=2.66, p=0.008$ ), Talker and Hearing Status ( $\beta=0.115, se=0.053, t(11010)=2.158, p=0.031$ ), and Talker and Emotion (significantly different effect of Talker when emotion changed from Angry to Sad,  $\beta=0.210, se=0.07, t(11010)=2.992, p=0.003$ ) were observed. The interaction between Talker and Congruency was modified in a three-way interaction between Talker, Congruency, and Emotion, specifically, when the emotion changed from Angry to Scared ( $\beta=-0.205, se=0.08, t(11010)=-2.609, p=0.009$ ), and this was modified by Hearing Status in a four-way interaction between Congruency, Hearing Status, Emotion and Talker ( $\beta=0.195, se=0.085, t(11010)=2.296, p=0.022$ ).

Overall, both Talkers' materials showed that CI users had a longer reaction time (mean=2.71s, s.d.=0.67s across talkers and emotion categories) than the NH listeners (mean=0.90s, s.d.=0.24s across talkers and emotion categories in the full-spectrum condition), and that incongruent stimuli produced a slightly longer reaction time (mean=1.61s, s.d.=1.02s across groups and emotion categories) than congruent stimuli (mean=1.4s, s.d.=0.89s across groups and emotion categories) overall in the full-spectrum condition. The effect of Emotion was non-significant overall, but modified effects of other variables in interactions. The NH listeners also showed a significantly longer reaction time in the full-spectrum condition (mean=0.90s, s.d.=0.24s as reported above) than in the 16-channel (mean=0.99s, s.d.=0.31s) or eight-channel (mean=0.98s, s.d.=0.27s) conditions. However, this difference of ~0.09s was smaller than the ~1.8s difference observed between NH listeners and CI listeners in the full-spectrum condition.

## DISCUSSION

The purpose of the current study was to examine the interaction between prosodic and lexical-semantic cues for emotion recognition in adult NH listeners and CI users. The results confirmed previous findings (Ben-David et al., 2016; Ben-David et al., 2019) showing that NH listeners rely predominantly on prosodic cues in full-spectrum speech. Analyses with the Linear and Categorical coding schemes produced consistent results, and post-hoc analyses added further insight into the findings. Novel contributions of the present study center on



the degraded listening conditions in the NH listeners and the full-spectrum condition in CI listeners. The results support our hypothesis that NH listeners have reduced reliance on prosodic cues and increased reliance on lexical-semantic cues in spectrally degraded conditions. Also, as hypothesized, CI users showed increased reliance on lexical-semantic cues compared to NH adults in the full-spectrum condition. The difference in cues used by NH listeners and CI users for emotion recognition was further examined by separating out the Congruent and Incongruent conditions. The NH listeners performed better than CI listeners in the Congruent condition – presumably because they had full access to the prosodic cues as well as the lexical-semantic cues. However, in the Incongruent condition, adult CI users performed better than NH listeners in the task. Recall that performance was scored on the correct identification of the lexical-semantic cues. The CI participants' results in the full-spectrum condition were similar to those obtained with NH listeners in the spectrally degraded conditions. As the speech became more degraded, the NH listeners' scores on the lexical-semantic emotion improved in the Incongruent condition, and decreased in the Congruent condition, indicating reduced reliance on prosodic cues and increased reliance on lexical-semantic cues to emotion.

The results showed significant effects of talker variability when separating and comparing the results by talker. As there were only two talkers, the ability to interpret effects of talker variability is limited. Further studies should include more than two talkers to better examine any influence of talker variability. For instance, Luo et al (2016) reported that NH listeners show significant effects of talker-variability on emotions based on prosodic cues when listening to speech in noise, but that this sensitivity to differences between talkers was not apparent when listening to CI-simulated versions. If this pattern holds in CI patients, then a study with a larger number of talkers may reveal differences between NH and CI listeners in how talker variability affects emotion perception in both Congruent and Incongruent conditions. The lexical-semantic cues to emotion would be expected to be perceived similarly across talkers, but if sensitivity to talker-variability in prosodic cues is greater for NH listeners than for CI users, then effects of incongruency might vary more strongly from talker to talker for NH listeners, but not as much for CI users.

While some differences in scores were observed between individual emotions, the effect of Congruency, and the interactions between the Spectral Condition and Congruency, or the interactions between Hearing Status and Congruency in the full-spectrum condition, did not depend strongly on the specific emotions. Additionally, excluding the Neutral emotion from analyses did not essentially alter the central findings. The effect of Emotion was observed in interactions with the other factors including Talker and Congruency in the NH listeners, supporting the idea that talker-variability influences the prosodic cues more than the lexical-semantic cues, which in turn influences performance in the Incongruent condition more than in the Congruent condition. However, when CI users' data were included in the analyses, the interaction between Emotion, Talker, and Congruency became less evident, suggesting that the effect size was too small to be adequately captured in the present design. This weaker effect of Emotion would be consistent with CI users showing reduced sensitivity to prosodic cues to emotion overall, and reduced sensitivity to talker variability as well.

The present results suggest that CI users may be at a disadvantage in everyday conversations that involve nuanced communication. When the prosodic and lexical-semantic cues converge in natural conversation, listeners with CIs might have somewhat reduced identification relative to NH listeners, but accuracy is still quite high. In natural conversations, prosodic and lexical-semantic cues do not often conflict – but when they do, the talker is likely attempting to convey sarcasm, or concealing some aspect of the communication deliberately. Further, the correct identification of the talker’s intent and emotional state depends critically on the listener’s perception of the prosodic cue in their speech. If the listener relies on the lexical-semantic cue to decipher the talker’s meaning, they run the risk of missing the point of the conversation altogether. Thus, the deficits demonstrated here likely reflect real-world communication problems faced by CI users.

Analysis of reaction times showed that in both the NH and CI populations, reaction time for the Incongruent condition were longer than the Congruent condition. In a comparison between NH and CI users, we found the reaction time for both the Congruent and Incongruent conditions was longer for CI users than NH listeners. However, as age was not controlled for in this study, the underlying impact of age on the reaction time data is unknown. The CI users were overall older than the NH listeners, so it is possible that their longer reaction times were partly due to aging. Previous research has found that both age and hearing status can influence reaction time data (Christensen et al., 2019), suggesting that our outcomes could be impacted by both of these factors as well. The longer reaction times in the Incongruent condition indicate that both groups were sensitive to the incongruence, and that the incongruence likely requires additional processing time. Previous studies have shown that listeners are able to detect vocal emotions based on prosodic cues quite rapidly, within the first one or two syllables of a spoken utterance (Pell & Kotz, 2011). On the other hand, the lexical-semantic cues to emotion are likely to require the processing of larger segments of speech. Thus, reaction times to prosodic and lexical-semantic cues to emotion may be quite different. Measuring these separately was not within the scope of the present study, however.

### Limitations of the present study

While this study confirmed and extended knowledge on emotion recognition in adult NH and CI users, a number of limitations must be acknowledged. First, the stimuli only included two female talkers, which limited our abilities to examine talker variability extensively. Future studies should include multiple talkers to better analyze the influence of different talkers on emotion recognition. Secondly, the NH and CI groups were not age matched, limiting our ability to investigate any impact of aging on for emotion recognition and differences in reaction time data. Ben David et al (2019) showed evidence for a slight shift toward lexical-semantic cues in older individuals listening to undistorted spoken emotions. In the present study, the shift toward lexical-semantic cues observed in the CI patients was echoed in the results obtained with the young NH listeners attending to CI-simulated speech, suggesting that age was not the factor. However, direct evidence for this remains to be obtained.

## Clinical implications

The findings of the present study have important implications for social communication by CI patients. The key finding of interest is that CI users, who have reduced access to prosodic information for vocal emotion perception, rely more on semantic cues to emotion than NH listeners. This can hinder their communication abilities, as prosodic cues dominate the communication of emotions in speech. Further, the ability to gather the talker's intended emotion from prosodic cues when prosody and semantic information conflict is an important requirement for sarcasm and humor perception. The inability to accurately identify these conflicting cues can negatively impact a person's ability to develop relationships, follow conversation, understand vocal emotion and intended message of a speaker, follow jokes, and everyday communication in general. Better understanding of how CI users understand emotion could potentially improve counseling and communication strategies provided to patients and their conversational partners. Additionally, an important subjective measure of CI outcomes is quality of life and Luo et al. (2018) reported a link between spoken emotion recognition and quality of life in CI users. Overall, improved understanding of emotional communication for CI users may lead to the development of improved intervention and counseling strategies, which may in turn improve emotion identification and quality of life in CI patients.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

We thank Mohsen Hozan for software support, Aditya M. Kulkarni, Shauntelle Cannon and Shivani Gajre for their help with this project. This work was supported by NIH/NIDCD grant nos. T35 DC008757, R01 DC014233, NIGMS grant no. P20 GM109023. We are grateful to the participants for their effort.

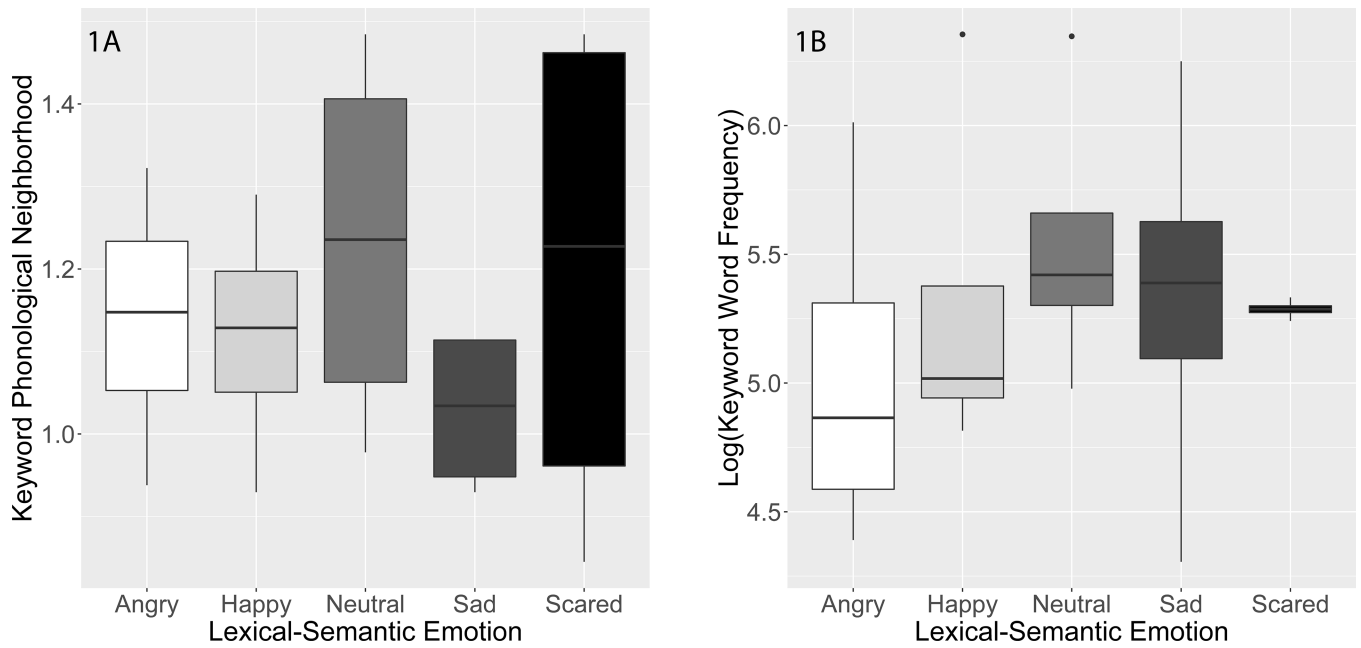
**Financial disclosures and conflicts of interest:** This study was funded by NIH/NIDCD grant nos. T35 DC008757, R01 DC014233, NIGMS grant no. P20 GM109023.

## REFERENCES

- Attardo S, Eisterhold J, Hay J, & Poggi I (2003). Multimodal markers of irony and sarcasm. *Humor*, 16(2), 243–260.
- Balota DA, Yap MJ, Cortese MJ, et al. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459 [PubMed: 17958156]
- Banse R, & Scherer KR (1996). Acoustic profiles in vocal emotion expression. *J. Personality & Social Psych*, 70(3), 614–636.
- Barrett KC, Chatterjee M, Caldwell MT, et al. (2020). Perception of Child-Directed Versus Adult-Directed Emotional Speech in Pediatric Cochlear Implant Users. *Ear Hear*.
- Bates D, Maechler M, Bolker B, et al. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J Stat. Software*, 67(1), 1–48.
- Ben-David BM, Multani N, Shakuf V, et al. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *J Speech Lang Hear Res*, 59(1), 72–89. [PubMed: 26903033]
- Ben-David BM, Gal-Rosenblum S, van Lieshout PH, et al. (2019). Age-Related Differences in the Perception of Emotion in Spoken Language: The Relative Roles of Prosody and Semantics. *J Speech Lang Hear Res*, 62(4S), 1188–1202. [PubMed: 31026192]

- Bosen AK, & Luckasen MC (2019). Interactions Between Item Set and Vocoding in Serial Recall. *Ear Hear*, 40(6), 1404–1417 [PubMed: 31033634]
- Cannon SA, & Chatterjee M (2019). Voice Emotion Recognition by Children with Mild to Moderate Hearing Loss. *Ear Hear*, 40(3), 477. [PubMed: 30074504]
- Chatterjee M, & Peng SC (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hear. Res*, 235(1–2), 143–156. [PubMed: 18093766]
- Chatterjee M, Zion DJ, Deroche ML, et al. (2015). Voice emotion recognition by cochlear-implanted children and their normally-hearing peers. *Hear Res*, 322, 151–162. [PubMed: 25448167]
- Christensen JA, Sis J, Kulkarni AM, & Chatterjee M (2019). Effects of Age and Hearing Loss on the Recognition of Emotions in Speech. *Ear Hear*, 40(5), 1069–1083. [PubMed: 30614835]
- Crew JD, & Galvin III JJ (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *J. Acoust. Soc. Am*, 132(5), EL429–EL435. [PubMed: 23145706]
- Deroche ML, Lu H, Limb CJ, et al. (2014). Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. *Front. Neurosci* 8:282. doi: 10.3389/fnins.2014.00282 [PubMed: 25249932]
- Deroche MLD, Kulkarni AM, Christensen, et al. (2016). Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants *Front. Neurosci.* 10:0007;. doi:10.3389/fnins.2016.00073.
- Deroche MLD, Lu HP, Kulkarni AM, et al. (2019). A tonal-language benefit for pitch in normally-hearing and cochlear-implanted children. *Scientific Reports* 9(1):109. doi:10.1038/s41598-018-36393-1. [PubMed: 30643156]
- Everhardt MK, Sarampalis A, Coler M, Baskent, et al. (2020). Meta-Analysis on the Identification of Linguistic and Emotional Prosody in Cochlear Implant Users and Vocoder Simulations. *Ear Hear*, 42, 2020 - Volume Publish Ahead of Print - Issue doi:10.1097/AUD.0000000000000863
- Fengler I, Nava E, Villwock AK, et al. (2017). Multisensory emotion perception in congenitally, early, and late deaf CI users. *PloS one*, 12(10), e0185821. [PubMed: 29023525]
- Geurts L, & Wouters J (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J. Acoust. Soc. Am* 109(2), 713–726. [PubMed: 11248975]
- Greenwood D (1990). A cochlear frequency-position function for several species—29 years later, *J. Acoust. Soc. Am* 87, 2592–2605 [PubMed: 2373794]
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat*, 6(2), 65–70.
- Hopyan-Misakyan TM, Gordon KA, Dennis M, et al. (2009). Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. *Child Neuropsychol*, 15(2), 136–146. [PubMed: 18828045]
- James Chris J., Karoui Chadlia, Laborde Marie-Laurence, et al. (2019). Early sentence recognition in adult cochlear implant users. *Ear Hear* 40, 4: 905–917. [PubMed: 30335668]
- Kirby BJ, Spratford M, Klein KE, et al. (2019). Cognitive Abilities Contribute to Spectro-Temporal Discrimination in Children Who Are Hard of Hearing. *Ear Hear*, 40(3), 645–650. [PubMed: 30130295]
- Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models, *J Stat. Software*, 82(13), 1–26.
- Lazic SE (2008). Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology. *BMC physiology*, 8(1), 16. [PubMed: 18644134]
- Luce PA, & Pisoni DB (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear*, 19(1), 1. [PubMed: 9504270]
- Luo X, Fu QJ, & Galvin JJ 3rd. (2007). Cochlear implants special issue article: Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif* 11(4), 301–315. [PubMed: 18003871]
- Luo X, Padilla M, & Landsberger DM (2012). Pitch contour identification with combined place and temporal cues using cochlear implants. *J. Acoust. Soc. Am* 131(2), 1325–1336. [PubMed: 22352506]

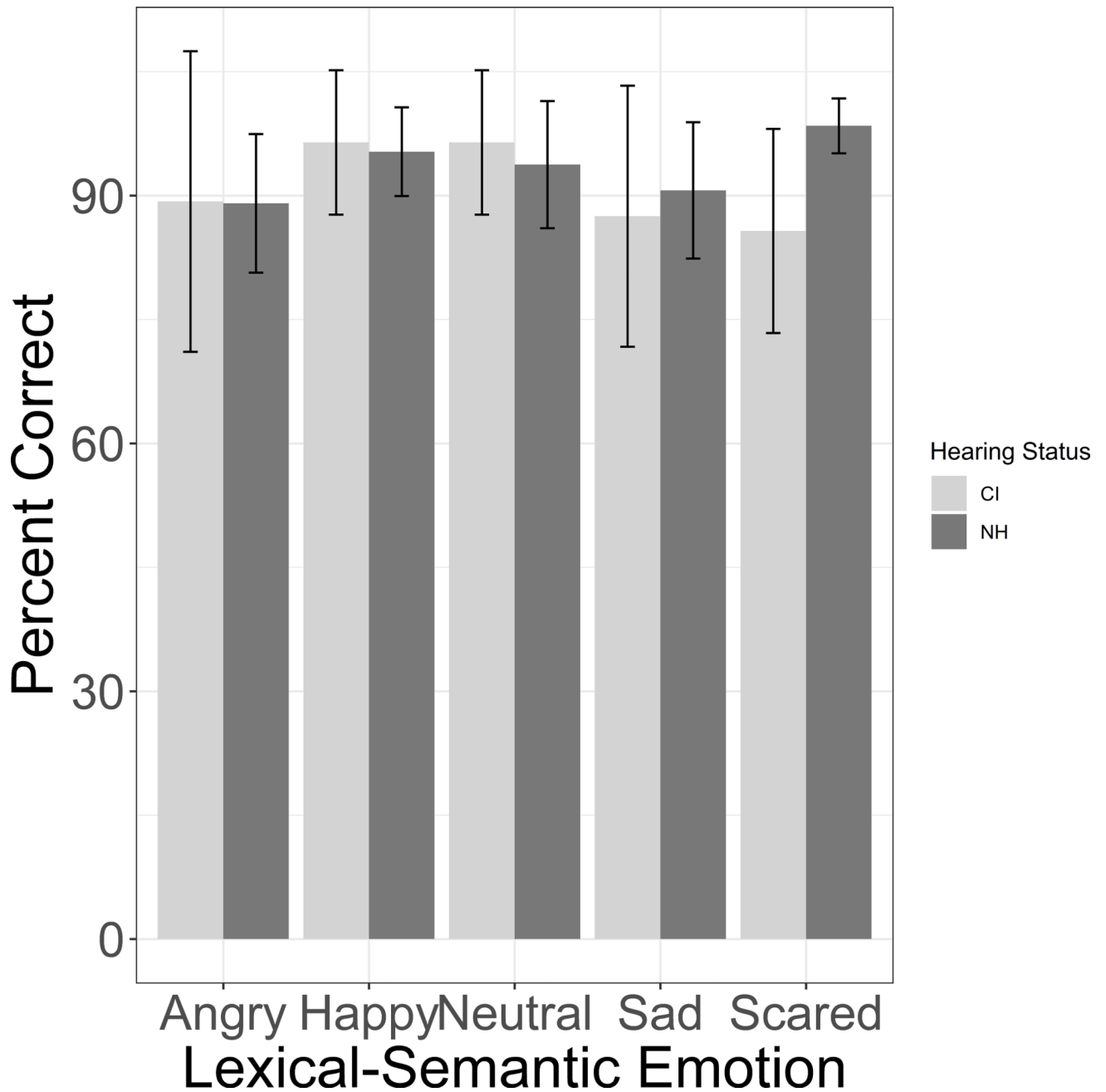
- Luo X (2016). Talker variability effects on vocal emotion recognition in acoustic and simulated electric hearing. *J. Acoust. Soc. Am*, 140(6), EL497–EL503. [PubMed: 28040001]
- Luo X, Kern A, & Pulling KR (2018). Vocal emotion recognition performance predicts the quality of life in adult cochlear implant users. *J. Acoust. Soc. Am*, 144(5), EL429–EL435. [PubMed: 30522282]
- Massaro DW, & Egan PB (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, 3(2), 215–222. [PubMed: 24213870]
- Milczynski M, Wouters J, & Van Wieringen A (2009). Improved fundamental frequency coding in cochlear implant signal processing. *J. Acoust. Soc. Am*, 125(4), 2260–2271. [PubMed: 19354401]
- Moberly AC, Harris MS, Boyce L, et al. (2017). Speech recognition in adults with cochlear implants: The effects of working memory, phonological sensitivity, and aging. *J Speech Lang Hear Res*, 60(4), 1046–1061. [PubMed: 28384805]
- Oxenham AJ (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends Amplif*, 12(4), 316–331. [PubMed: 18974203]
- Pals C, Sarampalis A, van Rijn H, et al. (2015). Validation of a simple response-time measure of listening effort. *J Acoust Soc Am*, 138, EL187–EL192. [PubMed: 26428811]
- Pasta DJ (2009, 3). Learning when to be discrete: continuous vs. categorical predictors. In *SAS Global Forum* (Vol. 248).
- Pell MD (1996) Evaluation of nonverbal emotion in face and voice: some preliminary findings on a new battery of tests. *Brain Cogn* 48(2–3), 499–504.
- Pell MD, & Kotz SA (2011). On the time course of vocal emotion recognition. *PLoS One*, 6(11), e27256. [PubMed: 22087275]
- R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: <https://www.R-project.org/>.
- Ritter C, & Vongpaisal T (2018). Multimodal and spectral degradation effects on speech and emotion recognition in adult listeners. *Trends Hear*, 22, 2331216518804966. [PubMed: 30378469]
- Roodenrys S, Hulme C, Lethbridge A, et al. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *J Exp Psychol Learn Mem Cogn*, 28(6), 1019. [PubMed: 12450329]
- Shannon RV, Zeng FG, Kamath V, et al. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304. [PubMed: 7569981]
- Stickney GS, & Assmann PF (2001). Acoustic and linguistic factors in the perception of bandpass-filtered speech. *J Acoust Soc Am*, 109(3), 1157–1165. [PubMed: 11303929]
- Takagi S, Hiramatsu S, Tabei KI, & Tanaka A (2015). Multisensory perception of the six basic emotions is modulated by attentional instruction and unattended modality. *Front. Integrative Neurosci*, 9, 1.
- Tao D, Deng R, Jiang Y, Galvin JJ III, Fu QJ, & Chen B (2015). Melodic pitch perception and lexical tone perception in Mandarin-speaking cochlear implant users. *Ear Hear*, 36(1), 102.
- 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 Hear*, 39(5), 874. [PubMed: 29337761]
- Tukey JW (1977). *Exploratory data analysis*. 1977. Reading Addison-Wesley, 1–688.
- Wang W, Zhou N, & Xu L (2011). Musical pitch and lexical tone perception with cochlear implants. *Int. J. Audiol*, 50(4), 270–278. [PubMed: 21190394]
- Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.



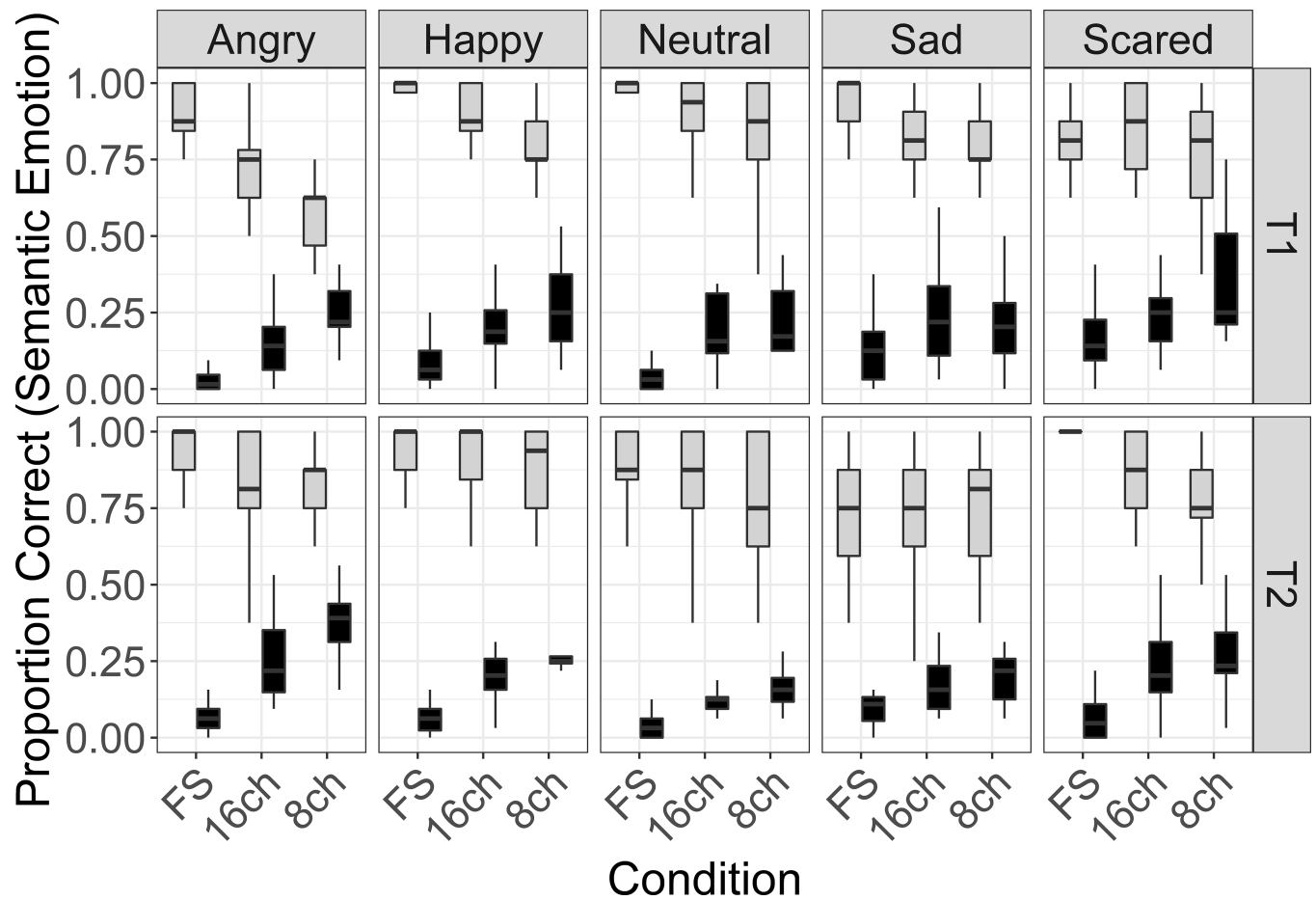
**Figure 1.**

A (left). Boxplots of phonological neighborhood density of keywords for the five categories of lexical-semantic. Each emotion is illustrated in the test stimulus set as a different color. Angry is shown in white, happy in light grey, neutral in medium grey, sad in dark grey, and scared in black fill.

Figure 1B (right). Boxplots of  $\log_{10}(\text{Word Frequency})$  of keywords for the five categories of lexical-semantic emotion in the test stimulus set. Each emotion is illustrated in a different color as in Figure 1A.

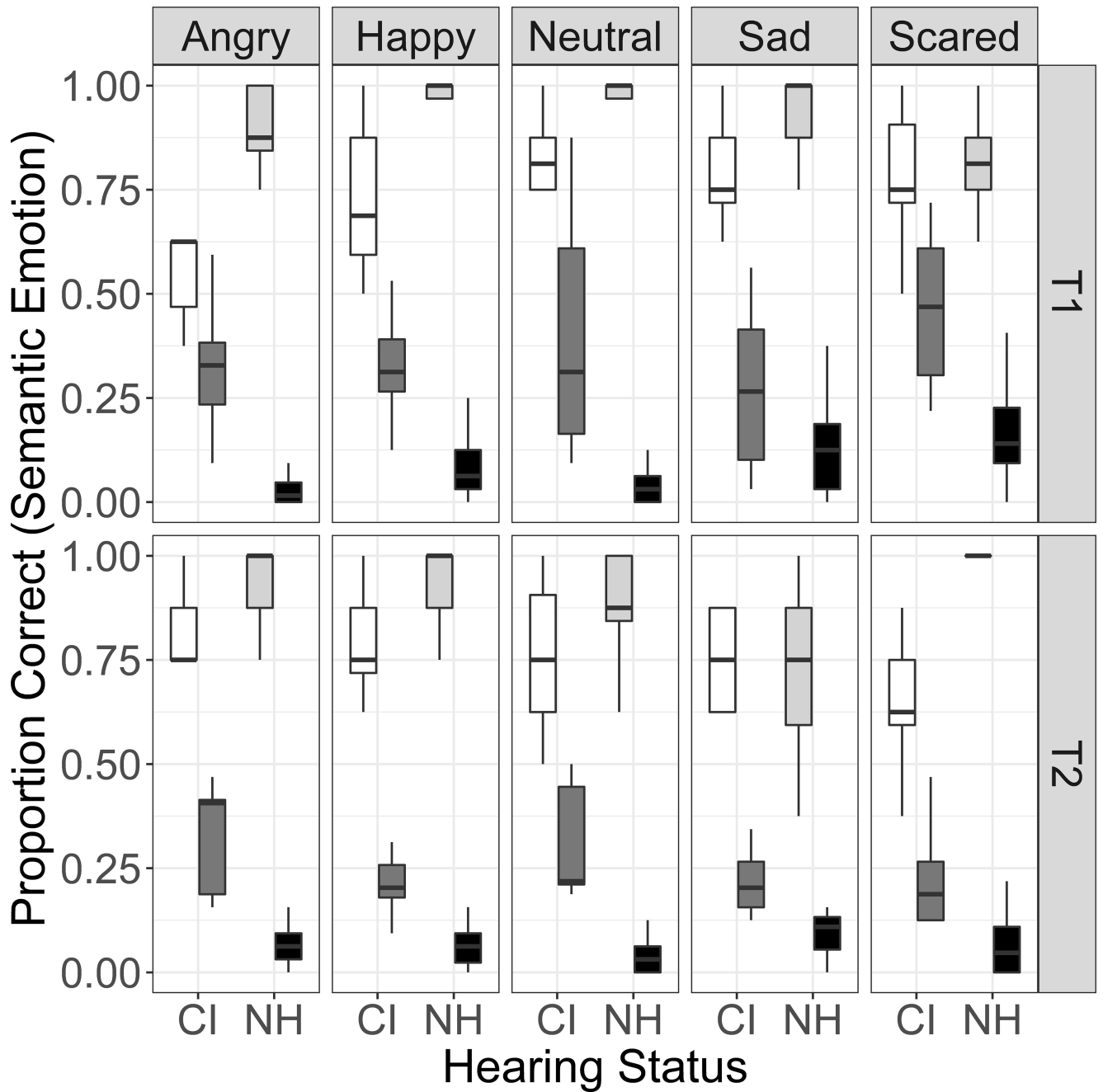


**Figure 2.** Results of written multiple-choice test of the perceived emotion for each sentence in the stimulus set. Proportion of emotions correctly identified is plotted against the lexical-semantic emotion category. Light and dark grey bars indicate CI and NH groups respectively. Error bars show  $\pm 1$  s.e. from the mean.

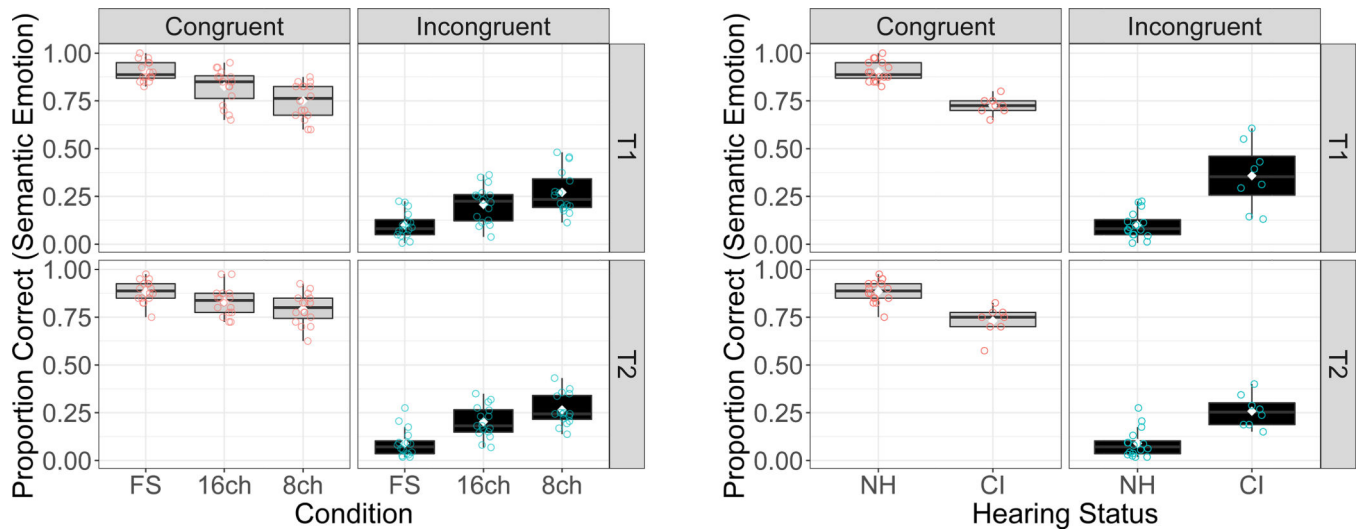


**Figure 3.** Proportion of trials in which lexical-semantic emotions were correctly identified by NH listeners plotted against the spectral degradation conditions (abscissa) for each semantic emotion in the Congruent and Incongruent conditions (light grey and black boxplots respectively), and for the two talkers T1 and T2 (upper and lower rows).





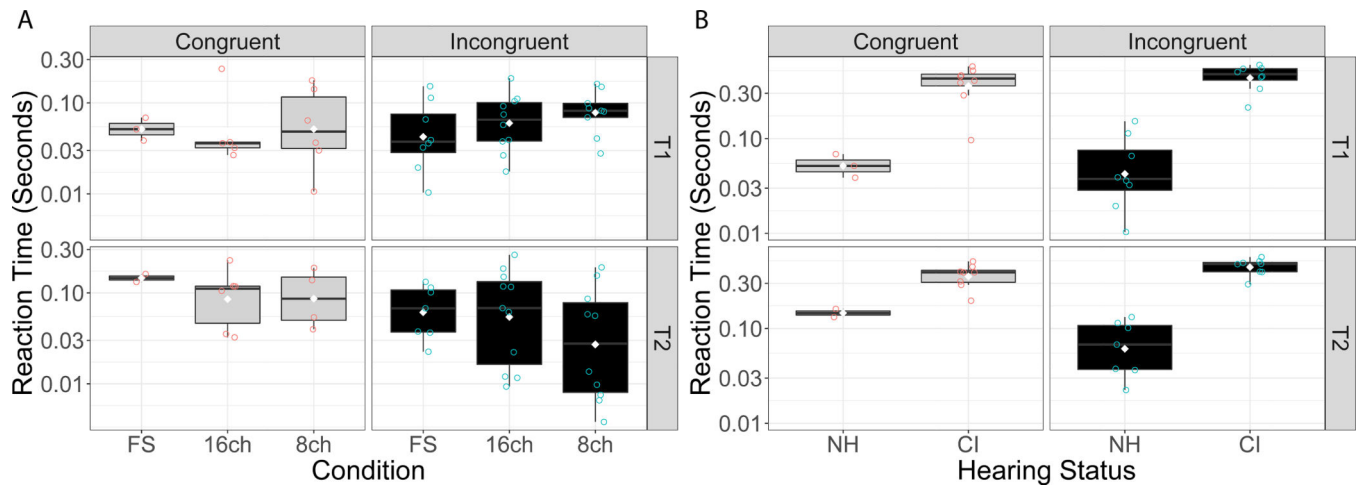
**Figure 4.** Results with CI and NH listeners (abscissa) in the full-spectrum condition for each semantic emotion in the Congruent (lighter boxplots: white and light grey respectively) and Incongruent (darker boxplots: dark grey and black respectively) conditions.



**Figure 5.**

A Similar to 3A, but with results averaged across semantic emotion categories. Left and right panels show Congruent and Incongruent conditions (grey and black boxplots respectively).

Figure 5B. Results with NH and CI data in the full-spectrum condition averaged across the semantic emotion categories in the Congruent (left panel) and Incongruent (right panel) conditions (light grey and black boxplots respectively).



**Fig. 6.**

A Reaction times in seconds (ordinate, in log scale) obtained in NH listeners plotted against the spectral degradation conditions and separated into Congruent (left panel) and Incongruent (right panel) conditions (light grey and black boxplots respectively). Figure 6B Reaction times in seconds (ordinate, in log scale) obtained in NH and CI listeners (abscissa) in the full-spectrum condition, separated into Congruent (left panel) and Incongruent conditions (upper and lower rows right panel) and for the two talkers T1 and T2 (left upper and lower rows). Congruent and Incongruent conditions are illustrated in back boxplots.

**Table 1:**

Detailed information for CI group

Subject	Previous HA Use	Etiology	Age at Testing	Gender	Age at Implantation	Duration of CI Use	Manufacturer
BT_N7	Yes	Unknown	64.58	Female	50.11	14.47	Cochlear
BT_C_03	Yes	Unknown	71.76	Male	55.65	16.11	Advanced Bionics
BT_N5	Yes	Unknown	59.49	Female	50.69	8.80	Cochlear
BT_N22	Yes	Unknown	60.43	Male	46.96	13.47	Cochlear
BT_N15	Yes	Unknown	65.03	Male	58.98	6.05	Cochlear
BT_N19	Yes	Unknown	68.50	Male	49.55	18.94	Cochlear
BT_N26	Yes	Unknown	34	Female	22.06	11.94	Cochlear
BT_C_10	Yes	Unknown	56.65	Male	40.03	16.62	Advanced Bionics

**Table 2.**

Sentences used for each lexical-semantic emotion category in Test and Practice conditions

Semantic Emotion	Sentence	Condition
Happy	I won the game	Test
	She loves school	Test
	My friend found a prize	Test
	My father is the best	Test
	I got a new job	Practice
Angry	She stole my car	Test
	You keep making us late	Test
	The dog ripped my carpet	Test
	I have waited hours	Test
	You broke my glasses	Practice
Neutral	I made a call	Test
	The car is red	Test
	You work at an office	Test
	The house is blue	Test
	The school is big	Practice
Sad	I lost a good friend	Test
	I am moving far away	Test
	My father is very ill	Test
	The oldest tree died	Test
	I never met my sister	Practice
Scared	I heard someone behind me	Test
	She saw someone following me	Test
	He might kill her	Test
	You need to hide	Test
	Someone strange is coming	Practice

**Table 3:**

Example of NH participant's test spectral conditions and order

<b>Talker</b>	<b>Spectral Condition</b>
Talker 1- female	Practice
Talker 1- female	Full-Spectrum
Talker 1- female	8ch NBV
Talker 1- female	16ch NBV
Talker 1- female	8ch NBV
Talker 1- female	Full-Spectrum
Talker 1- female	16ch NBV
BREAK	BREAK
Talker 2- female	Practice
Talker 2- female	8ch NBV
Talker 2- female	Full-Spectrum
Talker 2- female	16ch NBV
Talker 2- female	16ch NBV
Talker 2- female	Full-Spectrum
Talker 2- female	8ch NBV

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript