Title
Voice emotion perception and production in cochlear implant users

Permalink
https://escholarship.org/uc/item/8dq9h5sz

Authors
Jiam, NT
Caldwell, M
Deroche, ML
et al.

Publication Date
2017-09-01

DOI
10.1016/j.heares.2017.01.006

Peer reviewed

Voice Emotion Perception and Production in Cochlear Implant Users

NT Jiam[1], **M Caldwell**[1], **ML Deroche**[2], **M Chatterjee**[3], and **CJ Limb**[1]

[1]Department of Otolaryngology-Head and Neck Surgery, University of California San Francisco School of Medicine, San Francisco, California

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

[3]Auditory Prostheses and Perception Laboratory, Boys Town National Research Hospital, Omaha, Nebraska

Abstract

Voice emotion is a fundamental component of human social interaction and social development. Unfortunately, cochlear implant users are often forced to interface with highly degraded prosodic cues as a result of device constraints in extraction, processing, and transmission. As such, individuals with cochlear implants frequently demonstrate significant difficulty in recognizing voice emotions in comparison to their normal hearing counterparts. Cochlear implant-mediated perception and production of voice emotion is an important but relatively understudied area of research. However, a rich understanding of the voice emotion auditory processing offers opportunities to improve upon CI biomedical design and to develop training programs benefiting CI performance. In this review, we will address the issues, current literature, and future directions for improved voice emotion processing in cochlear implant users.

Keywords

Voice emotion; cochlear implant; speech prosody; voice emotion production; voice emotion perception

1. Introduction

Communicating emotion is a fundamental feature of human social interaction that transverses all cultures (Bryant & Barrett, 2008). In fact, some may argue that emotional cues formulate the very basis of human interaction and carry more valuable information that the actual words being spoken (Zajonc, 1980). There are many cues that come into play when communicating emotion, one of the most important being nonverbal cues (Skinner, 1935; Wallbott & Scherer, 1986). Among all types of nonverbal cues, humans frequently use prosodic vocal cues (e.g., voice pitch and tempo) to elicit emotive information in their interactions (Planalp, et al., 1996). So naturally, when prosodic vocal cues are degraded, voice emotion perception and production are often affected. Impairments in the perception and production of voice emotion usually result in serious ramifications on social interactions and social development, as in the case of infant-directed speech (Trainor, et al., 2000), underscoring the importance of this topic at hand.

Cochlear implants (CI) are surgically implanted electrical devices that allow people with severe-to-profound hearing loss to process sound. Over the past few decades, CI development has made remarkable ground such that most CI users have adequate speech perception in quiet environments. Despite this great success, limitations remain for present day CI systems including the transmission of spectro-temporal fine structure information (e.g. pitch and harmonics) (Kong, et al., 2004; Galvin, et al., 2007; Kang, et al., 2009; Kong, et al., 2011; Xu, et al., 2009). Forced to interface with highly degraded acoustic cues, CI users often demonstrate difficulty in perceiving prosodic cues. Limitations in the perception and production of prosody have adverse consequences for CI users, including the interpretation and communication of voice emotion. In this article, we will review the emerging body of work on CI-mediated perception and production of voice emotion.

2. Voice Emotion General Principles

2.1 Dimensions of Emotions in Relation to Speech Emotion Studies

Emotions are brief and strong reactions to goal-relevant changes in the environment. Historically, there are two main approaches towards studying emotion: discrete and dimensional. A discrete approach focuses on characteristics that distinguish emotional states from one another (Ekman, 1992) whereas a dimensional approach identifies emotions based on predetermined features underlying mood and affective states (Russell, 1980). Although there are many dimensions involved in emotion, the four most commonly referred-to dimensions of subjective feeling states are activation, valence, potency, and intensity (Smith & Ellsworth, 1985). Activation refers to the perceived sense of energy ranging from low to high (e.g. somnolence to feverish excitement) (Krumshansl, 1997; Gosselin, et al., 2007; Sammler, et al., 2007). Orthogonally, valence relates to the intrinsic evaluation of an event, object, or situation and ranges from positive to negative (e.g. joy to displeasure) (Krumhansl, 1997; Schubert, 1999; Dalla Bella et al., 2001). Potency is a dimension used to describe the degree of powerfulness or powerlessness an individual universally identifies with a particularly emotion (Russell & Mehrabian, 1977; Osgood, et al., 1957). Positive emotions almost always generate a high level of control or dominance. Thus, potency is specifically useful in differentiating between negative emotions such as fear and anger; where anger has

a high potency rating and fear has a low potency rating. Last but not least, emotional intensity is used to quantify the degree of emotion being felt (e.g. very happy or only a little bit happy). The valence-arousal model approaches emotion as two separable dimensions of valence and arousal (Russell, 1980). These two dimensions are commonly used in vocal expression studies (Bachorowski, 1999, Scherer, 1986) and capture the majority of the psychophysiological components of emotion, which is why some researchers reduce emotion theory down to only two components: valence and arousal.

2.2 Speech Prosody Cues & Voice Emotion

The origin of the term 'prosody' can be traced back to ancient Greek where it was used to indicate the tone or accent of a syllable. Over the years, the word prosody has evolved to govern the modulation of the human voice when uttering segmental sequences of phonemes. In modern phonology, prosody refers to elements of speech relating to the properties of syllables and larger units of linguistics, such as voice pitch, duration, intensity, spectral characteristics, nonverbal vocal expressions (e.g. crying), rhythm, and tempo. Prosodic features of speech often cannot be captured by conventional segmental phonetic transcriptions or orthography. These properties of speech play an important role in communication, such as informing a listener of the speaker's intent and affect. Overall, there are few articles on the acoustic correlates of emotion. Below, we chose to highlight the most common associations found between prosodic cues and voice emotion activation and valence. The general principles mentioned in this section are not steadfast rules particularly with findings concerning valence.

Strictly speaking, pitch is the perceptual correlate of the fundamental frequency (F_0) of a sound. Although pitch is a subjective attribute and fundamental frequency an objective acoustical parameter, the two terms are often used interchangeably in the literature. For the purposes of this review, we will also use the word 'pitch' to refer to F_0. In speech, the fundamental frequency is derived by the rate of vocal cord vibration. The fundamental frequency range varies between speakers and depends greatly on the length and mass of the vocal cords – As a result, male speakers (85 to 180 Hz) generally have a lower fundamental frequency range than female speakers (160 to 255 Hz). Within their individual ranges, speakers have a large degree of active control over voice pitch and can choose to speak in a high or low pitch with corresponding rises and falls.

As previously mentioned, pitch is a strong acoustic cue for voice emotion in both children and adults. In fact, many school-aged children perform to the same level as adults in pitch discrimination tasks, demonstrating that fine fundamental frequency cues in voice can be available at a young age (Figure 1) (Deroche, et al., 2012). It is yet unclear whether the ability of children to recognize emotion in voice develops because this fine sensitivity to pitch is available to them very early on, or on the contrary whether voice emotion processing is one of the causes driving the auditory system to refine its sensitivity to pitch. But it is clear that the two aspects are tightly connected. For example, high activation is often associated with a high mean fundamental frequency (Breitenstein, et al., 2001; Davitz, 1964; Levin & Lord, 1975; Pereira, 2000; Scherer & Oshinsky, 1977; Schröder, et al., 2001) and fundamental frequency variability (Breitenstein, et al., 2001; Pereira, 2000; Scherer &

Oshinsky, 1977). Studies involving valence and voice pitch, on the other hand, are much less consistent. Some authors observe positive valence with low mean fundamental frequencies and high levels of fundamental frequency variability (Scherer & Oshinsky, 1977; Uldall, 1960) whereas other investigators fail to find patterns of vocal cues for valence dimension (Apple, et al., 1979; Davitz, 1964; Pereira, 2000).

With most natural sounds, duration is determined by the time interval between an onset and offset. This is not only applicable for duration of sounds, but also for duration of silence between two sounds. In cases of clear and rapid changes in the sound stimuli, perception of the phonemic segmentation is more or less straightforward. With slower changes in the durations of sound segments or silence intervals, such as in glides and slurred speech, speech becomes subject to listener perception and interpretation. In general, shorter pauses are associated with high activation (Schröder, et al., 2001). Longer pauses are commonly observed with negative valence (Schröder, et al., 2001).

Tempo is a common acoustic cue used to convey emotion. Emotions with high levels of activation (e.g. excitement and anger) and positive valence are commonly associated with fast speech rates (Apple, et al., 1979; Breitenstein, et al., 2001; Davitz, 1964; Kehrein, 2002; Scherer & Oshinsky, 1977).

Vocal intensity is the sound pressure level (Isshiki, 1964) and is an important prosodic cue used in voice emotion. In general, high-frequency energy is highly predictive of perceived intensity in listeners (Juslin & Laukka, 2001). Previous studies frequently report increased voice intensity with high activation and negative valence (Davitz, 1964; Huttar, 1968; Pereira, 2000; Schröder, et al., 2001). However, vocal intensity is relatively overlooked in relation to voice emotion, and the relationship between intensity and other emotion dimensions is not well understood.

It is important to bear these associations in mind, particularly in the light of recent interest into speaking styles such as "clear speech". To facilitate conversations with listeners presenting hearing losses, speakers may raise their voice's intensity as observed in the Lombard effect (Lane & Tranel, 1971), may speak at a slower rate and mark longer pauses between words. Although this may provide notable benefits in terms of intelligibility, e.g. better signal-to-noise ratio and enhanced syllable segmentation, this speaking style is likely to carry more negative valence and in fact, recent evidence suggests that this is exactly the case (Morgan & Ferguson, 2014). Therefore, beyond audiologic considerations alone, there is a need for a better understanding of the emotions associated with different manipulations of prosodic cues in speech.

3. Review of Prosody and Voice Emotion Studies in Cochlear Implant Recipients

While speech recognition in CI users is relatively well studied, speech prosody and emotion are less so. Salient voice pitch information and pitch-based harmonic structures are demonstrably important components of speech prosodic cues. However, present day CI systems provide limited spectro-temporal fine structure information in speech (Geurt &

Wouters, 2001; Green, et al., 2004). This in turn impairs CI users' ability to perceive and produce important prosodic forms of communication such as question-statement contrasts, lexical tone recognition, and voice emotion (Shannon, 1983; Zeng, 2002; Chatterjee & Peng, 2008; Luo, et al., 2007; Peng, et al., 2004; Luo & Fu, 2004; Ciocca, et al., 2002; Wei, et al., 2004). Because voice emotion perception relies on many overlapping prosodic cues, the general consensus is that adults and children with CIs have significant difficulty in interpreting emotions. Critically, however, studies indicate wide inter-individual variability among CI users in emotion- and prosody- related tasks. In this section of this article, we will review the body of literature on CI171 mediated production and perception of prosody and voice emotion.

3.1 Voice Emotion and Prosody Perception

Many of the studies investigating CI users' perception of emotion in speech use professional actors' recordings of semantically neutral sentences (for example, "The coat is on the chair") to acoustically convey targeted emotions such as happy, sad, angry, anxious, relieved, or scared. Studies like these consistently observe significant deficits in target emotion recognition in CI users, both adults (Luo, et al., 2007; Pereira, 2000; Kalathottukaren, et al., 2015) and children (Hopyan-Misakyan, et al., 2009; Chatterjee, et al., 2015; Nakata, et al., 2012; Volkova, et al., 2012). Gilbers et al. (2015) also demonstrated poorer performance in CI users than NH listeners in an emotion identification task using nonce words. NH listeners performed better than CI users even when presented with CI simulations of stimuli.

In general, children and adult CI users can differentiate questions from statements, especially when envelope periodicity cues are used (Rosen, 1992). As a group, CI users' overall performance is well below normal hearing listeners (Peng, et al., 2008; Most & Peled, 2007; Van Zyl & Hanekom, 2013; Chatterjee & Peng, 2008; Meister, et al., 2009; Peng, et al., 2012) and severely-to-profoundly deaf pediatric hearing aid users (Most & Peled 2007). Statements typically end with a decrease in pitch relative to the remainder of the sentence as compared to an increase in pitch at the end of questions. The low CI performance on statement-question discrimination tasks is likely due to the poor representation of the pitch contour in CI users (See, Driscoll, Gfeller, Oleson, & Kliethermes, 2013). Crucially, this limitation can also affect representation of lexical tone languages, such as Mandarin, which is based on differences in the height and movement of pitch on a vowel phoneme. CI-mediated peer-to-peer interactions are often inhibited in lexical tonal languages because present day CI systems have difficulty detecting and transmitting rapid intonation changes within the syllable. This handicap ultimately stems from degraded pitch cues, and studies have observed significant impairment in identifying and producing lexical tones among CI users (He, et al., 2016, Peng, et al., 2004; Ciocca, et al., 2002; Holt & McDermott, 2013; Deroche, et al., 2016; Deroche, et al., 2014; Wang, et al., 2011; Wong & Wong, 2004).

Emphasis, or stress patterns in speech, is also important to vocal emotion communication and is dependent on the same prosodic cues. Among other findings, Meister and colleagues (2009) reported that CI users performed more poorly than NH listeners when asked to identify the stressed word (which varied between subject, verb, and object) in a series of

sentences. The authors demonstrated similar results using stimuli in which the speakers were instructed to stress a particular word and stimuli incorporating pitch manipulation to artificially stress or de-stress the word. This suggests that CI users are not particularly effective at using alternative strategies based on intensity and duration cues that are available in naturally uttered stresses. In other words, they rely heavily on pitch information despite its poor quality. Kalathottukaren et al. (2015) also found impaired stress identification ability in CI users using the Profiling Elements of Prosody in Speech-Communication (PEPS-C) test, a battery of four prosody tests measuring perception of contour, affect, stress, and "chunking", referring to utilization of prosodic cues to communicate intonational subsections of a phrase (for example, "fruit salad" as a compound word versus "fruit, salad" as a list). CI users performed more poorly than NH listeners in the prosodic subtests dependent on pitch representation (the contour, affect, and stress tests) but near normal on the "chunking" test, which relies more heavily on temporal cues. This is another manifestation of poor pitch perception in CI recipients but also suggests that they are able to utilize durational and timing cues as one means of extracting prosodic information (Figure 2, Kalathottukaren, et al., 2015).

Speech emotion perception in pediatric CI patients is of particular interest among researchers because the developmental implications of early auditory deprivation could have notable long-term effects; for example, infant-directed speech uses exaggerated prosodic cues to convey emotion, which help the infant understand the intended emotion of the adult and aid in development of communication skills (Trainor, et al., 2000; Soderstrom, et al., 2003). Early exposure to speech and prosody could play a significant role in the social development of CI children. The results of a study by Wiefferink et al. (2013) suggest that pre-school-aged children with CIs are delayed in facial affect recognition and emotional attribution abilities (Wiefferink, et al, 2013). This may be surprising given that facial emotion was conveyed visually to CI users who had normal vision and in fact, one might have expected these users to have learned to make a better use of visual cues to compensate for their auditory deficits. Interestingly, by school age, CI children seem to have established normal facial emotion recognition (Hopyan-Misakyan, et al., 2009). This finding suggests that, in the first few years of brain development, the connectivity between auditory information and emotion helps conceptualize human emotions. The loss of one emotion processing modality presumably leads to concepts that are less-well-defined which may then have repercussions on the emotional connectivity with the visual system itself. Ketelaar et al. (2013) measured empathic behaviors, emotion acknowledgement, and social competence in CI and NH children and found that, in contrast with other studies (Weifferink, et al., 2013), social competence and empathic behaviors were not impaired in the CI group. The authors also found that emotional acknowledgement predicted social competence in the CI group but not in the NH group, possibly suggesting that CI children are more attentive to visual emotion signals, perhaps to compensate for a diminished auditory input. Notably, the study reported language scores in their CI group higher than previously reported averages for CI children implanted prior to age two (Boons, et al, 2012). Furthermore, cognitive abilities were not measured in either group. It is therefore difficult to pinpoint the source of these findings.

Research investigating music emotion processing in CI users further supports the premise that CI users utilize alternative processes to process auditory emotion (Volkova, et al., 2012; Hopyan, et al., 2011; Shirvani, et al., 2015; Giannantonio, et al., 2015). Hopyan et al (2011) found that CI children performed significantly more poorly than their NH peers in distinguishing happy vs. sad music. The authors hypothesize that the CI group utilized tempo rather than pitch cues, which are fundamental to musical emotion, in making this distinction. Indeed, Caldwell et al. (2015) found that adult CI users rely on tempo rather than pitch in the processing of musical emotion, whereas NH listeners relied on both cues. The results of a study by Giannantonio et al. (2015) parallel these findings in children, additionally indicating that CI children's reaction times in musical emotion identification were affected more strongly by changes in tempo compared to mode, whereas those of NH children changed more significantly with concurrent changes in mode and tempo. These studies strongly support the notion of a unique processing strategy for emotion in auditory cues – in particular, by increased reliance on tempo-based modalities – in CI users compared to NH listeners.

More studies including adult populations are required to shed light on auditory input and long-term brain development in terms of emotion recognition (we will discuss the cortical reorganization effects of auditory deprivation later in this article). In a recent study (Chatterjee, et al., 2015), pediatric CI users performed similarly to adult CI users on tasks of speech emotion perception, and both performed comparably to NH adults listening to vocoded speech. However, children with NH performed more poorly than NH adults in these CI simulations (Figure 3, Chatterjee, et al., 2015). This could indicate that adults, both CI and NH, have sufficiently developed cognitive systems to be able to decipher emotion from degraded speech, while CI children may have found a way to adapt to the impoverished auditory signal. NH children listening to vocoded speech have neither fully developed cognitive systems nor learned adaptive methods of emotion, potentially explaining why they have trouble with deciphering voice emotion from CI simulations.

Acoustic analysis of questions/statements recordings by NH listeners demonstrates the importance of pitch contour, and to a much lesser degree intensity and duration (Peng, et al., 2012). When presented with many versions of the same contrasts manipulated incrementally along one dimension, NH listeners rely heavily on pitch contour. In contrast, listeners with poor spectral information (e.g. CI users or NH listeners attending to noise-vocoded stimuli) tend to modify their listening strategies to rely more on secondary cues such as intensity and duration (Figure 4, Peng, et al., 2012). As demonstrated, when F0 cues are degraded as it is the case in CI processing, other prosodic cues may be used to extrapolate emotion information, such as intensity and duration, although they are not as reliable (Pereira, et al., 2000; Luo, et al., 2007; Gilbers, et al., 2015). For example, intensity normalization has a greater impact on CI users' performance in emotion recognition than NH listeners'. This is consistent with the idea that CI users rely relatively more on intensity cues compared with NH listeners (Luo, et al., 2007). Duration may also be important to CI recipients for accessing emotion information in the absence of pitch and/or amplitude cues (Hegarty & Faulkner, 2013). Finally, CI users may utilize the limited pitch cues they have, but in a different manner than NH listeners. For example, Gilbers and colleagues (2015) asked speakers to produce salient nonce words. They analyzed the acoustic features of the

recordings in conjunction with emotion identification performance by CI users and NH listeners. They determined that *mean pitch* was more important to NH listeners, while CI users weighed *pitch range* more heavily. This could indicate that the impoverished acoustic signal in CI users leads to a lack of understanding of the relationship between mean pitch and emotion (for example, the association of high frequencies with high activation) and thus they tend to rely on large pitch differences and patterns to extrapolate emotional information.

3.2 Voice Emotion and Prosody Production

Few studies have examined voice emotion production in CI users. Prosody production in CI users is more studied in children than adults, though even this research has been sparse. Acoustic voice quality in CI children is generally better than that of hearing aid users (Guerrero Lopez, et al., 2013); however, there remain significant impairments in emotion production in the CI population. Studies report less accurate imitations of happy- and sad-sounding speech among CI children compared to NH children, particularly with regard to appropriate pitch modulation (Wang, et al., 2013). Furthermore, CI children demonstrate difficulty with perceiving more subtle emotions such as disappointment and surprise (Nakata, et al., 2012). Speech production studies suggest a lack of appropriate pitch contour in conveying questions and/or statements in CI children (Peng, et al., 2008; Peng, et al., 2007; Chin, et al., 2012). Recent preliminary data from Chatterjee et al. (2016) illustrates that CI children's productions of emotional speech incorporate smaller contrasts between happy and sad speech in mean pitch, pitch range, intensity, and spectral centroid compared to NH children. Post-lingually deafened CI adults' speech exhibited acoustic patterns more similar to the NH population.

3.3 Factors Influencing CI-Mediated Voice Emotion

CI-mediated voice emotion abilities are not uniform across the CI population. There is large inter-individual variability in speech prosody perception and production skills. In fact, many studies report a handful of "star" CI performers whose abilities are reportedly on par with their NH peers (Wang, et al., 2013; Chatterjee, et al., 2015). A myriad of factors contributes to variability in speech prosody skills including auditory deprivation prior to implantation, age at implantation, and technological factors.

As discussed above, the role of early CI implantation and auditory deprivation in appropriate development of speech emotion skills is highly complex and not well understood. Early implantation is consistently correlated to better performance in speech intelligibility tasks in congenitally and pre-lingually deaf children, demonstrating the importance of early auditory input in proper emotional speech development (Schorr, et al., 2009; Artières, et al., 2009). This is reinforced by neurophysiological evidence. Neurological differences between deaf and hearing people are well established, with functional reorganization occurring as a result of decreased auditory input. Electrophysiological studies of deaf children with and without CIs at specific periods of development suggest that a lack of auditory stimulation hinders auditory cortex development. CI usage can stimulate maturation at a rate close to normal, at least in parts of the auditory pathway; but this is limited by the degree of reorganization that occurred during the period of deafness (Ponton, et al., 2000; see Gordon, et al., 2011 for review).

Variations in technology can also influence CI users' prosody perception. Bimodal CI users, i.e. those using a CI and contralateral hearing aid simultaneously, are better at identifying intended emotion than deaf individuals using CIs alone (Straatman, et al, 2010). Processing strategies play a significant role in voice emotion as evidenced by both behavioral and electrophysiological data. Agrawal et al. (2013) used EEG to compare speech emotion perception and even-related potential (ERP) signals in NH listeners, CI recipients using the Psychoacoustic Advanced Combination Encoder strategy (also known as MP3000), and CI recipients using the Advance Combination Encoder strategy (i.e. ACE). M3000 functions to more precisely transmit spectral information by maximizing representation of relevant spectral information using psychoacoustic masking, whereas ACE focuses less on eliminating redundancies and instead on transmitting acoustically salient information (Wouters, et al., 2015). NH listeners' performance on the sentence emotion identification task was significantly more accurate than CI users, and their P200 ERP had a significantly higher amplitude than that of the CI group. CI users using MP3000 demonstrated more accurate identification of emotion in spoken sentences than those using ACE, and a more positive P200 ERP response (closer to NH performance) when assessing "happy" prosodic cues (Figure 5, Agrawal, et al., 2013). These results demonstrate the potential impacts of CI processing strategy parameters on emotion perception, and in particular that prosody processing might be optimized by strategies that target a focused representation of spectral information. In a similar study, Agrawal et al (2012) observed improved prosody perception when using MP3000 rather than standard CI simulations among NH listeners.

3.4 Rehabilitation in CI-Mediated Voice Emotion Perception

There is evidence that aural rehabilitation may improve CI-mediated detection of suprasegmental features of speech, such as intonation, pitch, intensity, and duration. A Chinese rehabilitation program with 28 prelingually deaf pediatric CI users demonstrated improved sentence recognition and story comprehension after two years of training (Wei, et al., 2000). Similarly, a pilot study implemented an 11-lesson psychoeducational program focused on improving emotional understanding among 14 deaf children and observed significant increases in emotion vocabulary and emotion comprehension from pretest to posttest. (Dyck & Denver, 2003). Although these findings imply that emotion recognition ability may be improved with training, the emotion training program heavily focused on using facial emotion cues and less so on vocal emotion cues. On the other hand, Zhang et al. (2013) did not find improved performance with voice gender and emotion identification tasks among 7 unilateral CI and contralateral acoustic stimulation users after 4 weeks of training, totaling 20 hours, suggesting that any improvement in emotion processing would require intense training over a sustained period of time.

Musical training is an increasingly popular avenue for speech emotion rehabilitation in CI users, as evidenced by studies indicating improvements in prosody perception as a result of musical engagement, training, or exposure. Longitudinal musical training studies consistently suggest that music training provides speech-related benefits, such as phonological awareness, perception of vowel duration, and speech segmentation, to NH listeners (Hausen, et al., 2013; Degé & Schwarzer, 2011) and benefits in melodic and pitch perception to CI users (Chen et al, 2010; Galvin, et al, 2007). Torppa et al. (2014) found that

CI children participating in music or dance activities over a period of 16 months performed similarly to NH children in tasks of pitch discrimination and word stress perception, whereas CI children participating in non-musical tasks over the course of the study performed more poorly. Patel (2014) provided preliminary evidence that melodic contour training with an emphasis on contour precision perception can improve intonation perception, and a recent study (Yhun Lo, et al., 2015) demonstrated improvements in question-statement discrimination and other speech perception parameters as a result of melodic contour training in adult CI users, both interval-based and duration-based (Figure 6, Yhun Lo, et al., 2015). These studies suggest that auditory training may serve as a rehabilitation mechanism for speech emotion perception skills in CI users. Nonetheless, the number of aural rehabilitation studies are meager and more investigation in this field of work is needed before conclusions can be drawn regarding the benefits of auditory training on voice emotion recognition.

3.5 Conceptualization of Voice Emotion and Musical Terminology in CI Users

It is important to note the difficulty CI users and children with NH face in conceptualizing auditory terms. Young children with NH often use terms associated with loudness to describe changes in pitch (Andrews & Diehl, 1970; Hair, 1981; Van Zee, 1976). In English, the words used to describe the pitch scale, "high" and "low", also have spatial, emotional, and loudness connotations. However, in other languages (e.g. Spanish and French) descriptor terms for pitch are specific and used only in reference to pitch, and this language factor alone results in an improved ability to label the direction of pitch change (Costa-Giomi & Descombes, 1996). This phenomenon is important to acknowledge because it suggests that a given subject may demonstrate deficits in a task simply because he or she did not fully understand the concept of the auditory cue within the research task.

In the case of CI users, despite scientists' best efforts to explain auditory terms in a simplistic but accurate manner, if the representation of the cue is poor (as in the case of fundamental frequencies), CI users' understanding of it may be extremely vague. This could be especially problematic with pre-lingually deaf CI subjects whom may have little reference and exposure to musical terms and sounds. In a sense, this effect could potentially confound study findings and thus, our understanding of CI-mediated processing because subjects cannot perform well in a task that they don't understand. These events may frequently occur in pitch and voice emotion; for example, a CI user may base his or her concepts of emotion on facial expressions and his or her own set of experiences and biases given a lack of dependability on impoverished auditory cues. However, when these visual cues are removed in a research experiment and the CI subject is asked to identify the voice emotion of a sound stimulus, a CI users' understanding of the task at hand may be akin to a NH subject being asked to use auditory cues to discriminate between a green sound and a blue sound. This analogy attempts to describe how the perception of emotion in music and speech involves cues that CI users do not have good access to and may not naturally associate with voice emotion. On the other hand, we recognize that despite these concerns of conceptualization, it may be necessary to use cues of interest within the auditory modality given that restoration of hearing is the ultimate goal of cochlear implantation. Nonetheless, this conceptualization issue should be acknowledged and may be addressed with careful

experimental implementation, standardization in descriptor terms for music, and music training, depending on a listener's ability to internalize and then verbalize these concepts, and later create personal cognitive meanings to them.

Although we emphasize the unique nuances of limited familiarity with voice emotion and musical terminology with respect to tasks involving emotion, there are many other factors that could confound study findings. Indeed, research subjects are often tested in study conditions that are not typical for them. For instance, bilateral listeners are often tested using only unilateral input, NH listeners interface with vocoded sounds, CI users grapple with new processors and programs, and test batteries involve unfamiliar and often meaningless stimuli. All of these conditions may confound study results, however, we would argue that these inherent constraints in study design may not be as specific to experiments involving emotion in speech and music. All in all, the significance of these issues involving unfamiliar study conditions remains of importance and may be more easily addressed than poor conceptualization secondary to limited exposure to emotional information.

4. Conclusion

The role of voice emotion perception and production in communication cannot be overstated. As evidenced by a combination of behavioral and electrophysiological data, CI users face significant deficits in prosodic cues due to significant limitations in pitch, intonation, and contour perception. Our review discusses how voice emotion processing, perception, and production may be hindered in CI users, and that the wide variability in CI performance is secondary to a number of factors including poor pitch representation and limited spectro436 temporal fine structure information. As a result, CI users may be forced to rely on cues other than frequency - such as intensity, tempo, and duration - to determine the intended emotion of a speaker (Kalathottukaren, et al., 2015). Other studies have shown that CI users are not particularly effective at using alternative strategies, and continue to rely most heavily on pitch information despite its poor quality for prosodic information (Meister et al., 2009).

Adult and pediatric CI users consistently demonstrate significant deficits in target emotion recognition ((Luo, et al., 2007; Pereira, 2000; Kalathottukaren, et al., 2015; Hopyan-Misakyan, et al., 2009; Chatterjee, et al., 2015; Nakata, et al., 2012; Volkova, et al., 2012; Gilbers, et al., 2015) and other prosodic-dependent components of communication (e.g. question-statement discrimination, lexical tone languages, stress patterns in speech). Voice emotion perception is of particular interest to researchers because of its implications on social interactions and auditory development in children. With the increasing population of implanted children and adult, our understanding of the impact of early auditory deprivation, CI use, and neuroplasticity in speech emotion development is improving. These are mixed findings regarding pediatric CI users and their competence in areas of emotion perception including facial affect recognition, empathetic behaviors, and emotional attribution abilities. However, the literature on this topic is severely limited and emotion processing in children using CIs is not well understood. Both adults and children with CIs, however, seem to perform comparably with NH adults in CI simulations, suggesting that this population of people may have sufficiently developed adaptive strategies to decipher emotion from

Author Manuscript

impoverished auditory signals. Among all the auditory cues, pitch cues appear to be the most poorly represented auditory stimulus among CI users for vocal emotion identification and production, and a focus on pitch perception rehabilitation and sound processing may improve performance in CI users.

Acknowledgments

References

1. Agrawal D, Thorne JD, Viola FC, et al. Electrophysiological responses to emotional prosody perception in cochlear implant users. Neuroimage Clin. 2013; 2:229–238. DOI: 10.1016/j.nicl. 2013.01.001 [PubMed: 24179776]

2. Agrawal D, Timm L, Viola FC, et al. ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. BMC Neurosci. 2012; 13(113):1–10. DOI: 10.1186/1471-2202-13-113 [PubMed: 22214384]

3. Andrews FM, Diehl NC. Development of a technique for identifying elementary school children's musical concepts. J Res Music Ed. 1970; 18:214–222.

4. Apple W, Streeter LA Krauss RM. Effects of pitch and speech rate on personal attributions. J Pers Soc Psychol. 1979; 37:715–27.

5. Artières F, Vieu A, Mondain M, Uziel A, Venail F. Impact of early cochlear implantation on the linguistic development of the deaf child. Otol Neurotol. 2009; 30(6):736–742. DOI: 10.1097/MAO. 0b013e3181b2367b [PubMed: 19638938]

6. Bachorowski J. Vocal expression and perception of emotion. Curr Dir Psychol Sci. Apr.1999 8:53–57. DOI: 10.1111/1467-8721.00013

7. Boons T, Brokx JP, Dhooge I, Frijns JH, Peeraer L, Vermeulen A, Wouters J, Van Wieringen A. Predictors of spoken language development following pediatric cochlear implantation. Ear Hear. Sep; 2012 33(5):617–39. DOI: 10.1097/AUD.0b013e3182503e47 [PubMed: 22555184]

8. Breitenstein C, Van Lancker D, Daum I, Waters CH. Impaired perception of vocal emotions in Parkinson's disease: influence of speech time processing and executive functioning. Brain Cogn. 2001 Mar; 45(2):277–314. [PubMed: 11237372]

9. Bryant GA, Barrett HC. Vocal emotion recognition across disparate cultures. J Cogn Cult. 2008; 8:135–148.

10. Caldwell M, Rankin SK, Jiradejvong P, Carver C, Limb CJ. Cochlear implant users rely on tempo rather than on pitch information during perception of musical emotion. Cochlear Implants Int. Sep; 2015 16(sup3):S114–20. DOI: 10.1179/1467010015Z.000000000265 [PubMed: 26561882]

11. Chatterjee, M., Christensen, J., Kulkami, A., Deroche, M., Damm, S., Bosen, A., Hozan, M., Limb, C. Voice emotion communication by listeners with cochlear implants. Poster 740 presented at: Association for Research in Otolaryngology 2016 39th Annual Midwinter Meeting; February 20-24, 2016; San Diego, California. http://c.ymcdn.com/sites/www.aro.org/resource/resmgr/Abstract_Archives/UPDATED_2016_ARO_Abstract_Bo.pdf

12. Chatterjee M, Peng S. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. Hear Res. 2008; 235(1):143–156. DOI: 10.1016/j.heares. 2007.11.004 [PubMed: 18093766]

13. Chatterjee M, Zion DJ, Deroche ML, et al. Voice emotion recognition by cochlear507 implanted children and their normally-hearing peers. Hear Res. 2015; 322:151–162. DOI: 10.1016/j.heares. 2014.10.003 [PubMed: 25448167]

14. Chen JK, Chuang AY, McMahon C, Hsieh JC, Tung TH, Li LP. Music training improves pitch perception in prelingually deafened children with cochlear implants. Pediatrics. 2010 Apr; 125(4):e793–800. DOI: 10.1542/peds.2008-3620 [PubMed: 20211951]

15. Chin SB, Bergeson TR, Phan J. Speech intelligibility and prosody production in children with cochlear implants. J Commun Disord. 2012; 45(5):355–366. DOI: 10.1016/j.jcomdis.2012.05.003 [PubMed: 22717120]

16. Ciocca V, Francis AL, Aisha R, Wong L. The perception of cantonese lexical tones by early-deafened cochlear implantees. J Acoust Soc Am. 2002; 111(5):2250–2256. DOI: 10.1121/1.1471897 [PubMed: 12051445]

17. Costa-Giomi E, Descombes V. Pitch labels with single and multiple meanings: A study with French-speaking children. J Res Mus Ed. 1996; 44(3):204–214.

18. Dalla Bella S, Peretz I, Rousseau L, Gosselin N. A developmental study of the affective value of tempo and mode in music. Cognition. 2001; 80:B1–B10. [PubMed: 11274986]

19. Davitz, Joel R. The communication of emotional meaning. New York: McGraw-Hil; 1964.

20. Degé F, Schwarzer G. The effect of a music program on phonological awareness in preschoolers. Front Psychol. 2011 Jun 20.2:124.doi: 10.3389/fpsyg.2011.00124 [PubMed: 21734895]

21. Deroche ML, Kulkarni AM, Christensen JA, Limb CJ, Chatterjee M. Deficits in the sensitivity to pitch sweeps by school-aged children wearing cochlear implants. Front Neurosci. 2016; 10(73):1–15. DOI: 10.3389/fnins.2016.00073 [PubMed: 26858586]

22. Deroche ML, Lu HP, Limb CJ, Lin YS, Chatterjee M. Deficits in the pitch sensitivity of cochlear-implanted children speaking English or Mandarin. Front Neurosci. 2014; 8(282):1–13. DOI: 10.3389/fnins.2014.00282 [PubMed: 24478622]

23. Deroche ML, Zion DJ, Schurman JR, Chatterjee M. Sensitivity of school-aged children to pitch-related cues. J Acoust Soc Am. 2012 Apr; 131(4):2938–47. DOI: 10.1121/1.3692230 [PubMed: 22501071]

24. Dyck MJ, Denver E. Can the emotion recognition ability of deaf children be enhanced? A pilot study. J Deaf Stud Deaf Educ. 2003; 8(3):348–56. [PubMed: 15448058]

25. Ekman P. Are there basic emotions? . Psychol Rev. 1992 Jul; 99(3):550–3. [PubMed: 1344638]

26. Galvin JJ 3rd, Fu QJ, Nogaki G. Melodic contour identification by cochlear implant listeners. Ear Hear. 2007 Jun; 28(3):302–19. [PubMed: 17485980]

27. Geurts L, Wouters J. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am. 2001; 109(2):713–26. DOI: 10.1121/1.1340650 [PubMed: 11248975]

28. Giannantonio S, Polonenko MJ, Papsin BC, Paludetti G, Gordon KA. Experience changes how emotion in music is judged: evidence from children listening with bilateral cochlear implants, bimodal devices, and normal hearing. PloS one. Aug.2015 10(8):e0136685.doi: 10.1371/journal.pone.0136685 [PubMed: 26317976]

29. Gilbers S, Fuller C, Gilbers D, et al. Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. IPerception. 2015; 6(5):1–19. DOI: 10.1177/0301006615599139 [PubMed: 26034566]

30. Gordon KA, Wong DDE, Valero J, Jewell SF, Yoo P, Papsin BC. Use it or lose it? Lessons learned from the developing brains of children who are deaf and use cochlear implants to hear. Brain Topogr. 2011; 24(3–4):204–219. DOI: 10.1007/s10548-011-0181-2 [PubMed: 21479928]

31. Green T, Faulkner A, Rosen S. Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J Acoust Soc Am. 2004; 116(4):2298–2310. DOI: 10.1121/1.1785611 [PubMed: 15532661]

32. Guerrero Lopez HA, Mondain M, de lB, Serrafero P, Trottier C, Barkat-Defradas M. Acoustic, aerodynamic, and perceptual analyses of the voice of cochlear-implanted children. J Voice. 2013; 27(4):523.e1–523.e17. DOI: 10.1016/j.jvoice.2013.03.005

33. Hair HI. Verbal identification of music concepts. J Res Mus Ed. 1981; 29:11–21.

34. Hausen M, Torppa R, Salmela VR, Vainio M, Särkämö T. Music and speech prosody: a common rhythm. Front Psychol. 2013 Sep 2.4:566.doi: 10.3389/fpsyg.2013.00566 [PubMed: 24032022]

35. He A, Deroche ML, Doong J, Jiradejvong P, Limb CJ. Mandarin tone identification in cochlear implant users using exaggerated pitch contours. Otol Neurotol. 2016; 37(4):324–331. DOI: 10.1097/MAO.0000000000000980 [PubMed: 26890043]

36. Hegarty L, Faulkner A. The perception of stress and intonation in children with a cochlear implant and a hearing aid. Cochlear Implants Int. 2013; 14(S4):S35–S39. DOI: 10.1179/1467010013Z. 000000000132 [PubMed: 24533762]

37. Holt CM, McDermott HJ. Discrimination of intonation contours by adolescents with cochlear implants. Int J Audiol. 2013; 52(12):808–815. DOI: 10.3109/14992027.2013.832416 [PubMed: 24053225]

38. Hopyan-Misakyan T, Gordon KA, Dennis M, Papsin BC. Recognition of affective speech prosody and facial affect in deaf children with unilateral right cochlear implants. Child Neuropsychol. 2009; 15(2):136–146. DOI: 10.1080/09297040802403682 [PubMed: 18828045]

39. Hopyan T, Gordon KA, Papsin BC. Identifying emotions in music through electrical hearing in deaf children using cochlear implants. Cochlear Implants Int. Feb; 2011 12(1):21–6. DOI: 10.1179/146701010X12677899497399 [PubMed: 21756455]

40. Huttar GL. Relations between prosodic variables and emotions in normal American English utterances. J Speech Hear Res. 1968 Sep; 11(3):481–7. [PubMed: 5722475]

41. Isshiki I. Regulatory mechanisms of voice intensity variation. JSLHR. 1964; 7:17–29. DOI: 10.1044/jshr.0701.17

42. Juslin PN, Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion. 2001 Dec; 1(4):381–412. [PubMed: 12901399]

43. Kalathottukaren RT, Purdy SC, Ballard E. Prosody perception and musical pitch discrimination in adults using cochlear implants. Int J Audiol. 2015; 54(7):444–452. DOI: 10.3109/14992027.2014.997314 [PubMed: 25634773]

44. Kang SY, Colesa DJ, Swiderski DL, Su GL, Raphael Y, Pfingst BE. Effects of hearing preservation on psychophysical responses to cochlear implant stimulation. J Assoc Res Otolaryngol. 2010 Jun; 11(2):245–65. DOI: 10.1007/s10162-009-0194-7 [PubMed: 19902297]

45. Kehrein, R. The prosody of authentic emotions; Proc. Speech Prosody Conference; 2002. p. 423-426.

46. Ketelaar L, Rieffe C, Wiefferink CH, Frijns JH. Social competence and empathy in young children with cochlear implants and with normal hearing. Laryngoscope. 2013; 123(2):518–523. DOI: 10.1002/lary.23544 [PubMed: 22996702]

47. Kong YY, Cruz R, Jones JA, Zeng FG. Music perception with temporal cues in acoustic and electric hearing. Ear Hear. 2004 Apr; 25(2):173–85. [PubMed: 15064662]

48. Kong YY, Mullangi A, Marozeau J, Epstein M. Temporal and spectral cues for musical timbre perception in electric hearing. J Speech Lang Hear Res. 2011 Jun; 54(3):981–94. DOI: 10.1044/1092-4388(2010/10-0196) [PubMed: 21060140]

49. Krumhansl CL. An exploratory study of musical emotions and psychophysiology. Can J Exp Psychol. 1997 Dec; 51(4):336–53. [PubMed: 9606949]

50. Lane H, Tranel B. The Lombard sign and the role of hearing in speech. JSLHR. 1971; 14:677–709. DOI: 10.1044/jshr.1404.677

51. Lane H, Tranel B. The Lombard sign and the role of hearing in speech. J Speech Hear Res. 1971; 14:677–709.

52. Levin H, Lord W. Speech pitch frequency as an emotional state indicator. IEEE T Syst Man Cyb. 1975; 5:259–273.

53. Luo X, Fu QJ. Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants. J Acoust Soc Am. 2004 Dec; 116(6):3659–67. [PubMed: 15658716]

54. Luo X, Fu QJ, GJ. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. Trends Amplif. 2007; 11(4):301–315. DOI: 10.1177/1084713807305301 [PubMed: 18003871]

55. Meister H, Landwehr M, Pyschny V, Walger M, von Wedel H. The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. International J Audiol. 2009; 48(1):38–48. DOI: 10.1080/14992020802293539

56. Morgan SD, Ferguson SH. Perceived emotional valence in clear and conversational speech. J Acoust Soc Am. 2014; 135:2224.

57. Most T, Peled M. Perception of suprasegmental features of speech by children with cochlear implants and children with hearing AIDS. J Deaf Stud Deaf Educ. 2007; 12(3):350–361. DOI: 10.1093/deafed/enm012 [PubMed: 17478417]

58. Nakata T, Trehub SE, Kanda Y. Effect of cochlear implants on children's perception and production of speech prosody. J Acoust Soc Am. 2012; 131(2):1307–1314. DOI: 10.1121/1.3672697 [PubMed: 22352504]

59. Osgood, CE., Suci, GJ., Tannenbaum, PH. The Measurement of Meaning. Urbana, IL: University of Illinois Press; 1957.

60. Patel AD. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. Hear Res. 2014 Feb.308:98–108. DOI: 10.1016/j.heares. 2013.08.011 [PubMed: 24055761]

61. Peng SC, Chatterjee M, Lu N. Acoustic cue integration in speech intonation recognition with cochlear implants. Trends Amplif. 2012; 16(2):67–82. DOI: 10.1177/1084713812451159 [PubMed: 22790392]

62. Peng SC, Tomblin JB, Cheung H, Lin YS, Wang LS. Perception and production of mandarin tones in prelingually deaf children with cochlear implants. Ear Hear. 2004; 25(3):251–264. DOI: 10.1097/01.AUD.0000130797.73809.40 [PubMed: 15179116]

63. Peng SC, Tomblin JB, Spencer LJ, Hurtig RR. Acquisition of rising intonation in pediatric cochlear implant recipients—a longitudinal study. Int Congr Ser. 2004; 1273:336–339. DOI: 10.1016/j.ics. 2004.08.046

64. Peng SC, Tomblin JB, Spencer LJ, Hurtig RR. Imitative production of rising speech intonation in pediatric cochlear implant recipients. J Speech Lang Hear R. 2007; 50(5):1210–1227. DOI: 10.1044/1092-4388(2007/085)

65. Peng SC, Tomblin JB, Turner CW. Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. Ear Hear. 2008; 29(3):336–351. DOI: 10.1097/AUD.0b013e318168d94d [PubMed: 18344873]

66. Pereira, C. Perception and expression of emotions in speech. Macquarie University; 2000. PhD thesis

67. Planalp S. Varieties of Cues to Emotion in Naturally Occurring Situations. Cognition Emotion. 1996; 10(2):137–154.

68. Ponton CW, Eggermont JJ, Don M, et al. Maturation of the mismatch negativity: Effects of profound deafness and cochlear implant use. Audiol Neurotol. 2000; 5(3–4):167–185. DOI: 10.1159/000013878

69. Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci. 1992 Jun 29; 336(1278):367–73. [PubMed: 1354376]

70. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res Pers. 1977; 11:273–294.

71. Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980; 39(6):1161–1178.

72. Scherer KR, Oshinsky JS. Cue utilization in emotion attribution from audiotory stimuli. Motiv Emotion. 1977; 1:331–346.

73. Scherer, KR. Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In: Weitz, S., editor. Nonverbal communication. New York: Oxford Univ. Press; 1874. p. 105-111.

74. Scherer KR. Vocal affect expression: A review and a model for future research. Psychol Bull. 1986; 99:143–165. [PubMed: 3515381]

75. Shirvani S, Jafari Z, Zarandi MM, Jalaie S, Mohagheghi H, Tale MR. Emotional Perception of Music in Children With Bimodal Fitting and Unilateral Cochlear Implant. Ann Otol Rhinol Laryngol. Jun; 2016 125(6):470–7. DOI: 10.1177/0003489415619943 [PubMed: 26681623]

76. Schorr EA, Roth FP, Fox NA. Quality of life for children with cochlear implants: Perceived benefits and problems and the perception of single words and emotional sounds. J Speech Lang Hear Res. 2009; 52(1):141–152. DOI: 10.1044/10924388(2008/07-0213) [PubMed: 18664684]

77. Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., Gielen, S. Acoustic Correlates of Emotion Dimensions in View of Speech Synthesis; Proc Eurospeech 2001, Aalborg, Vol. 1; p. 87-90.

78. Schubert E. Measuring Emotion Continuously: Validity and Reliability of the Two- Dimensional Emotion-Space. Australian Jnl of Psychology. 1999; 51:154–165. DOI: 10.1080/00049539908255353

79. See RL, Driscoll VD, Gfeller K, Kliethermes S, Oleson J. Speech intonation and melodic contour recognition in children with cochlear implants and with normal hearing. Otol Neurotol. 2013; 34(3):490–498. DOI: 10.1097/MAO.0b013e318287c985 [PubMed: 23442568]

80. Shannon RV. Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. Hear Res. 1983 Aug; 11(2):157–89. [PubMed: 6619003]

81. Skinner ER. A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness; and a determination of the pitch and force of the subjective concepts of ordinary, soft and loud tones. Speech Monogr. 2:81–137.

82. Smith CA, Ellsworth PC. Patterns of cognitive appraisal in emotion. J Pers Soc Psychol. 1985 Apr; 48(4):813–38. [PubMed: 3886875]

83. Soderstrom M, Seidl A, Nelson DGK, Jusczyk PW. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. J Mem Lang. 2003; 49(2):249–267. DOI: 10.1016/S0749-596X(03)00024-X

84. Straatman LV, Rietveld ACM, Beijen J, Mylanus EAM, Mens LHM. Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid. J Acoust Soc Am. 2010; 128(4):1884–1895. DOI: 10.1121/1.3474236 [PubMed: 20968360]

85. Torppa R, Huotilainen M, Leminen M, Lipsanen J, Tervaniemi M. Interplay between singing and cortical processing of music: a longitudinal study in children with cochlear implants. Front Psychol. 2014 Dec 10.5:1389.doi: 10.3389/fpsyg.2014.01389 [PubMed: 25540628]

86. Trainor LJ, Austin CM. Is infant-directed speech prosody a result of the vocal expression of emotion? Psychol Sci. 2000; 11(3):188–195. [PubMed: 11273402]

87. Udall E. Attitudinal meanings conveyed by intonation contours. Language and Speech. 1960; 3:223–34.

88. Van Zee N. Responses of kindergarten children to musical stimuli and terminology. J Res Mus Ed. 1976; 24:14–21.

89. Van Zyl M, Hanekom JJ. Perception of vowels and prosody by cochlear implant recipients in noise. J Commun Disord. 2013; 46(5–6):449–464. DOI: 10.1016/j.jcomdis.2013.09.002 [PubMed: 24157128]

90. Volkova A, Trehub SE, Schellenberg EG, Papsin BC, Gordon KA. Children with bilateral cochlear implants identify emotion in speech and music. Cochlear Implants Int. 2013; 14(2):80–91. DOI: 10.1179/1754762812Y.0000000004 [PubMed: 23453221]

91. Wallbott HG, Scherer KR. Cues and channels in emotion recognition. J Pers Soc Psychol. 1986; 51:690–699.

92. Wang DJ, Trehub SE, Volkova A, Lieshout P. Child implant users' imitation of happy708 and sad-sounding speech. Front Psychol. 2013; 4(351):1–8. DOI: 10.3389/fpsyg.2013.00351 [PubMed: 23382719]

93. Wang W, Zhou N, Xu L. Musical pitch and lexical tone perception with cochlear implants. Int J Audiol. 2011; 50(4):270–278. DOI: 10.3109/14992027.2010.542490 [PubMed: 21190394]

94. Wei C, Cao K, Zeng F. Mandarin tone recognition in cochlear-implant subjects. Hear Res. 2004; 197(1):87–95. DOI: 10.1016/j.heares.2004.06.002 [PubMed: 15504607]

95. Wei WI, Wong R, Hui Y, Au DK, Wong BY, Ho WK, Tsang A, Kung P, Chung E. Chinese tonal language rehabilitation following cochlear implantation in children. Acta Otolaryngol. 2000 Mar; 120(2):218–21. [PubMed: 11603776]

96. Wiefferink CH, Rieffe C, Ketelaar L, De Raeve L, Frijns JHM. Emotion understanding in deaf children with a cochlear implant. J Deaf Stud Deaf Edu. 2013; 18(2):175–186. DOI: 10.1093/deafed/ens042

97. Wong AO, Wong LL. Tone perception of cantonese-speaking prelingually hearing impaired children with cochlear implants. Otolaryngol Head Neck Surg. 2004; 130(6):751–758. DOI: 10.1016/j.otohns.2003.09.037 [PubMed: 15195063]

98. Wouters J, McDermott HJ, Francart T. Sound coding in cochlear implants: From electric pulses to hearing. IEEE Signal Process Mag. Mar; 2015 32(2):67–80. DOI: 10.1109/MSP.2014.2371671

99. Xu L, Zhou N, Chen X, Li Y, Schultz HM, Zhao X, Han D. Vocal singing by prelingually-deafened children with cochlear implants. Hear Res. 2009 Sep; 255(1–2):129–34. DOI: 10.1016/j.heares.2009.06.011 [PubMed: 19560528]

100. Yhun, Lo C., McMahon, CM., Looi, V., Thompson, WF. Melodic Contour Training and Its Effect on Speech in Noise, Consonant Discrimination, and Prosody Perception for Cochlear Implant Recipients. Behavioural Neurology. 2015; 2015:1–10. DOI: 10.1155/2015/352869

101. Zajonc RB. Felling and thinking: Preferences need no inferences. Am Psychol. 1980; 35:151–175.

102. Zeng FG. Temporal pitch in electric hearing. Hear Res. 2002 Dec; 174(1–2):101–6. [PubMed: 12433401]

103. Zhang T, Dorman MF, Fu QJ, Spahr AJ. Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation. Ear Hear. 2012; 33(6):e70–9. DOI: 10.1097/AUD.0b013e318259e5dd [PubMed: 22622705]

Highlights

- Voice emotion is a fundamental component of human social interaction and development.

- Cochlear implant (CI) users are forced to interface with highly degraded prosodic cues.

- CI users demonstrate significant difficulty in recognizing and producing voice emotions.

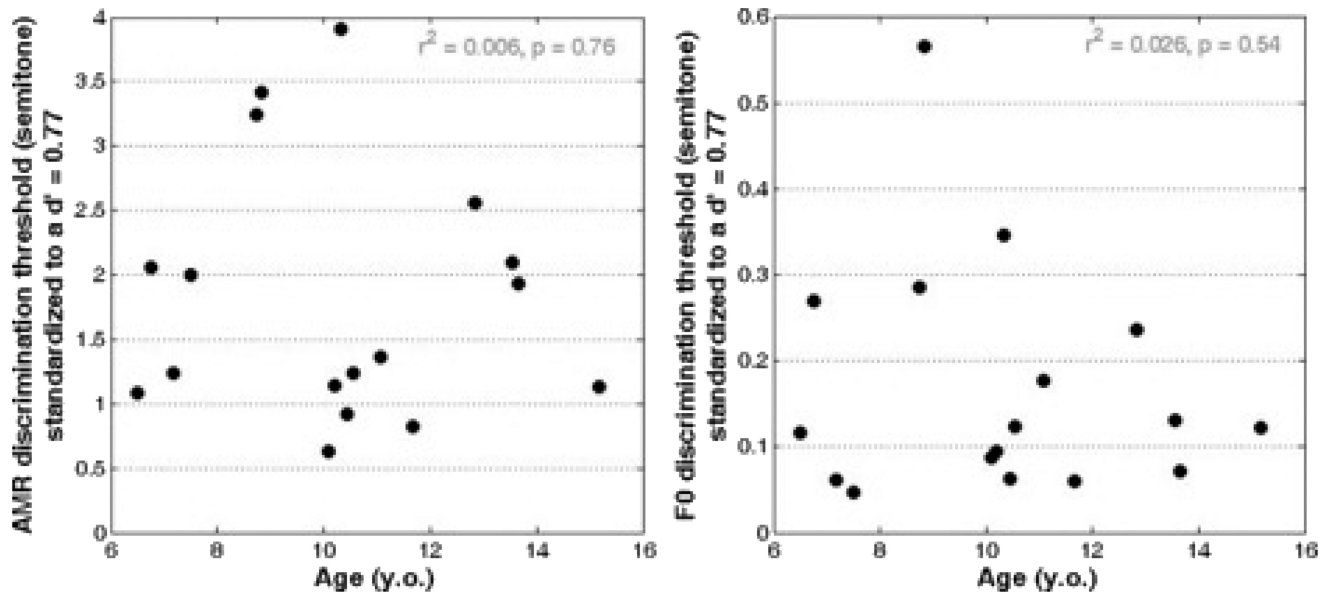- Aural rehabilitation and music training may improve prosody perception.

Figure 1.
17 school-aged children were asked to discriminate sinusoidal amplitude modulation rate (AMR) of broadband noise and to discriminate the fundamental frequency (F0) of broadband sine-phase harmonic complexes (Deroche, et al., 2012). Children who had adult-like sensitivity were not necessarily the oldest listeners. For example, the youngest research subject was 6.5 years old and demonstrated a standardized threshold of only 1.08 semitones for AMR discrimination and 12 cents for F0 discrimination, suggesting that children's sensitivity to pitch (regardless of the underlying cue) does not systematically improve beyond 6 years of age.
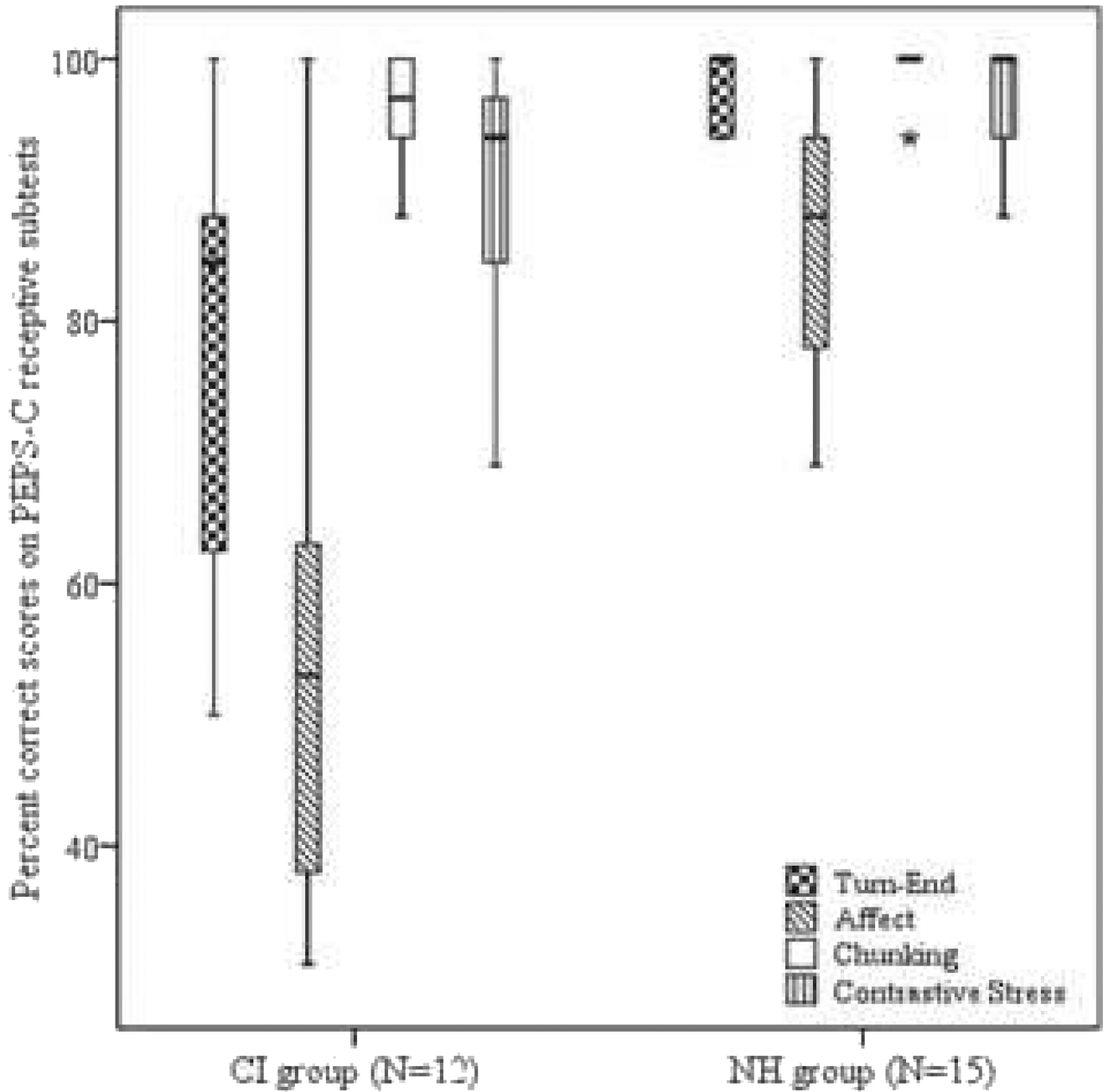
Figure 2.
Percent correct scores for CI users (left) and NH listeners (right) on four subtests of PEPS-C, including turn-end, affect, chunking, and contrastive stress (Kalathottukaren, et al., 2015). As indicated by the asterisk, chunking was the only task for which CI users did not perform significantly worse than NH listeners.
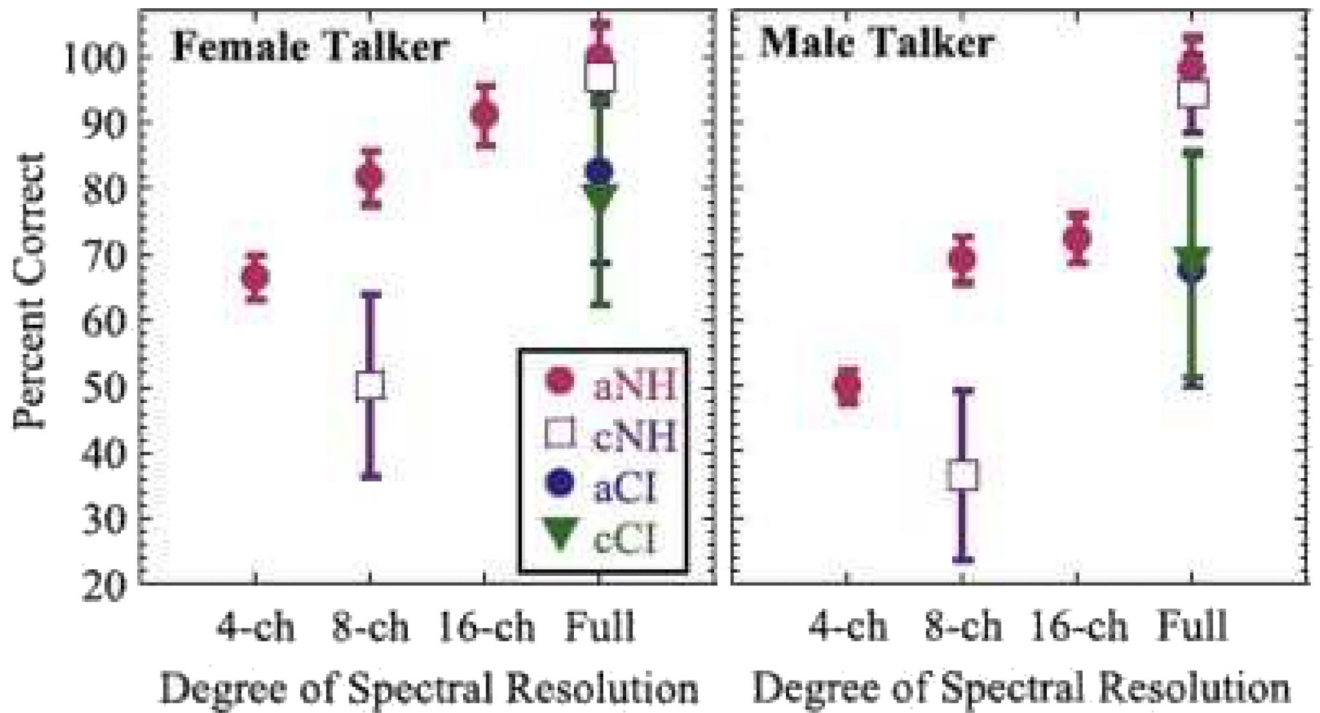
Figure 3.
Mean emotion recognition scores for adult NH, adult CI, child NH, and child CI study groups across full spectrum speech and speech presented with three levels of spectral degradation (Chatterjee, et al., 2015). Child and adult CI users' performance were similar to one another's and comparable to adult NH's performance in 8-channel degraded speech.
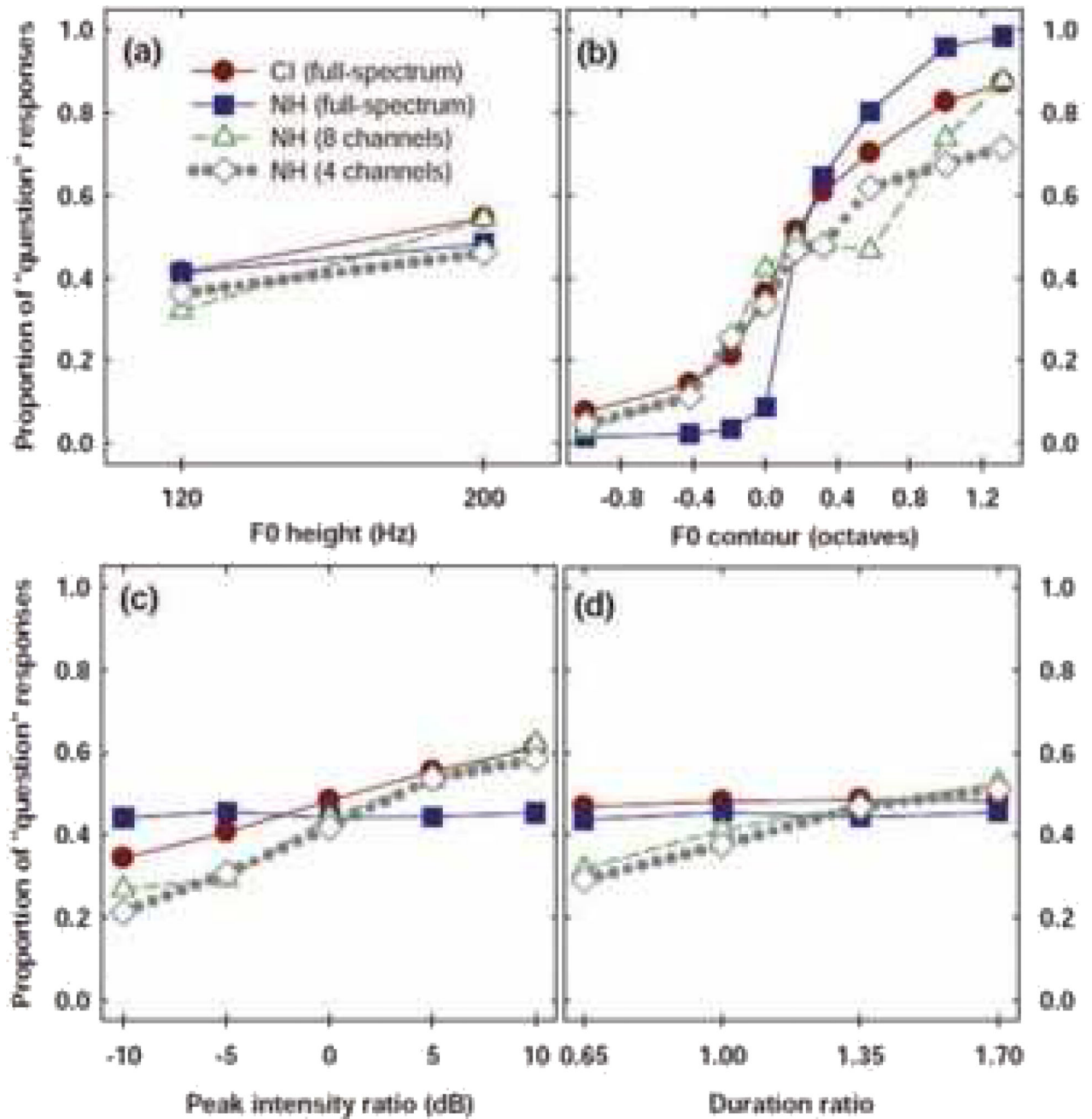
Figure 4.
Cochlear implant (CI) versus normal hearing (NH) group mean proportions of question judgements as a function of an acoustic dimension (F_0 height, F_0 contour, peak intensity ratio, and duration ratio) (Peng, et al., 2012). The NH group was also subject to different listening conditions (full-spectrum, 8-, and 4-channel conditions) as depicted by different symbols.
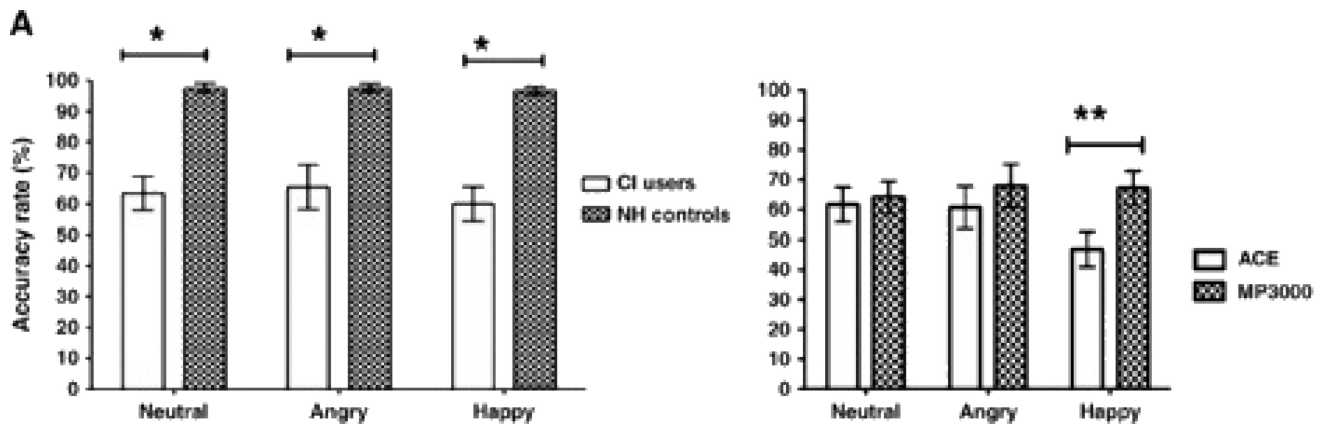
Figure 5.
Accuracy rate in identification of neutral, angry, and happy prosody in NH listeners and CI users using ACE and MP3000 strategies (Agrawal, et al., 2013).
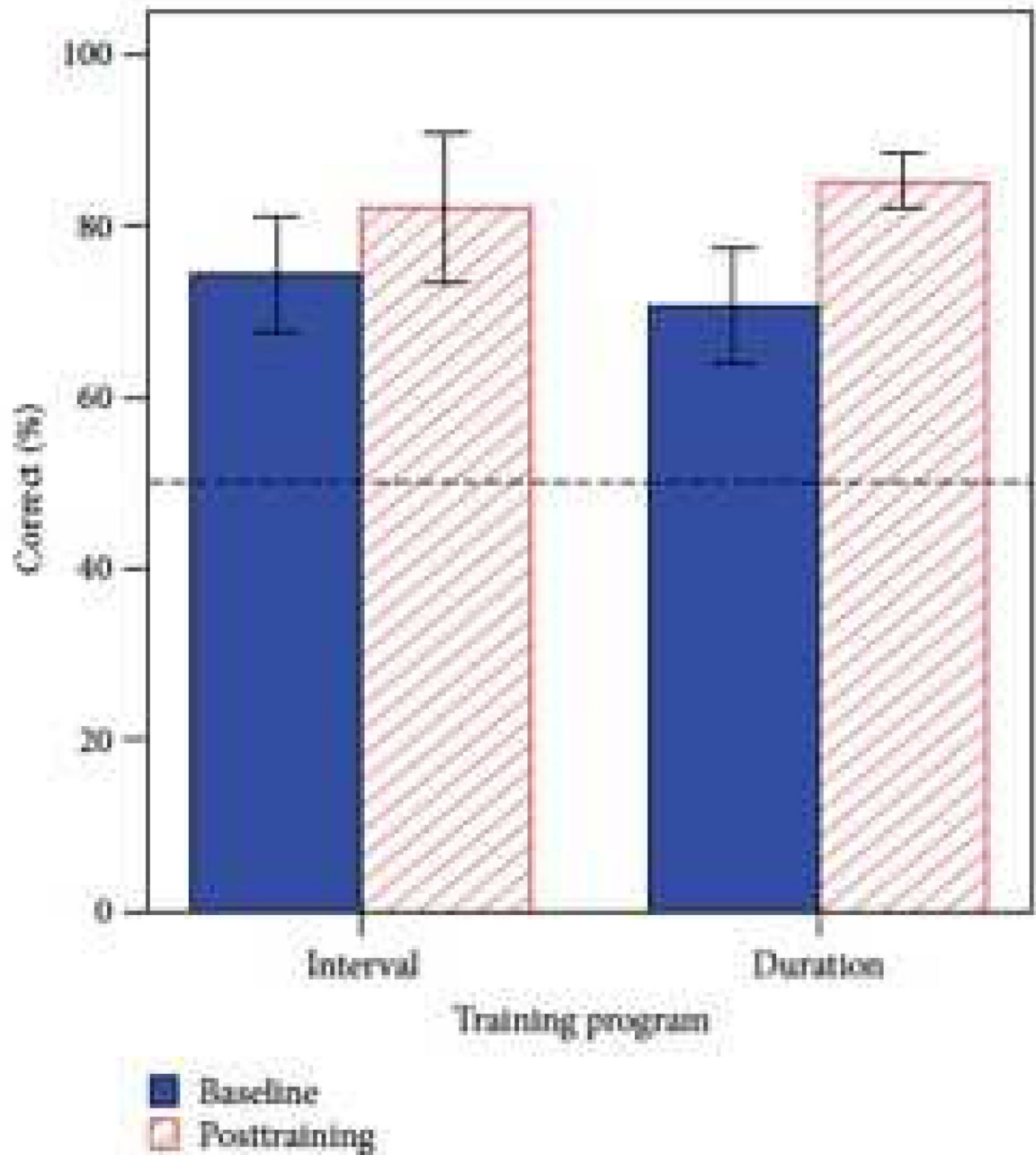
Figure 6.
Graphical representation of improvements in question-statement discrimination (as measured by the PEPS-C test) in CI users after participation in melodic contour training (Yhun Lo, et al., 2015). Two programs of melodic contour training were administered. The "Interval" program incorporated manipulations of note intervals, whereas the "Duration" program changed the note durations, in order to make the tests more challenging. The dashed line represents chance performance.