Processing of Musical and Vocal Emotions through Cochlear Implants

By

Duha G. Ahmed, MBBS

Department of Otolaryngology- Head and Neck Surgery

McGill University, Montreal

April, 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

TABLE OF CONTENTS

TABLE OF CONTENTS ii
ABSTRACT iv
RÉSUMÉ v
ACKNOWLEDGMENTS vi
CONTRIBUTION OF AUTHORS vii
CHAPTER 1: Thesis Introduction 8
1.1 Introduction 8
1.2 Emotional perception 8
1.3 Auditory emotions 8
1.4 Neuro-imaging studies of emotional processing 9
1.5 Cochlear implants users' perception of auditory emotions 11
1.6 Brain mechanisms of emotion processing in cochlear implants users 14
1.7 Thesis Rationale 15
CHAPTER 2: Recognition of Musical and Vocal Emotions through Cochlear Implants Simulation (Manuscript 1) 16
2.1. Introduction 18
2.2 Methods 22
2.2.1 Participants 22
2.2.2 Stimuli 22
2.2.3 Procedure 22
2.2.4 Analysis 23
2.3. Results 25
2.3.1 Accuracy 25
2.3.2 Arousal and Valence Ratings 27
2.3.3 Confidence judgments 30
2.3.4 Correlation of emotional judgments with acoustical features 31
2.3.4.1 Accuracy 31
2.3.4.2 Arousal 31
2.3.4.3 Valence 32
2.4. Discussion 33

2.5. Acknowledgement and funding sources 36
2.6 Connecting text 36
CHAPTER 3: Neural Processing of Musical and Vocal Emotions through Cochlear Implants (Manuscript 2) 37
3.1. Introduction 39
3.2. Methods 41
3.2.1. Participants 41
3.2.2. Stimuli 41
3.2.3. Task and procedure 42
3.2.4. Neurophysiological recording 42
3.2.5. Processing and artifact rejection 43
3.2.6. Statistical Analysis 44
3.3. Results 45
3.3.1. Behavioral results 45
3.3.2. ERP results 45
3.4. Discussion 52
3.4.1. The effect of cochlear implant simulation 52
3.4.2. Comparing vocal and musical emotional processing 54
3.5. Conclusion 55
CHAPTER 4: General Discussion & Conclusions 56
4.1 General discussion 56
4.2 General conclusion 57
REFERENCES 59
ABBREVIATIONS 72
LIST OF TABLES AND FIGURES 73

ABSTRACT

Cochlear implants (CI) partially restore hearing in the deaf. However, the ability to recognize emotions in speech and music is limited due to the implant's technological limitations and the impaired neural pathways that developed after sensorineural hearing loss. This leads to developmental and socioeconomic problems for CI-users and thus a decrease in quality of life. Behavioural and neural correlates of this deficit are not yet well established. This thesis aims to characterize the effect of CIs on auditory emotion perception and, for the first time, to directly compare vocal and musical emotion perception through a CI-simulator. The thesis investigated the ability of normal hearing individuals to perceive basic emotions in CI-simulated vocal and musical sounds, using a behavioural task and electroencephalography (EEG). In the behavioural study, the perception of musical and vocal emotions was impaired in the CI-simulated condition. Perception was correlated with timbral acoustic cues. In the EEG study, the averaged event-related potentials' components had reduced amplitudes and delayed latency as early as 50 milliseconds in the CI-simulated condition. Using this previously validated neuro-behavioural approach with CI-users can further enhance our knowledge and prove the importance of timbral acoustical cues for emotion recognition. It can lead to developing new processing strategies that capitalize on these cues, leading to better perception of auditory emotions and thus improving the quality of life for CI-users.

RÉSUMÉ

Les implants cochléaires (IC) restaurent partiellement l'ouïe chez les sourds. Cependant, la capacité à reconnaître les émotions dans la parole et la musique est limitée en raison des limites technologiques de l'implant et de la voie neurale altérée après perte d'audition neurosensorielle. Ceci conduit à des problèmes développementaux et socio-économiques pour les utilisateurs de IC et donc à une diminution de la qualité de vie. Les corrélats comportementaux et neuronaux de ce déficit ne sont pas encore bien établis. Cette thèse vise à caractériser l'effet des IC sur la perception des émotions auditives et, pour la première fois, à comparer directement la perception de l'émotion vocale et musicale à travers un simulateur IC. L'étude a investigué la capacité des individu avec une ouïe normale à percevoir les émotions fondamentales dans les sons vocaux et musicaux simulés par IC, en utilisant une tâche comportementale et l'électroencéphalographie (EEG). Dans l'étude comportementale, la perception des émotions musicales et vocales a été altérée dans l'état simulé par IC. La perception était corrélée avec les indices timbrales acoustiques. Dans l'étude EEG, les composantes des potentiels moyens liés à l'événement avaient des amplitudes réduites et une latence retardée dès 50 millisecondes dans les conditions simulées par IC. En utilisant cette approche neuro-comportementale validée précédemment avec les utilisateurs de IC, nous pouvons améliorer nos connaissances et prouver l'importance des signaux acoustiques timbrales pour la reconnaissance des émotions. Cela peut conduire à développer de nouvelles stratégies de traitement qui capitalisent sur ces indices, conduire à une meilleure perception des émotions auditives et ainsi améliorer la qualité de vie des utilisateurs de IC.

ACKNOWLEDGMENTS

I would first like to thank my thesis supervisors Dr. Alexandre Lehmann and Dr. Anthony Zeitouni, for their support and guidance.

I would also like to thank my colleagues and friends at BRAMS for their help and support. And a special thanks to Mihaela Felezeu for her help with EEG.

Finally, I must express my very profound gratitude to my mother and siblings for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Duha Ahmed

CONTRIBUTION OF AUTHORS

This thesis comprises 2 manuscripts. The first manuscript (Chapter 2), entitled, "Recognition of Musical and Vocal Emotions through Cochlear Implants Simulation" is by S Paquette, D Ahmed G., I Peretz, and Alexandre Lehmann, and was submitted to the Journal of Hearing Research. The work was originally conceived by A Lehmann, with contributions from S Paquette, D Ahmed, and I Peretz. All data was collected only by D Ahmed; S Paquette wrote the first draft of the manuscript; S Paquette, D Ahmed, and A Lehmann analyzed the data and interpreted the results; finally, D Ahmed, A Lehmann, S Paquette, and I Peretz critically reviewed and revised the manuscript.

The second manuscript (Chapter 3), entitled, "Neural Processing of Musical and Vocal Emotions through Cochlear Implants" is by D Ahmed, S Paquette, A Zeitouni, and A Lehmann, and was submitted to the Journal of Clinical EEG & Neuroscience. The work was originally conceived by A Lehmann, with contributions from D Ahmed, S Paquette, and A Zeitouni. All data was collected only by D Ahmed; D Ahmed wrote the first draft of the manuscript; D Ahmed, S Paquette, and A Lehmann analyzed the data and interpreted the results; finally, D Ahmed, S Paquette, A Lehmann, Anthony Zeitouni, critically reviewed and revised the manuscript.

CHAPTER 1: Thesis Introduction

1.1 Introduction

Cochlear implants (CI) are surgically implanted devices that partially restore hearing abilities in those with severe to profound hearing loss. However, the capacity to recognize emotions in speech and music is limited, because understanding both vocal and musical expressions require the perception of specific acoustic cues (Juslin and Laukka, 2003). CI users' perception of many of these cues is severely impoverished due to the limitations of the electrical signal transmitted by the implant and due to the impaired neural-pathway post-sensorineural hearing loss (Nakata et al., 2012; Wang et al., 2013; Volkova et al., 2013). Perceiving emotions is essential for social interaction and development. Its deficit leads to miscommunication, and eventually, it might lead to depression and a general decrease in the quality of life of cochlear implant users (Wiefferink et al., 2012).

1.2 Emotional perception

An important element of social interaction is communicating emotions (Ekman, 1992). It helps us to empathize and to rejoice with others, and to alert us to sarcasm or hazards. The communication of emotions occurs through vocal sounds of verbal & nonverbal expressions (Scherer, 1986) (e.g. growls, screams, and laughter), facial expressions, postures and gestures (Ekman and Friesen, 1969). It has been proposed that there is a group of basic emotions from which all other emotions are derived (Ekman, 1992). Basic emotions include fear, anger, sadness, enjoyment, disgust and surprise. They are discrete, innate, universal and easily conveyed and perceived through facial, verbal and nonverbal expressions.

1.3 Auditory emotions

Neuropsychological research indicates that both music and speech use acoustical-cue transmission to relay their messages through changes in pitch (fundamental frequency (F0)), intensity and duration (Juslin, 1998; Scherer, 2003). Moreover, Juslin and Lukka (2003) have suggested the presence of emotion-specific patterns of cues across voice and music for the three major cues (speech rate/tempo, vocal intensity/sound level, and high-frequency energy). Each of these cues is neither necessary nor sufficient, but the more cues used, the more reliable the communication is (Juslin, 2000).

Some musical features were found to be processed through the same pathways as speech (e.g., timbre); while others elicited some unique neurophysiological responses (e.g., tonality) (Patel and Peretz, 1997; Peretz, 2002). For these reasons, a close relationship has been always suggested between the processing of these two means of emotional communication.

The previous literature suggests that the recognition of basic emotions in auditory stimuli is dependent on the transmitted acoustical cues for both voice and music. The parameters of these acoustical cues changes for each basic emotion in a way that makes them specific and recognizable for that emotion.

1.4 Neuro-imaging studies of emotional processing

How the brain processes vocal emotions can be investigated in two ways, using two modalities. One is “the when”, which is the time course of decoding the acoustical cues. It is studied using electroencephalography (EEG) by measuring evoked event-related potentials (ERPs) and time-frequency analysis (TF). The second is “the where”, which are the anatomical regions of the brain that are involved in the processing of emotional perception. These are studied using brain lesions and neuroimaging employing functional magnetic radio imaging (fMRI).

EEG measures brain activity, and when averaged, it can be used to obtain ERPs, which are the neural responses associated with specific events. Recorded ERP waveforms consist of a sequence of positive and negative voltage deflections, which are called peaks, waves, or components. These waves are sometimes named after their peak’s position (N1, P2, N2, and P3) or their latency (e.g., P50 is a positive wave at 50 ms from stimulus onset, and N100 is a negative wave at 100 ms…) (Luck, 2014). TF measures brain oscillations at various frequencies (theta, alpha, beta, and gamma). Brain oscillations have been found to be a powerful measurto analyze the cognitive processes related to emotion processing (Başar et al., 1999; Krause, 2003).

Using ERPs, Schirmer and Kotz (2006) have described a model of three stages for the processing of emotionally-relevant acoustic cues in vocal stimuli. The first stage is the sensory processing stage. It occurs in a pathway that runs from the ears to the auditory cortex. In this phase, the frequency and intensity of the stimulus modulate the amplitude of N1 (occurs 100 ms after the onset of a stimulus). The second stage of processing is the integration of emotionally important acoustic cues and extracting a significant emotion. It happens 200 ms (P200) after stimulus onset, in a pathway between the auditory cortex to the lateral superior temporal gyrus (STG) and the superior temporal sulcus (STS). Moreover, linguistic auditory objects processing was found to be localized in the left hemisphere's STS and processing of paralinguistic aspects of vocal speech (speaker gender, age, prosody, emotional state) is lateralized to the right middle and anterior STS. The third stage is cognition, which is the integration of the vocal (the extracted emotion) with the verbal (speech semantics processing). Semantics processing occurs at 400 ms from word onset with activation of the left inferior frontal gyrus. Many scientists in their pursuit to explain auditory decoding have adopted this model.

However, in contrast to Schirmer and Kotz's findings, Iredale et al. (2013) reviewed the neural correlates of emotional prosody, and they found N1 to have a greater amplitude in response to words spoken in emotional prosody compared to neutral prosody. This might suggest that emotional prosody processing starts at an earlier stage (N1) than what was previously suggested (P200). Furthermore, Liu et al. (2012) investigated the processing of nonverbal vocal emotions and found that it starts as early as 50 ms (increased P50 amplitudes). Their findings demonstrate that the response to nonverbal emotional vocalization is an automatic and rapid process. Pell at al. (2015) further examined this topic and compared the processing of speech prosody (anger, happiness, and sadness) and nonverbal vocalizations. N100 amplitude was significantly reduced and P200 was significantly larger for vocalizations than speech. This suggests that nonverbal vocalizations are prioritized over speech due to their basic nature and are preferentially processed as early as N100 (Jessen and Kotz, 2011; Paulmann et al., 2013). In conclusion, these two studies support an evolutionary view of emotions (Izard, 1993; Keltner and Gross, 1999). For instance, a scream, which is a nonverbal primitive shout for help, is quicker to induce fear and subsequent flight than explaining the danger verbally.

The same is thought about processing music. A close relationship between voices and music has been a recurrent notion in the literature. The timber in music is thought to be as the prosody in a human voice. This resemblance led to the assumption that they are processed through the same neural circuit in the brain. Moreover, it was found that the processing of musical meaning information is reflected in (at least) in two negative ERP components: The N400 and the N5. The N400 reflects processing of semantical meaning in both language and music, the N5 has so far only been observed for the processing of musical information (Koelsch, 2011).

In summary, the stages of emotional processing are; (1) sensory processing of acoustical signals (P50: nonverbal, N100: verbal). (2) Integration of emotionally specific acoustic cues (P200); and (3) Cognition and processing of semantical and musical information and integrating it with emotional cues (N400 and N5 for music).

1.5 Cochlear implants users' perception of auditory emotions

Cochlear implants are surgically implanted devices that partially treat sensorineural hearing loss (SNHL) by electrically stimulating the auditory nervous system. A cochlear implant consists of an external microphone (capture sounds signals), an audio processor (amplify, compress, and filter sound signals), and a transcutaneous transmitter which transmits encoded signals to the implanted part of the device (magnet, telemetry coil, and a hermetically sealed electronics housing). In the electronics housing, the signal is decoded and sent through an electrode array to the spiral ganglion neurons, transduced into action potentials, and delivered to the brain (Carlson et al., 2012).

With SNHL, the auditory cortex can reorganize itself due to the deprivation of acoustic signals. After implantation, the brain adapts to the new stimulation and rewires itself to transmit signals to the auditory cortex (Fallon et al., 2008). Presumably this occurs because the device processor has limitations. The processed, transmitted signals do not contain all the fine spectro-temporal cues needed to perceive emotions as accurately as normal hearing (NH) individuals do. The transmission of temporal cues (e.g. rhythm and tempo) is determined by the sampling rate of the device, it is of high yield. Thus, there is minimal impairment in rhythm and tempo discrimination's tasks (Gfeller et al., 2000; Looi et al., 2008; Kong et al., 2004). On the other hand, spectral cues (pitch) are degraded due to the excessive spread of electrical current away from the electrode (Macherey & Carlyon, 2014). It suggests that CI users might be able to differentiate between different vocal emotions depending on the rhythm of verbal utterances rather than their pitch. In a study by Nakata et al. (2011), Japanese CI-children were asked to detect prosody in semantically neutral sentences. The sentences were spoken in happy (fast rhythm and high pitch), sad (slow rhythm and low pitch) and angry prosody (fast rhythm and high pitch). CI- children performed poorly compared to NH-children, but they were able to detect happy and sad prosody above chance levels. Their results suggest that it is possible they relied primarily on differences in speaking rhythm, which led to (a) the successful differentiation of happy from sad utterances and (b) unsuccessful differentiation of happy from angry (both had fast rhythm). The same could be said for musical emotions' perception. Hopyan et al. (2011) found that pre-lingually deaf children with unilateral implants distinguished happy from sad classical music excerpts. Volkova et al. (2013) found that children with bilateral cochlear implants are able to identify happy and sad emotions but they were poorer than their normal hearing peers. The ability of CI- children to recognize both happy and sad emotions in verbal and musical stimuli suggests that they can recognize emotions on the basis of articulatory timing cues and speaking rate (rhythm), even though normal hearing (NH) children do so largely on the basis of pitch and timbre cues (Vongpaisal et al., 2010).

In addition, Caldwell et al. (2016) found that CI-users were not affected by varying consonance and dissonance when appraising musical melodies. Consonance and dissonance are fundamental features of musical harmony that depend on pitch relationships between musical notes or chords. The first is usually perceived as pleasant and the second as unpleasant (Fishman et al., 2001). They suggested that this finding is another demonstration of decreased pitch perception, which may be partially responsible for CI-users' reduced musical enjoyment (Lassaletta et al., 2007).

Also, Giannantonio et al. (2015) further examined this in relation to access to auditory cues during development. They compared the perception of musical emotions by changing the tempo, mode and both tempo and mode of the musical excerpts in children bilaterally implanted versus unilaterally implanted with contralateral hearing aid use (bimodal devices) and normal hearing peers. They found that all CI-users - in contrast to the normal reliance on mode cues - changed their opinions mostly when tempo changed and adding mode changes to the tempo changes had little effect. This again points out the importance of tempo cues for perceiving emotions in music. However, bimodal CI users (who received a CI at older ages, and thus had longer access to acoustic sound in both ears) were found to be less swayed by tempo changes than children who only had access to hearing through the electrical pulses of the CI because of poorer hearing thresholds. This suggests that this "complementary acoustic information prevented the increased reliance on tempo by promoting the use of mode cues and/or reducing a developmental shift toward tempo cues" (Giannantonio et al., 2015).

These previous studies have used complex stimuli either vocal (spoken sentences) or musical (songs) that may have underestimated the deficit in auditory emotional processing. Factors such as semantics, tempo (music), duration and context may involve the interaction of non-affective processing systems and contain redundancy that could make emotional recognition easier (Paquette et al., 2013). To determine the extent of the deficit, basic emotions in their most primitive expressions must be tested (Lehmann and Paquette, 2015).

Cochlear-implant users rely on rhythm and tempo to perceive emotions in speech and music, not only because of the implant's processing mechanisms but also perhaps because the brain adapts and capitalizes on the cues most available in the signals. So presumably, pitch perception can be enhanced by either improving the processing strategies of the device or by persevering the ability to perceive pitch cues for patients with residual hearing by using bimodal devices.

1.6 Brain mechanisms of emotion processing in cochlear implants users

The brain mechanisms underlying the processing of auditory emotions in CI users are poorly understood. Only one study (Agrawal et al., 2013) examined this neural pathway using electroencephalography. They compared different implant processing strategies in regards to emotional prosody perception using ERPs and time-frequency analysis. Agrawal found CI users to exhibit prolonged ERP latencies and reduced amplitudes compared with NH controls. However, the structure of the ERPs was similar in both CI users and NH controls and happy prosody induced a larger P200 peak amplitude in one of the processing strategies. This suggests that the central auditory system in CI users processes the prosodic stimuli in a relatively normal fashion (Emotional identification occurs at P200). In regards to the time-frequency analysis, theta and gamma bands showed a significant power increase with the emotional stimuli, whereas alpha and beta frequencies showed no significant difference. In their conclusion, they advocated the use of ERPs and time-frequency analysis (Gamma band activity) as a good measure for comparing the performance of speech coding strategies for prosody recognition in CI users.

To assess which acoustic cues are accessible to CI-users, their ability to perceive music can be examined. Sandmann et al. (2010) measured the mismatched negativity (MMN) in response to the multi-deviant oddball paradigm. MMN is a component of ERP. It is the response to an odd stimulus in a sequence of stimuli. It is sensitive to small, nearly indiscriminable acoustic changes and is largely independent of attention, which makes it an objective index of auditory discrimination accuracy. The deviant stimuli in this study differed from the standard tone in either frequency, intensity or duration. They found that CI users had smaller amplitudes for frequency and intensity deviations than their normal hearing controls. This result showed that cochlear implant users have difficulties in perceiving small changes in acoustic properties of musical sounds, mainly because of the pitch's components (frequency & intensity). This in turn supports the notion that, not all cues are accessible through the implant. Currently however, there is only preliminary knowledge of which cues are transmitted, and of which cues are lost.

1.7 Thesis Rationale

In view of the reduced hearing-perceptual capabilities of CI-patients summarized in the preceding section, the goal of this thesis study will be to better characterize emotional processing in cochlear implant users and in normal-hearing subjects across the domains of music and voice. To do so, validated affective emotional bursts will be used that represent three basic emotions (happiness, sadness, fear), as well as neutral emotion, by employing (i) Musical Emotional Bursts (MEB) (Paquette et al., 2013) and (ii) Montreal Affective Voices (MAV) (Belin et al., 2008). These stimuli will be used because they are known to overcome the confounding factors that underestimate the deficit in auditory-emotional processing. The thesis goal will be implemented by validating this approach in normal hearing subjects, using two experimental studies. In the first study (chapter 2), a behavioural paradigm will be used to assess the effect of cochlear implant simulation on the perception of auditory emotions in normal-hearing individuals. Based on the current literature, and a well-controlled behavioural design, an accurate characterization of the deficit in auditory emotions' recognition is expected. In the second study (Chapter 3), an evaluation of the neural mechanisms of auditory emotional processing of natural and CI-simulated sounds will be made in normal-hearing individuals using electroencephalography and ERPs.

After validating the behavioural and EEG tasks with CI-simulation and normal hearing participants, it is suggested that this paradigm should be used with CI-users and matched normal-hearing controls in a future study.

The information gained from this research should improve knowledge of brain rewiring and neural plasticity following CI-implantation, as well as the effect of an implant's signal degradation on auditory perception of emotion. This should lead to better understanding of the implant's limitations, which would be critical for improving perceptual-hearing capabilities of CI-patients.

The next chapter will present the first study using the behavioural paradigm to assess the effect of cochlear implant simulation on the perception of auditory emotions in normal-hearing individuals.

CHAPTER 2: Recognition of Musical and Vocal Emotions through Cochlear Implants Simulation (Manuscript 1)

By

Paquette, S[1]., Ahmed, G.D[2]., Peretz, I[1]., & Lehmann, A[1,2].

(1) International Laboratory for Brain Music and Sound Research, Center for Research on Brain, Language and Music, Department of Psychology, University of Montreal, Montreal, Québec, Canada

(2) Department of Otolaryngology Head and Neck Surgery, McGill University, Montreal, Québec, Canada

Submitted to: Hearing Research (Manuscript number: HEARES_2016_502)

Corresponding author:

Sebastien Paquette

BRAMS-Pavillon 1420 boul. Mont-Royal

University of Montreal

Montréal (QC) Canada H2V 4P3

T: +1 514 343 6111 P3985 E: sebastien.paquette.1@umontreal.ca

Abstract: Recognition of musical and vocal emotions through cochlear implants simulation

Cochlear implants can successfully restore hearing in profoundly deaf individuals and enable speech comprehension. The acoustic signal provided is severely degraded and, as a result, many important acoustic cues for perceiving emotion in voices and music are unavailable. The deficit of cochlear implant users in auditory emotion processing has been clearly established and has negative consequences for development and socio-professional integration. However, the exact extent of this deficit and the specific cues that remain available to cochlear implant users are unknown due to several confounding factors.

Here we assessed the recognition of the most basic form of auditory emotion and aimed to identify which acoustic cues are most relevant to recognize emotions through cochlear implants.

To do so, we used specific stimuli that allowed an assessment of basic auditory emotions (vocal, musical) while controlling for confounding factors. These stimuli, together with a validated cochlear implant simulation approach, allowed testing of natural versus cochlear implant hearing in the same participants.

Results show that timbral acoustical cues (brightness, energy, and roughness) correlate with participant emotional rating for both vocal and musical emotion bursts in the cochlear implant simulation condition, suggesting that specific attention should be given to these cues in the design of cochlear implant processors and rehabilitation protocols, in order to improve emotion perception in this population. For instance, musical therapy focused on timbre could improve emotion perception and regulation, and thus improve social functioning in cochlear implant children during development.

Keywords:

Cochlear implants; emotional acoustical cues; cross-domain comparison; music; voice; timbre.

2.1. Introduction

Cochlear implants (CI) can successfully restore some sense of hearing in profoundly deaf individuals. After intensive rehabilitation, most users can reach a good level of speech comprehension. However, the acoustic signal provided by the device is severely degraded, and the frequency resolution is poor. As a result, CI users are less accurate at discriminating pitch patterns (Gfeller and Lansing, 1992, 1991; Hopyan et al., 2012) and identifying pitch changes or direction (Gfeller et al., 2002a; Laneau et al., 2004) than normal hearing controls.

Because understanding both vocal and musical emotional expressions require in part the processing of specific acoustic cues (Juslin and Laukka, 2003), most of which are based on pitch, CI users are impaired at perceiving those expressions. As opposed to normal hearing individuals who seem to rely on mean pitch (or mode: Hopyan et al., 2016), CI users are biased toward using pitch range cues (less salient cues: Gilbers et al., 2015). Perceiving emotion conveyed through vocal intonations (prosody) and through music is therefore highly challenging for CI users (Nakata et al., 2012; Wang et al., 2013). Fortunately, other non-pitch based cues can also convey emotions (Gabrielsson and Lindström, 2001; Gosselin et al., 2015) and these cues, such as temporal (e.g. rate, rhythm) and spectro-temporal (e.g. timbre) variations, are available to some extent to CI users (Gfeller et al., 2002c; Kong et al., 2004; Looi et al., 2012) and reported to be used for emotion discrimination (e.g. tempo: Hopyan et al., 2016). CI users can discriminate happy from sad auditory emotions or identify them above chance, but not as well as normal hearing controls (Hopyan et al., 2012; Stabej et al., 2012; Volkova et al., 2013). In a recent study by Ambert-Dahan and colleagues (2015), CI users identified musical emotions above chance level (happiness, sadness, threat and peacefulness), but demonstrated deficits in perceiving arousal (relaxing/stimulating) of musical excerpts, whereas ratings of valence (negative/ positive) remained unaffected. Judgments of emotional dimensions (arousal, valence) of stimuli do not seem to be uniformly impaired in CI users.

Accurately perceiving emotions (or emotional dimensions) in language is crucial for social integration, interpersonal communication and a general quality of life. Access to musical emotions is also paramount: the most common reason for listening to music is its rich emotional content (Juslin and Sloboda, 2001; Lonsdale and North, 2011), and many deaf individuals specify music enjoyment as a major motivation for getting an implant (Gfeller et al., 2002a). This is all the more critical for developing children receiving an implant, as a failure to perceive their parents' or teacher's emotional expressions will result in maladjusted behaviour and the incapacity to express emotions in their own voice. This auditory emotional deficit can extend to general emotion processing. Although Hopyan-Misakyan and collaborators (2009) have shown that 10-year-old children who had received their CI prelingually were just as accurate as their normal hearing controls in discriminating between different facial expressions, normal hearing preschool children performed significantly better on facial expression recognition than prelingually implanted CI children (Wang et al., 2011), suggesting that there is a delayed development in deaf children with CI. In contrast, Wiefferink and collaborators (2013, 2012) observed impaired emotional regulation (emotion expression and coping strategies) and social functioning (social competence and externalizing behaviours) in CI users compared to normal hearing individuals.

Whereas a deficit in auditory emotion processing has been clearly established in CI users, the extent of this deficit is unknown. A first issue is that several confounding factors may have caused an underestimation of the extent of the deficit because previous studies have used complex stimuli, either vocal (spoken sentences) or musical (songs). Factors such as semantics and language (voice), tempo (music), duration and context may involve the interaction of non-affective processing systems and contain redundancy that could make emotional recognition easier (Paquette et al., 2013). Hence, a necessary step to further study fundamental auditory emotion processing in CI users is to test basic emotions in their most primitive expressions (Lehmann and Paquette, 2015). Minimizing the role of confounding factors is also a key step in addressing two important issues in the literature. How CI users are able to perceive emotions (besides visual cues) despite their lack of access to many critical pitch-based cues has not been determined. Which cues do they rely on and what explains the large inter-individual variability? The second issue is concerned with the direct comparison of musical and vocal emotional processing in CI users. Music and voice are indeed hard to compare because of many confounding factors and also because voice has continuous pitch whereas most music studies on emotion employ the piano, which has a discrete pitch.

The goal of this study is to address these important questions by assessing the recognition of the most basic form of auditory emotion and comparing it between a cochlear implant and normal-hearing conditions. We aimed to identify which acoustic cues are most relevant for CI users to recognize emotion: correlations were made between emotional recognition and acoustical features to see which is used to extract auditory emotions when pitch perception is degraded. Finally, we sought to directly compare emotional recognition through cochlear implant simulations between the domains of music and voice.

To do so, we used validated sound stimuli that allow an assessment of the basic auditory emotions while controlling for confounding factors and allowing the most direct comparison between voice and music: the Montreal Affective Voices (MAV) and the Musical Emotional Bursts (MEB). The MAV (Belin et al., 2008) are non-speech vocalizations (e.g., screams, laughter) depicting basic emotions that are fundamental to human communication (Scherer, 1986). They consist of short vocal interjections on the vowel /a/ expressing basic emotions; they have minimal semantic information and minimal interaction with linguistic processes (Bestelmeyer et al., 2010). The MEB (Paquette et al., 2013) are designed to be the musical counterparts of the MAV and they consist of a few notes expressing basic musical emotion. They are all the more similar to vocal stimuli because they use continuous pitch instruments (clarinet or violin, which have a seamless progression between notes).

Because those stimuli have been previously validated, we can perform a well-controlled acoustical analysis and correlate the acoustical features with participants' emotional recognition. Acoustical cues (both pitch- and non-pitch based) known to affect emotional recognition in music and voice were quantified (computed) for each stimulus (Juslin and Laukka, 2003; Quarto et al., 2014): timbral cues (brightness, energy and roughness; Lartillot et al., 2007), pitch cues (minimum, maximum and average pitch; Boersma, 2002), and temporal cues (rate of Spectral Information Change (SIC), inspired by Rogalsky et al., 2011 ; Sharda et al., 2015: mean peak temporal modulation rate).

We use a validated cochlear implant simulation approach, allowing for testing of natural versus cochlear implant hearing in the same participants. Vocoding algorithms have been successfully employed to simulate the auditory signal degradation through a CI in normal hearing individuals; their use for perception research has been validated for both speech (Fu and Nogaki, 2005; Nogaki et al., 2007; Poissant et al., 2006; Qin and Oxenham, 2003) and music (Cousineau et al., 2010; Limb, 2006; Moore and Tan, 2003). Furthermore, this approach provides easier access to a larger number of homogenous participants (Driscoll et al., 2009; Poissant et al., 2006). Because participants hear through cochlear implants for the first time, this has the added benefit of limiting the variability commonly observed in CI users, which is partly related to training outcomes.

Based on previous studies, we expect reduced emotion recognition accuracy and impaired arousal judgments in the cochlear implant simulation condition, when compared to the normal hearing condition. In terms of relevant acoustical features, we expect that the quantified value of certain acoustical cues that do not entirely rely on pitch (temporal cues and timbral cues) will correlate with participant judgments values.

Assessing the extent of the deficit using basic emotions and identifying available acoustical cues enabling CI users to optimally identify emotions can inform the optimization of implant processors and rehabilitation strategies to facilitate the development of emotion regulation strategies and social competence in this population. If successful, the next logical step shall be to apply this approach to a CI user population.

2.2 Methods

2.2.1 Participants

Sixteen French and or English-speaking participants (10 females) between the ages of 23 and 52 (M=29) with no self-reported hearing problems were recruited through our laboratory participant mailing list and gave written informed consent. They had on average 18.5 years of education (SD 3.4) and 5.1 years of musical training (SD: 5.4). The ethics committee of the faculty of Arts and Sciences of the University of Montréal approved the following research protocol.

2.2.2 Stimuli

120 stimuli were used: 40 short (M=1.3 seconds) vocal emotional interjections from the MAV (Belin et al., 2008) and 80 (40 clarinet, 40 violin) short musical bursts (a few notes, M=1.6s) from the MEB (Paquette et al., 2013). Each category (voice, clarinet, and violin) contained ten stimuli per emotion (fear, sadness, happiness, neutrality). These 120 stimuli (original condition) were also presented in a CI-simulated version (CI-sim condition), processed through an 8-band noise vocoder developed to simulate CI-hearing in normal-hearing individuals. We used the cochlear implant simulation algorithm developed by Cousineau and collaborators (2010). This specific algorithm was selected because it equates the pitch discrimination capabilities of normal hearing individuals with those of CI users.

2.2.3 Procedure

Participants sat in a soundproof booth for a familiarization phase followed by the experimental protocol.

To familiarize participants with the hearing of CI-sim stimuli, they listened to vocal and music excerpts with and without CI-simulation. Participants were first presented with 11 spoken sentences from the IEEE database (Rothauser et al., 1969) if they spoke English (n=12), and from HINT database (Nilsson et al., 1994) if they spoke French (n=4). After hearing the sentences in their CI-sim version, they were asked to repeat what they had heard and were then presented with the sentence in the original condition. Participants then heard seven CI-sim one-minute-long excerpts of popular music (e.g. Michael Jackson – Thriller, 1982). They were asked how familiar the song was on a scale from "not familiar at all" to "extremely familiar", and then asked if they could give us the title of the song. They then heard the original version.

For the main task, participants first heard a pseudo-random presentation of the 120 stimuli in their CI-sim version and then in their original version, in order to avoid an effect of prior exposition to the original version of the stimuli. They were asked, after hearing each excerpt, to rate them on seven different visual analog scales using a computer mouse. On the first page displaying four scales, they rated from "Absent" to "Present" related to how much a stimulus expressed each emotion (happiness, sadness, fear and/or neutrality). They then rated how confident they were about their emotional ratings from "Not at all confident" to "Extremely confident". On the second page, they rated the stimulus’s emotional valence from "Extremely negative" to "Extremely positive", and its level of arousal from "Not at all arousing" to "Extremely arousing". This procedure has been adapted from Gosselin and colleagues (Gosselin et al., 2007). The experiment was self-paced and the complete session lasted on average 90 minutes. During a practice run just before the main task, participants heard six example stimuli to ensure they understood the different rating dimensions and were familiarized with the rating interface. To minimize fatigue, participants were given short breaks for every stimuli. The main task was programmed using the Psychtoolbox (Kleiner et al., 2007), the sound was delivered through BeyerDynamic DT 990 Pro headphones at 70 dB and rating scales were presented on an LG Flatron screen.

2.2.4 Analysis

Ratings on the four emotions scales (happy, sad, fear, neutral) were converted to accuracy scores, with the highest rating corresponding to the identified emotion (0.8% of all rating were excluded because two emotion were rated equally). To directly quantify the effect of the CI simulation on emotional ratings, difference scores were computed (original minus CI-simulated) for each participant. For accuracy scores, this subtraction was done between the average accuracy for each timbre and emotional categories. For arousal, valence and confidence ratings, this was done for each stimulus. Because the scores obtained for the confidence ratings showed different usage for each participant, these ratings were normalized using a Z-Score computed within participants to allow meaningful comparison. When applicable, a Holm–Bonferroni correction for multiple comparisons was applied.

The 240 stimuli (120 original, 120 CI-sim) were further analyzed and quantified for the presence of various acoustical features known to affect judgments (Juslin and Laukka, 2003; Quarto et al., 2014). Values for timber, defined by brightness (amount of energy above 3000 Hz; Juslin, 2000), energy (root mean square; RMS) and roughness (beating), as well as pitch values (min, mean, max) and rate of temporal spectral information change were extracted to allow us to identify the acoustical cues that were linked to the different emotional ratings (average values are presented in Table 1).

The acoustical analysis was performed on the stimuli using Matlab (Language and Computing, 2004), the MIRtoolbox (Lartillot and Toiviainen, 2007) and Praat (Boersma, 2002) Analyses were performed using SPSS (IBM Corp., Version 23.0. Armonk, NY) with a significance level of $p < .05$.

Table 1. Average value of extracted acoustical features presented for each stimuli type (voice, Music), emotion (happy, sad, fear neutral) and condition (original, CI-sim) standard deviation is presented in parentheses.

Original	Voice				Music			
Acoustical values	Happy	Sad	Fear	Neutral	Happy	Sad	Fear	Neutral
Brightness	0,40 (0,06)	0,37 (0,08)	0,45 (0,19)	0,36 (0,13)	0,50 (0,17)	0,40 (0,24)	0,64 (0,19)	0,44 (0,18)
RMS energy[1]	0,26 (0,07)	0,32 (0,15)	0,76 (0,15)	0,71 (0,16)	0,55 (0,17)	0,63 (0,13)	0,65 (0,27)	0,91 (0,20)
Roughness[n]	0,01 (0,01)	0,01 (0,00)	0,03 (0,04)	0,04 (0,02)	0,00 (0,00)	0,00 (0,00)	0,03 (0,05)	0,00 (0,01)
Min Pitch	150,89 (70,00)	148,46 (60,13)	235,93 (83,92)	153,65 (55,81)	279,34 (109,93)	261,47 (81,81)	281,41 (112,51)	351,25 (104,60)
Max Pitch	455,16 (156,19)	547,92 (78,96)	462,61 (110,46)	278,35 (177,93)	525,44 (133,88)	372,98 (140,77)	495,61 (130,60)	373,71 (115,35)
Mean Pitch	296,86 (97,20)	311,21 (83,15)	363,06 (81,37)	175,21 (70,62)	402,31 (116,12)	322,92 (97,40)	394,42 (105,61)	359,17 (106,54)
Rate of temporal SIC[2]	0,32 (0,07)	0,26 (0,10)	0,22 (0,04)	0,21 (0,03)	0,21 (0,05)	0,20 (0,05)	0,25 (0,11)	0,21 (0,04)
CI-simulated	Voice				Music			
Brightness	0,51 (0,05)	0,46 (0,05)	0,51 (0,19)	0,38 (0,09)	0,56 (0,16)	0,48 (0,21)	0,68 (0,16)	0,48 (0,20)
RMS energy[1]	0,26 (0,07)	0,32 (0,15)	0,76 (0,15)	0,71 (0,16)	0,55 (0,17)	0,63 (0,13)	0,65 (0,27)	0,91 (0,20)
Roughness[n]	0,03 (0,02)	0,06 (0,07)	0,34 (0,13)	0,28 (0,13)	0,18 (0,13)	0,20 (0,06)	0,33 (0,27)	0,48 (0,17)
Min Pitch	84,10 (8,77)	76,42 (6,08)	76,77 (2,92)	82,15 (21,60)	78,22 (6,70)	110,32 (113,67)	94,91 (58,97)	175,25 (159,30)
Max Pitch	550,12 (68,87)	576,93 (27,96)	323,18 (204,32)	325,41 (190,91)	481,08 (159,31)	533,28 (101,89)	463,95 (143,50)	438,74 (173,57)
Mean Pitch	252,48 (65,62)	266,44 (72,88)	164,11 (82,31)	164,84 (107,28)	244,50 (106,55)	259,82 (134,75)	235,14 (126,78)	266,69 (169,54)
Rate of temporal SIC[2]	0,31 (0,02)	0,27 (0,07)	0,30 (0,04)	0,26 (0,03)	0,32 (0,04)	0,30 (0,06)	0,31 (0,05)	0,31 (0,06)

[n] Roughness was normalised with minimum and maximum values of all stimuly.
[#] RMS energy (Xe^{-1}) and Rate of spectral information change (Xe^{-2}) values are presented using scienfic notation

2.3. Results

2.3.1 Accuracy

Stimuli were well recognized in their original condition (Table 2), except for the musical fear stimuli that were less accurately identified, similarly to what was observed in the validation study (Paquette et al., 2013). As expected, CI-sim stimuli were less well recognized than their original version. The CI simulation had a significant effect on all emotion identification; all differences (original minus CI-sim) were significantly ($p < 0.05$) different from 0 (null hypothesis). When presented through CI-simulation, vocal and musical fear stimuli were not recognized above chance (voice: $t(15) = 1.88$, $p > .05$; music: $t(15) = 1.92$, $p > .05$), neither were musical-sad stimuli ($t(15) = -.81$, $p > .05$; see Table 2).

In order to evaluate the effect of CI-simulation on emotion recognition accuracy, an ANOVA with Category (voice, music) and Emotion (happy, sad, fear, neutral) as between subjects factors was computed with the difference in accuracy between the original and CI-sim ratings for each participant. A significant Category x Emotion interaction was observed ($F(3, 45)$ =

13.88, $p < .001$, $\eta^2_{partial} = .48$) as well as a main effect of Emotion ($F(3, 45) = 8.75$, $p < .001$, $\eta^2_{partial} = .37$) but not of Category ($F(1, 15) = 0.98$, $p = .34$, $\eta^2_{partial} = .06$).

Post hoc tests using the Bonferroni correction were used to break down the interaction and revealed which category of stimuli was most affected by the CI-simulation for each emotion.

Musical happy stimuli were more affected by the CI-simulation than their vocal equivalent ($p < .001$) and the opposite was found for the fear stimuli as vocal stimuli were more affected their musical equivalent ($p < .001$; see Figure 1). Sad and neutral stimuli were not differently affected by the category of stimuli (all $p > .05$).

Table 2. Confusion matrix of emotion recognition (accuracy %) for the vocal and musical sounds in both the CI-sim and original conditions. When two emotions were rated equally, identification was rated undefined. Framed values were not significantly different from chance level (25%).

Accuracy: Original	Voice					Music				
Intended emotions	Happy	Sad	Fear	Neutral	*Undefined*	Happy	Sad	Fear	Neutral	*Undefined*
Happy	96.9	1.3	0.6	0.6	0.6	86.6	1.6	2.2	9.1	0.6
Sad	5.6	92.5	1.3	0.6	-	5.9	80.0	2.2	11.3	0.6
Fear	1.3	4.4	90	4.4	-	25.6	8.4	47.8	17.5	0.6
Neutral	3.1	1.9	1.9	93.1	-	6.3	18.4	1.6	73.4	0.3
Accuracy: ci-simulated	Voice					Music				
Happy	81.9	-	6.9	8.1	3.1	51.3	2.2	17.8	27.5	1.3
Sad	32.5	41.3	10.6	15.0	0.6	11.9	20.9	36.6	29.1	1.6
Fear	10.0	7.5	36.3	45.0	1.3	22.2	4.1	32.2	41.6	-
Neutral	6.3	11.3	16.9	63.8	1.9	4.4	13.1	36.6	44.7	1.3

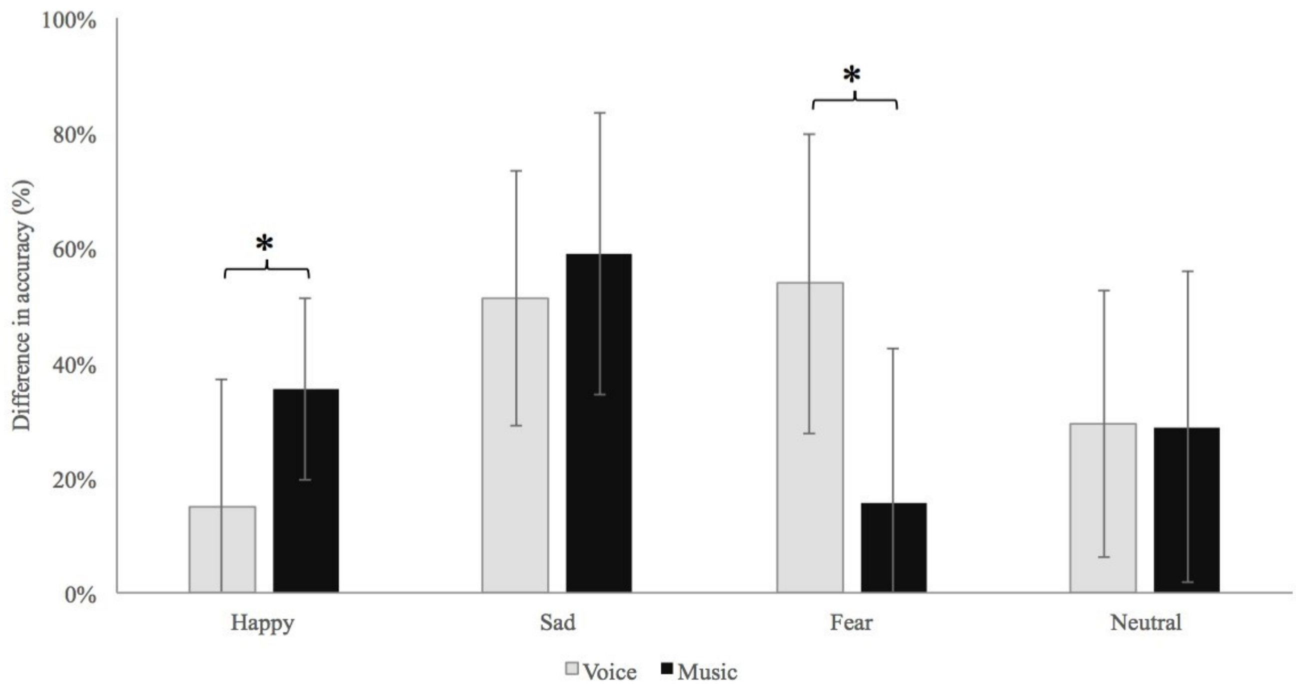

*Figure 1. Average effect (difference) of CI simulation on accuracy scores for each emotion (happy, sad, fear neutral) and category (voice, music). Error bars represent the standard deviation. *= p < .05*

2.3.2 Arousal and Valence Ratings

Valence and arousal ratings are less distinguishable when presented through CI- simulation; happy, sad and fear emotions are rated as more similar to neutral (Figure 2).

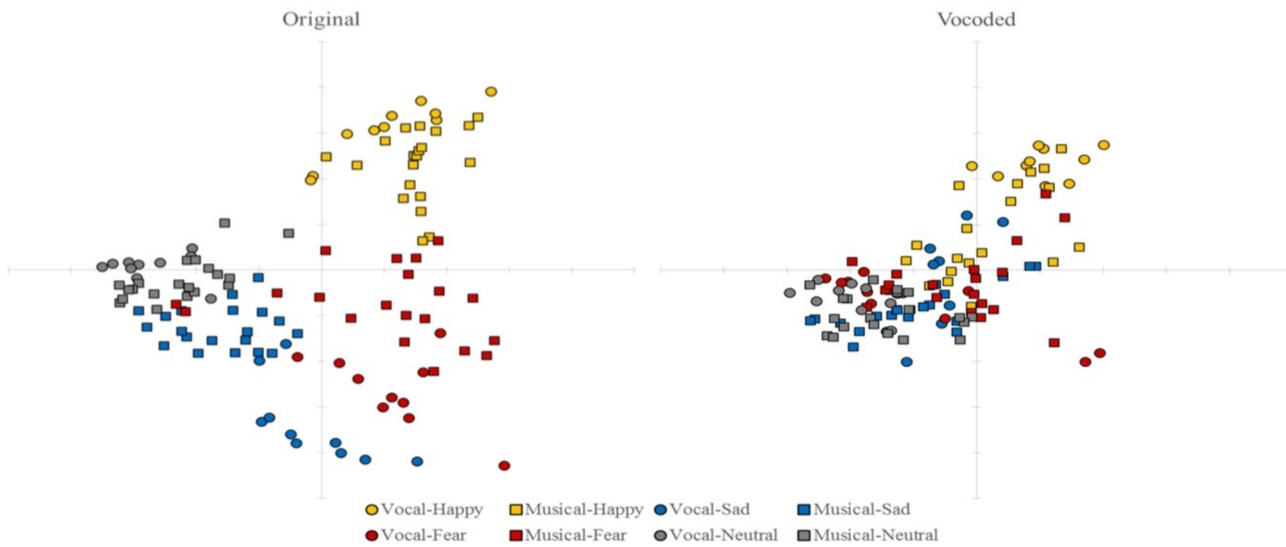

Figure 2. Raw arousal and valence ratings for both the vocal and musical stimuli presented for each emotion (happy, sad, fear neutral) and condition (original, CI-sim).

Separate ANOVAs were computed for the arousal and valence ratings by considering the subtracted ratings (original minus CI-sim). For arousal, the ANOVA with Category (voice, music) and Emotion (happy, sad, fear, neutral) revealed a significant Category x Emotion interaction ($F(3, 45) = 8.82$, $p < .001$, $\eta^2_{partial} = .37$) as well as a main effect of Emotion ($F(3, 45) = 14.43$, $p < .001$, $\eta^2_{partial} = .49$), but not of Category ($F(1, 15) = 0.33$, $p = .57$, $\eta^2_{partial} = .02$). By computing post hoc tests using the Bonferroni correction to compare the differences in arousal, we identified which category was more affected by the CI-simulation for each emotion. As can be seen in Figure 3, the arousal ratings for the music stimuli expressing happiness were significantly more affected by the CI-simulation than their vocal equivalent ($p < .01$). The other vocal and musical emotional stimuli were similarly affected (all p > .05).

The CI-simulation generally had no significant effect on the arousal ratings; only the differences (original minus CI-sim) for the fear stimuli (vocal: $t(15) = 4.76$, $p < .001$; musical 8 $t(15) = 3.47$, $p < .05$) and musical-happy stimuli ($t(15) = 4.76$, $p < .001$) were significantly different from 0 (null hypothesis) whereas all others were not significantly affected (all $p > .5$).

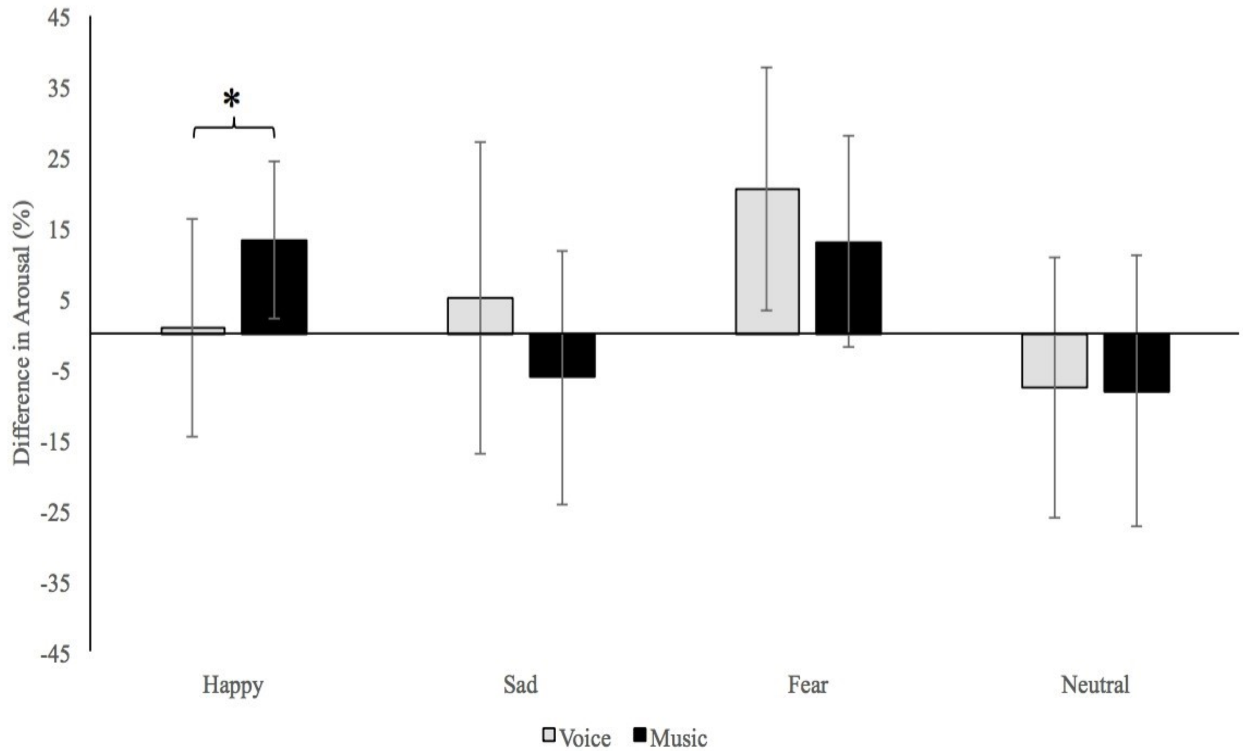

*Figure 3. Average effect of the CI simulation on arousal ratings presented for each emotion (happy, sad, fear neutral) and stimuli type (voice, music). Error bars represent the standard deviation. *= p < .05*

On valence ratings, the ANOVA yielded two main effects: Category: $F(1, 15) = 101.78$, $p < .001$, $\eta^2_{partial} = .87$, and Emotion: $F(3, 45) = 53.63$, $p < .001$, $\eta^2_{partial} = .78$), which were significantly modulated by an interaction ($F(3, 45) = 13.26$, $p < .001$, $\eta^2_{partial} = .47$). Post hoc tests using the Bonferroni correction revealed that ratings of valence for negative vocal stimuli were significantly more affected by the CI-simulation than their musical counterpart (fear: $p < .001$; sad: $p < .001$) and the opposite was found for the happy stimuli as musical stimuli were more affected than their vocal equivalent ($p < .05$; see figure 4). Neutral vocal and musical stimuli were similarly affected (all $p = .75$).

The CI-simulation only significantly affected some of the valence ratings. Only the negative vocal stimuli (sad: $t(15) = -11.04$, $p < .001$; fear: $t(15) = -7.30$, $p < .001$) and the happy ($t(15) =$

7.04, $p < .001$) and neutral ($t(15) = 3.75$, $p < .05$) music stimuli were significantly different from 0 (null hypothesis), whereas all other differences (original minus CI-sim) in valence were not different from 0 (all $p > .07$).

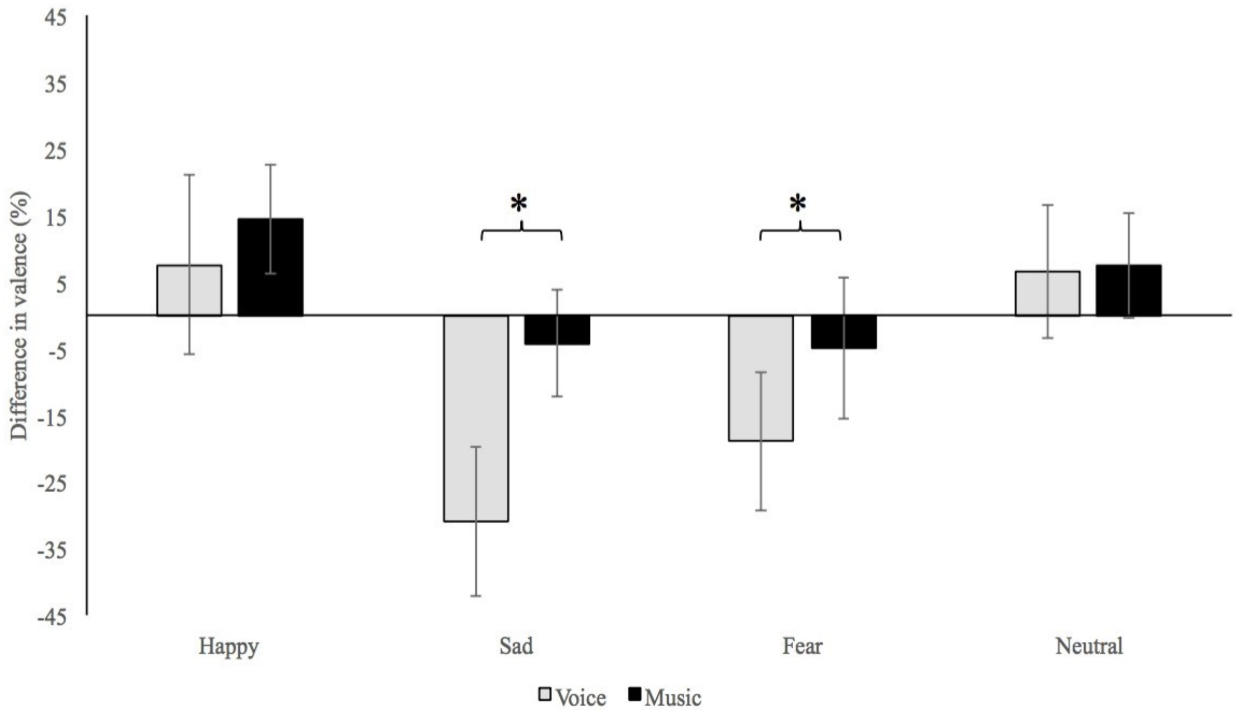

*Figure 4. . Average effect of the CI simulation on valence ratings presented for each emotion (happy, sad, fear neutral) and stimuli type (voice, music). Error bars represents the standard deviation. * = p < .05*

2.3.3 Confidence judgments

When performing a similar analysis to the ones above on participants' confidence judgment of their emotional ratings, only a main effect of Category (voice, music) was observed ($F(1, 15) = 22.00$, $p < .001$, $\eta^2_{partial} = .60$), indicating that participants were generally less confident in their emotional ratings for the musical stimuli. This said, all confidence ratings for every emotion from every stimuli category were significantly affected by the procedure (all $p < 0.005$) and different from 0 (null hypothesis).

2.3.4 Correlation of emotional judgments with acoustical features

In order to determine which acoustical cues participants used to make their emotional judgment during the CI-sim condition, average ratings (accuracy, arousal, and valence) of each stimulus were correlated with each stimulus's previously extracted acoustic features: brightness, energy, roughness, pitch (min, mean, max) and rate of temporal spectral information (average values are presented in Table 1). Correlations by item (stimulus) were done separately for each category (voice, music) and version (original (O), CI-sim (CI)), but not by emotions in order to utilize the variety of valence and arousal values provided by different emotions.

2.3.4.1 Accuracy

Surprisingly, accuracy ratings correlated with no particular acoustical features of the stimuli, perhaps because emotion recognition is dependent on both arousal and valence appraisal; emotion recognition could be a result of a multidimensional analysis including both arousal and valence ratings (Bigand et al., 2005) presented below.

2.3.4.2 Arousal

Brightness positively correlated with arousal ratings: the brighter a stimulus was (amount of energy above 3000 Hz), the more arousing the sound was rated, regardless of category and version (voice-O: $r(40) = .37$, $p < .05$; voice-CI: $r(40) = .34$, $p < .05$; music-O: $r(80) = .53$, $p < .001$; music-CI: $r(80) = .45$, $p < .001$).

Arousal ratings of the original musical stimuli were negatively correlated with their average amount of energy ($r(80) = -.41$, $p < .005$). This can be explained by the fact that the stimuli with the less variation in energy (neutral), had on average the highest energy (see Table 1). Similarly, when sounds (vocal or musical) were presented through the CI-simulation, stimuli energy (voice: $r(40) = -.46$, $p < .005$; music: $r(80) = -.33$, $p < .005$) negatively correlated with participants' arousal ratings.

Pitch cues mostly positively correlated with arousal judgments of the original vocal and musical stimuli, with maximum pitch (voice-O: $r(40) = .48$, $p < .005$; music-O: $r(80) = .62$, $p < .001$) and average pitch (voice-O: $r(40) = .67$, $p < .001$; music-O: $r(80) = .39$, $p < .001$). Interestingly, pitch values were also used to judge the CI-sim version; they positively correlated with the arousal judgements of the vocal (maximum pitch: $r(40) = .40$, $p < .05$; average pitch (40) 6 = .34, $p < .05$;) stimuli and correlated negatively with the musical stimuli (minimum pitch: r(80) 7 = -.28, $p < .05$).

The rate of temporal spectral information change only correlated positively with the arousal ratings of the musical stimuli in their original ($r(80) = .35$, $p < .001$) and CI-sim (r(80) = 10 .27, $p < .05$) version, but surprisingly not the vocal stimuli (original: $r(40) = .27$, $p > .05$; CI-sim: 11 r(40) = .31, $p > .05$).

Roughness (negatively correlated) was used to rate the specific version of each stimulus category. It was used to make arousal judgment of the original musical stimuli ($r(80) = .34$, $p < .005$) and the CI-sim vocal stimuli ($r(40) = -.40$, $p < .05$).

2.3.4.3 Valence

When sounds (vocal or musical) were presented through the CI-simulation, the stimuli's energy (voice: $r(40) = -.56$, $p < .001$; music: $r(80) = -.45$, $p < .001$) and roughness (voice: r(40) = 18 -.51, $p < .005$; music: $r(80) = -.37$, $p < .001$)) negatively correlated with participant valence ratings.

As for the pitch cues, in this case, they were only used to rate the musical stimuli: maximum pitch ($r(80) = .31$, $p < .001$) and average pitch ($r(80) = .27$, $p < .05$) positively correlated with the valence ratings for the original musical stimuli and minimum pitch ($r(80) = -.23$, $p < .05$) for their CI-sim versions.

The rate of temporal spectral information change only positively correlated with valence 25 ratings of the original vocal stimuli ($r(40) = .39$, $p < .05$).

In summary, pitch cues were used to make both arousal and valence judgments, but few were common to both voice and music in their CI-sim version. Furthermore, brightness and energy were related to arousal ratings, and roughness and energy to the valence ratings for musical and vocal stimuli in their CI-sim versions (for a summary of the correlations, see Table A.1 in Appendix A).

2.4. Discussion

We directly compared the perception of vocal and musical emotions through a cochlear implant simulation in healthy subjects using comparable short bursts of emotions. Our goal was to examine how emotion perception (recognition, valence, and arousal) is affected by cochlear implant listening to identify relevant cues accessible to CI users and directly compare how vocal and musical emotions are impacted. This careful examination also allowed us to identify that timbral (brightness, energy, and roughness) acoustical features are available to CI users to perceive both vocal and musical emotions.

All stimuli's emotion recognition accuracy was affected by the CI-simulation. Happy vocal stimuli were the most preserved. Musical fear stimuli were not severely affected by the CI simulation but were generally poorly recognized.

As observed in the CI literature (e.g. Nakata et al., 2012), not all emotions were identified above chance under CI-simulation. In our case, fear and musical-sad CI-sim stimuli were not identified above chance by participants. Sad musical stimuli were often confused with another negative valence stimuli (fear) or perceived as neutral. Fear stimuli were perceived as neutral or confused with happy, as was sometimes the case for the vocal-sad stimuli. The confusion with happy (Volkova et al., 2013), as well as the high recognition accuracy for the happy stimuli (Hopyan et al., 2012; Stabej et al., 2012), is similar to what has been observed in the CI user literature, thus validating our CI-simulation approach.

The stimuli used in our study allowed to measure variable degrees of valence (even if we only had one positive emotion) and arousal. All emotions could be distinguished along these dimensions. As reported by Paquette and collaborators (2013), for these stimuli, the arousal and valence ratings obtained here for the original stimuli fit well with this dimensional representation of emotions, with happy stimuli as conveying positive and arousing emotions, fear stimuli as conveying negative and arousing emotions (for the most part), sad stimuli as conveying moderately arousing and negative emotions, and the neutral stimuli as conveying an emotional valence that is neither positive nor negative with little arousal.

Perceived through CI simulation, all emotions were less differentiable; they tended to cluster more around the neutral position than in their original version. This neutralization effect was more pronounced for the valence ratings of the vocal-sad and music-neutral stimuli but seemed to generally affect the most arousing stimuli: music-happy, vocal fear stimuli (valence and arousal rating) and musical-fear stimuli (arousal ratings). This neutralization effect could explain the results of Ambert-Dahan and collaborators (2015), who reported that CI users had difficulty perceiving arousal in musical excerpts.

Some acoustical properties were used for both music and voice in their original form (brightness, mean-pitch, max-pitch, rate of temporal SIC). Again here the use of normal hearing individuals could account for the fact that participants still use pitch despite its degradation; our results cannot account for effects of deafness or adaptation to the device, such effects could be measured by applying our protocol to a population of CI users. Of interest here is that some acoustical properties of the stimuli were used for both categories (voice, music) in their CI simulated form (timbral/spectro-temporal properties: brightness, energy, roughness) to make arousal and valence judgments. Timbral cues seem to be important when pitch perception is degraded, the same cues are known to be used by amusic individuals, who also experience a congenital pitch perception deficit, albeit in a milder form (e.g. Hyde and Peretz, 2004), to realize emotional judgments (Gosselin et al., 2015). Teaching CI users to focus on these acoustical properties could help emotion discrimination in the CI user population.

It is documented that CI users can to some extent differentiate instrumental timbres on qualitative scales (e.g. dull to brilliant; Gfeller et al., 2002c). Of most relevance here is that CI users who undergo music rehabilitation demonstrate improved timbre identification from baseline with increases in their subjective appraisal of music (Gfeller et al., 2002b). The current study proposes the idea that timbre discrimination (musical) training could also help emotion perception/discrimination in both the musical and vocal domains. Hence, musical therapy focused on timbre could help promote the improvement of optimal tools for emotional regulation and social functioning in the CI-population during development.

The research strategy of using normal hearing individuals allowed us to compare the effect of CI processing within participants. However, it might have created a bias towards neutral. Although researchers have emphasized the benefit of cochlear implant stimulation (see Driscoll et al., 2009; Poissant et al., 2006), CI-simulated sounds can be perceived as quite neutral by normal-hearing individuals in comparison to their everyday perception of auditory emotions. Similarly, the fact that the participants were less confident about their judgment with the CI-sim stimuli could be attributed to the fact that distorted sound (musical and vocal) can be perceived by normal hearing individuals as less natural (Moore and Tan, 2003), such that their novelty could have lessened participant judgments. Perhaps reduced confidence in judgments to musical stimuli can be explained by the fact that we are, first and foremost, voice experts (Latinus and Belin, 2011; in addition to speech perception, we routinely extract from voices a wealth of socially-relevant information), and by our choice of stimuli, which had greater variability of expression/rendition for a specific emotion for music versus the same expression of an emotion for vocal sounds (e.g. laughter for joy).

Further neurobehavioural studies should be done with CI users, using a similar approach to compare musical and vocal emotions, first to assess how the mechanisms (neural markers) involved in vocal and musical emotional processing are affected by the implantation, and secondly, to directly evaluate the effects of timbre discrimination training on auditory emotion perception.

In sum, our results contribute to the characterization of the emotional perception impairment of cochlear implant users and bring forward common timbral acoustical cues (brightness energy and roughness) as instrumental for emotion perception (regardless of stimulus category) when pitch perception is degraded. This suggests that specific attention should be given to these cues in the design of cochlear implant processors and rehabilitation protocols in order to improve emotion perception in this population. For instance, musical therapy focused on timbre could improve emotion perception and regulation, thus bettering social functioning in CI children during development.

2.5. Acknowledgement and funding sources

The project described was supported by a Research Incubator Award from the Centre for Research on Brain, Language and Music to AL and IP and by Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Awards from the Canadian Institute of Health Research to SP. We would like to thank Marion Cousineau, Olivier Dussault and Gabrielle Laufer for their help.

2.6 Connecting text

The next chapter will present an evaluation of the neural mechanisms of auditory emotional processing of natural and CI-simulated sounds in normal-hearing individuals using electroencephalography and ERPs.

CHAPTER 3: Neural Processing of Musical and Vocal Emotions through Cochlear Implants (Manuscript 2)

By

Duha G. Ahmed [a, b, c], Sebastian Paquette [a, d], Anthony Zeitouni [b], Alexandre Lehmann [a, b]

[a] International Laboratory for Brain Music and Sound Research, Center for Research on Brain, Language and Music, Department of Psychology, University of Montreal, Montreal, Québec, Canada H2V 4P3

[b] Department of Otolaryngology Head and Neck Surgery, McGill University, Montreal, Québec, Canada H4A 3J1

[c] Department of Otolaryngology, Head and Neck Surgery, King Abdulaziz University, Rabigh Medical College, Jeddah, Saudi Arabia

[d] Neurology department, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States, 02215

Submitted to: Clinical EEG & Neuroscience (Manuscript ID EEG-17-0024)

Corresponding author:

Duha G. Ahmed

BRAMS-Pavillon 1420 boul. Mont-Royal

University of Montreal

Montréal (QC) Canada H2V 4P3

E: duha.ahmed@mail.mcgill.ca

Abstract

Cochlear implants (CI) partially restore the sense of hearing in the deaf. However, the ability to recognize emotions in speech and music is reduced due to the implant's degraded electrical signal and impaired neural pathway, post sensorineural hearing loss. Electrophysiological correlations of this deficit are not yet well established. We aim to characterize the effect of CIs on auditory emotion processing and, for the first time, directly compare vocal and musical emotion processing through a CI-simulator. We recorded 16 normal hearing participants' electroencephalographic activity while listening to vocal and musical emotional bursts in their original condition and in a degraded (CI-simulated) condition. We found prolonged P50 latency and reduced N100-P200 complex amplitude in the CI-simulated condition. When comparing the processing of vocal and musical bursts, we found a delay in latency with the musical bursts compared to the vocal bursts. Our results show that with CI-simulation there is a delay in processing auditory signals, as much as 50 milliseconds. This points to a defect in encoding sound signals through the implant's processor. Further neurophysiological studies are needed to characterize the deficit in CI-users. A better understanding of the problem may lead to developing rehabilitation programs or device processing strategies that can improve the quality of life of CI-users.

Key words

Cochlear implants, emotional acoustical cues, cross-domain comparison, music, voice, event-related potentials

3.1. Introduction

Perceiving emotions in language and music is an important component of communication, social development, and general quality of life. With cochlear implants (CI), sensorineural hearing loss can be partially treated and speech comprehension can be achieved to some extent. However, CI users' ability to recognize emotions in speech and music is limited. Several acoustical cues (emotional prosody) that are relevant for emotional perception (Juslin and Laukka, 2003) are severely impoverished due to the degraded signal transmitted by the implant and the neural reorganization following deafness (Nakata et al., 2012; Wang et al., 2013; Volkova et al., 2013). A systematic characterization of this impairment in emotional perception is still lacking, and little is known of the brain mechanisms underlying this deficit.
Previous behavioral studies have shown that CI-users cannot perceive emotions in speech and music as well as normal hearing individuals can (Hopyan et al., 2012; Stabej et al., 2012; Volkova et al., 2013; Ambert-Dahan et al., 2015). However, the extent of the deficit is not well characterized. Most studies on emotional perception of CI users utilized complex stimuli (sentences or songs), introducing several confounding variables (semantics, tempo, duration and context) that may have made emotional recognition easier, and therefore may have underestimated the extent of the deficit (Paquette et al., 2013). In addition, these confounding variables make it difficult to compare musical and vocal emotional processing in CI users.

A systemic description of brain mechanisms underlying emotional processing in CI users is needed. Using electroencephalography (EEG) to measure event-related potentials (ERPs) is a validated method to examine the neurophysiological pathways of emotional processing. ERPs are an important objective measurement of auditory emotional prosody differentiation and recognition (Pinheiro et al., 2011). Only one study examined emotional prosody perception of CI users by utilizing ERPs of spoken sentences (Agrawal et al., 2013). Thus, this study falls into the same confounding flaw that makes it difficult to define the deficit and also underestimate it.

Previous literature comparing the neural processing of voice and music is scarce. Some studies found no differences in ERPs for vocal and musical sounds (Levy et al., 2001, 2003; Kaganovich et al., 2013), and others found vocal (speech and vocalizations) processing took precedence over music (Meyer et al., 2007; Rigoulot et al., 2015). This preferential precedence is also seen with the emotional processing of vocalizations over speech (Liu et al., 2012; Pell et al., 2015). Some stimuli may be preferentially processed; priority in emotional processing may be given first to the primitive and basic vocalizations, then to speech, and then given to more complex stimuli, like music. However, Studies controlling for confounding factors and directly comparing voice and music are still lacking. Such studies are crucial to characterize the deficit in emotional processing.

We recently designed a novel approach to overcome this issue and better estimate the emotional deficit in CI users. We used Montreal Affective Voices (MAV) (Belin et al., 2008) and Musical Emotional Bursts (MEB) (Paquette et al., 2013), which are validated affective bursts of basic auditory emotions that control for confounding factors and allow the most direct comparison between voice and music. We presented the MEB and MAV through CI-simulation and in their original condition to normal hearing individuals and found emotion recognition accuracy to be affected by CI-simulation (Paquette et al., under revision). This behavioral experiment characterizes the deficit. However, little is known about the brain rewiring post-implantation and its effect on the neural mechanisms of emotional processing. Therefore, to more precisely describe the defect in emotional processing with CIs, the neural markers must be assessed.

Here, we aimed to assess how cochlear implants impact brain mechanisms of emotional processing and to directly compare the processing of voice and music. We looked at the brain activity while implementing the same paradigm used in our previous behavioural experiment (Paquette et al., under revision); presenting the MAV and MEB in their original and in a CI-simulated condition while recording EEG activity. We measured the ERPs' components P50, N100, and P200, associated with stages of emotional processing (Schirmer & Kotz, 2006; Liu et al., 2012; Iredale et al., 2013).

We used CI-simulation, a validated method to test cochlear implant hearing in normal hearing individuals (Shannon et al, 1995; Lakshminarayanan et al, 2003; Moore and Tan, 2003; Qin and Oxenham, 2003; Fu and Nogaki, 2004; Burkholder et al, 2005; Fu et al, 2005; Poissant et al, 2006). CI-simulation gave us the advantage of comparing how the same brain processes the original and CI-simulated sounds. Furthermore, it limited the variability commonly seen with CI users (due to training outcomes), because our participants heard through CIs for the first time.

Based on previous literature, the use of CI-simulation and well-controlled vocal, and musical stimuli, we hypothesized ERPs to be affected at earlier stages of emotional processing (P50, N100) due to the brief, basic emotional nature of our stimuli. Regarding the comparison between voice and music, we expected a bigger difference with music due to the complexity of its auditory cues.

3.2. Methods

3.2.1. Participants

Sixteen participants completed the study (5 males, 11 females). They were between 20 and 38 years (mean age = 24.75). Informed written consent was obtained before initiating the study. Individuals with previously diagnosed hearing problems or psychiatric disorders were excluded from the study. Participants received compensation for their time. Participants were recruited through our laboratory participant mailing list. The study was approved by the Ethics Committee of Research in Arts and Sciences (CERAS) at University of Montreal.

3.2.2. Stimuli

Forty short (mean duration= 1.3 seconds) vocal emotional interjections on the vowel /a/ were selected from the Montreal Effective Voices (MAV; Belin et al., 2008) as well as 40 short (mean duration=1.6 seconds) Musical Emotional Bursts (MEB; 18 clarinet and 22 violin; Paquette et al., 2013). In our previous study, these 40 musical excerpts were the best-recognized musical stimuli with CI-simulation for each emotion (Paquette et al. under review). There were 80 stimuli in total: 40 stimuli per timbre (voice, music) and 10 stimuli from each timbre expressing one of four emotions (sadness, happiness, fear and neutrality). There was no significant difference in the mean duration of the vocal and musical sounds ($t(78) = 1.76$, $p = 0.082$). An 8 band vocoder (Cousineau et al., 2010) was used to simulate CI; this specific algorithm was selected because it equates the pitch discrimination capabilities of normal hearing individuals with those of CI users. Square waves (click sound) were inserted 200 milliseconds into randomly chosen stimuli using MATLAB, to create targets for the participants.

3.2.3. Task and procedure

The experiment was carried out in a soundproof room where participants were seated in a comfortable armchair that was 100 cm from a screen with a fixation cross. Stimuli were presented using MATLAB 2007 at 75 dB through insert-style earphones (Etymotic ER-2). The 80 stimuli were repeated 11 times (880 trials) and presented in a pseudo-random order with an Inter-stimulus interval of 2000 milliseconds. Targets were presented every 7-12 stimuli and were 7% of the trials. The experiment consisted of 8 blocks (6 minutes each), to allow the participants to move and relax in-between. Participants were instructed to listen attentively to the stimuli and press the spacebar on the keyboard when they heard a sound with a click (target). Participants started with a training block to familiarize them with the clicks and to check their recognition of target trials. The entire experiment, including breaks and EEG preparation, lasted around 3 hours.

3.2.4. Neurophysiological recording

Continuous EEG was recorded using 64 active sintered Ag-AgCl electrodes placed on the scalp, according to the International 10/10 system (ActiveTwo, BioSemi, The Netherlands). Active electrodes contain the first amplifier stage within the electrode cover and provide impedance transformation on the electrode. This prevents interference currents from generating significant impedance-dependent nuisance voltages. We, therefore, did not control electrode impedances, but rather kept direct-current offset close to zero during electrode placement. Vertical and horizontal eye movements were monitored using three additional electrodes placed on the outer canthus of each eye and on the inferior area of the left orbit. Reference-free electrode signals were amplified, sampled at 1024 Hz (ActiveTwo amplifier, BioSemi, The Netherlands), and stored for offline analysis using BioSemi ActiView software.

3.2.5. Processing and artifact rejection

The recorded brain activity was analyzed offline using EEGLAB version 12.0.2.6b (Delorme and Makeig, 2004) and its plugin ERPLAB version 5.0 (Lopez-Calderon and Luck, 2014), running under the MATLAB environment. The data was re-referenced to the average mastoids, band-pass filtered from 0.1 to 30 Hz (Butterworth), resampled at 256 Hz, and epoched from -200 milliseconds to 2300 milliseconds relative to stimulus onset. Epochs containing artifacts were rejected by EEGLAB`s automatic iterative rejection procedure with an initial threshold of three standard deviations (Delorme and Makeig 2004). Independent component analysis (ICA) was performed and ocular artifacts (blinks and horizontal eye movement) were rejected (Jung et al., 2000). Epochs were then averaged for each condition. There were no significant differences between the number of trials per emotion in both conditions (NH, CI) (Happy voice ($t(15) = -0.07$, $p = .95$), sad voice ($t(15) = -0.79$, $p = .44$), fear voice ($t(15) = -0.33$, $p = .75$), neutral voice ($t(15) = -0.32$, $p = .76$), happy music ($t(15) = -0.45$, $p = .66$), sad music ($t(15) = 0.89$, $p = .39$), scary music ($t(15) = 0.63$, $p = .54$), and neutral music ($t(15) = -0.77$, $p = .45$)). ERPs waveforms and topographical maps for each emotion were computed and compared for latency (P50, N100, and P200) and amplitude of peak voltage activity (P50, N100-P200 complex) at the onset of the stimulus. Visual inspection showed that distribution of ERP effects was predominantly frontocentral. Therefore, peak amplitude and latency analyses were conducted at the Cz electrode (Figure 5).

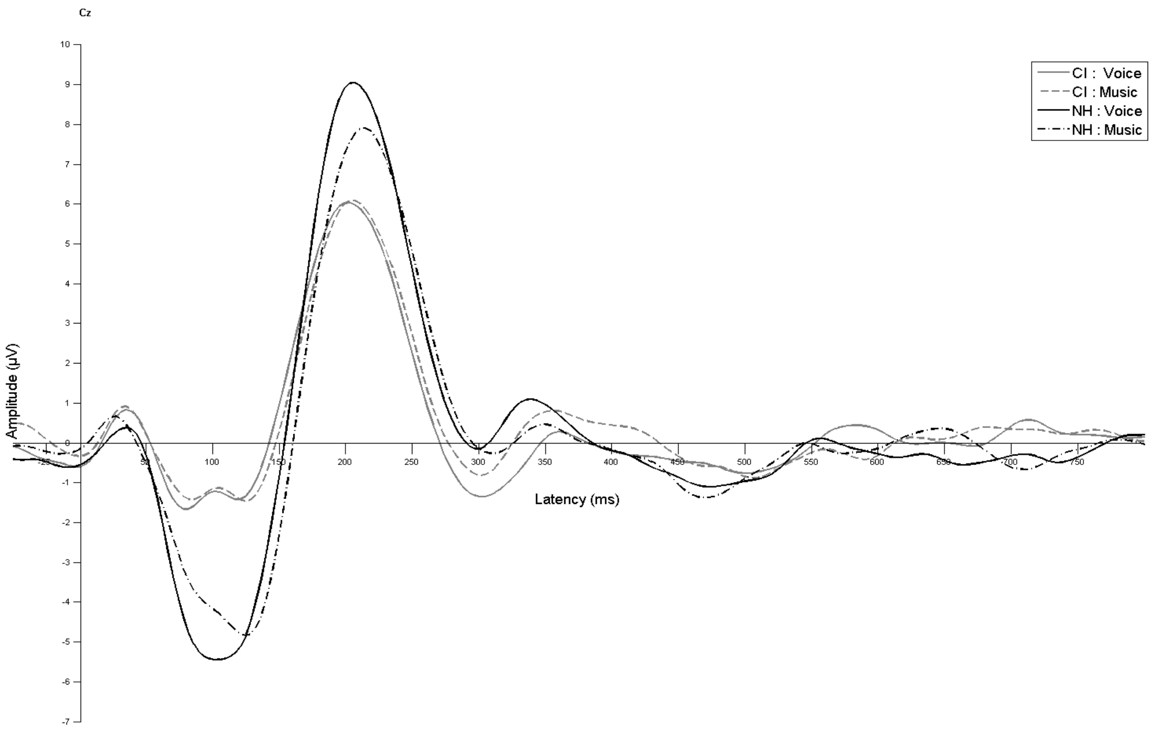

Figure 5. Average ERP waveforms recorded at Cz electrode in original (NH) and CI-simulated (CI) conditions for the vocal and musical stimuli from 50 ms before onset to 800 ms after onset of the stimuli (X-axis: latency in milliseconds, Y-axis: amplitude in µV).

3.2.6. Statistical Analysis

Statistical analysis was performed using SPSS 22.0 (IBM, Armonk, NY, USA). For the behavioral data (target hit rate and reaction time), paired t-tests comparing the NH and CI conditions were performed. For the peaks' amplitude analysis, the difference between the original and CI-simulated conditions was calculated and compared to zero (paired t-test) to measure the effect of CI-simulation on the P50 peak and the N100-P200 complex. The P50 amplitude value used was; P50 peak amplitude minus the peak at zero after baseline correction. Analysis of peak latencies and amplitudes was performed using repeated measure ANOVAs [Type 2 (NH, CI-simulation) X Timbre 2 (voice, music) X Emotion 4 (happy, sad, fear and neutral)]. When a triple interaction was found (type by timbre by emotion), we performed repeated measure ANOVAs for each type (NH and CI) [Timbre 2 (voice, music) X Emotion 4 (happy, sad, fear and neutral)]. If a timber by emotion interaction was found, we collapsed the type (separate types in the case of a triple interaction) and performed posthoc tests; paired t-test comparing the vocal and musical emotions and repeated measure ANOVAs [Timbre 1 (voice or music) X Emotion 4 (happy, sad, fear and neutral)], followed by t-tests to explore the effect of emotions on each timber. The significance level was set at $p < 0.05$. When relevant, the Greenhouse-Geisser or the Huynh-Feldt corrections were used to correct for violations of sphericity in the performed ANOVAs. A Bonferroni-Holm correction was performed on the p-values when a paired t-test was performed, to account for multiple comparisons.

3.3. Results

3.3.1. Behavioral results

Task difficulty was matched across the original and CI-simulated conditions. Both target detection accuracy (Mean accuracy; NH 99.29%, CI 98.79%) and reaction times did not differ between conditions (Accuracy $t(30) = 0.79$, $p= 0.43$; RT: $t(30) = 0.23$, $p= 0.82$).

3.3.2. ERP results

3.3.2.1. P50 latency

The P50 latency ANOVA with condition (NH, CI-simulation), timbre (voice, music) and emotion (happy, sad, fear and neutral) revealed the main effects of condition (NH, CI) ($F(1, 15) = 12.53$, $P= .003$, $^{\eta 2}$partial = .46), emotion ($F(2.7, 40.5) =3.16$, $P=.039$, $^{\eta 2}$partial =.17), and a significant interaction between timber and emotion ($F(2.36, 35.45) = 3.89$, $p= .02$, $^{\eta 2}$partial = .21).

The P50 latency is overall longer in the CI-simulation (Figure 6). We performed planned posthoc t-tests using the Bonferroni correction to explore the timber by emotional interaction regardless of stimuli type (NH, CI); there was no difference in the P50 latency between all the vocal and musical stimuli ($p > .05$). In addition, for each timbre, we performed an ANOVA with timbre (voice or music) and emotion (happy, sad, fear, and neutral). There was no effect of emotion on the vocal stimuli ($F(2.59, 80.33) = 2.690$, $p=.06$, $^{\eta 2}$partial =.08). But we found a main effect of emotion on the musical stimuli. The post hoc t-test showed P50 latency was longer with the happy musical stimuli than the fearful ones ($t(31) = 3.43$, $p < .05$) (Figure 7).

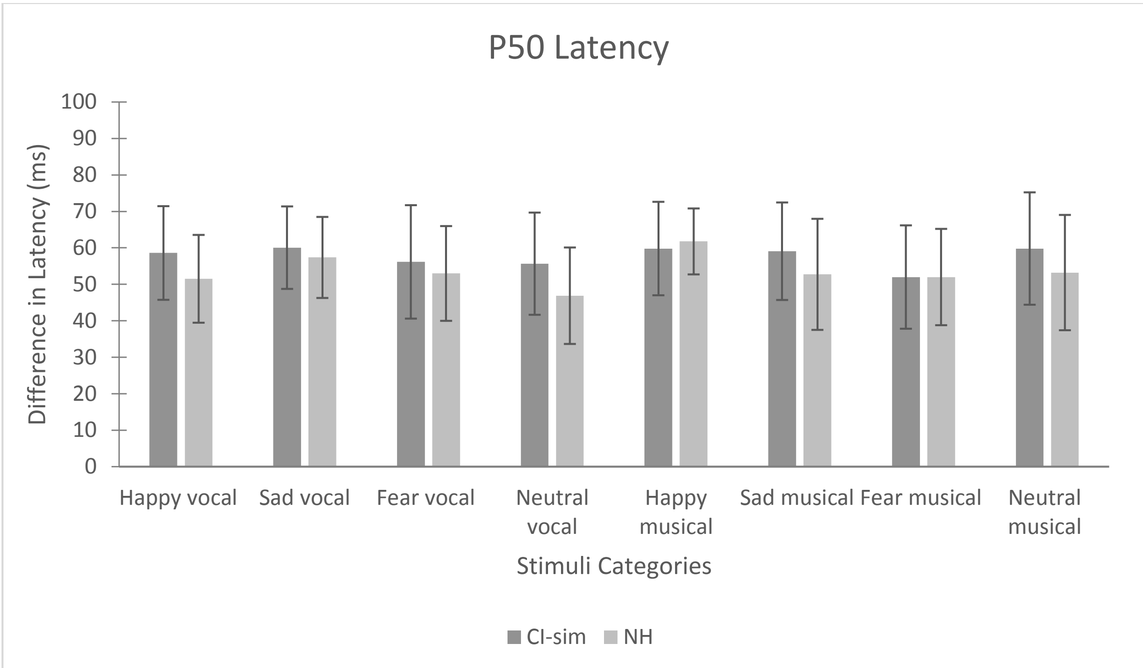

Figure 6. P50 latencies in the original (NH) and CI-simulated (CI-sim) conditions (error bars represent standard deviation).

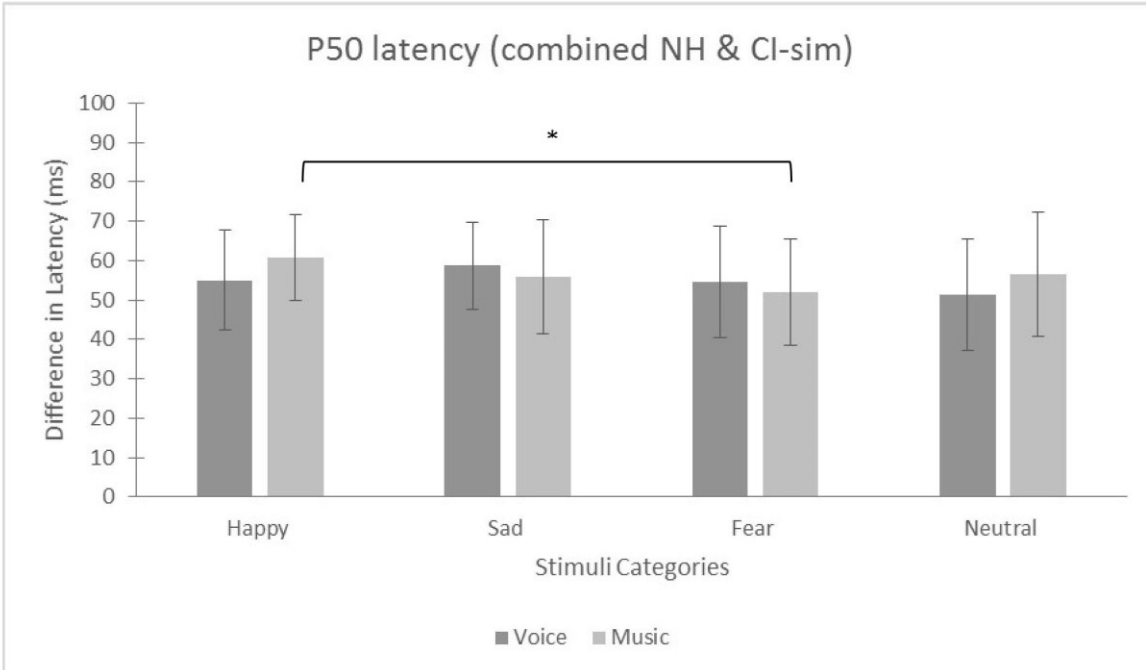

*Figure 7. A comparison between the P50 latency for all the vocal and musical stimuli (NH & CI-sim) (error bars represent standard deviation). *Happy musical stimuli have longer latencies than fearful stimuli (p < .05).*

3.3.2.2. N100 latency

The N100 latency ANOVA with condition (NH, CI-simulation), timbre (voice, music) and emotion (happy, sad, fear and neutral) revealed the main effects of timbre ($F(1,15)=21.72$, $P=.00$, η^2partial =.59), emotion ($F(2.481, 37.21)=19.80$, $P=.00$, η^2partial =.57) and a significant interaction between timber and emotion ($F(3 , 45)= 7.98$, $p= .000$, η^2partial = .35). We then performed planed posthoc t-tests to explore the interaction regardless of stimuli type (NH, CI). We found an earlier N100 latency with happy and neutral vocal stimuli compared to their musical counterparts (Happy $t(31) = -4.00$, $p < .05$, Neutral $t(31) = -4.35$, $p < .05$). In addition, for each timbre, we performed an ANOVA with timbre (voice or music) and emotion (happy, sad, fear and neutral). With the vocal stimuli, we found a main effect of emotion ($F(2.63, 80.59) = 8.26$, $p=.007$, η^2partial =.21). The posthoc t-test showed N100 for sad vocal stimuli were longer than the happy ($t(31) = -3.94$, $p < .05$), fearful ($t(31) = 3.8$, $p < .05$) and the neutral stimuli ($t(31) = 5.67$, $p < .05$). Similarly, we also found a main effect of emotion on the musical stimuli ($F(2.72, 84.29) = 14.15$, p=.000, η^2partial =.31). The posthoc t-test showed N100 for fearful musical stimuli were earlier than the happy ($t(31) = 4.98$, $p < .05$), sad ($t(31) = 4.76$, $p < .05$) and neutral ($t(31) = -5.48$, $p < .05$) musical stimuli (Figure 8).

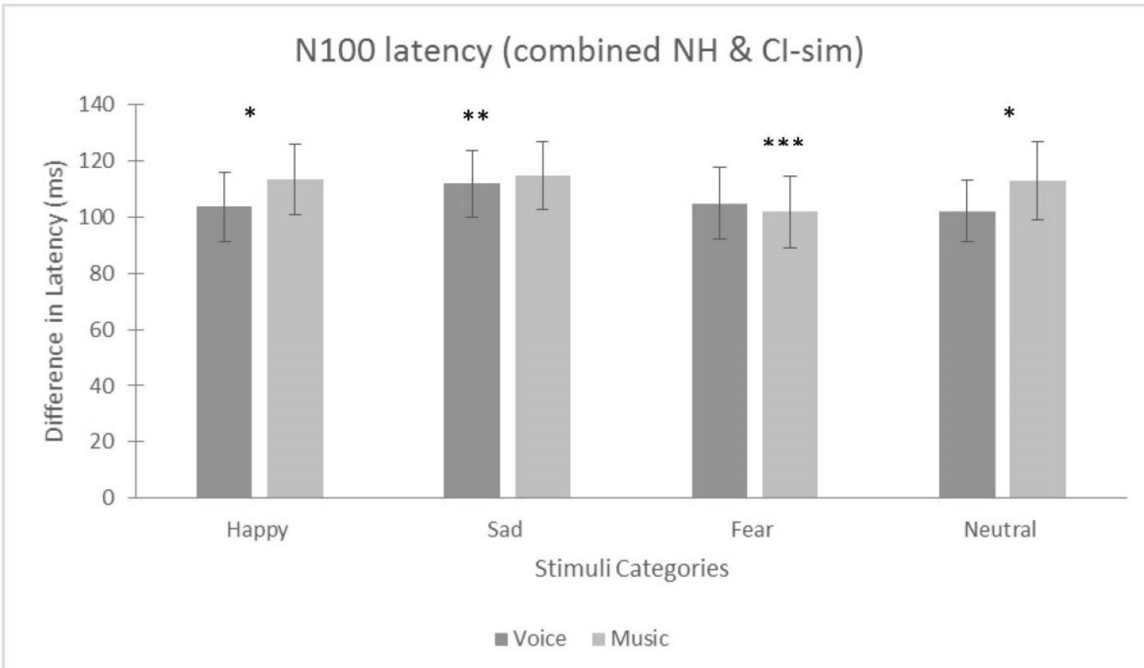

*Figure 8. A comparison between the N100 latency for all the vocal and musical stimuli (NH & CI-sim) (error bars represent standard deviation). * Vocal stimuli (happy and neutral) have significantly shorter latencies than their musical counterparts (p < .05). ** Sad vocal stimuli have significantly longer latencies than the other vocal emotions (p < .05). *** Fear musical stimuli have shorter latencies than the other musical emotions (p < .05).*

3.3.2.3. P200 latency

The P200 latency ANOVA condition (NH, CI-simulation), timbre (voice, music) and emotion (happy, sad, fear and neutral) revealed main effects of timbre ($F(1, 15) = 9.53$, P=.008, η^2partial

=.39), emotion (F(2.36, 35.44) = 6.5, P =.003, η^2partial =.3), and significant interactions; timber by emotion (F(2.14, 32.13) = 7.98, p= .02, η^2partial = .22) and condition by timbre by emotion (F(2.37, 35.49) = 4.011, p= .02, η^2partial = .21). We explored the triple interaction by performing another ANOVA for each type (NH and CI) [timbre (voice, music) X emotion (happy, sad, fear and neutral)]. In the original condition, we found main effects of timbre (F(1, 15) =9.71, P=.007, η^2partial =.39), emotion (F(15, 33.25) = 4.28, P=.02, η^2partial =.22) and a significant timbre by emotion interaction (F(2.296, 34.44) = 8.27, P=.001, η^2partial =.36). When we explored the interaction by performing posthoc t-tests, we found the neutral musical stimuli to have a longer latency than their vocal counterparts (t(15) = -4.35, p < .05). In addition, we performed a repeated measure ANOVA for each timbre with Emotion (happy, sad, fear and neutral). We found no effect of emotion on the vocal original stimuli. However, there was an effect on the musical stimuli F(3, 45) =12.27, P=.000, η^2partial =.45). The posthoc t-test showed musical fear stimuli had an earlier P200 latency than the sad and neutral ones (Sad and fear: t(15) = 5.4, p < .05; Fear and neutral: t(15) = -5.68, p < .05) (Figure 9).

When we explored the interaction in the CI-simulated condition, the ANOVA revealed a main effect of emotion (F(2.41, 36.18) = 3.21, P=.04, η^2partial =.18). No other significant differences were observed.

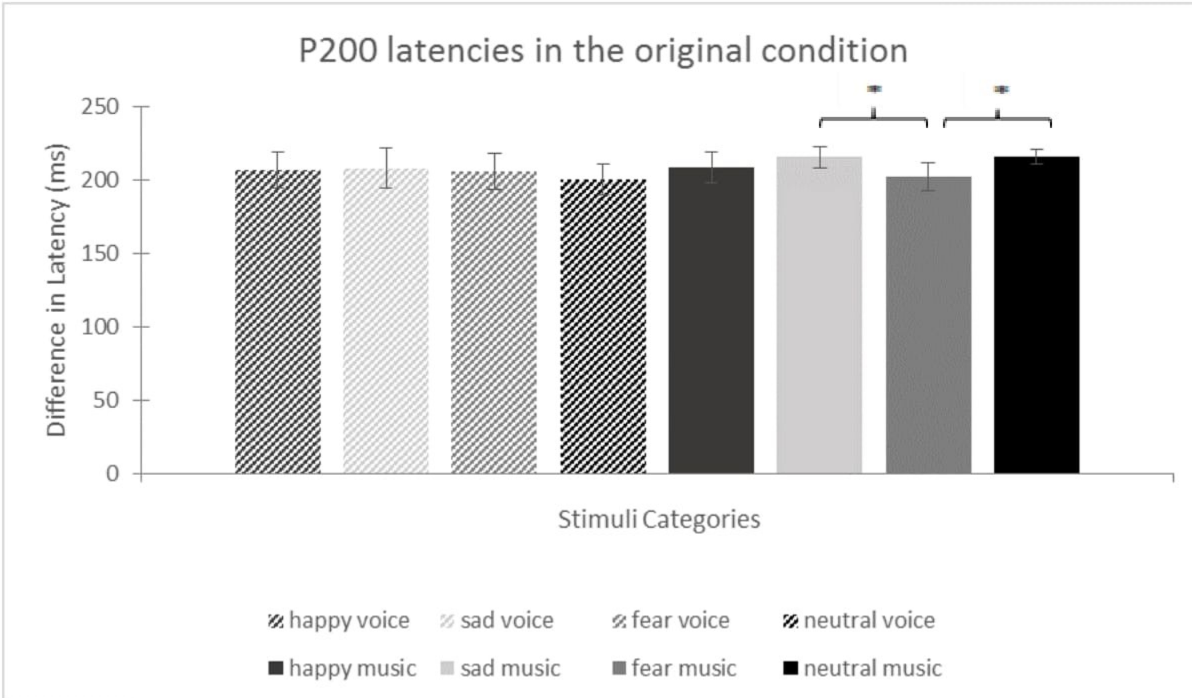

*Figure 9. P200 latency in the original condition (error bars represent standard deviation). * Musical fear have an earlier latency than sad and neutral (p < .05).*

3.3.2.4. P50 amplitude

The CI simulation had no effect on all the stimuli. All differences (original – CI-simulation) were not different from zero ($p > 0.05$).

3.3.2.5. N100 – P200 Peak to Peak Amplitude

The CI simulation had a significant effect on all stimuli, all differences (original – CI-simulation) were significantly ($p < 0.05$) different from zero.

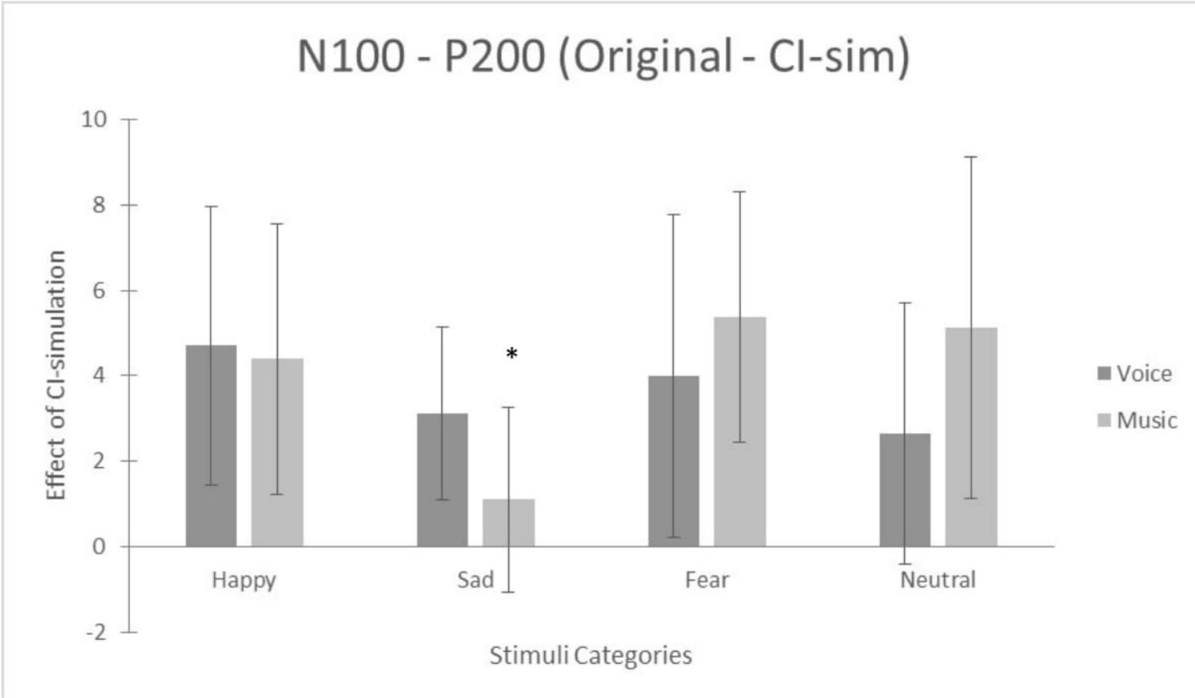

*Figure 10. Effect of CI-simulation on the N100-P200 complex (error bars represent standard deviation). All the stimuli were affected by the CI-simulation (p < 0.05). * Sad musical stimuli were the least affected by the CI-simulation than the other emotions (p < 0.05).*

To compare how much the N100 –P200 complex was affected by the CI simulation for the different timbres and emotions, we performed an ANOVA with timbre (voice, music) and emotions (happy, sad, fear, neutral). It revealed a main effect of emotion ($F(3, 45) = 7.14$, $P = .001$, η^2partial =.32) and a significant interaction of timbre by emotion ($F(3, 45) = 6.01$, $p = .005$, η^2partial = .29). When we explored the interaction using planned posthoc t-tests, we found no differences in the amplitude between the vocal and musical emotions ($p > 0.05$). In addition, we performed a repeated measure ANOVA for each timbre (voice or music) with emotion (happy, sad, fear and neutral). We found no effect of emotions on the vocal original stimuli. However, there was an effect on the musical stimuli ($F(1.93, 28.91) = 8.87$, $p = .001$, η^2partial =.37). The posthoc t-test showed the sad musical stimuli were significantly less affected by CI-simulation than the other musical stimuli (Happy and sad $t(15) = 3.6$, $p < 0.05$; Sad and fear $t(15) = -4.85$, $p < 0.05$; Sad and neutral: $t(15) = -6.05$, $p < 0.05$) (Figure 10).

3.4. Discussion

We investigated the effect of cochlear implants on neural emotional processing. For the first time, we directly compared the processing of vocal and musical emotions by measuring event-related potentials of comparable short bursts of emotions through CI-simulation in normal hearing individuals. We found prolonged P50 latency and reduced N100-P200 complex amplitude in the CI-simulated condition. Our results demonstrate that the auditory signals are affected by cochlear implant signal, as there is a delay in the early stages of emotional processing. There was no difference in amplitude with CI-simulation when comparing vocal versus musical bursts. However, in accordance with our hypothesis, we found a delay in latency with the musical bursts compared to vocal bursts.

3.4.1. The effect of cochlear implant simulation

Our ERP data showed an effect of CI-simulation at P50 latency and N100-P200 complex amplitude. We found an early effect at 50ms, where the simulated bursts (both vocal and musical) had longer P50 latency than the original ones. The auditory P50 is a neurophysiological indicator of preferential attention to sensory inputs and is thought to reflect the general level of arousal (Key et al., 2005). In addition, it is used as a biomarker of auditory cortical development in child CI-users (P50 decreases in latency with age) (Eggermont et al., 1997; Eggermont & Ponton, 2003). The P50 latency delay found here could be related to the longer time required for sensory processing when the signal is difficult to discriminate (Beynon et al., 2005; Agrawal et al., 2012). The P50 was also found to be modulated by nonverbal emotional vocalizations (Liu et al., 2012). Thus, the delay here can be due to the neutralizing effect of CI-simulation on the arousal levels of the stimuli in normal hearing individuals (Paquette et al, under review). In addition, our participants were not familiar with CI-listening, and hence, CI-simulated sounds might not be preferentially processed. This early effect on the P50 latency is evidence that the input of cochlear implants is defective.

We also found the CI-simulated bursts to have reduced N100-P200 peak to peak amplitude. The sad bursts were the least affected by CI-simulation. The reduction in the N100 and P200 amplitudes is constantly reported in the literature with CI-simulation and actual CI-users (Groenen et al., 2001; Sandmann et al., 2009; Agrawal et al., 2012). This was explained by the decreased amount of sensory information when sound is processed through a CI, resulting in reduced synchronization of neuronal activity; this synchronization is needed for generating ERPs (Groenen et al., 2001). A reduction in ERP amplitudes is a consistent effect observed in EEG studies with CIs. Because implant processors amplify sounds differently than the normal ear and often have automatic gain control and compression strategies, it is difficult to know whether the reduction in ERP amplitude is related to a difference in perceived loudness. Our approach allowed us to disentangle this effect for the first time, by presenting the sounds with the exact same loudness to the same participants under two conditions. Under those circumstances, we also observed a reduction in ERP amplitude in the CI-simulated condition, therefore ruling out a possible effect of loudness. It is known that the P200 and N100 amplitudes are sensitive to physical parameters of the stimuli, such as pitch (Novak et al., 1992) and loudness (Hegerl & Juckel, 1993; Hillyard & Picton, 1987). Some alternative explanations may be either peripheral (i.e. a reduction in frequency content by the vocoder, thereby lowering the number of responding auditory neurons) and/or central (i.e. a decrease in saliency for vocoded sounds).

The main advantage of our design was to compare the electrophysiological data of the same brain in the original and the CI-simulated conditions. However, because our participants had normal hearing and did not suffer the effect of hearing deprivation, we could only assess the effect of cochlear implants in terms of sound input degradation, and we could not evaluate the impact of plasticity and brain rewiring, post-implantation. To further assess the impact of cochlear implantation on brain emotional processing, the next step is to use this well-controlled paradigm to compare CI-users and normal hearing controls.

3.4.2. Comparing vocal and musical emotional processing

We found a difference in the neural processing of vocal and musical emotions at N100 and P200 latencies.

The P200 latency in the original condition was longer for the neutral musical stimuli than their vocal counterparts, which is similar to what Rigoulot et al. (2015) found: later P200 latency for the musical than the vocal stimuli. The P200 is the stage of processing where the integration of emotionally significant acoustical cues and the extraction of emotional significance happen. This delay in the musical latency can be due to the more complex musical spectral profile (Meyer et al. 2007).

We also found an earlier N100 latency with happy and neutral vocal stimuli than their musical counterparts, which is in contrast to what Rigoulot et al. (2015) found, where the N100 latency did not differ between musical and vocal stimuli. The N100 is associated with sensory processing (from the ear to the auditory cortex). It is modulated by attention (controlled in our study) (Hillyard et al., 1973; Knight et al., 1981; Mangun, 1995; Ritter et al., 1988). It is also found to be modulated by emotional prosody (Iredale et al., 2013), which might explain why we found an effect on the N100 latency with the emotional sad vocal stimuli (longer latencies than the other vocal stimuli). Similarly, Pell et al. (2015) found sad vocalizations to have longer N100 and P200 latencies than happy vocalizations and explained this by the higher arousal level of happy emotions. This is also the case with our stimuli (sad, moderate arousal level; happy, higher arousal level) (Belin et al., 2008, Paquette et al., under review).

With the ERPs amplitudes (P50 and N100-P200 complex), we did not find a difference between the vocal and musical emotional bursts. This is in contrast to Rigoulot et al. (2015), as they found larger amplitudes in response to musical than to vocal sounds. In agreement with what we found, Levy et al. (2001, 2003) and Kaganovich et al. (2013) found no differences between the musical and vocal sound amplitudes (and also latencies).

3.5. Conclusion

This study shows that with CI-simulation, there is a deficit in processing the auditory signal at earlier stages of emotional processing (P50). In addition, the results support our hypothesis that emotional processing starts at earlier stages (P50 and N100), especially for vocal and musical bursts compared to more complicated speech and music. This study adds to the scarce literature differentiating between vocal and musical processing, where music has prolonged latency (N100 & P200). Further neuro-behavioral studies should be done with actual CI-users to assess how the mechanisms involved in vocal and musical emotional processing are affected by the implantation.

Acknowledgements

We would like to thank Mihaela Felezeu, Bernard Bouchard, Olivier Dussault and Gabrielle Laufer for their help.

Funding

The project described was supported by a Research Incubator Award from the Centre for Research on Brain, Language and Music to Alexandre Lehmann and Isabelle Peretz.

CHAPTER 4: General Discussion & Conclusions

4.1 General discussion

The goal of this thesis was to carefully characterize vocal and musical emotional processing in cochlear implant users and normal-hearing subjects. In two studies, an evaluation was made of how cochlear implants affect the perception of auditory emotions and of how they impact neural emotional processing by using a well-controlled behavioural approach and by measuring previously-validated brain markers.

In the first study, the effect of CIs on the perception of vocal and musical emotions was assessed. Affective emotional bursts were used that controlled confounding factors in complex stimuli (sentences and songs; semantics, tempo, duration, and context) that masked the effect of CIs on emotional recognition, thus avoiding difficulties experienced in previous behavioural studies where it was hard to directly compare voice and music. (Hopyan et al., 2012; Stabej et al., 2012; Volkova et al., 2013; Ambert-Dahan et al., 2015). The stimuli were presented to normal-hearing participants in their original condition and in a validated CI-simulated condition using a vocoder. An assessment was made of the emotional recognition and the acoustical cues needed for recognition with cochlear implants. The results successfully showed a deficit and a correlation between timbral acoustical cues (brightness, energy and roughness; Lartillot et al., 2007) with the emotional rating for both vocal and musical emotional bursts in the CI-simulated condition. Finding the known deficit in emotional perception with cochlear implants with the employed well-controlled behavioural task, validated this approach and led the way to the next step - characterizing the brain mechanisms involved in emotional processing with cochlear implants.

In the second study, the same approach was employed while measuring EEG activity in order to assess the neural processing of vocal and musical emotion in their original and CI-simulated conditions by comparing ERPs. This approach had the advantage of comparing the neural processing of both conditions (natural versus CI-simulation) in the same brains. Results showed a reduction in amplitude with CI-simulation, which is a consistent finding in EEG studies with cochlear implants. Implant processors amplify sounds differently than the normal ear; thus it's hard to know whether the reduction in ERP amplitude was related to a difference in perceived loudness. The approach employed allowed, for the first time, the capability to disentangle this effect by presenting the sounds with the same loudness to the same participants under two conditions. The fact that a reduction in ERP amplitude was observed allows elimination of any effect of loudness. More interestingly, for the first time, a delay at P50 in the CI-simulated condition was found. This delay could be attributed to the longer time required for sensory processing when the signal is difficult to discriminate (Beynon et al., 2005; Agrawal et al., 2012). Thus, this implies a defect in processing the input of the auditory signal due to the implant processor's properties. From this and the results of the behavioural experiment, a clinical solution can be found for CI-users by focusing on timbral acoustical cues in rehabilitation programs and timbre discrimination musical training as it may enhance the perception of vocal and musical emotion. In addition, this information might aid in developing new implants' processing algorithms that preserve timbral acoustical cues and thus achieve better emotional recognition.

The study participants were not CI-users; thus, this approach only allowed the evaluation of the processing of sound signals - and not an evaluation of the effect of brain rewiring and plasticity post implantation. These results suggest future experiments: a neuro-behavioural study employing both CI-users and normal-hearing controls. It will be possible to associate neural markers with behaviour, and with clinical outcomes. The project should also assess the effect of brain reorganization on emotional processing by comparing pre- and post-lingually deaf populations, which would permit evaluation of various factors affecting plasticity (such as the duration of deafness, age at implantation, and language exposure).

4.2 General conclusion

In summary, this thesis characterizes the deficit in auditory emotional perception and helps clarify the neural mechanisms involved in processing CI-simulated sounds in normal hearing subjects by using well-controlled stimuli. In addition, timbral acoustical cues were found to be correlated with more accurate perception of auditory emotions. Also, the neural markers showed a delay in processing CI-simulated sounds, suggesting a possible defect in the implant processor. These results can be translated to clinical situations. The results of this thesis suggests capitalizing on timbral cues in the design of rehabilitation protocols, or by including them in the implant’s processors. This might improve the recognition of auditory emotions, thus improving the social and professional outcomes of CI-users.

REFERENCES

Agrawal, D., Thorne, J. D., Viola, F. C., Timm, L., Debener, S., Büchner, A. … Wittfoth, M. (2013). Electrophysiological responses to emotional prosody perception in cochlear implant users. NeuroImage : Clinical, 2, 229–238.

Agrawal, D., Timm, L., Viola, F. C., Debener, S., Büchner, A., Dengler, R., & Wittfoth, M. (2012). ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies. BMC Neuroscience, 13, 113.

Ambert-Dahan, E., Giraud, A.-L., Sterkers, O., & Samson, S. (2015). Judgment of musical emotions after cochlear implantation in adults with progressive deafness. Frontiers in Psychology, 6.

Başar, E. (1999). Brain Function and Oscillations. Berlin, Heidelberg: Springer Berlin Heidelberg.

Başar, E., Başar-Eroğlu, C., Karakaş, S., & Schürmann, M. (1999). Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG? Neuroscience Letters, 259(3), 165–168.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. Behavior research methods, 40(2), 531-539.

Bestelmeyer, P., Rouger, J., DeBruine, L., Belin, P., 2010. Auditory adaptation in vocal affect perception. Cognition.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. Cognition and Emotion, 19(8), 1113–1139.

Boersma, P., 2002. Praat, a system for doing phonetics by computer. Glot Int. 5.9/10 5, 341–345. doi:10.1097/AUD.0b013e31821473f7

Burkholder, R. A., Pisoni, D. B., & Svirsky, M. A. (2005). Effects of a cochlear implant simulation on immediate memory in normal-hearing adults. International Journal of Audiology, 44(10), 551–558.

Cacioppo, J. T. (2004). Feelings and emotions: roles for electrophysiological markers. Biological Psychology, 67(1-2), 235–243.

Caldwell MT, Jiradejvong P, Limb CJ. Impaired perception of sensory consonance and dissonance in cochlear implant users. Otology & Neurotology. 2016 Mar 1;37(3):229-34

Carlson, M. L., Driscoll, C. L., Gifford, R. H., & McMenomey, S. O. (2012). Cochlear implantation: current and future device options. Otolaryngologic Clinics of North America, 45(1), 221-248.

Cousineau, M., Demany, L., Meyer, B., & Pressnitzer, D. (2010). What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant. Hearing Research, 269(1–2), 34–41. https://doi.org/10.1016/j.heares.2010.07.007

Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. Motivation and Emotion, 12(4), 399–413.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. Cognition, 80(3), B1-B10.

Daly, I., Malik, A., Hwang, F., Roesch, E., Weaver, J., Kirke, A., … Nasuto, S. J. (2014). Neural correlates of emotional responses to music: An EEG study. Neuroscience Letters, 573, 52–57.

Davidson, R. J. (1998). Affective Style and Affective Disorders: Perspectives from Affective Neuroscience. Cognition and Emotion, 12(3), 307–330.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods, 134(1), 9-21.

Dolgin, K. G., & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. Psychology of Music, 18(1), 87-98.

Driscoll, V.D., Oleson, J., Jiang, D., Gfeller, K., 2009. Effects of Training on Recognition of Musical Instruments Presented through Cochlear Implant Simulations. J. Am. Acad. Audiol. 20, 71–82.

Eggermont, J. J., & Ponton, C. W. (2003). Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: correlations with changes in structure and speech perception. Acta Oto-Laryngologica, 123(2), 249–252.

Eggermont, J. J., Ponton, C. W., Don, M., Waring, M. D., & Kwong, B. (1997). Maturational delays in cortical evoked potentials in cochlear implant users. Acta Oto-Laryngologica, 117(2), 161–163.

Ekman, P. (1992). An argument for basic emotions. Cognition & emotion, 6(3-4), 169-200.

Fallon, J. B., Irvine, D. R., & Shepherd, R. K. (2008). Cochlear implants and brain plasticity. Hearing research, 238(1), 110-117.

Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, Arezzo JC, Howard MA, Steinschneider M. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. Journal of Neurophysiology. 2001 Dec 1;86(6):2761-88

Flom, R., Gentile, D., & Pick, A. (2008). Infants' discrimination of happy and sad music. Infant Behavior And Development, 31(4), 716-728.

Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. Journal of the Association for Research in Otolaryngology: JARO, 6(1), 19–27. https://doi.org/10.1007/s10162-004-5024-3

Fu, Q.-J., Nogaki, G., & Galvin, J. J. (2005). Auditory Training with Spectrally Shifted Speech: Implications for Cochlear Implant Patient Auditory Rehabilitation. JARO: Journal of the Association for Research in Otolaryngology, 6(2), 180–189. https://doi.org/10.1007/s10162-005-5061-6

Gabrielsson, A., Lindström, E., 2001. The influence of musical structure on emotional expression. Oxford University Press.

Gfeller, K., & Lansing, C. (1992). Musical perception of cochlear implant users as measured by the Primary Measures of Music Audiation: an item analysis. Journal of Music Therapy, 29(1), 18-39.

Gfeller, K., & Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. Journal of Speech, Language, and Hearing Research, 34(4), 916-920.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., ... & Stordahl, J. (2002a). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. Cochlear implants international, 3(1), 29-53.

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., & Ringgenberg, S. (2002b). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. Journal of the American Academy of Audiology, 13(3), 132-145.

Gfeller, K., Witt, S., Mehr, M. A., Woodworth, G., & Knutson, J. (2002c). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. Annals of Otology, Rhinology & Laryngology, 111(4), 349-356.

Gfeller, K., Witt, S., Stordahl, J., Mehr, M., & Woodworth, G. (2000). The effects of training on melody recognition and appraisal by adult cochlear implant recipients. Journal-Academy of Rehabilitative Audiology, 33, 115-138.

Giannantonio, S., Polonenko, M. J., Papsin, B. C., Paludetti, G., & Gordon, K. A. (2015). Experience Changes How Emotion in Music Is Judged: Evidence from Children Listening with Bilateral Cochlear Implants, Bimodal Devices, and Normal Hearing. PLoS ONE, 10(8), e0136685.

Gosselin, N., Paquette, S., & Peretz, I. (2015). Sensitivity to musical emotions in congenital amusia. cortex, 71, 171-182.

Gosselin, N., Peretz, I., Johnsen, E., & Adolphs, R. (2007). Amygdala damage impairs emotion recognition from music. Neuropsychologia, 45(2), 236-244.

Groenen, P. A., Beynon, A. J., Snik, A. F., & Broek, P. V. D. (2001). Speech-evoked cortical potentials recognition in cochlear implant users and speech. Scandinavian audiology, 30(1), 31-40.

Hegerl, U., & Juckel, G. (1993). Intensity dependence of auditory evoked potentials as an indicator of central serotonergic neurotransmission: a new hypothesis. Biological psychiatry, 33(3), 173-187.

Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. Comprehensive Physiology.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. Science, 182(4108), 177-180.

Hopyan, T., Peretz, I., Chan, L. P., Papsin, B. C., & Gordon, K. A. (2012). Children using cochlear implants capitalize on acoustical hearing for music perception. Frontiers in psychology, 3, 425.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. Psychological Science, 15(5), 356-360.

Iredale, J. M., Rushby, J. A., McDonald, S., Dimoska-Di Marco, A., & Swift, J. (2013). Emotion in voice matters: Neural correlates of emotional prosody perception. International Journal of Psychophysiology, 89(3), 483–490. https://doi.org/10.1016/j.ijpsycho.2013.06.025

Izard, C. E. (1993). Organizational and motivational functions of discrete emotions.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. NeuroImage, 58(2), 665–674. http://doi.org/10.1016/j.neuroimage.2011.06.035

Juslin, P. N. (1998). A functionalist perspective on emotional communication in music performance.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. Journal of Experimental Psychology: Human perception and performance, 26(6), 1797.

Juslin, P. N., & Laukka, P. (2000). Improving Emotional Communication in Music Performance through Cognitive Feedback. Musicae Scientiae, 4(2), 151–183. http://doi.org/10.1177/102986490000400202

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin, 129(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Juslin, P. N., & Sloboda, J. A. (2001). Music and emotion: Theory and research. Oxford University Press.

Kaganovich, N., Kim, J., Herring, C., Schumaker, J., MacPherson, M., & Weber-Fox, C. (2013). Musicians show general enhancement of complex sound encoding and better inhibition of irrelevant auditory change in music: an ERP study. European Journal of Neuroscience, 37(8), 1295–1307. https://doi.org/10.1111/ejn.12110

Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. Cognition & Emotion, 13(5), 467-480.

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking Brainwaves to the Brain: An ERP Primer. Developmental Neuropsychology, 27(2), 183–215. https://doi.org/10.1207/s15326942dn2702_1

Kleiner, M., Brainard, D.H., Pelli, D.G., Broussard, C., Wolf, T., Niehorster, D., 2007. What's new in Psychtoolbox-3? Perception 36, S14. doi:10.1068/v070821

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1981). The effects of frontal cortex lesions on event-related potentials during auditory selective attention. Electroencephalography and clinical neurophysiology, 52(6), 571-582.

Koelsch, S. (2011). Toward a Neural Basis of Music Perception – A Review and Updated Model. Frontier in Psychology, 2. http://doi.org/10.3389/fpsyg.2011.00110

Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nature Reviews Neuroscience, 15(3), 170–180. http://doi.org/10.1038/nrn3666

Kong, Y. Y., Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. Ear and hearing, 25(2), 173-185.

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J.-D. (2013). Predicting vocal emotion expressions from the human brain. Human Brain Mapping, 34(8), 1971–1981. http://doi.org/10.1002/hbm.22041

Krause, C. M. (2003). Brain electric oscillations and cognitive processes. In Experimental methods in neuropsychology (pp. 111-130). Springer US.

Lakshminarayanan, K., Ben Shalom, D., van Wassenhove, V., Orbelo, D., Houde, J., & Poeppel, D. (2003). The effect of spectral manipulations on the identification of affective and linguistic prosody. Brain and Language, 84(2), 250–263.

Laneau, J., Wouters, J., & Moonen, M. (2004). Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. Journal of the Acoustical Society of America, 116(6), 3606-3619.

Language, T., Computing, T., 2004. MATLAB The Language of Technical Computing. Components 3, 750. doi:10.1007/s10766-008-0082-5

Lartillot, O., Lartillot, O., Toiviainen, P., Toiviainen, P., 2007. A matlab toolbox for musical feature extraction from audio. Int. Conf. Digit. Audio … 1–8.

Lassaletta L, Castro A, Bastarrica M, Pérez-Mora R, Madero R, De Sarriá J, Gavilán J. Does music perception have an impact on quality of life following cochlear implantation?. Acta oto-laryngologica. 2007 Jan 1;127(7):682-6

Latinus, M., & Belin, P. (2011). Human voice perception. Current Biology, 21(4), R143-R145.

Lehmann, A., & Paquette, S. (2015). Cross-domain processing of musical and vocal emotions in cochlear implant users. Frontiers in neuroscience, 9.

Levitin, D. J. and Tirovolas, A. K. (2009), Current Advances in the Cognitive Neuroscience of Music. Annals of the New York Academy of Sciences, 1156: 211–231.

Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. Neuroreport, 12(12), 2653-2657.

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology, 40(2), 291–305. https://doi.org/10.1111/1469-8986.00031

Limb, C. J. (2006). Cochlear implant-mediated perception of music. Current opinion in otolaryngology & head and neck surgery, 14(5), 337-340.

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations: NeuroReport, 23(2), 108–112. https://doi.org/10.1097/WNR.0b013e32834ea757

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. British Journal of Psychology, 102(1), 108-134.

Looi, V., Gfeller, K., & Driscoll, V. D. (2012, November). Music appreciation and training for cochlear implant recipients: a review. In Seminars in hearing (Vol. 33, No. 04, pp. 307-334). Thieme Medical Publishers.

Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). The effect of cochlear implantation on music perception by adults with usable pre-operative acoustic hearing. International journal of audiology, 47(5), 257-268.

Luck, S. (2005). An introduction to the event-related potential technique. Cambridge, Mass.: MIT Press.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press.

Luo, X., Fu, Q. J., & Galvin, J. J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. Trends in Amplification, 11(4), 301-315. Xin Luo, X., Fu, Q.-J., Galvin, J.J., 2007. Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users. Trends Amplif. 11, 301–315. doi:10.1177/1084713807305301

Macherey, O., & Carlyon, R. P. (2014). Cochlear implants. Current Biology, 24(18), R878-R884.

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. Psychophysiology, 32(1), 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Marsella, P., Scorpecci, A., Vecchiato, G., Maglione, A. G., Colosimo, A., & Babiloni, F. (2014). Neuroelectrical imaging investigation of cortical activity during listening to music in prelingually deaf children with cochlear implants. International Journal of Pediatric Otorhinolaryngology, 78(5), 737–743. http://doi.org/10.1016/j.ijporl.2014.01.030

Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: Differential neural responses to perception and imagery of speech and music. Restorative Neurology and Neuroscience, 25(3–4), 411–431.

Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. The Journal of the Acoustical Society of America, 114(1), 408–419.

Most, T., & Aviner, C. (2009). Auditory, visual, and auditory–visual perception of emotions by individuals with cochlear implants, hearing aids, and normal hearing. Journal of Deaf Studies and Deaf Education, enp007.

Nakata, T., Trehub, S. E., & Kanda, Y. (2012). Effect of cochlear implants on children's perception and production of speech prosody. Journal of the Acoustical Society of America, 131(2), 1307–1314. http://doi.org/10.1121/1.3672697

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. The Journal of the Acoustical Society of America, 95(2), 1085-1099.

Nogaki, G., Fu, Q. J., & Galvin III, J. J. (2007). The effect of training rate on recognition of spectrally shifted speech. Ear and hearing, 28(2), 132.

Paquette, S., Peretz, I., & Belin, P. (2013). The “Musical Emotional Bursts”: a validated set of musical affect bursts to investigate auditory affective processing.

Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deli & J. Sloboda (Eds.), Perception and cognition of music (pp. 191–215). Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. Frontiers in Psychology, 4. http://doi.org/10.3389/fpsyg.2013.00345

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. Biological psychology, 111, 14-25.

Peretz, I. (2002). Book Review: Brain Specialization for Music. The Neuroscientist, 8(4), 372–380. http://doi.org/10.1177/107385840200800412

PERETZ, I. (2010). TOWARDS A. Handbook of Music and Emotion: Theory, Research, Applications, 99.

Peretz, I., Aubé, W., & Armony, J. L. (2013). Towards a neurobiology of musical emotions. The evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man, 277-299.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. Cognition, 68(2), 111–141. http://doi.org/10.1016/S0010-0277(98)00043-2

Pinheiro, A. P., Galdo-Álvarez, S., Rauber, A., Sampaio, A., Niznikiewicz, M., & Gonçalves, Ó. F. (2011). Abnormal processing of emotional prosody in Williams syndrome: An event-related potentials study. Research in Developmental Disabilities, 32(1), 133–147. https://doi.org/10.1016/j.ridd.2010.09.011

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. Brain and Language, 140, 24–34. http://doi.org/10.1016/j.bandl.2014.10.009

Poissant, S. F., Whitmal, N. A., & Freyman, R. L. (2006). Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. The Journal of the Acoustical Society of America, 119(3), 1606–1615.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. The Journal of the Acoustical Society of America, 114(1), 446–454.

Quarto, T., Blasi, G., Pallesen, K. J., Bertolino, A., & Brattico, E. (2014). Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits. PloS one, 9(7), e103278.

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. Neuroscience, 290, 175-184.

Ritter, W., Simson, R., & Vaughan, H. G. (1988). Effects of the amount of stimulus information processed on negative event-related potentials. Electroencephalography and Clinical Neurophysiology, 69(3), 244-258.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. The Journal of Neuroscience, 31(10), 3843-3852.

Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., & Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. IEEE Trans. Audio Electroacoust, 17(3), 225-246.

Sandmann, P., Eichele, T., Buechler, M., Debener, S., Jäncke, L., Dillier, N., ... & Meyer, M. (2009). Evaluation of evoked potentials to dyadic tones after cochlear implantation. Brain, 132(7), 1967-1979.

Sandmann, P., Kegel, A., Eichele, T., Dillier, N., Lai, W., Bendixen, A., … Meyer, M. (2010). Neurophysiological evidence of impaired musical sound perception in cochlear-implant users. Clinical Neurophysiology, 121(12), 2070–2082. http://doi.org/10.1016/j.clinph.2010.04.032

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. Psychological bulletin, 99(2), 143.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. Speech Communication, 40(1–2), 227–256. http://doi.org/10.1016/S0167-6393(02)00084-5

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends in Cognitive Sciences, 10(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schmidt, S., & Zimmermann, E. (2013). The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man. OUP Oxford.

Scorpecci, A., Zagari, F., Mari, G., Giannantonio, S., D'Alatri, L., Di Nardo, W., & Paludetti., G. (2012). Investigation on the music perception skills of Italian children with cochlear implants. International Journal of Pediatric Otorhinolaryngology, 76(10), 1507–1514. http://doi.org/10.1016/j.ijporl.2012.07.005

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science (New York, N.Y.), 270(5234), 303–304.

Sharda, M., Midha, R., Malik, S., Mukerji, S., & Singh, N. C. (2015). Fronto-Temporal Connectivity is Preserved During Sung but Not Spoken Word Listening, Across the Autism Spectrum. Autism Research, 8(2), 174-186.

Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., & Vatovec, J. (2012). The music perception abilities of prelingually deaf children with cochlear implants. International Journal of Pediatric Otorhinolaryngology, 76(10), 1392–1400. http://doi.org/10.1016/j.ijporl.2012.07.004

Terwogt, M. M., & Van Grinsven, F. (1988). Recognition of emotions in music by children and adults. Perceptual and Motor Skills, 67(3), 697–698. http://doi.org/10.2466/pms.1988.67.3.697

Volkova, A., Trehub, S. E., Schellenberg, E. G., Papsin, B. C., & Gordon, K. A. (2013). Children with bilateral cochlear implants identify emotion in speech and music. Cochlear Implants International: An Interdisciplinary Journal, 14(2), 80–91. http://doi.org/10.1179/1754762812Y.0000000004

Wang, D.J., Trehub, S.E., Volkova, A., van Lieshout, P., 2013. Child implant users' imitation of happy- and sad-sounding speech. Front. Psychol. 4, 351. doi:10.3389/fpsyg.2013.00351

Wiefferink, C. H., Rieffe, C., Ketelaar, L., & Frijns, J. H. (2012). Predicting social functioning in children with a cochlear implant and in normal-hearing children: The role of emotion regulation. International journal of pediatric otorhinolaryngology, 76(6), 883-889.

Wright, R., & Uchanski, R. M. (2012). Music Perception and Appraisal: Cochlear Implant Users and Simulated CI Listening. Journal of the American Academy of Audiology, 23(5), 350–379. http://doi.org/10.3766/jaaa.23.5.6

Yuan-Pin Lin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, & Jyh-Horng Chen. (2010). EEG-Based Emotion Recognition in Music Listening. IEEE Transactions on Biomedical Engineering, 57(7), 1798–1806. http://doi.org/10.1109/TBME.2010.2048568

ABBREVIATIONS

CI Cochlear Implant

EEG Electroencephalography

ERP Event-related potential

fMRI Functional Magnetic Radio Imaging

MAV Montreal Affective Voices

MEB Musical Emotional Bursts

MMN Mismatched Negativity

NH Normal hearing

STG Superior Temporal Gyrus

STS Superior Temporal Sulcus

SNHL Sensorineural Hearing Loss

LIST OF TABLES AND FIGURES

Table 1. Average value of extracted acoustical features presented for each stimuli type (voice, Music), emotion (happy, sad, fear neutral) and condition (original, CI-sim) standard deviation is presented in parentheses. 25

Table 2. Confusion matrix of emotion recognition (accuracy %) for the vocal and musical sounds in both the CI-sim and original conditions. When two emotions were rated equally, identification was rated undefined. Framed values were not significantly different from chance level (25%).......... 26

Figure 1. Average effect (difference) of CI simulation on accuracy scores for each emotion (happy, sad, fear neutral) and category (voice, music). Error bars represent the standard deviation. *= $p < .05$.......... 27

Figure 2. Raw arousal and valence ratings for both the vocal and musical stimuli presented for each emotion (happy, sad, fear neutral) and condition (original, CI-sim). 28

Figure 3. Average effect of the CI simulation on arousal ratings presented for each emotion (happy, sad, fear neutral) and stimuli type (voice, music). Error bars represent the standard deviation. *= $p < .05$.......... 29

Figure 4. Average effect of the CI simulation on valence ratings presented for each emotion (happy, sad, fear neutral) and stimuli type (voice, music). Error bars represents the standard deviation. * = $p < .05$.......... 30

Figure 5. Average ERP waveforms recorded at Cz electrode in original (NH) and CI-simulated (CI) conditions for the vocal and musical stimuli from 50 ms before onset to 800 ms after onset of the stimuli (X-axis: latency in milliseconds, Y-axis: amplitude in µV).......... 44

Figure 6. P50 latencies in the original (NH) and CI-simulated (CI-sim) conditions (error bars represent standard deviation).......... 46

Figure 7. A comparison between the P50 latency for all the vocal and musical stimuli (NH & CI-sim) (error bars represent standard deviation). *Happy musical stimuli have longer latencies than fearful stimuli ($p < .05$). 47

Figure 8. A comparison between the N100 latency for all the vocal and musical stimuli (NH & CI-sim) (error bars represent standard deviation). * Vocal stimuli (happy and neutral) have significantly shorter latencies than their musical counterparts ($p < .05$). ** Sad vocal stimuli have significantly longer latencies than the other vocal emotions ($p < .05$). *** Fear musical stimuli have shorter latencies than the other musical emotions ($p < .05$). 48

Figure 9. P200 latency in the original condition (error bars represent standard deviation). * Musical fear have an earlier latency than sad and neutral ($p < .05$). 50

Figure 10. Effect of CI-simulation on the N100-P200 complex (error bars represent standard deviation). All the stimuli were affected by the CI-simulation ($p < 0.05$). * Sad musical stimuli were the least affected by the CI-simulation than the other emotions ($p < 0.05$). 51